# COMPUTER-AIDED DIAGNOSIS IN MAMMOGRAPHY: CORRELATION OF REGIONS IN MULTIPLE STANDARD MAMMOGRAPHIC VIEWS OF THE SAME BREAST

By

Jayanethie Padayachee

Submitted in fulfilment
of the academic requirements for the degree of
Doctor of Philosophy
in the
School of Physics,
University of KwaZulu-Natal
Durban, South Africa

September 1, 2006

# Abstract

According to the Cancer Association of South Africa, breast cancer is currently the most common cancer among women worldwide and second to cervical cancer in South Africa. Although much progress has been made in the treatment of breast cancer, the key is early detection. Mammography is currently the most effective method of detecting breast cancer in its early stages, but the analysis of mammograms is sometimes difficult due to the complex and varying structure of the human breast. Adding to the complexity is the fact that abnormalities appear very rarely and radiologists that are tired or distracted may miss the signs of breast cancer, especially if the signs are very subtle.

Computer-aided diagnosis (CAD) systems were introduced to consistently highlight those features that may be missed by a radiologist and studies have shown that a radiologist's performance is enhanced when "prompted" by a CAD-system. However, the high number of false-positive areas highlighted by current CAD-systems provides further distraction and wastes the radiologist's time. CAD-systems generally emulate the actions of a radiologist in interpreting a mammogram, however no CAD-system uses both standard mammographic views of the same breast to confirm the presence of abnormalities and reduce false-positives.

This study uses texture-based image processing methods to investigate a method of analysis that matches a suspicious feature from one standard mammographic view to the same feature in the other mammographic view of the same breast. The matching algorithms employ grey-level co-occurrence matrices (GLCMs), texture measures and similarity metrics (Euclidean and Mahalanobis distance, mutual information) for matching the information between mammographic views of the same breast. The algorithms are applied to 68 pairs of cranio-caudal and mediolateral-oblique mammograms as well as stereotactic biopsy mammograms. Results are evaluated in terms of the area under the receiver operating characteristic curve ($A_{ROC}$) and contrast ($C_{fb}$).

The best results for the pairs of mammograms were obtained for matching using texture measures and a Euclidean distance similarity metric, which achieved an average $A_{ROC}$=0.80±0.17 with an average $C_{fb}$=0.46±0.26 while mutual information with GLCMs achieved an average $A_{ROC}$=0.77±0.25 with an average $C_{fb}$=0.50±0.42. Mutual information with GLCMs performed remarkably well with the matching of malignant masses and achieved an average $A_{ROC}$=0.96±0.05 with an average $C_{fb}$=0.90±0.21. Results of applying the algorithms to match regions between stereotactic biopsy mammograms were significantly poorer than the results for the standard mammograms and it was not possible to adequately demonstrate that the matching algorithms can improve localisation accuracy for biopsies.

# Preface

The work described in this thesis was carried out in the Applied Physics Group of the School of Physics, University of KwaZulu-Natal, Durban, South Africa, from 1 April 2002 to 31 December 2005, under the co-supervision of Prof. Michael J. Alport (School of Physics) and Dr. William I. D. Rae (Inkosi Albert Luthuli Central Hospital).

These studies represent original work by the author and have not otherwise been submitted in any form for any degree or diploma to another tertiary institution. Where use has been made of the work of others, it is duly acknowledged in the text.

Padayachee
08/09/2006

# Dedication

For Kevin. For being by my side for the eight years, four projects, three supervisors and two cities, that has been this degree, and for putting up with me during my few highs and many, many lows. Thank you.

# Acknowledgements

I would like to express my appreciation to the following.

**For advice and assistance during this research project:**

- My supervisors - Prof. M. J. Alport and Dr. W. I. D. Rae
- Addington Hospital and Inkosi Albert Luthuli Central Hospital staff - Ms. P. Baxter, Dr. S. Chick, Dr. F. Malan, Dr. L. Marishane, Ms. L. Moodley, Dr. I.J. Movson, Dr. J. Sarawan and Dr. J. Zietkiewicz
- Prof. C. Herbst (Department of Medical Physics, U. Free State) and Mrs. M.A. Sweetlove (Independent Consultant Medical Physicist)
- University of KwaZulu-Natal, Department of Labour (South Africa), National Research Foundation and Medical Research Council (South Africa) for funding

**For support of body, mind and computer(s):**

- My mother and brother (Mrs. L. Padayachee, Mr. T. Padayachee)
- Mrs. G.R.A. Browne
- Fellow PhD students (Mr. M. Gar-Elnabi, Mr. A. Hamza)
- Members of the School of Physics (Prof. M. Hellberg, Mr. R. Piasecki, Ms. C. Webster)
- Members of the SAIP Council (Prof. N. Chetty, Prof. D. Grayson, Prof. J.A.A. Engelbrecht, Prof. H. Moraal, Dr. P.A. Whitelock, Prof. E.C. Zingu)
- Dr. G. Hough

# Ode to a Mammogram

For years and years they told me, Be careful of your breasts.
Don't ever squeeze or bruise them. And give them monthly tests.
So I heeded all their warnings, And protected them by law.
Guarded them very carefully, And always wore my bra.
After 30 years of astute care, My gyno, Dr. Pruitt,
Said I should get a Mammogram. "O. K," I said, "let's do it. "
"Stand up here real close" she said, (She got my boob in line),
"And tell me when it hurts," she said, "Ah yes! Right there, that's fine. "
She stepped upon a pedal, I could not believe my eyes!
A plastic plate came slamming down, My hooter's in a vice!
My skin was stretched and mangled, From underneath my chin.
My poor boob was being squashed, To Swedish Pancake thin.
Excruciating pain I felt, Within its vicelike grip.
A prisoner in this vicious thing, My poor defenceless tit!
"Take a deep breath" she said to me, Who does she think she's kidding?!?
And woozy I am getting. "There, that's good," I heard her say, (The room was slowly swaying. )
"Now, let's have a go at the other one." Have mercy, I was praying.
It squeezed me from both up and down, It squeezed me from both sides.
I'll bet SHE'S never had this done, To HER tender little hide.
Next time that they make me do this, I will request a blindfold.
I have no wish to see again, My knockers getting steamrolled.
If I had no problem when I came in, I surely have one now.
If there had been a cyst in there, It would have gone "kerpow!"
This machine was created by a man, Of this, I have no doubt.
I'd like to stick his balls in there, And see how THEY come out.
–Anon.

# Table of Contents

# List of Symbols

| | |
|---|---|
| $A_{ROC}$ | area under the receiver operating characteristic curve determined by the trapezoidal rule |
| $A_z$ | area under the receiver operating characteristic curve determined by binormal model |
| $\mathbf{C}$ | co-variance matrix |
| $C_{fb}$ | contrast with foreground grey-level, $f$, and background grey-level, $b$ |
| $D_{ES}$ | Euclidean distance with standardised variables |
| $D_E$ | Euclidean distance |
| $D_M$ | Mahalanobis distance |
| $d$ | distance between two pixels in grey-level co-occurrence matrix calculation |
| $FN$ | number of false-negatives |
| $FPF$ | false-positive fraction |
| $FP$ | number of false-positives |
| $\mathbf{G}_{\theta,d}$ | grey-level co-occurrence matrix at angle, $\theta$, and distance, $d$ |
| $g$ | grey-level |
| $H$ | entropy |
| $h_{ACF}$ | scale height of image texture obtained from autocorrelation function |
| $\kappa$ | matching accuracy $= A_{ROC} \cdot C_{fb}$ |
| $nbins$ | number of bins used to construct grey-level histogram |
| $nbits$ | number of grey-levels in image expressed in bits |
| $P(X)$ | probability density function of random variable $X$ |
| $p(x_i)$ | probability associated with $i^{th}$ state of random variable $X$ |
| $p$ | significance obtained from paired $t$-test |
| $\rho$ | Pearson's linear correlation coefficient |
| $\sigma$ | standard deviation |
| $\theta$ | angle between two pixels in grey-level co-occurrence matrix calculation |
| $TNF$ | true-negative fraction |
| $TN$ | number of true-negatives |
| $TPF$ | true-positive fraction |
| $TP$ | number of true-positives |
| $w_{ACF}$ | scale width of image texture obtained from autocorrelation function |
| $w_{sf}$ | sampling window step fraction |
| $w_{step}$ | sampling window step size |
| $w$ | sampling window size |

# List of Abbreviations

| | |
|---|---|
| ACF | autocorrelation function |
| B | benign (diagnosis of region of interest) |
| c | calcification (mammographic feature) |
| CAD | computer-aided diagnosis |
| CC | cranio-caudal |
| CR | computed radiography |
| CT | computed tomography |
| dpi | dots per inch |
| FDA | United States Food and Drug Administration |
| FFDM | full-field digital mammography |
| FROC | free response receiver operating characteristic |
| GLCM | grey-level co-occurrence matrix |
| GLH | grey-level histogram |
| I | indeterminate (diagnosis of region of interest) |
| m | mass (mammographic feature) |
| M | malignant (diagnosis of region of interest) |
| MI | mutual information |
| MIAS | Mammographic Image Analysis Society |
| MLO | mediolateral-oblique |
| MRI | magnetic resonance imaging |
| N | normal (diagnosis of region of interest) |
| PET | positron emission tomography |
| ROC | receiver operating characteristic |
| ROI | region of interest |
| SAAPMB | South African Association of Physicists in Medicine and Biology |
| SAIP | South African Institute of Physics |
| SB | stereotactic biopsy |
| SGLD | spatial grey-level dependence (matrix) |
| TM | texture measure |

# List of Tables

# List of Figures

# List of Algorithms

# Chapter 1

# Introduction

## 1.1 A World of Pictures

*A picture is worth a thousand words.* From drawing in the sand to cave paintings to the Sistine Chapel, Man has always expressed himself in pictures. Pictures have also served an important role in recording information, and even writing is an evolved form of capturing information in pictures.

Pictures, moving and still, form a vital part of society from entertainment to security to health. In the field of medicine where the early diagnosis of diseases is crucial, medical imaging has become a standard part of the diagnostic process, often giving physicians a non-invasive glimpse inside the human body. In the past few decades, medical imaging has evolved from only being used for visualisation and inspection of anatomic features, to becoming an important tool for surgical and radiotherapy planning, intra-operative navigation and to track the progress of a disease. The *modalities* or *methods of imaging* have become equally diverse. Radiography, using x-rays to image internal body structures, is still the most common modality, but Magnetic Resonance Imaging (MRI), Computed Tomography (CT), Positron Emission Tomography (PET) and ultrasonography are being increasingly used as complementary modalities to radiography [McInerney & Terzopoulos 1996].

The increased use of imaging for medical diagnosis, especially in the early detection of cancer, has led to an increase in the amount of information that has to be processed by physicians. This in turn, has led to the development of computer-based methods to assist physicians to consistently and efficiently process the extra information from images.

## 1.2 Breast Cancer

According to the Cancer Association of South Africa[1], breast cancer is currently the most common cancer among women worldwide, and is second to cervical cancer among South African women. Whilst sometimes fatal, breast cancer can be successfully treated, provided it is detected early. The most common method of detecting breast cancer in its early stages is mammography, which uses low energy x-rays to image the human breast. X-rays were first used to study diseases of the breast in 1913, but mammography as a method of detecting breast cancer only started in the 1970's [Tabar & Dean 1987, Gold et al. 1990, Elmore et al. 2005].

---

[1] `http://www.cansa.org.za/`

## 1.3 Mammography

### 1.3.1 The Mammography Process

During mammography, the breast is compressed between two parallel plates to an average thickness ranging approximately between 4 cm and 6 cm [Highnam et al. 1998*b*] and is exposed to low energy x-rays along the direction of compression. The image that is created in a plane perpendicular to the direction of compression is a projection of the compressed breast. This image is known as a *mammogram* and is visually analysed by a radiologist for the signs of breast cancer. A schematic of the mammography process is shown in **Figure 1.1**.

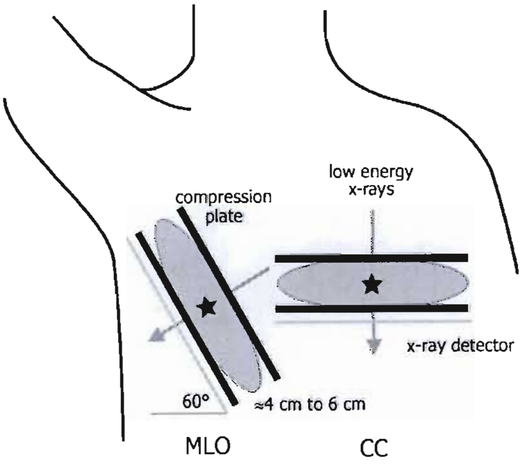


**Figure 1.1**: Schematic of the mammography process showing two standard mammographic views: mediolateral-oblique (MLO) and cranio-caudal (CC). The breast, compressed between two parallel plates, is exposed to x-rays along the direction of compression (indicated by arrows) and is imaged in a plane perpendicular to the direction of compression. The resulting image is known as a mammogram.

### 1.3.2 Anatomy of the Human Breast and the Mammogram

The human breast has a very complex structure that is composed of varying amounts of fibrous, glandular, fatty and lymphatic tissues (**Figure 1.2**). The structure of the breast varies with age and from patient to patient and this variability makes the analysis of mammograms a very difficult task. An important point to note is that there is no definitive normal, healthy breast, but that there is a wide range of features which make up a normal, healthy breast.

The different tissues in the breast have different densities and consequently different x-ray attenuation factors. The differences in tissue density are seen as variations in brightness (or intensity) on the mammogram because the high attenuation factor of a region with dense tissue causes fewer x-rays to reach the mammographic film resulting in a bright region on the film. Similarly, the lower attenuation factor of an area with less dense tissue causes more x-rays to reach the mammographic film resulting in a dark region on the film [Huynh et al. 1998, Mudigonda et al. 2001, Bushberg et al. 2002].

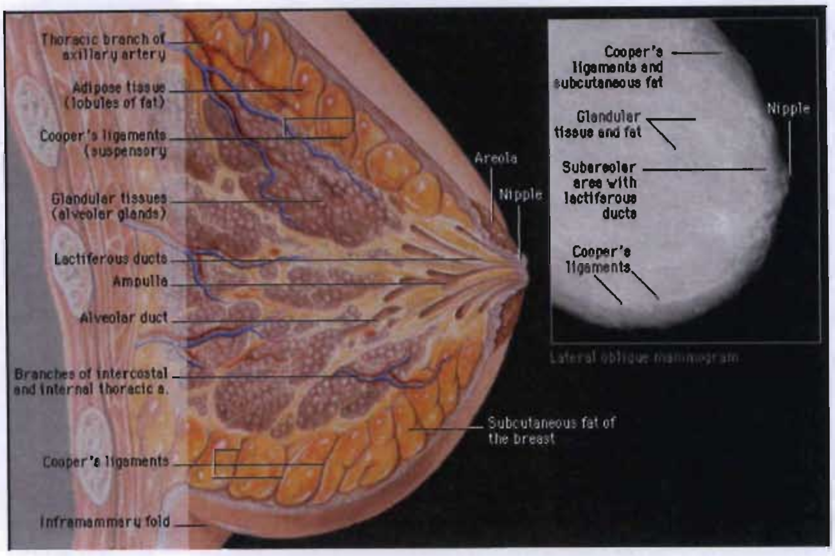


**Figure 1.2**: Anatomy of the breast with a mammogram inset [Yale 2002]. The human breast is composed of various tissue types that result in varying intensities in the mammogram.

### 1.3.3 Screening and Diagnostic Mammography

There are two main uses for mammography: screening and diagnostic. Screening mammography is routinely performed on women not exhibiting symptoms of breast cancer to detect a potential cancer. Diagnostic mammography is performed on women exhibiting physical symptoms consistent with breast cancer, or to further evaluate a specific finding, by obtaining mammograms at additional angles, magnifications and compressions. Diagnostic mammography is often complemented by ultrasonography to differentiate between benign cysts and malignant masses that have similar mammographic appearances [Bushberg et al. 2002, Majid et al. 2003].

Biennial mammograms are recommended for women older than 40 years and it has been shown that periodic screening can reduce breast cancer mortality [US Preventive Services Task Force 2002, Tabar et al. 2002, Feig 2005].

### 1.3.4 Standard Mammographic Views

**Figure 1.1** shows a schematic of the technique used to obtain two of the standard mammographic views, mediolateral-oblique (MLO) and cranio-caudal (CC). Two views are routinely taken as it has been shown that breast cancer is more effectively diagnosed with two views than with a single view [Wald et al. 1995, Chan et al. 1999, Paquerault et al. 2002]. **Table 1.1** details the various mammographic views and the direction of compression for each view. Example CC- and MLO-view mammograms are shown in **Figure 1.4**.

### 1.3.5 X-ray Detector Technologies

Three x-ray detector technologies are currently used in mammography: screen-film mammography, computed radiography and full-field digital mammography. Traditionally, x-ray film was

**Table 1.1**: Possible mammographic views and directions of compression for each view

| View | Compression Direction | Angle of compression |
|---|---|---|
| cranio-caudal (CC) | top to bottom | 0° |
| mediolateral-oblique (MLO) | upper middle to lower side | 60° |
| mediolateral (ML) | middle to side | 90° |
| lateromedial (LM) | side to middle | 90° |
| compression/spot | local | - |

directly exposed to x-rays, but film insensitivity meant that the patient received a very high dose of radiation. The current technologies all require significantly lower x-ray doses to achieve a clinically useful contrast between the different tissue types.

#### 1.3.5.1 Screen-film Mammography

The film is placed in direct contact with a phosphor screen in screen-film mammography. The screen-film combination is placed inside a light-tight cassette made of a low x-ray attenuation material and the cassette is placed in the mammography machine such that the x-rays pass through the cassette and film before interacting with the screen. The interaction between the phosphor and the x-rays causes light, in direct proportion to the intensity of the incident x-rays, to be emitted. This light exposes the film. The advantages of screen-film mammography (compared to radiographic film only) are the significantly lower dose, its good spatial resolution (which is only dependent on the crystal size in the screen and film and these can be made very small) and its low cost (which makes it widely available). However, the film has a limited exposure dynamic range and the contrast of the mammogram depends on the speed of the film that is used. Also, poor screen-film contact and film processing parameters adversely affect the quality of the mammogram [Bushberg et al. 2002].

#### 1.3.5.2 Computed Radiography

Computed radiography (CR) is the second most-common detector technology used in mammography. The main components of a CR system are the image plate and the image plate reader. The image plate, which is placed in a light-tight enclosure, consists of a photostimulable or storage phosphor screen. X-rays incident on the screen cause charge to be trapped in metastable traps, creating a latent image, which is read out by raster scanning with a laser. When the screen is optically stimulated, the trapped charge is released from the metastable traps, triggering a process known as photostimulated luminescence where short wavelength light is emitted in an amount proportional to the original incident x-ray intensity. The emitted light is collected with a light guide and detected by a photomultiplier tube. The photomultiplier tube signal is digitised to form the image on a pixel-by-pixel basis. The advantages of CR are that the detector has a large exposure dynamic range and that it has a digital nature. However, CR has a limited spatial resolution, which is dependent on the sampling frequency used during readout [Rowlands 2002].

#### 1.3.5.3 Full-Field Digital Mammography

The third x-ray detector technology used in mammography is full-field digital mammography (FFDM) where a full field-of-view digital detector captures the transmitted x-rays. FFDM systems have only been in use since 2002. However, prior to that, small field-of-view detectors were used on biopsy units. The advantages of FFDM are that it has a large exposure dynamic range and that it is a fully digital technology. However, because it is a relatively new technology, it is very expensive and is, therefore, not widely used. The spatial resolution of FFDM-systems is determined by the size of the pixel elements in the detector and this can be constructed to be close to that of a screen-film system. However, small pixel sizes means higher storage requirements for the digital images [Bushberg et al. 2002].

### 1.3.6 Breast Compression

Compression of the breast during mammography reduces overlapping tissue and decreases the effective thickness of the breast. The effect of the latter is to reduce the x-ray dose to the patient, to reduce the scatter of x-rays passing through the breast and to reduce blurring on the mammogram. A uniformly thick breast also means that a higher contrast film can be used because the dynamic range required for the exposure can be reduced [Huynh et al. 1998, Bushberg et al. 2002].

## 1.4 How a Radiologist Interprets a Screen-film Mammogram

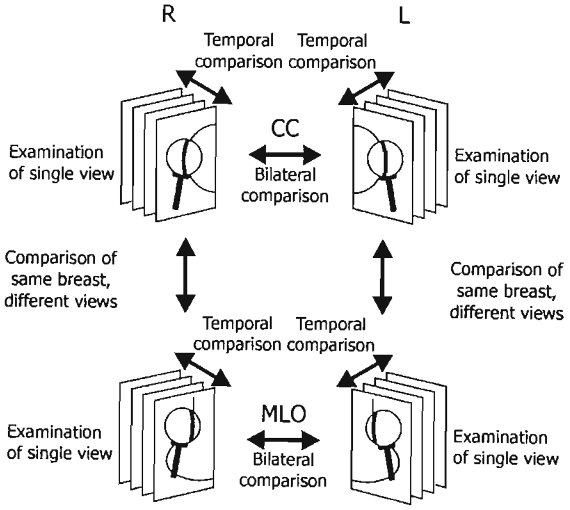
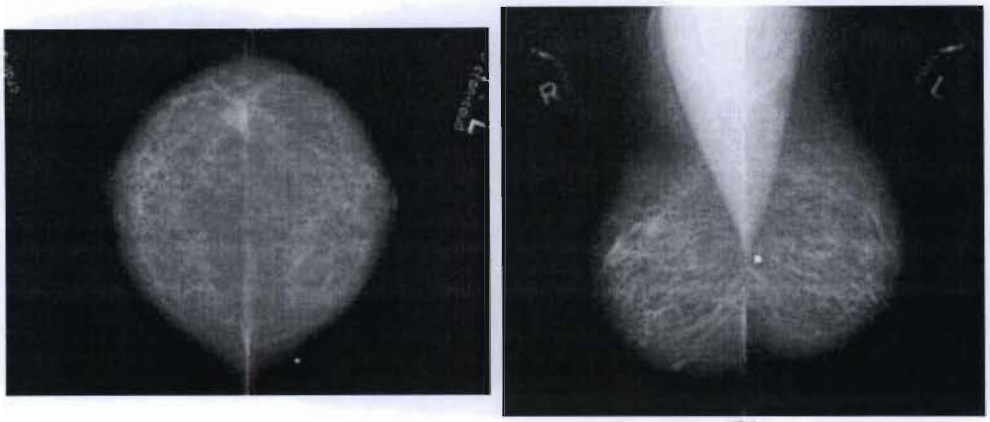


**Figure 1.3**: Summary of how a radiologist interprets a mammogram. Symmetry of the human breast forms the basis of mammogram interpretation. Comparison of corresponding left and right views highlights bilateral differences. Comparison of CC and MLO views of each breast eliminates artefacts such as overlapping dense tissue that may mimic a mass in one view or confirms the presence of a suspicious region if it is present in both views. Comparison between prior and current mammograms identifies temporal changes.

The discussion of how a radiologist interprets a mammogram is based on screen-film mammography since this is the most common mammography detector technology and most studies have focussed on it.



(a) (b)

**Figure 1.4**: A method of placing mammograms during analysis is shown, with bilateral mammograms placed as mirrored pairs, to facilitate easy comparison between breasts. The mammogram of the right breast is placed on the left and that of the left breast, on the right. **(a)** Placement of CC views. **(b)** Placement of MLO views. The pectoral muscle is visible in the upper left and right corners of the respective left and right MLO views. The bright white spot on the MLO view of the left breast is a radio-opaque marker indicating the position of a palpable mass.

Symmetry of the human breast forms the basis for the analysis of a mammogram (**Figure 1.3**). Bilateral CC views and bilateral MLO views are placed as mirrored pairs on the view-box (**Figure 1.4**). These views are analysed simultaneously with corresponding regions on each image compared to identify any asymmetries between left and right breasts. If a suspicious region is found in one view, the radiologist will attempt to find the same object in other available views of the same breast to identify the object as a true or false mass. Radiologists consider the distance from the nipple to the centroid of the suspicious regions in one view and then search an annular region in the second view at about the same radial distance from the nipple. If prior mammograms are available, these are compared with the current set to identify any temporal changes in the breast [Huynh et al. 1998, Paquerault et al. 2002, Majid et al. 2003].

To further aid interpretation, radiologists also use magnifying glasses and spot lights to highlight features. (Movson 2005, pers. comm.[2]). Radiologists look for the following signs of breast cancer [Martin et al. 1979, Goodsitt et al. 1998]:

1. Direct signs
    - ◇ masses (spiculated and irregular) and microcalcifications (**Figure 1.5**),
    - ◇ focus of asymmetric breast tissue density,
    - ◇ architectural distortion,
    - ◇ skin or nipple thickening and/or retraction,
    - ◇ spiculation, and
    - ◇ ulceration.
2. Indirect signs
    - ◇ a solitary dilated duct or an unusual complex of dilated ducts that extends 3 cm or more

---

[2] Dr. I. J. Movson, Department of Radiology, Addington Hospital, P. O. Box 977, Durban, 4000

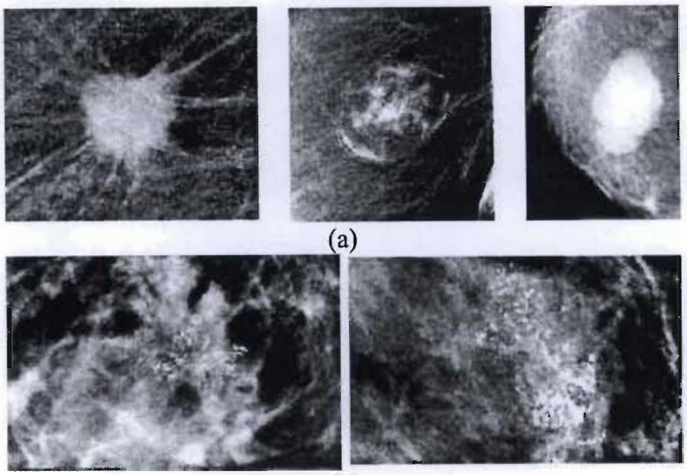


**Figure 1.5**: Some common signs of breast cancer are shown. **(a)** Examples of masses. Malignant masses (spiculated at top left and lobular at top right) are usually brighter than the surrounding tissue, have a uniform density, are approximately circular in shape and have fuzzy edges. Benign masses (top middle) can have a structure that is composed of different amounts of normal compared to that of the surrounding tissue. **(b)** Examples of microcalcification clusters. Most microcalcification clusters are benign, but some clusters can indicate malignancy.

within the breast,

- intraductal and intralobular calcifications,
- a progressive density in a specific area,
- bilateral asymmetry, and
- a benign appearing mass in a peri- or post-menopausal woman.

Mammograms are classified as normal if there are no signs of breast cancer or abnormal if there are suspicious features. If the mammogram is classified as abnormal then the features can either be *malignant*, which is cancerous or *benign*, which is not cancerous. These signs of breast cancer manifest as the following radiographic and mammographic features [O'Doherty 1999, te Brake et al. 2000, Thurfjell et al. 2002]:

1. brightness and contrast - the brightness and contrast of a mass is usually higher than that of surrounding tissue,
2. isodensity - a mass is usually isodense and opaque,
3. location of the mass,
4. texture - lines radiating from a central area is suspicious,
5. deformation of the skin-air interface or of the glandular tissue,
6. appearance in both MLO and CC views,
7. bilateral asymmetry, and
8. temporal changes.

## 1.5 Problems with Visual Analysis of Mammograms

The analysis of mammograms is a difficult task that involves identifying small features of low contrast superimposed on a non-uniform background. While there are many standard techniques that radiologists use to interpret a mammogram, part of the diagnosis is very subjective, often relying on the experience of the radiologist. This sometimes means that a mammogram that has been diagnosed as normal is not necessarily free of breast cancer. The causes of an incorrect diagnosis may be grouped into two categories: technical and interpretive [Yankaskas et al. 2001].

Problems are discussed with respect to screen-film mammography since most studies have focussed on problems associated with this mammography detector technology.

### 1.5.1 Technical Problems

There are two main groups of technical problems. The first group can be described as *quality of the mammogram* and includes problems inherent to screen-film mammography (discussed in §1.3.5.1 on page 4), poor mammographic technique on the part of the radiographer, problems with the mammography machine and problems with processing of the film. The second group can be described by *viewing conditions* and includes problems with viewing equipment. This author is not aware of any studies that examine the effect of technical problems on diagnostic accuracy although Huynh et al. [1998] lists some examples of how poor mammographic technique resulted in missed cancers.

The quality of the mammogram critically depends upon proper positioning of the breast. Ideally, the breast should be positioned in a manner that allows the maximum amount of breast tissue to be imaged. The CC and MLO views should also contain as much complementary information as possible. The breast must be adequately compressed to reduce scatter, radiation exposure and blurring due to motion. Adequate compression is particularly important in women with dense breasts. Exposure of the film also affects the quality of the mammogram. An under-exposed mammogram prevents differentiation within areas of dense tissue while an over-exposed image lacks information on the subcutaneous or fatty tissues. Lack of sharpness of the mammogram can also result from poor screen-film contact. Poor positioning of the breast during mammography or poor exposure and processing of the radiographic film can mean that the cancer is not clearly visible on the mammogram [Huynh et al. 1998].

The quality of viewing conditions is crucial to the analysis of a mammogram. A view-box with a high luminance is necessary. The radiologist must be able to adequately mask unnecessary view-box light around the mammograms as well as be able to mask bright regions on the edges of mammograms, and the ambient lighting of the analysis environment must be low [Huynh et al. 1998].

### 1.5.2 Interpretive Problems

Interpretive problems account for a significant amount of the research conducted into problems with mammography. The first set of interpretive problems occurs because subtle, unusual or small lesions can be easily missed. The second set of interpretive problems occurs when an abnormality is observed, but is misclassified. This can be as a result of radiologist inexperience, fatigue, inattention, not using all available views, not using prior images or judging an abnormality by its benign features only. Another common cause of misinterpretation is overlooking a subtle lesion in the presence of an obvious lesion [Huynh et al. 1998, Majid et al. 2003].

The main reason cited for cancers being missed in a screening mammography programme is the low occurrence rate of breast cancer in such an environment. Beam et al. [2002] estimates about 2 to 6 cancers are seen per 1 000 cases in a screening programme, and Yankaskas et al. [2001] quotes similar figures of 18 to 24 cancers per 6 000 cases per year. This means that radiologists who are tired or distracted can easily miss the signs of breast cancer.

There is also significant variability in radiologists' performances when interpreting mammograms. Quantitative information on the variability of radiologists' performances is unfortunately just as varied because results from these studies depend on the type of cases selected as well as the number and experience of the radiologists involved in the studies. Wirth et al. [1999] reported that between 10% and 20% of cancers are missed in current interpretations, and Astley & Gilbert [2004] reported that a typical breast-screening radiologist is likely to miss between 4% and 38% of cancers. Berg et al. [2002] reported on two studies in 1995 and 1996 that highlighted the variability in radiologists' performance. The 1995 study had ten radiologists (interpreting 150 mammograms, 23 with cancer) recommending immediate clinical studies for 74% to 96% of women with cancer and 11% to 65% of women without. The 1996 study had 108 radiologists (interpreting 79 mammograms, 45 with cancer) recommending clinical studies for 47% to 100% of women with cancer and 1% to 64% for those without. Yankaskas et al. [2001] re-examined 339 mammograms with cancers originally missed from a screening mammography programme, and found that 71% of these were still missed. The main reason was the low occurrence rate of cancers in a screening mammography programme. Elmore et al. [2002] conducted a study on 24 radiologists. Masses were noted in 0% to 7.9% of the films examined and calcifications in 0% to 21.3%, while 55.1% to 83.6% of the films were classified as normal. This study also found that films interpreted by younger radiologists or those who had graduated within the past 15 years were more likely to incorrectly identify a feature as abnormal.

The contributing factors to misdiagnosis can be summarised by [Berg et al. 2002, Elmore et al. 2002, Astley & Gilbert 2004]:

1. search pattern might have missed part of the film,
2. detected abnormality, but misclassified as normal,
3. failed to detect because of:

- ◇ low conspicuity,
- ◇ poor image quality,
- ◇ eye fatigue,
- ◇ oversight,
- ◇ distraction,
- ◇ mental fatigue,
- ◇ boredom, and
- ◇ inexperience.

### 1.5.3 Overcoming the Problems

Problems with mammography occur because reading a mammogram is a highly demanding task involving a detailed visual search for signs that are often subtle or small and appear infrequently. Problems inherent to screen-film mammography may be overcome with CR mammography or digital mammography. Interpretive problems can be overcome with double reading of mammograms (i.e. having two radiologists independently interpret the mammogram or the same radiologist interpreting the mammogram at different times), especially with arbitration where a third radiologist makes the decision when there is a lack of agreement between the first two [Astley & Gilbert 2004]. Double reading has been shown to be more effective at detecting abnormalities than a single interpretation by a single radiologist and can reduce recall rates [Warren-Burhenne et al. 2000, Dinnes et al. 2001]. In a review article, Bassett [2000] stated that double reading had the potential to effectively increase the sensitivity of mammography by about 10%. However, double reading means an even greater workload for (most likely, already overworked) radiologists.

### 1.5.4 Consequences of a Misdiagnosis

There are two problems with a misdiagnosis: if a cancer is missed then early treatment may be unsuccessful while if a benign mass is incorrectly diagnosed as malignant, then the patient has to face the trauma of a biopsy to confirm the diagnosis. While the former is more costly, the latter has the consequence of women not returning for screening mammograms and this non-return could be detrimental over the long term.

## 1.6 Computer-aided diagnosis (CAD) in Mammography

### 1.6.1 Advantages of CAD-systems

Since the consequences of errors in analysing mammograms are costly, there has been considerable interest in developing computer-based methods to assist radiologists. Computers have the following advantages [Chan et al. 1999, Bassett 2000, Astley & Gilbert 2004]:

1. a well-defined objective,

2. no tiredness, boredom or distraction, and
3. can consistently process images over a long period of time.

Computer-aided diagnosis (CAD) was developed to consistently *prompt* or draw radiologists' attentions to suspicious regions in a mammogram that may be missed [Bassett 2000]. Commercial CAD-systems have been designed to be consulted *after* the radiologist has made an initial assessment of the mammogram. So, provided the radiologist maintains the same unprompted performance, the overall detection performance should not decrease with CAD [Astley & Gilbert 2004] and it has been shown that prompting by a CAD-system *improves* radiologists' detection performance [Kegelmeyer et al. 1994, Chan et al. 1999, Warren-Burhenne et al. 2000, Freer & Ulissey 2001, Taft & Taylor 2001, Zheng et al. 2001, Marx et al. 2004].

### 1.6.2 Evaluating Detection Performance

The most important aspect in the development of CAD-systems is the evaluation of the results. CAD algorithms are evaluated by comparing the outputs of the algorithms to corresponding histological evidence of exact locations of malignant or benign masses and microcalcification clusters on these mammograms. Keeping in mind that the aim of the CAD-system is to identify malignant regions, the first step of the evaluation is to label each mark made by the algorithm as belonging to one of the following four categories [Bushberg et al. 2002]:

1. true-positive: identified as malignant and is malignant
2. false-negative: identified as benign and is malignant
3. false-positive: identified as malignant and is benign
4. true-negative: identified as benign and is benign

The second step of the evaluation is to combine the number of marks in each category as follows:

1. True-Positive Fraction (*TPF*): ratio of number of true-positives to number actually malignant
2. False-Positive Fraction (*FPF*): ratio of number of false-positives to number actually benign

Detection performance can then be described in terms of:

1. *sensitivity* defined by the *TPF*
2. *specificity* defined by 1-*FPF* or
3. the area under the curve of *TPF vs. FPF*.

Sensitivity is the fraction of truly malignant regions in the mammogram that are diagnosed as malignant. Specificity is the fraction of truly benign regions on the mammogram that are diagnosed as benign. Both these quantities are usually quoted at a specific decision threshold [Bushberg et al. 2002].

A plot of *TPF vs. FPF* using varying thresholds is referred to as a *receiver operating characteristic* (ROC) curve. A ROC curve has the advantage of examining the evaluation results at all possible decision thresholds, compared to sensitivity and specificity, which only examine the results at a single decision threshold. The area under the ROC curve, $A_{ROC}$, is accepted as a complete description of detection performance. Since $0 \leq TPF \leq 1$ and $0 \leq FPF \leq 1$, this implies that $0 \leq A_{ROC} \leq 1$. Detection performance is directly proportional to $A_{ROC}$, with higher values indicating better performance and $A_{ROC}=1$ for the perfect detection algorithm [Bakic & Brzakovic 1997]. The use of ROC curves for evaluating algorithms is detailed in **Chapter 7**.

The aim is to develop a CAD-system with a high sensitivity and a high specificity or a high value of $A_{ROC}$.

### 1.6.3 Detection Performance of Commercial CAD-systems

The Food and Drug Administration (FDA) in the United States of America has approved three commercial CAD-systems (ImageChecker™, Second Look™, and MammoReader™) for use as second readers in that country. These systems are also being used in many other countries like the United Kingdom, Canada and Germany. Currently, to our knowledge, there are no mammographic CAD-systems in use in South Africa.

ImageChecker™ by R2 Technologies[3] was the first CAD-system to receive FDA approval in April 2000. Prior to this, ImageChecker™ had been used in Germany, Norway, Sweden, Finland and France. ImageChecker™ marks microcalcification clusters (with a sensitivity of 98%) and masses (with a sensitivity of 74.7%) [R2 Technology 1998]. Vyborny et al. [2000] reported that ImageChecker™had a sensitivity of 86% in identifying spiculated masses.

Second Look™ by CADx Medical Systems Inc. (now iCAD[4]) received FDA approval in January 2002. Prior to this, Second Look™ had been marketed in Europe, Asia, Australia and Canada. Second Look™ detects calcification clusters and masses, and marks potential malignancies based on all the clusters or masses detected in the mammogram [CADx Medical Systems 2002]. According to its FDA *Summary of Safety and Effectiveness* document, Second Look™ had an overall sensitivity of 85% and also marked 62.7% of previously missed cancers. Taft & Taylor [2001] conducted a study with Second Look™ and reported a sensitivity of 88% with 1.5 false-positive marks per image. Marx et al. [2004] compared diagnostic results with and without use of Second Look™ and reported a 2% improvement in sensitivity with CAD. Most importantly, there was a decrease in the number of unnecessarily recommended biopsies.

MammoReader™ by Intelligent Systems Software, Inc. (now iCAD) received FDA approval in January 2002, and highlights microcalcification clusters, masses, architectural distortions and asymmetric densities. According to its FDA *Summary of Safety and Effectiveness* document, MammoReader™ had a sensitivity of 91.0% ± 2.2% for calcification clusters and 87.4% ± 1.9% for malignant masses with 3.32 false-positive marks per normal case and 2.32 false-positive marks

---

[3] `http://www.r2tech.com/`

[4] `http://www.icadmed.com/`

per abnormal case [Intelligent Systems Software 2002].

Overall, these CAD-systems improve the detection performance of radiologists, and microcalcification clusters are detected with a greater accuracy than masses. However, there is a relatively high false-positive rate, which means that radiologists have to spend more time on these regions that have been incorrectly identified as warranting further attention.

The mass detection performances and false-positive rates of ImageChecker™ and Second Look™ is summarised in **Table 1.2** for a few studies and shows that there is scope for improvement. No studies could be found evaluating the detection performance of MammoReader™.

**Table 1.2**: Summary of mass detection performance of commercial CAD-systems

| Year | Reference | CAD-system | Mass Sensitivity | False-Positive Rate (average per image) |
|---|---|---|---|---|
| 2000 | Warren-Burhenne et al. [2000] | ImageChecker™ | 75% | 1.0 |
| 2000 | Vybomy et al. [2000] | ImageChecker™ | 86% | 0.24 |
| 2001 | Freer & Ulissey [2001] | ImageChecker™ | 67% | – |
| 2001 | Taft & Taylor [2001] | Second Look™ | 88% | 1.5 |
| 2004 | Marx et al. [2004] | Second Look™ | 88.9% | 1.04 |

### 1.6.4 Detection Performance of Non-Commercial CAD-systems

Studies in which the detection performances of non-commercial, research CAD-systems were evaluated also indicated an improvement in radiologist performance as well as that the mass detection algorithms were not as efficient as calcification detection algorithms.

Chan et al. [1999] reported on the evaluation of a non-commercial CAD-system by six radiologists analysing 253 mammograms from 103 patients. Detection performance was evaluated with ROC analysis. The CAD-system alone achieved $A_{ROC}$=0.92 from analysing a single view. Equivalently, the radiologists achieved $A_{ROC}$=0.87 without a CAD-system and $A_{ROC}$=0.91 with a CAD-system. Using two standard views, the radiologists achieved $A_{ROC}$=0.92 without and $A_{ROC}$=0.96 with a CAD-system. The CAD information from the two standard views was merely displayed together and it was up to the radiologist to best use the information.

Zheng et al. [2001] examined how the sensitivity of the CAD algorithm affects detection performance. The sensitivity and number of false-positives per image was varied on a non-commercial CAD-system (analysing a single view) and was tested on 209 mammograms from 120 patients. It was found that while a high-performance system can significantly improve overall detection performance, a low performance system could be highly detrimental to radiologist performance.

Lauria et al. [2003] described CALMA, a non-commercial CAD-system that detected microcalcifications and lesions. The mass detection algorithm was evaluated on 180 images of normal breasts and 145 images of abnormal breasts and yielded a sensitivity of 90% and a specificity of 85%. The microcalcification detection algorithm was evaluated on 500 images of normal breasts and 306 images containing microcalcifications. This test yielded a 92% sensitivity and 92% specificity. The detection performances of three radiologists were also tested, with and without the CAD-system. The radiologists' performances, summarised in **Table 1.3**, show that there is an improvement of sensitivity with the CAD-system, but that the specificity is better without the CAD-system. This most likely means that the radiologists are being misled into believing that the incorrectly marked malignant areas are truly malignant, which is a disadvantage of a CAD-system with a high false-positive rate.

**Table 1.3**: Detection performance of radiologists on a non-commercial CAD-system, CALMA

| Radiologist | Sensitivity | | Specificity | | Experience (years) |
|---|---|---|---|---|---|
| | with CAD | without CAD | with CAD | without CAD | |
| A | 94.3% | 82.8% | 87.5% | 87.5% | 5 |
| B | 90.0% | 80.0% | 88.4% | 91.7% | 3 |
| C | 87.1% | 71.5% | 70.9% | 74.2% | 2 |

Helvie et al. [2004] evaluated a non-commercial system (using single-view analysis) on mammograms from 2 389 patients, 11 of whom had been diagnosed with cancer. The CAD-system detected 10 out of 11 of the cancers, indicating a sensitivity of 91%. No details of false-positives or specificity were given.

### 1.6.5 Shortcomings of Current CAD-systems

Despite the interpretive problems of a radiologist visually analysing a mammogram, it is important to note that this combination of human vision and the ability to analyse the morphology and texture of the structures in the mammogram to render a diagnosis works very well under ideal conditions [Paquerault et al. 2002]. The secret to successfully implementing a completely automated diagnostic algorithm relies on an exact emulation of the radiologist's methodology and ability.

While current CAD methods can achieve sensitivities up to 100% in identifying microcalcification clusters, the variable appearance of masses and their similarity to normal tissue means that masses are detected with a lower sensitivity [Astley & Gilbert 2004]. The danger of CAD algorithms with low sensitivities is that radiologists could become complacent and could start using the lack of CAD marks as an assurance that the mammogram is normal [Alberdi et al. 2004].

On the other hand, CAD algorithms with high sensitivities have an associated increase in the number of false-positives per mammogram analysed (e.g. ranging from an 2.2 per image to an 5.3 per image [Yin et al. 1991, Petrick et al. 1998, Petrick et al. 1999, Liu et al. 2001]). If these CAD-systems were to be used in a screening environment with $\approx$6 cancers per 1 000 cases screened, then there would be an average ranging between 367 and 883 false marks for each true cancer and all these additional marks also have to be evaluated by a radiologist. Such systems would not be very practical.

Bassett [2000] states that experienced radiologists should be able to quickly differentiate between those marked areas that warrant further attention from those that do not, but the radiologist would still spend more time analysing each mammogram than if there were no false-positives.

## 1.7 Overview of Thesis

The aim of any mass detection CAD-system should be to detect all masses with no false-positives. Based on evaluation results of current CAD-systems, mass detection performances can be improved upon and the number of false-positives per image can be reduced. This study investigates a method that could be used to reduce the number of false-positive masses detected.

As discussed in §1.4 (page 5), radiologists use many methods to eliminate false-positives, including comparison between both standard mammographic views of the same breast. Most CAD algorithms analyse single mammographic views independently of others and multiple views are usually used for comparison between left and right breasts, and for comparison of the same breast over time. Information from multiple views of the same breast is usually combined at the end of the single-view analysis. While there have been a few studies indicating the usefulness of using two standard mammographic views for false-positive reduction [Chang et al. 1999, Paquerault et al. 2002, Sun et al. 2004], these algorithms have not been incorporated into the systems described in §1.6.3 (page 12) and §1.6.4 (page 13). Most importantly, these dual-view algorithms do not *simultaneously* use information from both standard mammographic views to perform the analysis.

The aim of this study is to apply standard image processing techniques to two standard mammographic views of the same breast for template matching: *matching* a suspicious region of interest (ROI), identified in one mammographic view, to the same region in another mammographic view of the same breast. Grey-level co-occurrence matrices, texture measures and grey-level histograms are used to quantify the textural information in the image, while various similarity metrics (Euclidean distance, Mahalanobis distance, mutual information) are used to quantify the similarity between textural regions in multiple mammographic views of the same breast. Texture measures and distance metrics have been applied to the texture analysis of mammograms on numerous occasions. Literature surveys have indicated that the full grey-level co-occurrence matrix has not been used with mutual information for purposes of texture matching in mammograms. Results are evaluated using the area under the ROC curve and a measure of contrast.

The algorithms that are developed will also be applied to matching regions between stereotactic biopsy mammograms to improve selection of points on each stereoscopic view to ultimately improve the localisation accuracy for the biopsy.

The algorithms developed in this project have the advantage of slotting into existing CAD-systems as a method of providing further information to reduce false-positives and to improve the localisation accuracy for stereotactic biopsies.

### 1.7.1 Summary of Thesis Structure

**Chapter 1** introduces mammography and its use in the diagnosis of breast cancer. CAD is introduced as a means to assist radiologists by consistently highlighting abnormal regions for the radiologist to further analyse. The shortcomings of CAD led to the motivation for this study.

**Chapter 2** serves as a background for the remainder of the study and gives the reader a short overview of image processing methods that have been used in mammographic CAD.

**Chapter 3** places the goals of this study within the context of mammographic CAD. The matching problem is introduced together with the approach (image processing methods and materials) used in this study.

**Chapter 4** describes the pre-processing methods used in this study to remove the background, pectoral muscle and unneeded breast tissue from mammograms.

**Chapter 5** details the texture analysis image processing methods used. The calculation of grey-level histograms and co-occurrence matrices are described, together with the various texture measures used.

**Chapter 6** describes the distance metrics and mutual information, which are used as similarity metrics.

**Chapter 7** describes the theory behind ROC analysis and defines the measure for contrast.

**Chapter 8** details the algorithms and parameters used.

**Chapters 9** and **10** contain the results of applying the matching algorithms to mosaic images and pairs of mammograms.

**Chapter 11** describes the application of the matching algorithms to matching similar regions in stereotactic biopsy mammograms.

**Chapter 12** summarises the research and provides directions for future research based on the outcomes of this study.

An overview of the structure of the thesis is shown in **Table 1.4**.

## 1.8 Summary

Breast cancer is the most common form of cancer among women, worldwide. Successful treatment relies on early detection. Currently, mammography is the most widely available method of detecting breast cancer, but suffers from the problem that radiologists, in their visual interpretation of the resulting mammograms, sometimes miss the subtle signs of breast cancer. Computer-aided diagnosis was introduced as a means to assist radiologists by consistently highlighting abnormal regions for the radiologist to further analyse. The shortcomings of CAD i.e. high false-positive rates, and the current lack of systems that fully utilise all available mammographic information, form the basis of the motivation for this research, which investigates various texture-based image processing methods of correlating information between two mammographic views of the same

**Table 1.4**: Overview of thesis structure

| Chapter | Content |
|---|---|
| 1 | Introduction to mammography and CAD. |
| 2 | Review of image processing methods in mammography. |
| 3 | Background to this project & introduction to the matching problem. |
| 4 | Mammogram pre-processing. Description of a novel method to detect breast edge. |
| 5 | Details of the texture quantification methods to be used. |
| 6 | Details of the various similarity metrics to be used. |
| 7 | Description of ROC analysis & evaluation of results. |
| 8 | Summary of the methods and details of the matching algorithms. |
| 9 | Results, discussion and comparison of matching methods for mosaic images. |
| 10 | Results, discussion and comparison of matching methods for mammograms. |
| 11 | Application of matching algorithms to stereotactic biopsy mammograms. |
| 12 | Summary, conclusions & directions for future research. |
| A | Detailed results from matching methods |
| B | Basic information about images |
| C | Information theory basics |
| D | Description of a selection of IDL's functions |

breast. The algorithms developed in this study can ultimately be used to reduce false-positive detections in any CAD-system.

# Chapter 2

# Image Processing Techniques in Mammographic CAD

Many image-processing techniques have been used in mammographic CAD. The overall methodology for the detection and classification of masses and some examples of the methods used to address this mammographic CAD problem are described in this chapter. While research directly relevant to this research is discussed in the chapters to follow, some important concepts are briefly introduced in this chapter, notably grey-level co-occurrence matrices, texture measures and mutual information. The terminology used to describe images is given in **Appendix B** (page 246).

The detection of calcifications is not specifically considered in this study. For information on the processing methods used for detection and classification of calcifications see, for example, Shen et al. [1993, 1994], Ema et al. [1995], Qian et al. [1995], Hume & Thanisch [1996], Gurcan et al. [2001], Jiang et al. [2001], Lado et al. [2001], Nunes et al. [2001], Veldkamp et al. [2001], Gurcan et al. [2002] and Bocchi et al. [2004].

## 2.1 Detection and Classification of Masses

One of the main aims of mammographic CAD is to consistently identify malignant masses. This requires that the mass is firstly detected in the breast tissue and is secondly classified as malignant. However, the detection accuracy for all masses is much lower than that for calcifications, and some masses do not get detected. This is most likely because [Wei et al. 1995, Qian et al. 2001]:

1. Masses are of variable size, shape and density.
2. Masses sometimes have poor image contrast.
3. Masses are highly connected to surrounding tissue, especially with spiculated lesions.
4. The non-uniform background tissue surrounding masses are often similar to the mass.

In surveying the literature, it was found that the image processing techniques used in the detection and classification of masses in a mammographic CAD-system could be summarised by the five categories shown in **Figure 2.1**. Features that are not essential to the detection of masses are removed and the contrast of the remaining features are enhanced during *pre-processing*. The suspected masses are extracted from the breast tissue region during *detection*. Characteristics of the suspicious regions (texture, shape, size, location) are determined during *quantification of image information* and are used to discriminate between masses and normal tissue during *false-positive reduction*. Characteristics from the remaining regions are then used to classify the mass as malignant or benign during *classification and diagnosis*.

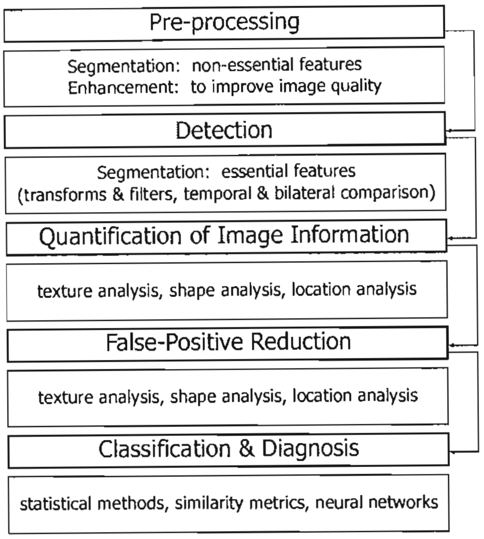


**Figure 2.1**: A summary of the image processing techniques used in mammographic CAD. Examples of some image processing methods, within each of the five categories are shown.

## 2.2 Pre-processing

The first step in mammographic CAD is the removal of non-essential features like the mammogram background and pectoral muscle, to reduce the effective region to be analysed. This is followed by enhancement of features in the remaining tissue.

### 2.2.1 Removal of Non-essential Features: Background

The many methods used to detect the breast border are reviewed in detail in §4.1.2 (page 44). These methods include:

1. thresholding [Tahoces et al. 1995, Mendez et al. 1996, Faizon & Sun 2000, Masek et al. 2000, Blot & Zwiggelaar 2001];
2. thresholding and iso-intensity contours [Mudigonda et al. 2001];
3. tracking or connectivity algorithms [Yin et al. 1991, Bick et al. 1995, Tahoces et al. 1995, Mendez et al. 1996, Faizon & Sun 2000, Ferrari et al. 2001];
4. artificial neural networks [Suckling et al. 1995];
5. modelling background as a surface [Chandrasekhar & Attikiouzel 1997];
6. model guided edge-tracking [Morton et al. 1996, Goodsitt et al. 1998];
7. active contours [Ojala et al. 2001, Wirth & Stapinski 2003];
8. $B$-splines [Ferrari et al. 2001, Ojala et al. 2001];
9. Fourier transforms [Ojala et al. 2001].

A novel method of detecting the breast edge using iso-intensity contours is presented in **Chapter 4**.

### 2.2.2 Removal of Non-essential Features: Pectoral Muscle

The pectoral muscle edge is essentially a straight line and straight-line detection methods, e.g. the Radon transform or the Hough transform are most often used to detect the edge of the muscle. The method of Karssemeijer [1998] for detecting the pectoral muscle using the Hough Transform is detailed in §4.2 (page 57).

### 2.2.3 Enhancement to Improve Image Quality

Image enhancement is necessary to improve the poor signal-to-noise ratio in most mammograms. Image quality can suffer as a result of the imaging procedure or breast motion during compression or positioning between the plates. Image details are also lost when the radiographic film is digitised because the spatial resolution of the scanning device seldom matches that of the film [Bovis & Singh 2002].

Image enhancement can be global (over the entire image) or local (over regions of the image). Global enhancement methods include those based on modifying the grey-level histogram [Morrow et al. 1992, Nunes et al. 2001, Bovis & Singh 2002] and for mammograms, image correction using the x-ray attenuation coefficients of the different type of breast tissue [Nunes et al. 2001].

Contrast enhancement expands the range of grey-levels over the full bit-depth of the image, and is a popular method of enhancing features in a mammogram. Adaptive contrast enhancement is preferred to global contrast enhancement and has been based on region-growing and local statistics [Morrow et al. 1992], the wavelet transform (using dyadic, $\phi$ and hexagonal functions as wavelets) [Laine et al. 1994], local entropy and the fractal dimension [Bovis & Singh 2002].

## 2.3 Detection

Once the non-essential features have been removed and the remaining features have been enhanced, the next step is to extract potential masses. This is done by *segmenting* or isolating the potential masses from the background breast tissue.

### 2.3.1 Segmentation of Essential Features: Transforms and Filters

Transforms (e.g. Fourier and wavelet) are types of filters that highlight features that are similar to the basis functions of the particular transform. Fourier functions are localised in frequency, but not in space and are appropriate for analysis of periodic structures, while wavelet functions are localised in frequency and space and can be used to analyse structures at different scales [Sonka et al. 1999]. An overview of the use of wavelets in temporal and spatial processing of biomedical images is given in Laine [2000]. In mammographic CAD, the wavelet transform has been used to:

1. detect spiculated masses [Liu et al. 2001];
2. extract a directional texture image with linear features to detect masses [Li et al. 2002].

3. characterise linear oriented components for use in the bilateral comparison of mammograms [Ferrari et al. 2001]
4. to detect and classify masses [Qian et al. 2001].

### 2.3.2 Segmentation of Essential Features: Bilateral & Temporal Comparison

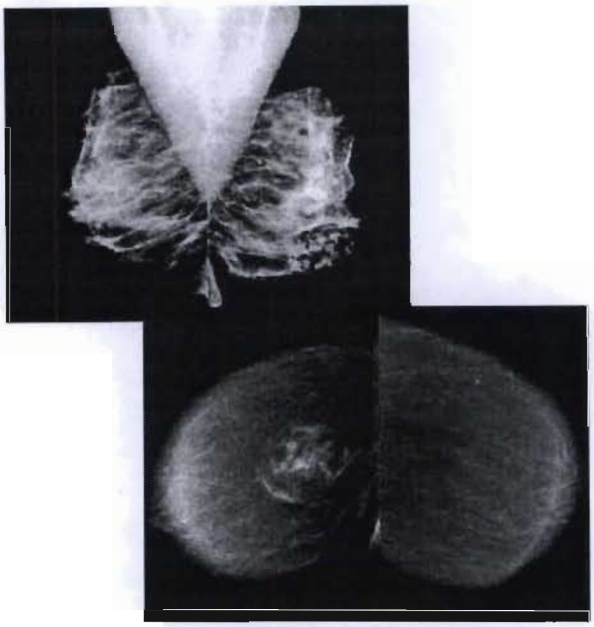


**Figure 2.2**: Bilateral mammographic views (top and bottom from different patients) showing asymmetry. Comparison of bilateral views highlights signs of abnormality and is used to extract suspicious regions from mammograms

Normal human breasts are highly symmetric and radiologists use bilateral asymmetry as a sign of abnormality (**Figure 2.2**). Mammograms can also be compared temporally, to identify changes in the breast over time. The general method of bilateral and temporal comparison is identical: accurately align or register the corresponding mammographic views and subtract one from the other to highlight differences. The most important and most difficult aspect of bilateral and temporal comparison is registration.

The difficulty with registering mammograms is mainly due to the differences in positioning and compression of the breast between mammograms. The non-rigid and inhomogeneous physical properties of breast tissue would have to be accurately modelled to take into account how the breast tissue changes with changes in patient positioning and with compression. This would enable more accurate registration [Wirth et al. 1999].

Methods used to register mammograms have included:

1. manual registration of the nipples and skin lines of bilateral mammograms [Yin et al. 1991];
2. establishing correspondence between temporal mammograms by using the intersection of elongated structures as control points for registration [Vujovic & Brzakovic 1997];
3. using thin-plate spline interpolation on boundary landmark points to register temporal mammograms [Marias et al. 1999];

4. using automatically identified landmarks internal to the breast to register temporal mammograms [Marias et al. 1999];
5. using a regional registration method to align temporal mammograms [Sanjay-Gopal et al. 1999];
6. using a multi-quadratic radial basis function in a nonrigid-body approach to automatically align bilateral and temporal mammograms [Wirth et al. 1999];
7. using thin plate splines and mutual information as a similarity metric to register bilateral and temporal mammograms [Wirth et al. 2002];
8. using Gabor filters to characterise linear oriented structures in the breast for use in bilateral registration [Ferrari et al. 2001];
9. using linear structures in the breast (extracted from local scale, orientation and position) to perform a point-by-point registration of bilateral and temporal mammograms [Marti et al. 2001*a*];
10. using matched control points and a temporal tracking algorithm to track structures over time [Marti et al. 2001*b*];
11. automatically registering bilateral mammograms by registering the nipple by only using translation and then registering the entire breast by only using rotation about the nipple [Bovis et al. 2000].

### 2.3.3 Other Segmentation Methods

#### 2.3.3.1 Active Contours and Region Growing

An active contour is an example of a deformable model that is fitted to the boundary of features in an image. External forces, based on the model, keep the contours moving outwards while internal forces, based on the data, keep the contours moving inwards. The solution is found by balancing the internal and external forces [McInerney & Terzopoulos 1996].

Region growing uses a pixel as a seed and expands the region according to specified criteria about neighbouring pixels.

Active contours and region growing have primarily been used to segment masses from mammograms [te Brake et al. 2000, Hadjiiski et al. 2001, Timp & Karssemeijer 2004].

#### 2.3.3.2 Contrast Enhancement

Contrast enhancement has also been used to segment features from a mammogram.

Anguh & Silva [1997] used thresholding based on grey-level histogram moments to automatically segment mammographic features. The enhanced features were then presented as a pseudo-colour breast map to the radiologist for further examination.

Petrick et al. [1998, 1999] used global density-weighted contrast enhancement to segment features of interest.

Mudigonda et al. [2001] segmented masses by sub-sampling and thresholding the image to identify isolated regions, in the form of closed iso-intensity contours. These contours were grouped or eliminated, yielding a segmented region corresponding to a mass.

#### 2.3.3.3 Removal of Normal Tissue

Abnormalities in a mammogram are enhanced when the surrounding normal breast tissue is removed.

Liu et al. [1998] identified and removed normal tissue by assuming that normal regions in a mammogram contain quasi-parallel linear markings, which were approximately linear over short segments, between 1 mm and 2 mm long and between 0.1 mm and 1.0 mm wide. The algorithm used a set of correlation filters in 16 radial orientations to detect lines.

Zwiggelaar & Rubin [1999] used the Fourier transform and fractal measures to remove normal mammographic texture from an image to enhance abnormalities.

#### 2.3.3.4 Miscellaneous Segmentation Methods

te Brake & Karssemeijer [1999] compared single- and multi-scale mass detection methods by using a difference of Gaussians with the ratio between the $\sigma$ of each successive Gaussian being $\approx 1.6$.

Zwiggelaar & Boggis [2001] used scale-orientation signatures for the detection of linear structures, to identify spiculated masses.

Tourassi et al. [2003] developed a knowledge-based CAD-system for detection of masses. The database contained 455 biopsy-proven malignant masses, 354 benign masses and 655 normal regions, stored as ROIs of $512 \times 512$ pixels. Mutual information based on grey-level histograms was used as a similarity metric. The mutual information between two quantities is a measure of the amount of information one quantity contains about the other. Mutual information is a maximum when the quantities are identical and zero when the quantities have nothing in common.

## 2.4 Quantification of Image Information

Once the regions of suspicion have been extracted, the information contained therein must be quantified. Characteristics such as texture, morphology and location are used to quantify image information.

### 2.4.1 Texture Analysis

Texture, as an image characteristic (§5.1 on page 63), forms the basis of most image processing applications. The aim of using texture for image processing is either to segment areas of similar texture (texture segmentation) or to identify similar textures (texture recognition) [Sonka et al. 1999].

The texture segmentation problem consists of identifying the boundaries between similar textures in a multi-textured image, using only the information contained in the image. In general, the aim of image segmentation, using texture, is to identify regions in an image that are uniform and homogeneous with respect to texture. The interiors of the identified regions should be simple with only a few small holes, while adjacent regions should contain considerably different textures. Boundaries of each segmented region should be simple, smooth and spatially accurate [Haralick & Shapiro 1985]. A simple example of a texture recognition problem is shown in **Figure 2.3(a)**.

The texture recognition problem can be summarised as finding the best match to an unknown single texture from a fixed set of known single textures. For example, a CAD-system that classifies an unknown sample of breast tissue by comparison to a set of previously classified breast tissues would use a texture recognition algorithm. A simple example of a texture recognition problem is shown in **Figure 2.3(b)**.

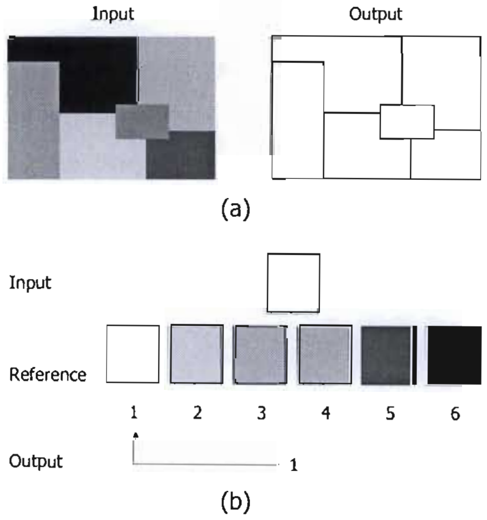


**Figure 2.3**: Examples demonstrating the main applications of texture in image processing. **(a)** Example of perfect texture segmentation. The borders between texturally homogeneous regions in the image have been identified. **(b)** Example of perfect texture recognition. The input texture has been matched with reference texture 1.

Both texture recognition and segmentation require that the textural information in the image be quantified in some manner. Once the information has been quantified, then a classifier is used to differentiate between textures. **Table 2.1** lists some methods of texture quantification [Haralick et al. 1973, Haralick 1979, Gonzalez & Wintz 1987, Russ 1995, Sonka et al. 1999].

Texture measures are the most common characteristics used in image processing and are commonly extracted from grey-level co-occurrence matrices, Laws' texture filters or run-length matrices. Texture measures summarise textural characteristics such as homogeneity, contrast, presence of spots, or information on linear structures as a function of grey-level.

#### 2.4.1.1 Grey-Level Co-occurrence Matrices

The most common texture measures are those of Haralick et al. [1973], which are usually calculated from grey-level co-occurrence matrices (GLCMs). GLCMs contain information on the

**Table 2.1**: Some methods of quantifying texture

| Texture Quantification Methods |
|---|
| statistical |
| grey-level histograms |
| structural |
| grey-level co-occurrence matrices |
| local covariance |
| run-length matrices |
| fractal dimension and lacunarity |
| Laws' texture energy |
| spectral |
| Fourier transforms |
| Wavelet transforms |

distributions of the grey-levels between two pixels in an image, positioned a distance *d* apart at a relative angle $\theta$. Haralick's texture measures (e.g. correlation, entropy, angular second moment, inertia, inverse difference moment, sum average, sum entropy and difference entropy) generally contain information about the homogeneity, contrast, complexity, presence of organised structures and grey-level transitions within the image. It should be noted that a specific texture measure cannot be uniquely related to a specific image characteristic [Chan et al. 1995]. The calculation of GLCMs and Haralick's texture measures are detailed in **Chapter 5**.

GLCM-based texture measures have been used:

1. to discriminate between masses and normal tissue [Chan et al. 1995, Bovis & Singh 2000, Bovis et al. 2000];
2. as inputs to artificial neural networks for segmentation of mammograms into background, pectoral muscle, fibro glandular tissue and adipose tissue [Suckling et al. 1995];
3. as inputs to a genetic algorithm to reduce false-positives [Sahiner et al. 1996, Wei et al. 1997];
4. to characterise masses as benign or malignant [Sahiner et al. (1998a, 1998b, 2001), Chan et al. 1999];
5. as inputs to artificial neural networks to classify ROIs as normal or abnormal [Christoyianni et al. 1999];
6. to search for corresponding regions in temporal mammograms [Marias et al. 1999];
7. to analyse the region around the margins of masses to discriminate between malignant and benign masses [Mudigonda et al. (2000, 2001), Sahiner et al. 2001];
8. for multi-scale analysis using wavelets [Wei et al. 1995, Petrick et al. 1998].

### 2.4.1.2 Laws' Texture Energy Measures

K.I. Laws [Laws 1980] derived Laws' texture energy measures to extract secondary features [average grey-level (L), edge (E), ripple (R), spot (S) and wave (W)] from natural microstructure characteristics of an image that could then be used for segmentation and classification. Laws derived five labelled vectors that could be combined to form matrices. The five vectors (L5, E5, R5, S5, W5) were derived from three vectors: $L3 = [1,2,1], E3 = [-1,0,1]$ and $S3 = [-1,2,-1]$. When these three vectors are convolved with themselves and each other, the following result is obtained: $L5 = [1,4,6,4,1]$, $E5 = [-1,-2,0,2,1]$, $R5 = [1,-4,6,-4,1]$, $S5 = [-1,0,2,0,-1]$ and $W5 = [-1,2,0,-2,1]$. Multiplying these five vectors with each other yields $5 \times 5$ matrices which when convolved with an image, yields texture energy measures [Gupta & Undrill 1995, Sonka et al. 1999].

Laws' filters have been used by Miller & Astley [1992] to classify glandular tissue, by Kegelmeyer et al. [1994] to detect spiculated masses and by Gupta & Undrill [1995] and Undrill et al. [1996] to segment masses from mammograms.

#### 2.4.1.3 Run-Length Statistics

Many neighbouring pixels of constant grey-level represent a coarse texture while a few neighbouring pixels of constant grey-level represent a fine texture. The length of these *runs* of constant grey-level pixels can then be used as a measure of texture. A *run-length* is defined as a set of pixels of constant grey-level, $g$, located in a straight line at an angle, $\theta$. If $P_\theta(g,r)$ represents the probability of finding features with length, $r$, of grey-level, $g$ and in the direction, $\theta$, then the following texture measures based on run-length statistics can be computed [Sonka et al. 1999]: long run emphasis, short run emphasis, run-length non-uniformity, run-length percentage and grey-level non-uniformity.

Run-length statistics have been used by Chan et al. [1999], Hadjiiski et al. [2001] and Sahiner et al. [2001] to discriminate between malignant and benign masses.

#### 2.4.1.4 Fractal Analysis

Fractal-based texture analysis relies on a correlation between the fractal dimension and texture coarseness [Sonka et al. 1999]. The fractal dimension is the rate at which the perimeter or surface area of an object increases as the measurement scale gets smaller [Russ 1995]. A small fractal dimension implies a fine texture while a large value implies a coarse texture. Unfortunately, the fractal dimension alone is insufficient to uniquely describe textures, and lacunarity measures have to be used. Lacunarity measures describe characteristics of textures that have the same fractal dimension, and are small for fine textures and large for coarse textures.

Marchette et al. [1997] used the fractal dimension to improve detection of masses by automatically determining the sizes of sampling windows so that the windows contained homogeneous textures, while Zwiggelaar & Rubin [1999] used information contained in the Fourier space of an image combined with a fractal measure to segment normal mammographic texture out of an image to enhance abnormalities.

### 2.4.2 Morphological Analysis

Shape and size information also help characterise masses. Morphological analysis of boundaries is used to characterise malignant and benign masses because malignant masses generally have complex boundaries while benign masses generally have smooth boundaries [Menut et al. 1997, Rangayyan et al. 1997*a*, Rangayyan et al. 2000]. Morphological information has been used:

1. to classify regions as masses or normal tissue [Petrick et al. (1996, 1999)];
2. as inputs to a genetic algorithm to discriminate between masses and normal breast tissue and between malignant and benign masses [Sahiner et al. (1996, 2001)];
3. to classify benign and malignant masses based on shape and acutance characteristics. Acutance is a measure of the gradient between light and dark regions in an image. [Rangayyan et al. (1997, 2000)];
4. as inputs to an artificial neural network to discriminate between malignant and benign masses [Huo et al. 1999].

### 2.4.3 Location Analysis

Chang et al. [1999], Paquerault et al. [2002] and Sun et al. [2004] used the correlated locations of potential masses in the MLO and CC views of the same breast to reduce false-positives.

### 2.4.4 Fourier Transforms

Transformed images have also been used to quantify information. The Fourier spectrum is ideal for describing the directionality of periodic or almost periodic two-dimensional patterns in an image, which are difficult to determine using spatial techniques as these techniques are usually of a local nature [Gonzalez & Wintz 1987].

#### 2.4.4.1 The Fourier Spectrum

The Fourier transform of a function $f(x,y)$ is defined by [Bracewell 1965]:

$$F(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)e^{-i2\pi(xu+yv)}\,dx\,dy \qquad (2\text{-}1)$$

$F(u,v)$ is referred to as the Fourier spectrum and is a useful tool in texture analysis. The following properties of the Fourier spectrum are useful for texture description [Gonzalez & Wintz 1987].

1. Prominent peaks in the spectrum give the principal direction of texture patterns
2. Location of peaks in the frequency plane gives the fundamental spatial period of the patterns
3. By filtering out periodic components, the remaining non-periodic image components can be analysed using statistical approaches

Josso et al. [2005] described a method of extracting the orientation of texture in images with a confidence indicator that provided some insight to the randomness of the arrangement of structures in the image, using the Fourier transform and principal component analysis.

Zwiggelaar & Rubin [1999] used the information contained in the Fourier space of an image combined with a fractal measure to describe texture in an image. This was then used to segment normal mammographic texture out of an image to enhance abnormalities.

## 2.5 False-Positive Reduction

The classification of a suspected mass as malignant or benign is made simpler if the total number of potential masses is as low as possible. The elimination of normal tissue regions from the set of potential masses is known as *false-positive reduction* and the following methods have been used to discriminate between masses and normal tissue:

1. GLCM texture measures [Chan et al. 1995, Bovis & Singh 2000, Bovis et al. 2000];
2. morphological filtering and eliminating features smaller than a specified size [Yin et al. 1991];
3. morphological features [Petrick et al. (1996, 1999)];
4. genetic algorithms & artificial neural networks with GLCM texture measures as inputs [Sahiner et al. 1996, Wei et al. 1997, Christoyianni et al. 1999];
5. the arc and Cartesian methods between the MLO and CC views of the same breast [Chang et al. 1999, Paquerault et al. 2002, Sun et al. 2004];
6. region specific features (size, average grey-level, contrast, compactness) as inputs to an artificial neural network [te Brake et al. 2000];
7. size, shape and difference in homogeneity and entropy in the regions identified by bilateral comparison [Bovis et al. 2000];
8. a knowledge base of previously classified suspicious regions [Chang et al. 2001].

Apart from the work of Chang et al. [1999], Paquerault et al. [2002] and Sun et al. [2004] (discussed in detail in §3.5 on page 36), which use two mammographic views, all the other studies use a single mammographic view for false-positive reduction.

## 2.6 Classification and Diagnosis

The textural, morphological and location features extracted from the regions remaining after false-positive reduction form a feature vector that summarises the various characteristics of each region. The feature vector is subject to feature selection where only those features that provide unique information are retained. The reduced feature vector is passed to a classifier to make the final distinction between the region being malignant or benign. Some examples of classification methods are shown in **Table 2.2**. Popular classification methods in mammographic CAD include linear discriminant analysis, binary classification trees and artificial neural networks.

**Table 2.2**: Some methods of classification

| Classification Methods |
|---|
| linear and quadratic classifiers |
| nearest neighbour classifiers |
| minimum distance |
| decision trees |
| artificial neural networks |
| mutual information |
| maximum likelihood |
| Bayes classifier |

### 2.6.1 Feature Selection

In general, the feature vectors are multi-dimensional with only a few features providing useful information for any given image. An important step before classification is *feature selection*, which is the selection of those few features from the feature vector that provide unique information about the region. The main idea behind this step is to reduce the dimension of the classification problem. The following methods have been used for feature selection:

1. step-wise feature selection with linear discriminant analysis [Petrick et al. 1998, Petrick et al. 1999, Wei et al. 1997, Sahiner et al. 1996, Sahiner et al. 1998*a*, Sahiner et al. 1998*b*, Sahiner et al. 2001, Chan et al. 1999, Bovis & Singh 2000, Mudigonda et al. 2000, Hadjiiski et al. 2001];
2. binary classification trees [Kegelmeyer et al. 1994, Liu et al. 2001];
3. genetic algorithms [Sahiner et al. 1996, Sahiner et al. 1998*b*];
4. mutual information [Tourassi et al. 2001];
5. principal component analysis [Bovis et al. 2000].

### 2.6.2 Linear Discriminant Analysis

Linear discriminant analysis is a statistical technique for classifying samples as one of a set of predefined classes, using a discriminant function that is a linear combination of the features of each sample. The classifier must first be trained on a set of samples with known classes to determine the coefficients or weights of the discriminant function.

The disadvantage of linear discriminant analysis is that *only* linear combinations are considered and the classifier requires training.

Linear discriminant analysis has been used as a classifier:

1. for mass segmentation [Petrick et al. 1998, Zwiggelaar et al. 1997, Petrick et al. 1999];
2. to discriminate between masses and normal tissue [Chan et al. 1995, Wei et al. 1995, Sahiner et al. 1996, Wei et al. 1997, Zwiggelaar et al. 1997, Sahiner et al. 1998*a*, Sahiner et al. 2001, Bovis et al. 2000, Helvie et al. 2004];
3. to classify masses as malignant or benign [Chan et al. 1999, Mudigonda et al. 2000, Hadjiiski et al. 2001, Mudigonda et al. 2001].

### 2.6.3 Binary Classification Trees

Binary classification or decision trees use the concept of branches in a tree to separate objects into classes. The classes are the tips of the branches and decisions are made at the branch points by examining various object properties [Hand 1981]. In a binary classification tree, only two branches are allowed at each node. At the first node, all input data is divided into two classes. At the next node, each of the first two classes is further divided into two classes. This continues until the maximum number of allowed classes is reached. The branching is defined according to a training set of data. A simple example of a binary classification tree is shown in **Figure 2.4**.

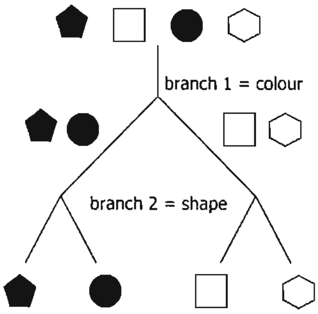


**Figure 2.4**: Example of a binary classification tree. The aim is to place the four objects into four classes. The decision at the first branch point is made according to colour, and at the second branch point, according to shape.

Binary classification trees have the following advantages: automatic feature selection, is robust with respect to outliers and misclassified points in the training set, the final classifier can be compactly stored, new data is efficiently classified and provides easily understood and interpreted information regarding the predictive structure of the data [Liu et al. 2001]. The disadvantage is that binary classification trees require training [Kegelmeyer et al. 1994].

Binary classification trees have been used:

1. to differentiate between normal and abnormal tissue [Kegelmeyer et al. 1994]; and
2. to identify spiculated lesions [Liu et al. 2001].

### 2.6.4 Artificial Neural Networks

#### 2.6.4.1 Background to Artificial Neural Networks

Artificial neural networks represent one of many approaches to create a machine capable of human-like thought processes. The most important components of an artificial neural network are neurons, which, like their biological counterparts, can accept inputs and generate an appropriate output based on the value of the input and the nature of the connections between the inputs.

An artificial neural network has to be first trained on a set of known data, which should span the range of data of interest. This process yields the values of the weights between the inputs and outputs obtained by minimising some error function. There are two types of training: supervised and unsupervised. In supervised training, the network is presented with inputs and the corresponding known outputs. The weights are determined based on this information. In unsupervised training the network is presented with inputs and the network is allowed to discover to which (of a number of classes) each input belongs [Masters 1993].

An important part of the training process is cross-validation. Cross-validation is necessary to prevent over-training (i.e. creating a model that over fits the data). The set of data used for cross validation must be separate to the training data and like the training data, must span the range of the data for it to be meaningful. Once the artificial neural network has been trained, it can be applied to the general population of data [Masters 1993].

### 2.6.4.2 Application to Mammographic CAD

A fundamental problem of using standard image processing algorithms in mammography is that a description of what is normal and abnormal is needed. However, because of the significant variability between both, general-purpose algorithms are not sufficiently adaptable to cope successfully. Since artificial neural networks are non-parametric pattern recognition systems that can generalise by learning from examples, they are useful in problems where decision rules are vague and there is no explicit knowledge about the probability density functions governing sample distributions. This makes them ideal for application to mammographic CAD[Bakic & Brzakovic 1997].

Artificial neural networks have been used:

1. for segmentation [Suckling et al. 1995]
2. to discriminate between normal and abnormal tissue [Christoyianni et al. 1999, Bovis & Singh 2000, Bovis et al. 2000]
3. to assign measures of malignancy based on characteristic features extracted from the mammogram [Huo et al. 1999, te Brake et al. 2000]
4. to detect and classify masses [Qian et al. 2001, Lauria et al. 2003]
5. for false-positive reduction [Sun et al. 2004]

## 2.7 Summary

The detection accuracy of CAD-systems for microcalcification clusters is higher than that for masses. The appearance of masses in mammograms is very varied and many image-processing approaches have been used to detect and classify masses as malignant or benign. Methods of segmenting suspicious regions from the surrounding breast tissue include removal of normal tissue structures, contrast enhancement, the Wavelet transforms, active contours, symmetry between left and right breasts and changes in the breast over time. Once these suspicious regions have been identified, morphological, geometrical and textural information are extracted and used to reduce false-positives, i.e. separate any normal tissue regions from truly abnormal regions. This reduced set of regions is then passed to a classifier like linear discriminant analysis, a binary classification tree or an artificial neural network that performs the final discrimination between whether the region is benign or malignant.

# Chapter 3

# Multiple Mammographic-View Analysis

One of the concerns regarding current CAD algorithms is the high false-positive rate, which arises as a consequence of the requirement that the algorithms have a high sensitivity [Astley & Gilbert 2004]. When a radiologist interprets a mammogram, all available views of the patient are examined in conjunction with each other. The radiologist uses the inherent symmetry between left and right breasts to identify abnormalities and then uses the MLO and CC views of the same breast to confirm presence of the abnormality or to eliminate a false-positive detection.

This chapter details the motivation behind this study and outlines the image processing methods, based on texture analysis, that are applied to the problem of matching a suspicious ROI in one mammographic view to the same region in the other mammographic view of the same breast. The output of the matching algorithm can then either be used to confirm the presence of an abnormality or to indicate a false-positive detection.

## 3.1 How CAD-systems Treat Multiple Views of the Same Patient

The various CAD algorithms that have been listed in **Chapters 1** and **2** handle different views of the same patient in one of the following methods:

1. Each view of the same patient is analysed independently as if there is no correlation between them. The CAD outputs, for each view, are then independently presented to the radiologist. If the output of the classifier is a number, which is related to the likelihood of malignancy, then taking the average, minimum or maximum of the classifier output yields a combined output for each patient [Liu et al. 2004, Sun et al. 2004].
2. For bilateral comparison algorithms, CC views of left and right breasts and MLO views of left and right breasts are analysed together.
3. For temporal comparison algorithms, current CC views of a single breast and previous CC views of the same breast are analysed together. Similarly for the MLO views.

These methods show that apart from bilateral and temporal comparison algorithms, multiple views of the same patient are not analysed simultaneously.

## 3.2 Mass Detection Performance of CAD-Algorithms

**Table 3.1** summarises the sensitivity and/or specificity or the area under the ROC curve, $A_{ROC}$, for a few mass detection studies. The false-positive rates, where available, have also been listed. Two points are clear from this data: the mass detection accuracy is not very high and where it is very high, the false positive rate is also very high. It is therefore logical for mass detection algorithms to have as high sensitivities as possible, and to then reduce the false-positive detections. This would at least ensure that all masses *are* detected.

**Table 3.1**: Summary of sensitivity and/or specificity or $A_{ROC}$ and false positive rates (if reported) for a few mass segmentation studies in mammographic CAD.

| Reference | Sensitivity | Specificity | $A_{ROC}$ | False-positives/image |
|---|---|---|---|---|
| Yin et al. [1991] | 100% | - | - | 5.3 |
| Chan et al. [1995] | - | - | 0.82 | - |
| Petrick et al. [1998] | 90% | - | - | 4.4 |
| Petrick et al. [1998] | 80% | - | - | 2.3 |
| Petrick et al. [1999] | 90% | - | - | 4.2 |
| Bovis & Singh [2000] | - | - | 0.74 | - |
| Warren-Burhenne et al. [2000] | 75% | - | - | 1 |
| te Brake et al. [2000] | 70% | - | - | 0.1 |
| Liu et al. [2001][a] | 84.2% | - | - | $< 1$ |
| Liu et al. [2001][a] | 100% | - | - | 2.2 |
| Qian et al. [2001] | - | - | 0.93 | - |

[a] Liu et al. [2001] quote two separate sensitivities for two different false positives per image.

## 3.3 Multiple Views and False-Positive Reduction

In contrast to current CAD algorithms, radiologists often use images of the same patient from different orientations or different modalities to provide complementary information. The use of all available information to perform a computer-aided diagnosis has been shown to improve sensitivity and reduce false-positives [Wen et al. 2004] and spot compression magnification views have been used to improve classification between malignant and benign over using MLO and CC views alone [Huo et al. 2001].

If all methods described in **Chapter 2** were to be incorporated into a single CAD-system, this system would still not completely emulate the actions of a radiologist, especially regarding the use of multiple views and false-positive reduction.

Most of the methods of false-positive reduction described in §2.5 (page 28) compare abnormal and normal regions from the same view. Very few false-positive reduction methods use both standard views of the same breast. Similarly, very few methods use both standard views of the same breast to confirm the presence of an abnormality. The lack of research into the use of multiple views of the same breast is due entirely to the difficulty with finding correspondence between multiple views of the same breast.

## 3.4 Correspondence between Multiple Views of the Same Breast

### 3.4.1 Problems

An x-ray intensity image can be described as a transmission image while a light intensity image can be described as a reflective image. The main difference between these two types of images relates to objects that are placed in a line between the light/x-ray source and the detector. In a reflective image, objects (or parts of objects) obscured by the object closest to the detector, are not seen in the image, while all objects (depending on x-ray energies or x-ray attenuation factors) can be seen in an x-ray image. For all images, photographs of an identical scene from slightly different angles, allows for three-dimensional information to be extracted. This is known as stereovision.

The main complication with finding correspondence between different mammographic views of the same breast is that the mammogram is a two-dimensional projection of a compressed, variably elastic three-dimensional structure, with the exact geometry of each view being determined by how the radiographer has positioned the breast. Conventional stereovision theory cannot be applied to find correspondence between objects in standard mammographic views, because the compression of the breast changes for each view.

In summary, problems associated with determining correspondence between multiple views of the same breast are [Vujovic & Brzakovic 1997, Highnam et al. 1998*b*, Wirth et al. 1999]:

1. breast tissue has a complex, inhomogeneous, anisotropic nature,
2. compression may obscure abnormalities,
3. compression may cause dense tissue to overlap, creating apparent abnormalities,
4. compression may distort the variably elastic breast tissue,
5. there may be differences in positioning and compression between views,
6. there is a lack of clearly defined landmarks, and
7. there is a changing geometry between views.

### 3.4.2 Current Approaches

There have been two approaches to determine the correspondence between two mammographic views of the same breast: breast compression models and geometric models.

#### 3.4.2.1 Breast Compression Models

Kita et al. [1998, 2001, 2002] described a model-based method to find the distorted epi-polar line in one view corresponding to a point in the other view as a precursor to using information available in two views of the same breast to improve detection and diagnosis of breast cancer. The model used information on compressed and uncompressed breast shapes to predict the movement of points on the surface of the breast. These movements were then interpolated to within the breast. This model relies on patient specific information for its implementation.

Other breast compression models are based on finite element analysis, which divides the breast into finite elements. Physical properties of each element are defined according to the type of tissue the element represents and the elements are connected together in a manner representative of how the tissues are connected together. Kinetic theory defines the motion of the elements. Finite element analysis has been used:

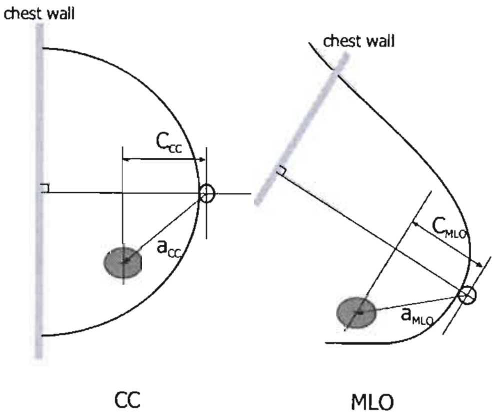
1. to model deformation of the breast during mammography [Samani et al. 1999];

2. in a virtual reality system for guiding breast biopsy with MRI [Azar et al. 2001];
3. to study deformation of the breast for registration of mammograms and MRI images [Ruiter et al. 2002];
4. to model compression to fuse information from different modalities and to assist with biopsies [Pathmanathan et al. 2004].

The breast compression models of Kita et al and the finite element models have the disadvantage of being patient-specific. Finite element models have the further disadvantages of depending greatly on the physical characteristics of the various tissue types present in each breast and being computationally intensive. These models are therefore not practical for use in a CAD-system.

#### 3.4.2.2 Geometric Models

There are two geometric models to determine correspondence between mammographic views of the same breast: the arc method and the Cartesian straight-line method. For an ROI, either automatically identified by a CAD-system or manually marked by a radiologist, the arc method is based on the distances between the nipple and the centroid of the ROI ($a_{CC}$ and $a_{MLO}$ in **Figure 3.1**). Similarly, the Cartesian straight-line method is based on the distances between the nipple and the centroid of the ROI along a line perpendicular to the chest wall ($c_{CC}$ and $c_{MLO}$ in **Figure 3.1**). Paquerault et al. [2002] found that there was a linear correlation between $a_{CC}$ and $a_{MLO}$.



CC MLO

**Figure 3.1**: Geometry of the arc method (a) and the Cartesian straight-line method (c). [$a_{CC}$,$a_{MLO}$] represents the location, in the CC and MLO-views respectively, of the ROI as determined by the arc method and [$c_{CC}$,$c_{MLO}$] represents the location, in the CC and MLO-views respectively, of the ROI as determined by the Cartesian straight-line method.

The arc and Cartesian straight-line method have been used to determine the arc and Cartesian distances in the MLO and CC views. These distances are then used together with other morphological and textural features as inputs to e.g. an artificial neural network for false-positive reduction [Chang et al. 1999, Sun et al. 2004]. Paquerault et al. [2002] used the arc method to define a smaller search region to reduce false-positives, while radiologists apply the arc method during analysis of mammograms by considering the distance from the nipple to the centroid of the suspicious region in one view and then searching an annular region in the second view at about the same radial distance from the nipple for a similar suspicious feature.

While the geometric models of finding correspondence are more general than the breast compression models, the best that can be done is to define a smaller region of the breast within which the ROI most likely occurs. The geometric models cannot be directly used for determining correspondence between multiple views of the same breast.

## 3.5 Review of Multiple View Analysis to Reduce False-Positives

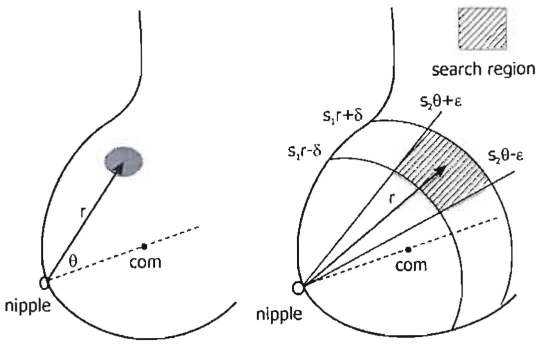
The three studies investigating the use of multiple mammographic views for false-positive reduction are described.

### 3.5.1 Chang et al. [1999]

Chang et al. [1999] applied the arc method and the Cartesian straight-line method to two standard mammographic views to reduce false-positives. Each view was independently analysed to search for potential suspicious regions and then values of $[a_{CC}, a_{MLO}]$ and $[c_{CC}, c_{MLO}]$, as detailed in **Figure 3.1** were extracted. The method was tested on 571 pairs of CC and MLO mammograms, containing 290 masses on both views. A mass detection algorithm identified all masses and 3 992 false-positive objects. Identified objects from both views were paired with each other. ROC analysis was used to evaluate performance levels for each method in determining, based solely on location, whether a pair of suspicious regions represented a true mass or a false-positive combination. $A_{ROC}$=0.79 was obtained for the arc method and $A_{ROC}$=0.78 was obtained for the Cartesian straight-line method. At 90% sensitivity, the arc method eliminated ≈48% of false-positives while the Cartesian method eliminated ≈47%. Results were comparable, but the arc method was preferred because it only needed nipple position and was easier to implement.

### 3.5.2 Paquerault et al. [2002]

Based on an object's location in one mammographic view, Paquerault et al. [2002] used the arc method to define an annular search region in the other view, in an extension of the work of Sanjay-Gopal et al. [1999], to reduce false positives. Sanjay-Gopal et al. used a regional technique to register temporal mammograms. The registration method was based on registering a small region containing a suspected mass with a previous mammogram. The border and nipple position were used to globally register the current and previous mammograms, via a series of translations and rotations of the previous mammogram with respect to the current mammogram. Once the mammograms were globally aligned, the nipple and centroid of the breast region were used to set up a frame of reference (**Figure 3.2**). The positions of suspicious features identified in the current mammogram (either automatically or manually) were then used to zoom in on the corresponding region in the previous mammogram. The template search, performed on the zoomed-in region to identify the same feature in the previous mammogram, identified 85% of corresponding regions in prior mammograms.



current previous

**Figure 3.2**: Frame of reference used for regional comparison of temporal mammograms. The position of the mass on the current view is used to define a reduced search region in the previous view. The shaded region on the previous mammogram indicates the reduced search region. [Sanjay-Gopal et al. 1999]

Paquerault et al. compared one and two view algorithms for false-positive reduction. The algorithms were tested on 169 pairs of mammograms with masses on CC and MLO views from 117 patients. The one-view algorithm used a density weighted contrast enhancement filter [Petrick et al. 1998] to enhance mammographic structures. False-positives were reduced by use of morphological features, overlap of detected regions and texture features in a three-step process, which also restricted the number of objects per image to three. These methods had the unfortunate side effect of reducing the sensitivity of the detection scheme.

The two-view algorithm used the results of the one-view algorithm as the input image. A geometric model, based on the arc method (§4.3.1 on page 59), was used to define an annular search region in one mammographic view, based on a reference object selected in the other mammographic view. The reference object was then paired with all objects that were located within the annular region, because an object located in one view cannot be uniquely paired with a single object in the other view. Morphological and textural features were determined for each object and similarity measures (absolute difference and mean) were determined for each pair.

Available cases were divided into training and testing sets in a 3:1 ratio. Two separate linear discriminant analysis classifiers with step-wise feature selection were trained to differentiate between the true and false pairs using the morphological and textural similarity measures, respectively. The scores from each classifier were then averaged to get a single "correspondence" score for each pairing. This yielded a case-based sensitivity of 75% for all masses and the number of false-positives per image was reduced from 1.5 for one-view analysis to 1.13 for two-view analysis. When the subset of malignant mass mammograms was examined, false-positives were reduced from 1.5 to 0.5 at a case-based sensitivity of 85%.

### 3.5.3 Sun et al. [2004]

Sun et al. [2004] used the same basic idea as Chang et al. [1999]. Potential masses were automatically segmented from each view by fuzzy C-means clustering. Various characteristic features describing texture, shape and location, including the arc and Cartesian straight-line distances, were extracted from each segmented area. Concurrent features were extracted by examining characteristics of ROIs segmented from one view paired with all ROIs segmented from the other view. A genetic algorithm was used for feature selection and the selected features were used as inputs to an artificial neural network, trained on 60 images and tested on 40. It was not clear as to whether the images were paired or not for input to the artificial neural network. Free-response ROC analysis was used to test performance of the CAD-system. A clear increase in specificity was reported.

Since the characteristics of the detected ROIs in one view were paired with the characteristics of every detected ROI in the other view, the analysis was not truly concurrent. However, the analysis of the paired regions did provide information on similarity.

### 3.5.4 Discussion

These studies show that there is potential in using multiple mammographic views to correlate the complementary information that is present in these views. Since CAD algorithms often analyse multiple views independently of each other, the complementary information is not considered during the processing, but is usually examined at the end [Liu et al. 2004]. For these studies, each view is still independently analysed and then objects detected in both views are paired with each other in a multiple-view analysis. Classification also relies on a trained system to discriminate between the true mass pairs and the false pairs for false-positive reduction.

The pitfalls of using a trained classification system are numerous. The first significant problem arises from the cases used to train the classification system. If the training set does not cover the full range of possible input data, then the classifier will be inadequately trained. Such a classifier will not be able to handle those cases that fall outside the scope of the training. This is most important for mammography where there is significant variation between mammograms. The second significant problem arises from the requirement of a large set of input data in order to get a single result, since most of the cases are used for training the classifier.

This study was inspired by these shortcomings of current multiple-view analysis methods.

## 3.6 Description of This Project

### 3.6.1 Basis

The minimum requirements of a CAD-system should be to completely emulate the actions of a radiologist. As shown in **Figure 3.3**, most methods used by radiologists in interpreting mammograms have been implemented in many CAD algorithms, mostly for the detection of abnormalities. Very few studies have compared the similarity of features between standard mammographic views of the same breast, as used by radiologists to confirm presence of abnormalities and to eliminate false-positives.

This project is concerned with the application of texture-based image processing methods

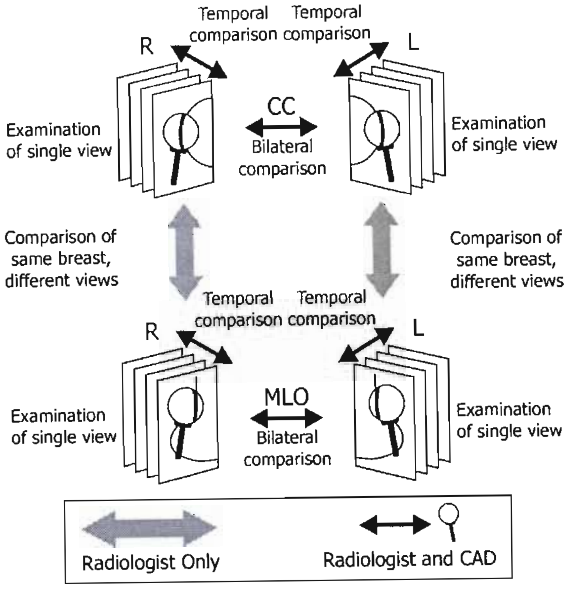


**Figure 3.3**: Schematic of the methods used by a radiologist during the interpretation of a mammogram. The methods indicated by black arrows have been implemented in CAD-systems. Very little research has been done into implementing the methods indicated by the grey arrows. The implementation of the latter methods will complete the basic emulation of how a radiologist examines a mammogram, in software.

to match a suspicious feature found in one mammographic view to the same feature in other mammographic views of the same breast, to further emulate the actions of a radiologist in software.

### 3.6.2 Hypothesis

Texture analysis methods used with suitable similarity metrics will allow a suspicious feature from one mammographic view to be matched with the same suspicious feature in other mammographic views of the same breast.

### 3.6.3 Assumptions

For this project, the following are assumed to hold true [Paquerault et al. 2002]:

1. At least two mammographic views of the breast are available.
2. A mass is visible in at least two mammographic views.
3. A mass has similar image characteristics in all mammographic views.

Wellman et al. [1999] measured mechanical properties of normal and abnormal breast tissue, under compression immediately after excision and investigated the stiffness behaviour of tissues at various strain rates. It was found that the stiffness behaviour of cancerous tissue varies non-linearly with strain and that cancerous tissue was much stiffer than fat or normal glandular tissue.

This means that a mass is not very compressible, as is widely known by radiologists reading mammograms, and the image characteristics of a mass should be similar in each mammographic view as required by Assumption 3. Any change in orientation that may exist from changing the direction of compression is addressed in the textural analysis.

### 3.6.4 Description of the Matching Method

The template matching method described here can be summarised as a hybrid of texture classification and texture segmentation (§2.4.1 on page 23). Regions of interest (ROIs) are manually extracted from a *reference* image (as for texture classification) and are matched to a portion of the breast in a *test* image (as for texture segmentation).

Radiologists consider the distance from the nipple to the centroid of the suspicious region in one view and then search an annular region in the second view at about the same radial distance from the nipple. This basic idea is implemented by using the location of the reference ROI and the arc method to identify an annular region of the breast in the test image that is searched for a match. This reduces the number of potential matches by reducing the search area. Textural characteristics of the ROIs are extracted using grey-level histograms, GLCMs and GLCM-based texture measures and are compared using distance metrics and mutual information as measures of similarity. This is done using two methods.

The first method uses GLCM-based texture measures to quantify textural information and distance metrics as measures of similarity, and is referred to as *texture measure matching*. The second method uses probability density functions (grey-level histograms and GLCMs) to quantify textural information and mutual information as a measure of similarity, and is referred to as *mutual information matching*.

### 3.6.5 Output of the Matching Algorithm: The Matching Map

The results of using the similarity metrics are arrays of distance and mutual information values, and are referred to as distance and mutual information maps, respectively. Distance is inversely proportional to similarity since similarity increases as distance decreases, while mutual information is proportional to similarity. Therefore the distance maps indicate an optimal match at a minimum of distance while the mutual information maps indicate an optimal match at a maximum of mutual information. To avoid confusion, only *matching maps* are presented as results of the matching algorithm, in this thesis. A matching map is defined as a map, which has an optimal match at a maximum intensity. Distance and mutual information maps are appropriately converted to matching maps.

**Figure 3.4** illustrates the differences between a distance map, a mutual information map and their corresponding matching maps for an 8-bit checkerboard image. The reference ROI and test image are also shown. The maximum image/map intensity is indicated by white and the minimum image/map intensity by black. It can be seen that the distance map (**Figure 3.4(c)**) has minimum intensities (i.e. the best match) for the same positions that the corresponding matching map (**Figure 3.4(e)**) has maximum intensities. It can also be seen that the mutual information map (**Figure 3.4(d)**) has maximum intensities for the best match and this map is identical to the corresponding matching map (**Figure 3.4(f)**).

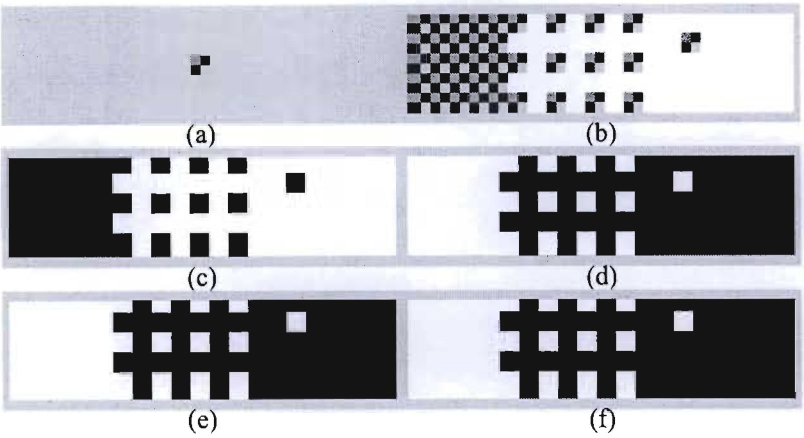


**Figure 3.4**: Example of the output of the matching algorithm for an 8-bit checkerboard test image created in Microsoft Paint. For all images and maps shown, the maximum intensity is indicated by white and the minimum intensity is indicated by black. **(a)** 2 pixels × 2 pixels reference ROI extracted from bottom left of test image. **(b)** Checkerboard test image. **(c)** Distance maps showing minima where the reference ROI matches exactly with the test image. **(d)** Mutual information map showing maxima where the reference ROI matches exactly with the test image. **(e)** Matching map from distance map showing maxima where the reference ROI matches exactly with the test image. **(f)** Matching map from mutual information map showing maxima where the reference ROI matches exactly with the test image.

### 3.6.6 Evaluation of Matching Results

The accuracy of the matching algorithms is evaluated by comparing the matching maps to ground-truth maps. In this study, ground-truth maps are generated from regions manually marked by a radiologist. Matching accuracy ($\kappa$) is defined as a combination of two quantities: the area under the ROC curve ($A_{ROC}$) and contrast ($C_{fb}$). For ROC analysis, the matching and ground-truth maps are compared at various decision thresholds and values for the *TPF* and the *FPF* are computed at each threshold. The sets of *TPF* and *FPF* values are used to generate the ROC curve and the area under the curve is used as an indication of what proportion of the matched region was actually matched. Contrast gives an indication of how well the matched area stands out from the background of the matching map. Ideally, the matched regions should be the brightest objects in the matching map.

### 3.6.7 Where this Project Fits into Current CAD Algorithms

The algorithms developed in this study can be easily incorporated into existing CAD-systems. Current CAD-systems analyse both standard mammographic views, independently, and then combine the outputs. The matching algorithms developed in this study can slot in between the independent outputs and the combination of the results, as a method of providing more information for false-positive reduction.

## 3.7 Summary of Methods to be Used

A schematic of the template matching algorithm is shown in **Figure 3.5**. A reference ROI (or template) is identified in the reference image. The template is compared to sub-images extracted

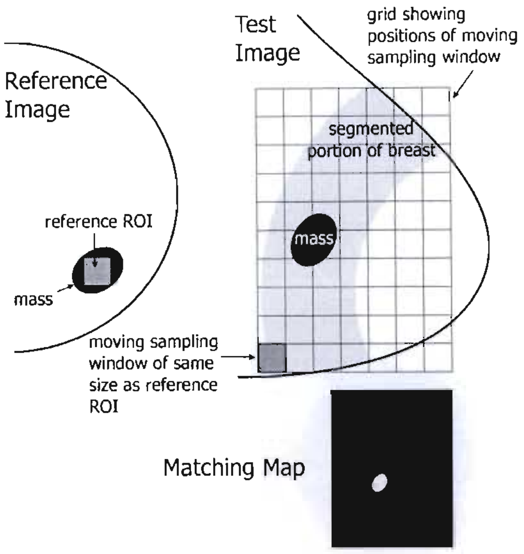


**Figure 3.5**: Schematic of the matching algorithm. The location of the reference ROI is used to define an annular search region in the test image. Textural characteristics of the reference ROI are compared to textural characteristics of equivalently sized sub-images in the reduced search region in the test image. The result of this comparison process is a similarity map referred to as a *matching map*. The brighter the regions on the matching map, the greater the similarity.

by sampling windows that move over a reduced search region in the test image and computing a similarity metric at each position. The position with the highest similarity corresponds to the region in the test image that best corresponds to the reference template. The steps and methods used in the matching algorithm are summarised, following the methodology used in **Chapter 2**.

1. Pre-processing of test image
    (a) Removal of mammogram background using iso-intensity contours.
    (b) Removal of pectoral muscle using the Hough transform.
    (c) Definition of annular segment of breast tissue, containing ROI, using the arc method.
2. Detection
    (a) ROIs identified by a radiologist are used to define the reference region.
3. Quantification of image information using:
    (a) grey-level histograms,
    (b) GLCMs, and
    (c) GLCM-based texture measures.
4. ROIs between views are compared using:
    (a) distance metrics, and
    (b) mutual information

## 3.8 Computational Environment

A PC with an AMD Athlon XP 2.4GHz processor, with 500 Mb of RAM, running Microsoft Windows 2000 was used for the software development. All algorithms were developed in IDL 6.1[1], an interactive programming environment providing mathematical functionality and a graphical interface. Graphs were created in IDL and figures were created in CorelDraw10[2]. The thesis document was typeset using MikTex[3], a TEX implementation for Microsoft Windows.

## 3.9 Images and Ground Truth Data

The matching algorithms were applied to three sets of images. The first set of images was constructed from a set of single texture images. These mosaic images were used to evaluate the matching algorithms under matching conditions where the exact position and transformation of the reference image in the test image was known. The second set of images consisted of pairs of mammograms (CC- and MLO-views) and the third set of images consisted of stereotactic biopsy mammograms. The image sets, ground truth data and selection of reference sub-images are described in detail in the respective results chapters.

## 3.10 Summary

One of the concerns regarding current mass detection CAD algorithms is the high false-positive rate, which arises as a consequence of the requirement that the algorithms have a high sensitivity. Radiologists use all available mammographic views of a single breast for diagnosis, but very little research has been done into the use of multiple single-breast mammograms for confirmation of abnormalities and false-positive reduction in CAD-systems. This study is concerned with the development of an analysis technique that uses both standard mammographic views of the same breast with the aim of confirming the presence of abnormalities and ultimately to eliminate false-positives. The algorithms developed employ standard image processing methods based on texture analysis, including grey-level co-occurrence matrices, texture measures and the use of an information theory measure, mutual information. All algorithms developed can be easily incorporated into existing CAD-systems.

---

[1] `http://www.rsinc.com/`

[2] `http://www.corel.com/`

[3] `http://www.miktex.org/`

# Chapter 4

# Mammogram Pre-processing

One of the first steps in CAD is the segmentation of the mammogram into background, breast tissue and pectoral muscle. This has the advantage of simplifying further processing of the image, by reducing the area of the image to be processed, and also provides a reference for the alignment of views when two views are being compared.

Additionally, this study uses the arc method to define an annular region in one view of the breast based on a reference region selected in the other view of the same breast, to further reduce the area of the breast tissue to be processed.

## 4.1 Detection of the Breast Edge

### 4.1.1 Introduction

Although many methods have been developed to detect the breast edge in mammograms, very few researchers use borders drawn by radiologists to evaluate the automated fits and even fewer quantitatively evaluate the results. A method using areas enclosed by iso-intensity contours is presented as an improvement to the basic grey-level thresholding algorithm. The effect of various pre-processing methods on the accuracy of automated borders is investigated. The algorithms developed are tested on 25 mammograms, for which automated borders are quantitatively compared to manual borders drawn by three radiologists.

### 4.1.2 Overview of Breast Border Detection Methods

Many methods have been used to detect the breast border in mammograms, including thresholding, tracking, artificial neural networks and modelling (of background and breast border). These methods are briefly described.

#### 4.1.2.1 Thresholding

For mammograms, thresholding usually involves selecting a single grey-level from an analysis of the grey-level histogram, to segment the mammogram into background and breast tissue. All pixels with grey-levels less than the threshold are marked as background and the rest as breast. Thresholding uses only the grey-level histogram and no spatial information is considered. Therefore, the major shortcoming of thresholding is that there is often an overlap between the grey-levels of objects in the breast and the background.

Blot & Zwiggelaar [2001] described a global thresholding algorithm where a peak detection method was used to automatically determine the threshold from the grey-level histogram. An evaluation of the results was not described.

Masek et al. [2000] described how to combine the outputs of two breast border detection algorithms to improve the results. The method, based on a minimum cross entropy thresholding algorithm, was tested on 161 pairs of images from the Mammographic Image Analysis Society[1] (MIAS) database. Although results using the multi-algorithm approach were described as being better than using any single algorithm, no quantitative evaluation of results was presented.

#### 4.1.2.2 Tracking

Tracking the breast border involves implementing a tracking algorithm that marks a pixel as a border pixel if it satisfies certain criteria.

Yin et al. [1991] used a 4-connectivity tracking algorithm to identify the border. The results were not evaluated.

Bick et al. [1995] identified unexposed and direct-exposure regions in the mammogram and generated a border surrounding the valid breast border by combining grey-level histogram analysis with morphologic filtering. A closed, 8-connected border was generated from this processed image. The algorithm was tested on 740 digitised mammograms. The two radiologists and two medical physicists, who visually evaluated the results, rated 97% of the results as acceptable for use in a CAD scheme. Problems with the remaining mammograms were attributed to digitisation artefacts or poor mammographic technique. The merit of this research was that the algorithm was tested on a large set of mammograms, but since the evaluation results were not quantified, this algorithm cannot be compared with any other breast border detection algorithm.

Mendez et al. [1996] and Tahoces et al. [1995] described a semi-automatic method of detecting the breast border, utilising the gradient of grey-levels in 3 user-selected regions. The algorithm used thresholding and pair-wise pixel differences in specific directions to detect the breast border. Two radiologists and one physicist evaluated results and were required to categorise the fits as follows:

1. Automated border agreed exactly or almost exactly with a radiologist's estimated border
2. Automated border did not agree exactly and small deviations may be observed
3. Automated border clearly disagreed with a radiologist's border although it may be acceptable for future purposes

Of the 156 mammograms tested, 89% were classified as either category 1 or 2.

Faizon & Sun [2000] described a method using thresholding and tracking to identify the breast border, but no discussion of the accuracy of the results was presented.

Ferrari et al. [2001] used the chain-code method and cubic $B_3$ splines to detect an approximate border. The true border was found by further examining a region around the points making up the approximate border. A total of 66 images from the miniMIAS database were analysed and

---

[1] http://www.wiau.man.ac.uk/services/MIAS/MIAScom.html

borders were accurately detected in 61 images. The results were evaluated according to a protocol established by an expert radiologist, but no details of the protocol or quantitative results were given.

Mudigonda et al. [2001] used a fixed-width Gaussian kernel to smooth the mammogram and generated iso-intensity contours by thresholding this image at a grey-level close to zero. The contour enclosing the largest area was selected as the breast border. The method was tested on 56 images and all borders were described as being "successfully detected". However, neither details of threshold selection nor evaluation of results was given.

#### 4.1.2.3 Artificial Neural Networks

Suckling et al. [1995] used multiple, linked self-organising neural networks to segment the breast into four components: background, pectoral muscle, fibro glandular tissue and adipose tissue. This method had the advantage of simultaneously identifying the background and pectoral muscle, but no evaluation of the background segmentation results were given.

#### 4.1.2.4 Models of the Background

Chandrasekhar & Attikiouzel [1997] used the Weierstrass approximation theorem as a basis for fitting a surface to the background [Lancaster & Šalkauskas 1986]. The method was tested on 58 images and results were evaluated by visual comparison with the original images using pseudo-colour. The algorithm [Chandrasekhar & Attikiouzel 2000] was further tested on 28 images from the MIAS database, where all images gave clear breast-background segmentation. Results were again not evaluated by a radiologist. The fully automated method [Chandrasekhar & Attikiouzel 2001] was tested on the full MIAS database. The results, evaluated by non-radiologists, indicated acceptable segmentation in 95% of the MIAS images, but results were not quantitatively evaluated.

#### 4.1.2.5 Models of the Breast Border

Morton et al. [1996] and Goodsitt et al. [1998] reported on a breast border detection algorithm using a two-pass, model-guided edge-tracking algorithm which, when compared to manually traced out borders, yielded an average root-mean-square difference of 1.4 mm. The algorithm was tested on more than 1 000 mammograms and the border was accurately found in about 95% of the images. It was not stated whether radiologists traced out the manual borders.

Ojala et al. [2001] used grey-level histogram thresholding and morphological filtering to obtain an initial estimate of the breast border. A smooth border was obtained by three methods: active contours, Fourier transforms and B-splines. The models were tested on two sets of 10 images each and the results were compared to manually drawn borders. The error, *ME*, between the manual and automated borders was as:

$$ME = \frac{1}{n}\sum_{i=1}^{n}\sqrt{(x_{m_i}-x_{a_i})^2+(y_{m_i}-y_{a_i})^2} \tag{4-1}$$

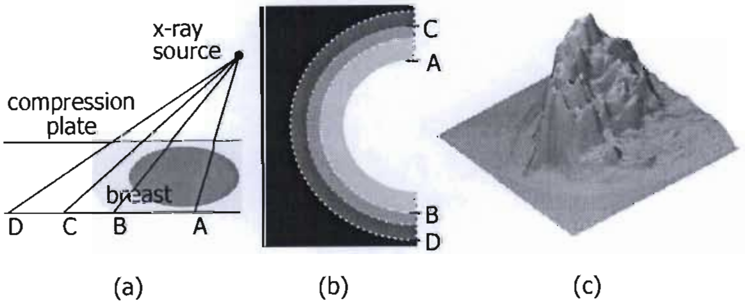
where $[x_m, y_m]$ and $[x_a, y_a]$ were points on the manual and automated breast borders, respectively. $n = 10$ was the number of points used for the error calculation. The best overall results were obtained for the active-contour modelling method, which had an error of 1.8 mm±2.0 mm and 2.6 mm±1.4 mm for the first and second sets of images, respectively. It was not stated whether radiologists drew the manual borders.

Wirth & Stapinski [2003] used active contours to identify the breast border, but no details of accuracy were given.

### 4.1.3 Iso-intensity Breast Edge Detection

Very few researchers have used borders drawn by radiologists to quantitatively evaluate the results of the automated borders. Of the two research articles [Goodsitt et al. 1998, Ojala et al. 2001] encountered where quantitative evaluation results were quoted, it was not specified whether or not radiologists drew the reference borders. This is important if the breast edge in a mammogram is not clear, as is the case of most of the images used in this study.

#### 4.1.3.1 Basis of the Method



(a) (b) (c)

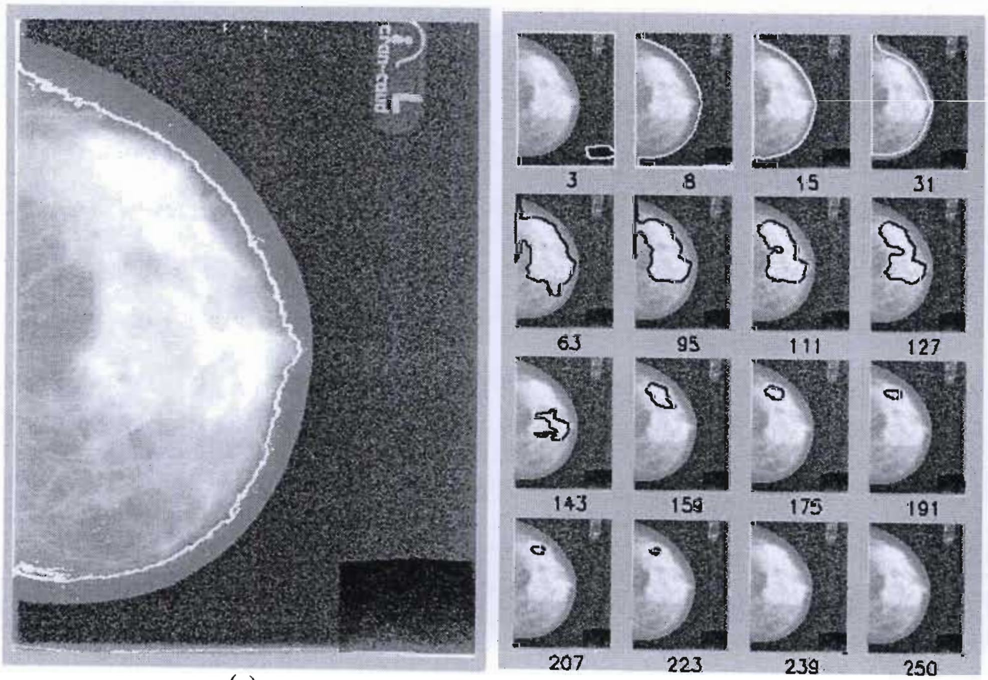
**Figure 4.1**: Schematic of CC-view mammography showing the breast edge. **(a)** The breast during mammography with dark shaded area representing the dense interior. The light shaded area represents the generally fatty, outer edges of the breast. X-rays along line D image the breast edge. **(b)** The mammogram shows that the grey-level in the image is proportional to x-ray attenuation in the breast. There is very little attenuation of the x-rays along arc D. **(c)** A surface plot of a mammogram with grey-levels represented as height. The breast edge can be clearly seen as the transition between the background and the breast. [Adapted from [Highnam et al. 1998*a*]]

The breast and background form the two largest contiguous regions on a mammogram, with the background dominating at the low grey-levels. Highnam et al. [1998*a*] (**Figure 4.1**) provides a good description of the *breast edge* based on how the breast is compressed during mammography. The portion of the breast between the compression plates is of equal thickness, but some of the breast bulges out towards the edges. This bulge is mostly composed of fat, except near the nipple, and does not form a straight vertical edge. **Figure 4.1** shows a schematic of this for a CC mammogram. An x-ray along line D (**Figure 4.1(a)**) experiences very little attenuation, but the attenuation should be uniform and arc D (**Figure 4.1(b)**) should be a smooth iso-intensity curve, representing the breast edge with respect to the x-ray source [Highnam et al. 1998*a*].

As the x-ray attenuation increases, the grey-levels in the mammogram increase. Therefore the areas of regions enclosed by iso-intensity contours would decrease as grey-level increases. This is the main idea behind the breast border detection method presented here. The optimal grey-level threshold is selected by analysing the area enclosed by iso-intensity contours at various grey-levels. While this method may be categorised as thresholding, spatial information about structures in the mammogram is taken into account, thereby overcoming some of the disadvantages of thresholding discussed in §4.1.2.1.

#### 4.1.3.2 Iso-Intensity Contouring

This novel method of selecting the optimal grey-level threshold for the breast border is based on the fact that there is often a sharp transition in grey-levels between the background and the breast in a mammogram. If this transition grey-level, $g_o$, can be identified, then a contour at $g_o$ will yield the breast border. $g_o$ is determined by analysing the areas enclosed by iso-intensity contours at various grey-levels (**Figure 4.2**). The areas enclosed by the contours should decrease sharply at the breast edge.



(a) (b)

**Figure 4.2**: **(a)** Examples of iso-intensity contours at a single grey-level indicated by the white solid lines, including many small contours in the background. **(b)** Contours drawn at various grey-levels (g=5, ..., 250) have different enclosed areas. For each image, the contour with the largest area at the indicated grey-level is shown on contrast-enhanced images.

#### 4.1.3.3 Analysis of Area *vs.* Grey-Level

A set of closed iso-intensity contours (**Figure 4.2(a)**) is generated at a single grey-level. There is a *set* of contours because there are many non-homogeneous regions in the background and in the interior of the breast that correspond to any particular grey-level. The aim of this algorithm is

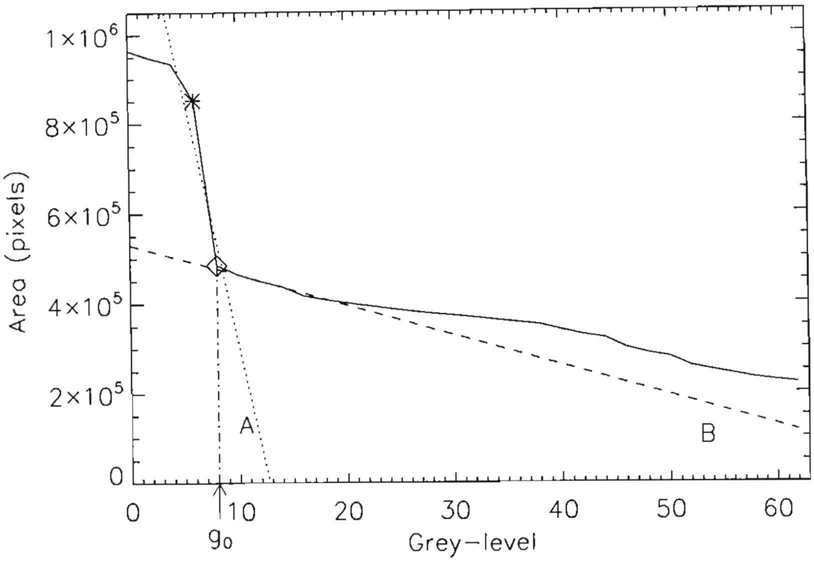


**Figure 4.3**: The plot shows the largest area enclosed by a contour as a function of that contour's grey-level. The contours were generated at grey-level increments of 2. Two straight lines (dotted, A, and dashed, B) are fitted to specific straight-line regions of the curve. The intersection of A and B (indicated by a ◇) determines $g_o$, the grey-level threshold corresponding to the breast edge. The grey-level values were incremented in steps of 2.

to identify the breast edge by examining the change in area enclosed by a contour as grey-level changes, only the contours with the largest areas are selected from the set of contours. A typical graph of the areas enclosed by iso-intensity contours, plotted as a function of the contour grey-level, is shown in **Figure 4.3**. The transition grey-level, $g_o$, is chosen as the point at which the area stops its sharp decrease and is determined by calculating the intersection (◇ in **Figure 4.3**) of two straight lines (A,B) that are automatically fitted to portions of area curve, as follows:

1. Determine the derivative of the area with respect to grey level value. Identify the grey-level that corresponds to the derivative of the area with the steepest slope ($*$ in **Figure 4.3**).
2. A straight line (dotted line A in **Figure 4.3**) is fitted to a region on either side of $*$. The method used to select the optimal number of data points used for this fit is described in §4.1.3.4. A minimum of three data points is used.
3. A second straight line (dashed line B in **Figure 4.3**) is fitted to the approximately horizontal region immediately following the region used to fit line A. The number of data points used for this fit is described in §4.1.3.4.

The derivative was determined by using IDL's `DERIV` function, which implements a 3-point Langrangian interpolation to determine the derivative numerically.

The number of points used for the fitting of the two straight lines was automatically determined for each image, because the characteristics of the step in the plot of area *vs.* grey-level varied from image to image.

#### 4.1.3.4 Optimising the Number of Data Points

Often data follows some general behaviour (e.g. linear) for a certain amount of data points and then deviates (e.g. becomes non-linear). It is therefore sometimes desirable to fit the model to that small subset of the data which best describes the model. A novel method is presented that selects a subset of data points to be used for the fit. The method assumes that the data points are ordered in a manner such that the subset is located at the beginning or end of the data. **Figure 4.4** illustrates an example of fitting a linear model to an arbitrary set of data. Start with a minimum number of data points (4 in the example), perform a least squares fit which yields an $\varepsilon^2$ goodness-of-fit statistic. Increase the number of data points and determine a value of $\varepsilon^2$ for each fit. Normalise the $\varepsilon^2$-values by dividing by the number of data points used for the fit. The minimum of the normalised $\varepsilon^2$, as a function of the number of data points, yields the optimal number of data points (5 in the example) that should be used to perform the fit. The method can be applied to any model or data.

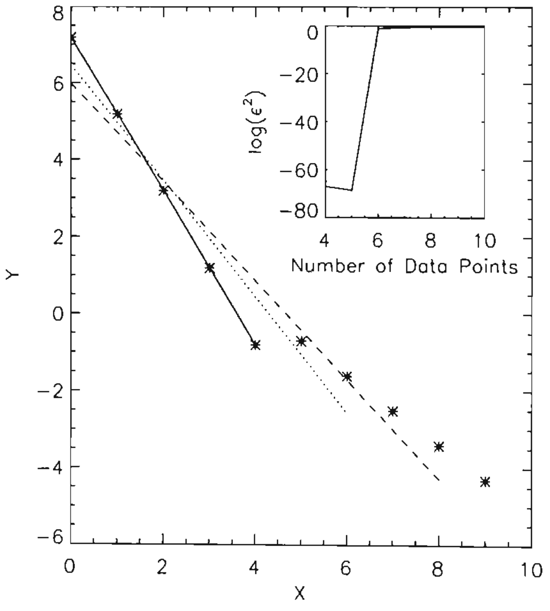


**Figure 4.4**: Example of fitting a linear model (using least-squares minimisation) to a set of arbitrary data points to illustrate method used to optimise the number of data points used for a fit. The solid line represents the best fit with 5 points, the dotted line represents the best fit with 7 points and the dashed line represents the best fit with 9 points.

#### 4.1.3.5 Pre-processing Methods

Pre-processing is necessary to remove the effects of noise and artefacts that might adversely affect the automatic border detection algorithm. Popular pre-processing methods include convolution or smoothing, median filtering and morphological operators. However, the choice of pre-processing method can also affect the accuracy of the automated breast borders. For the set of mammograms used in this study, preliminary investigations showed that convolution (or smoothing) with a top hat or Lorentzian kernel improved the performance of the automatic border detection algorithm, but results were dependant on the smoothing widths used. Mammograms with clear breast edges required less smoothing than those with grainy breast edges. In the latter case, the smoothing served to merge the grey-levels of the breast edge pixels into a contiguous region as required for the iso-intensity contour analysis.

A more detailed investigation was, therefore, carried out into the effects of varying the widths of the smoothing kernels, according to the clarity of the breast edge, on accuracy of the automated iso-intensity contour borders.

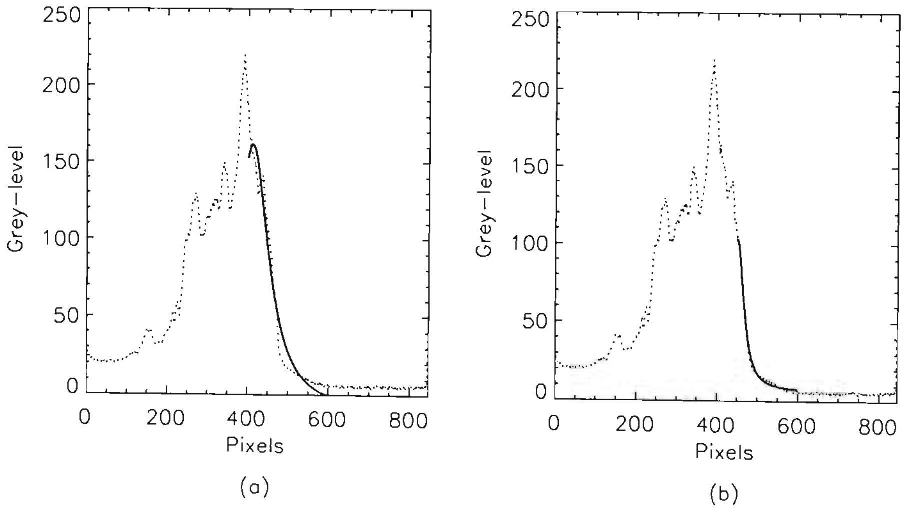
### 4.1.3.6 Determination of Widths of Smoothing Kernels

The defining characteristic of a smoothing kernel is its width. Since the breast edges of the mammograms used in this study varied from very clear to very grainy, the concept of *clarity of the breast edge* was used to define the width of the smoothing kernel. This was based on the assumption that a profile of the breast edge is adequately modelled by the tail of a Lorentzian, $L_{profile}$ as defined in **Equation 4-2**, with $a_2$, the width of the Lorentzian, quantifying the clarity of the breast edge. $a_2$ should be small for a clear breast edge and large for a grainy breast edge.

$$L_{profile} = \frac{a_0}{1+\left(\frac{x-a_1}{a_2}\right)^2} + a_3 \tag{4-2}$$

The parameters for $L_{profile}$ were determined by fitting $L_{profile}$ to a profile of the breast edge, over an optimised number of pixels as described in §4.1.3.4. The first point for the fit is determined manually. An example of a breast edge profile with $L_{profile}$ overlaid is shown in **Figure 4.5**. The Lorentzian has a better fit for the optimised number of pixels (**Figure 4.5(b)**) than the fixed number of pixels (**Figure 4.5(a)**).

The profile of the breast edge is selected as the average of the 21 rows centred on that row at half the image height. The average of 21 rows were used, as this corresponded to approximately 0.5 cm and this was assumed to be a sufficient distance for the edge of the breast to be approximated by a straight line.



(a) (b)

**Figure 4.5**: Breast profile (dotted) with a Lorentzian $L_{profile}$, overlaid (solid). **(a)** $L_{profile}$ is fitted to a fixed number of pixels. **(b)** $L_{profile}$ is fitted to an optimised number of pixels. The breast profile is selected as the average of 21 rows centred on that row at half the image height.

#### 4.1.3.7 Details of the Pre-processing Methods

The width of the Lorentzian, $L_{profile}$, ($a_2$ in **Equation 4-2**) was then used to define two smoothing kernels, a top-hat kernel (**Equation 4-3**) and a Lorentzian kernel (**Equation 4-2**), in one and two dimensions. The one-dimensional kernels were applied in the horizontal direction only, and the two-dimensional kernels were applied in the horizontal and vertical directions. The various pre-processing methods are detailed in **Table 4.1**.

$$T_\sigma = \begin{cases} 1 & \text{if } |x| \le \frac{\sigma}{2} \text{ for width=}\sigma \\ 0 & \text{otherwise} \end{cases} \tag{4-3}$$

**Table 4.1**: Pre-processing methods, using two smoothing kernels, for detection of the breast edge. Kernel widths are calculated from a Lorentzian, $L_{profile}$, fitted to a profile of the breast edge. In all cases, $d = 1$ refers to the kernel applied to the horizontal image plane only, while $d = 2$ refers to a two-dimensional kernel applied to the image.

| Method | Kernel | Description |
|---|---|---|
| $\mathrm{T}_{d,\sigma_1}$ | top-hat | $T$ with $\sigma_1 = a_2$. |
| $\mathrm{T}_{d,\sigma_2}$ | top-hat | $T$ with $\sigma_2 = \frac{1}{2}a_2$. |
| $\mathrm{L}_{d,\sigma_1}$ | Lorentzian | $L$ with $\sigma_1 = a_2$ |
| $\mathrm{L}_{d,\sigma_2}$ | Lorentzian | $L$ with $\sigma_2 = \frac{1}{2}a_2$. |

#### 4.1.3.8 The Algorithm

The algorithm to determine the breast edge is described in **Algorithm 4.1**.

---

**Algorithm 4.1**: Determination of breast edge

**Data**: Mammogram, $I$

**Result**: Mask, $P$, corresponding to breast tissue

1. Orient mammogram such such that the breast points to the right.
   This is done manually.
2. Take the average of 21 rows centred on the row at half the mammogram height.
   This is the breast profile.
3. Fit a Lorentzian, $L_{profile}$, to this breast profile to determine the width of the top-hat, $T$ ($a$ in **Equation 4-3**),
   and the Lorentzian, $L$, smoothing kernels.
4. Pre-process image using these smoothing kernels to enhance the breast edge.
5. Determine the largest area enclosed by the various contours, at each grey-level.
6. Fit straight lines to the area vs. grey-level data (**Figure 4.3**) and determine $g_o$.
7. Contour image at $g_o$ and extract contour with largest area to create a binary mask.
8. Post process mask (according to Masek et al. [2000]) to remove non-breast regions.
   The result is a mask, $P$, corresponding to the breast tissue.

---

#### 4.1.3.9 Materials

The algorithm, applying the various pre-processing methods, was tested on 25 mammograms (13 CC and 12 MLO), obtained from Addington Hospital (Durban, South Africa). The radiographs, were digitised at a bit-depth of 14 bits and a spatial resolution of 0.042 mm per pixel on an Epson Expression 1640XL scanner. The images were resampled to a bit-depth of 8 bits and a spatial resolution of 0.254 mm per pixel for analysis. The resulting images were approximately 800 pixels × 1200 pixels with a file size of 1.8Mb.

Automated borders were compared with borders drawn independently by three radiologists in their daily working environment. The borders were drawn on a transparent sheet placed over the mammogram. The radiologists were allowed to use all methods and equipment used during routine mammographic diagnosis to identify the breast border. The radiologists' borders were digitised and manually registered with the digitised mammograms. The accuracy of the automated borders was quantified by calculating the average root-mean-square difference ($\chi_{rms}$) between the automated and manual borders for each mammogram using:

$$\chi_{rms} = \sqrt{\frac{1}{n}\sum_{i=0}^{n-1}\left[r_a(\theta_i) - r_m(\theta_i)\right]^2} \tag{4-4}$$

with $n$ being the number of points on the border used for the evaluation. For this study, 50 evenly spaced points along the border were used. $r$ is the radial distance from the origin, for the automated (a) and manual (m) borders, at specific angles $\theta$, as shown in **Figure 4.6**. The origin of the polar co-ordinate system was selected as $[0, y_{dim}/2]$ where $y_{dim}$ is the height of the image of the breast. The start and ends points for the evaluation were determined by the detail available in the radiologist's borders. In the cases of the poor quality images, the borders did not extend to the edge of the image. Therefore, 180° was only covered if the radiologist data was available. If the data was not available, then the maximum angular extent was selected.

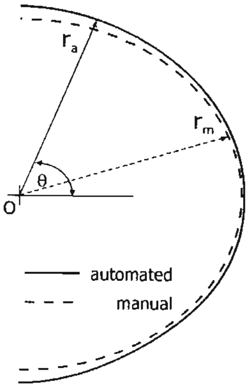


**Figure 4.6**: The polar co-ordinate system used to evaluate the accuracy of the automated borders (solid) compared to borders drawn by radiologists (dashed). $r$ is the radial distance from the origin, for the automated (a) and manual (m) borders, at specific angles $\theta$. The origin was selected as $[0, y_{dim}/2]$ where $y_{dim}$ is the image height.

The 25 images were divided into two sets: Set 1 contained 10 images with clear breast edges (clearly seen) and Set 2 contained 15 images with poor breast edges (grainy, poorly defined as a result of poor contrast between tissue and background). The allocation of images to Sets 1 and 2 was based on an analysis of the variation between the manual borders of the three radiologists for each image. A large variation in manual borders was taken to indicate an indistinct breast edge and manual borders that varied only slightly were taken to indicate a clear breast edge. The average $\chi_{rms}$ ($\overline{\chi}_{rms}$) for each method was determined for Set 1 and Set 2, for each of the radiologists.

A poorly defined breast edge implies poor contrast between the breast tissue and the background signal. The following factors affect contrast, and therefore the clarity of the breast edge (Rae 2005, pers. comm.[2]):

1. Patient related factors
    (a) The patient is very thin, but has relatively dense breasts, with gradual tapering to skin.
    (b) The breast has very dense central objects or breast tissue overlying automatic exposure control thus causing overexposure of the skin edge.
    (c) The patient has very thick breasts thus causing overexposure of the skin edge.
2. Technical factors
    (a) Any poor setting on the mammography machine causing overexposure
    (b) Positional, with a portion of the breast edge off the image
    (c) Poor setting of the processor with overexposure
3. Faulty equipment
    (a) Degeneration of old film
    (b) Poor processing of film
    (c) Increased background fog on film

### 4.1.4 Results and Discussion

#### 4.1.4.1 Results

The algorithm was completely automatic and running times depended on the pre-processing method used. Running times varied from 20s per image (for a one-dimensional top-hat kernel) to 500s per image (for a two-dimensional Lorentzian kernel). Results are presented in **Figure 4.7** and the evaluation results for the 3 best pre-processing methods applied to Set 1 and Set 2 can be seen in **Table 4.2**.

**Figure 4.7** shows that automated borders for Set 1 were generally more accurate for pre-processing with a top-hat kernel than with a Lorentzian kernel, although the most accurate result for Set 1 of 3.0 mm±0.3 mm was obtained by using a Lorentzian kernel (for method $L_{1,\sigma 2}$). Also for Set 1, the one-dimensional kernels generally yielded more accurate results than the two-dimensional kernels. The results for the two-dimensional Lorentzian kernel yielded results as accurate as those without pre-processing because the Lorentzian kernel most likely smoothed the edge more than was required.

**Figure 4.7** shows that automated borders for Set 2 were generally more accurate for pre-processing with a Lorentzian kernel than a top-hat kernel. The most accurate result for Set 2 of

---

[2] Dr. W.I.D. Rae, Department of Medical Physics, Addington Hospital, P. O. Box 977, Durban, 4000

**Table 4.2**: Evaluation results of the 3 best pre-processing methods applied to Set 1 containing 10 images with clear breast edges and Set 2 containing 15 images with indistinct breast edges. The average root-mean-square differences are shown for each radiologist (A,B,C) and the average for the three radiologists.

| Pre-processing | Radiologists |  |  | Average |
|---|---|---|---|---|
|  | A (mm) | B (mm) | C (mm) | (mm) |
| Set 1 |  |  |  |  |
| None | 5.5±1.5 | 4.9±1.5 | 4.9±1.4 | 5.1±0.8 |
| $L_{1,\sigma 2}$ | 3.2± 0.6 | 2.9± 0.5 | 2.8± 0.5 | 3.0± 0.3 |
| $T_{2,\sigma 1}$ | 3.3± 0.6 | 2.9± 0.5 | 2.9± 0.5 | 3.1± 0.3 |
| $T_{2,\sigma 2}$ | 3.7± 0.7 | 3.2± 0.6 | 3.0± 0.5 | 3.3± 0.4 |
| Set 2 |  |  |  |  |
| None | 7.9±1.5 | 8.3±1.6 | 7.6±1.5 | 7.9±0.9 |
| $L_{2,\sigma 2}$ | 5.1± 0.9 | 5.0± 0.9 | 4.4± 0.8 | 4.8± 0.5 |
| $L_{2,\sigma 1}$ | 5.2± 1 | 5.1± 0.9 | 4.9± 0.9 | 5.0± 0.5 |
| $L_{1,\sigma 1}$ | 6.6± 1.1 | 6.8± 1.2 | 6.1± 1 | 6.5± 0.6 |

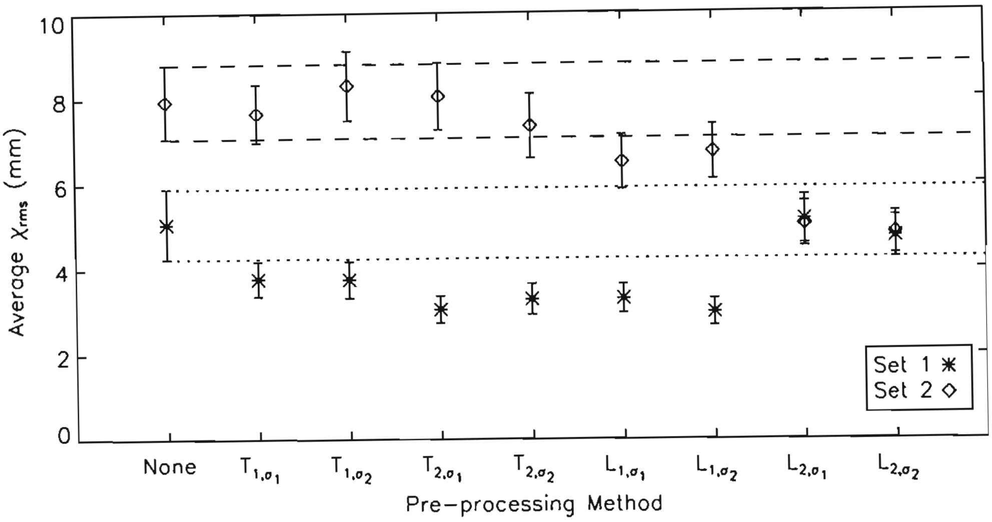


**Figure 4.7**: Average $\chi_{rms}$ for different pre-processing methods for smoothing with a top-hat kernel (T) and a Lorentzian kernel (L). The horizontal dotted and dashed lines indicate the extent of the average $\chi_{rms}$ with no pre-processing.

4.8 mm±0.5 mm was obtained for method $L_{2,\sigma 2}$. For Set 2, there was no clear dependence on the dimensionality of the kernel on the accuracy of the results.

#### 4.1.4.2 Discussion

Using the areas enclosed by iso-intensity contours to select a threshold aims to improve upon those thresholding methods that only use the grey-level histogram to threshold a mammogram for purposes of segmentation. The advantages of the iso-intensity contour method are that it is simple, requires no complex models of the breast or background and takes very little time to compute. The method works acceptably well on mammograms with clear breast edges, but performs less successfully on mammograms with unclear breast edges. The effects of different smoothing pre-processing methods were investigated to improve the results for images with unclear borders. Two smoothing kernels (top-hat and Lorentzian) were used with the smoothing widths determined by fitting a Lorentzian to a profile of the breast edge.

The automated borders with pre-processing were at worse, as accurate as those without pre-processing, but were often more accurate. Pre-processing with a Lorentzian smoothing kernel yielded the most accurate borders. The best results for Set 1, containing clear borders, of 3.0 mm±0.3 mm was obtained for method $L_{1,\sigma_2}$. The best results for Set 2 of 4.8 mm±0.5 mm was obtained for method $L_{2,\sigma_2}$. If the results are averaged over both image sets, method $L_{2,\sigma_2}$ performs best at 4.8 mm±0.3 mm.

Results were more accurate for pre-processing with a Lorentzian kernel because the spread in the breast edge is more similar to the shape of the Lorentzian than to the shape of the top-hat kernel.

Examples of the fits can be seen in **Figure 4.8**. There is significant variation between the radiologists' borders with unclear breast edges, while those for clear breast edges are quite similar. Automated borders are shown as solid lines.

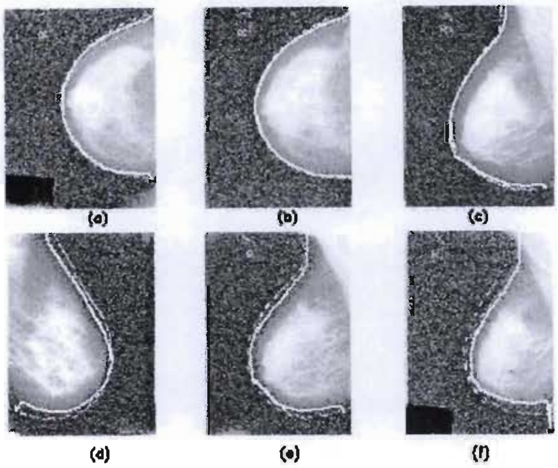


**Figure 4.8**: Examples of borders. Automated borders are shown as solid lines while those of the 3 radiologists are shown as dotted, dashed and dotted-dashed lines. The mammograms have been histogram equalised for display purposes. **(a-c)** Examples of 3 radiologists' borders for mammograms with clear breast edges. **(d-f)** Examples of 3 radiologists' borders for mammograms with unclear breast edges. There is significant variation between the radiologists' borders with unclear breast edges, while those for clear breast edges are quite similar.

A significant factor affecting the accuracy of automated borders is the quality of the ground truth data, which can strongly affect the evaluation results. This is most obvious when looking at radiologists' borders for Sets 1 and 2. There are minor variations between the radiologists' borders for clear breast edges (**Figure 4.8(a-c)**) and significant variations for unclear breast edges (**Figure 4.8(d-f)**). This probably means that the error bars for the results of Set 2 should be adjusted to take into account the variability in the radiologists' borders. This might bring these results into line with those for image Set 1.

Another factor, which might affect the accuracy of the automated borders, is the assumption that the clarity of the breast edge is uniform around the entire breast. The clarity of the breast edge is determined at one point, selected as the average of 21 breast edge profiles centred on that row at half the image height. Therefore, any variation in the clarity of the breast edge profile around the breast would affect the overall accuracy of the automated borders, since $\chi_{rms}$ is determined over most of the breast.

The best results (3.0 mm±0.3 mm) for mammograms with clear breast edges can only meaningfully be compared to the algorithms of Goodsitt et al. [1998] and Ojala et al. [2001] by testing the iso-intensity border algorithm on the identical data sets used by Ojala et al. [2001] and Goodsitt et al. [1998].

## 4.2 Detection of the Pectoral Muscle

If a radiographer follows the accepted guidelines for positioning the breast during mammography, then the pectoral muscle should be seen in the mediolateral and mediolateral-oblique views. Since the muscle contains no features of interest to this study, it was removed to reduce the region of computation. The edge of the pectoral muscle can be approximated by a straight-line and a straight-line detection method based on the Hough transform (**Figure 4.9**) was used to identify the edge of the pectoral muscle. The algorithm is based on that described by Karssemeijer [1998], where full details can be obtained. The algorithm is semi-automatic and is summarised in **Algorithm 4.2**. Examples of the various steps of the algorithm are shown in **Figure 4.10**.

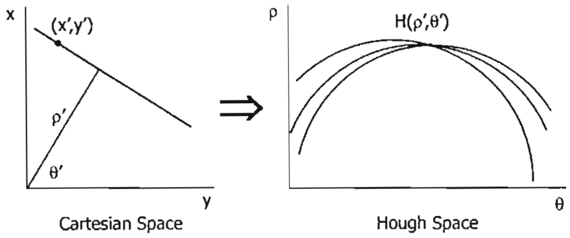


**Figure 4.9**: The Hough transform is used as a straight-line detector. Each point $(x',y')$ on a straight line oriented at an angle $\theta'$, placed a distance $\rho'$ away from the origin, in Cartesian space, generates a parametric curve in Hough space, corresponding to $\rho = x'\cos\theta + y'\sin\theta$. The set of parametric curves intersects at the point $(\rho',\theta')$ which corresponds to the location of the straight line in Cartesian space.

The mammogram is oriented such that the breast points to the right. The breast border mask is applied to the image to remove the background of the mammogram. The breast region to the upper left of the centroid of the breast is used as a reduced search area for the detection of the pectoral muscle. The Hough transform is applied to a gradient image of the reduced search region. The set of straight lines in the original image that correspond to maxima in Hough space are determined. The line that best corresponds to the edge of the pectoral muscle is manually selected and is used to generate a mask that excludes the pectoral muscle from the mammogram.

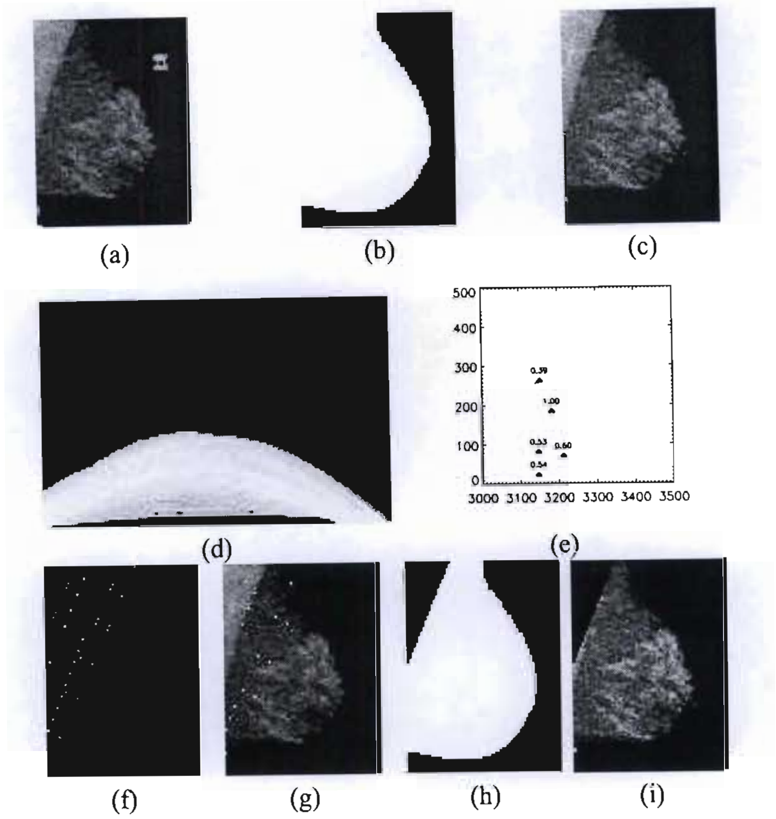


**Figure 4.10**: Example of implementation of detection of pectoral muscle edge. **(a)** MLO mammogram with pectoral muscle in top left corner. **(b)** Breast tissue mask excluding background. **(c)** Mammogram with breast tissue mask applied. **(d)** Hough transform of mammogram. **(e)** 5 highest peaks in the Hough transform. Heights relative to the highest peak are shown above each point. **(f)** Set of possible lines that correspond to pectoral muscle edge. **(g)** Set of possible lines that correspond to pectoral muscle edge overlaid on mammogram. **(h)** Mask with pectoral muscle removed. **(i)** Mammogram with pectoral muscle and background removed.

**Algorithm 4.2**: Determination of pectoral muscle edge

```
 Data: MLO mammogram, $I$
Result: Mask, $P$, corresponding to breast tissue
   1. Ensure that mammogram is oriented such that the breast points to the right.
   2. Apply breast border mask to $I$
   3. Find centroid of breast tissue
   4. Find region above line of gradient 1.5 passing through centroid of breast
   5. Find gradient image using, e.g. Sobel operator
   6. Use amplitudes of gradient image as weighting factors to normalise gradient image
   7. Find Hough transform, $H$, of normalised gradient image
   8. Find maxima in moving sampling windows in Hough space, $H$.
   9. Determine set of lines on mammogram that correspond to identified peaks in $H$.
  10. Manually select the line that (visually) best corresponds to pectoral muscle edge, from the set of
      lines.
  11. Use line to generate a mask of breast tissue, $P$, that excludes pectoral muscle.
```

## 4.3 Application of the Arc Method to Refine the Search Region

The geometric method developed by Paquerault et al. [2002], based on the relationship between the arc distances in the CC and MLO views, was used to define an annular search region in the test image, based on a reference region in the reference image. The positions of the nipple in both standard mammographic views and the centroid of the selected ROI in one view were used to extract that portion of the breast in the other view where the ROI could possibly lie.

### 4.3.1 Paquerault Geometric Model

Paquerault et al. [2002] developed the geometric model to reduce the search region (for false-positive reduction) on 177 objects identified on the CC and MLO views by a radiologist. The locations of the nipples on both views were also marked. The polar co-ordinate system detailed in **Figure 4.11** was determined for each mammogram. The location of each object in polar co-ordinates was determined for each mammographic view. An analysis of the correlation of the radial and angular components in each view showed that there was a high linear correlation between the radial components, but poor correlation for the angular components. The 177 objects were then used to determine the parameters for a linear model to predict the radial position of an object in one view based on its radial location in another view. The error associated with the determined radial value resulted in an annular search region defined by $r \pm \Delta r$. $r$ is the distance of the object from the nipple in one view and $\Delta r$ the error determined from the training set. Paquerault et al. fixed $\Delta r$ at 80 pixels, which reduced the search region in large breasts, but not to the same degree for small breasts. It was suggested that $\Delta r$ be chosen as a percentage of the breast area.

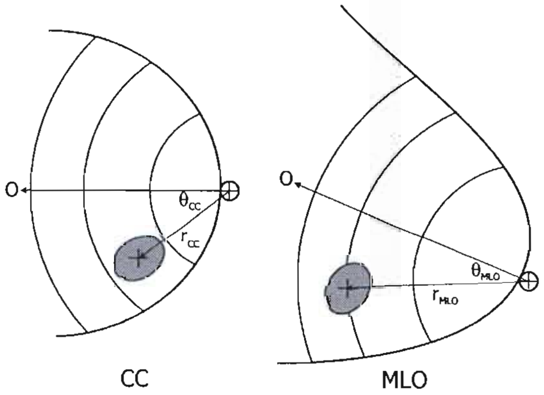


**Figure 4.11**: Geometry used by Paquerault et al. [2002] to study the relationship between the arc distances in the CC and MLO views. Drawing concentric circles to intersect with the breast border, with the nipple as origin, determined the polar co-ordinate system used for the analysis. The locus of the midpoints of the concentric arcs with the nipple as origin determined the co-ordinate system. The location of each object in polar co-ordinates was determined for each mammographic view.

### 4.3.2 Applicability of Paquerault Geometric Model to this Study

To test the applicability of the Paquerault model to the mammograms used in this project, the position of the nipples were identified and the centroids of the ROIs were determined from the borders drawn by the radiologist. The arc distances (distance from nipple to ROI centroid) were determined in the CC and MLO views. These distances are plotted in **Figure 4.12** to investigate whether the correlation between the arc distances in the CC and MLO views is similar to the correlation obtained by Paquerault et al.

Two linear models were fitted to data to determine correlation: the first model had a $y$-intercept of 0 (dotted line) and the second had a non-zero $y$-intercept (solid line). The correlation coefficients for these models are 0.93 and 0.94 respectively, which compare well to the value of 0.94 obtained by Paquerault et al. It was therefore concluded that the Paquerault geometric model could be applied to the mammograms used in this study, to reduce the region searched for a match.

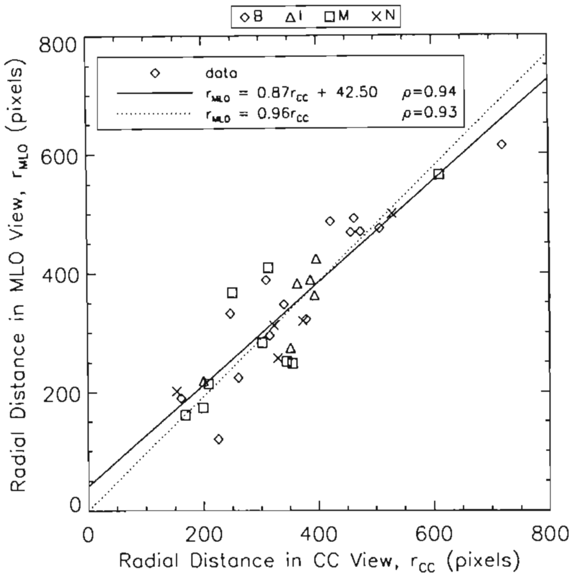


**Figure 4.12**: Justification of use of Paquerault model for the mammograms in this study. The correlation coefficient, ρ, between the distances from nipples to ROI centroids in the CC and MLO views was determined to investigate the applicability of the Paquerault geometric model to the set of mammograms used in this study. ρ compares well to the 0.94 obtained by Paquerault et al. [2002]. Data points correspond to mammograms diagnosed as: benign (B), indeterminate (I), malignant (M) and normal (N).

### 4.3.3 Geometric Model used in this Study

The Paquerault model was modified slightly for this study. Since the correlation coefficient was close to 1, a one-to-one correspondence was assumed between the arc distances in the CC and MLO views. Any error in determining the radial distance was taken into account with a variable $\Delta r$, that was extracted from the size of the ROI in the reference view. This meant that the area of the annular region would depend on the size of the ROI selected. The algorithm is described in **Algorithm 4.3** and in **Figure 4.13**. In summary, given the position of the nipple in the CC view, $[nx_{CC}, ny_{CC}]$, the position of the nipple in the MLO view, $[nx_{MLO}, ny_{MLO}]$, the position of the centroid of the ROI, $[x_m, y_m]$ and the maximum extent of the ROI, $\Delta r$, in the CC view, draw two arcs (of radii $r + 2\Delta r$ and $r - 2\Delta r$) in the MLO-view, using the nipple as origin, with the breast

**Algorithm 4.3**: Definition of annular region to reduce area of computation. The algorithm is described for a known ROI in the CC-view, but can be applied for a known ROI in any view.

```
Data: MLO mammogram, $I$ and breast border mask, $B$
Result: Mask, $P$, corresponding to breast tissue
1. Find the positions of the nipple in the MLO $[nx_{MLO}, ny_{MLO}]$, and CC views, $[nx_{CC}, ny_{CC}]$.
2. Find the position of the centroid of the ROI in the CC-view, $[x_m, y_m]$
3. Determine the distance, $r$, from the nipple to the centroid of the ROI, in the CC-view.
4. Determine the maximum extent of the ROI from the centroid, $\Delta r$, in the CC-view.
5. In the MLO-view, define the annular search region by using the nipple as centre and drawing two
   arcs of radii $r + 2\Delta r$ and $r - 2\Delta r$, bounded by the breast border.
6. Generate a mask for the MLO-view, based on the region enclosed by the arcs.
```

border bounding each. The region enclosed between the arcs and the breast border defines the annular, reduced search region that contains the ROI in the MLO-view.

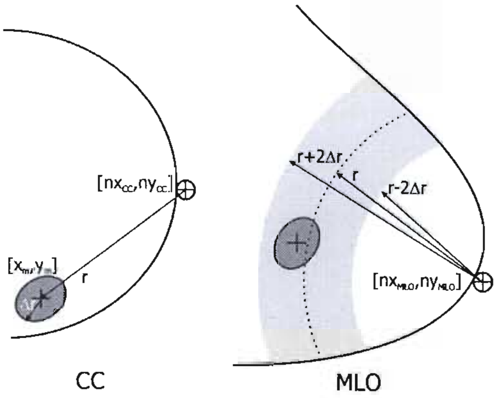


**Figure 4.13**: Geometry of the arc method used to reduce the search region in the test image. The position of the nipple in the CC view, $[nx_{CC}, ny_{CC}]$, the position of the nipple in the MLO view, $[nx_{MLO}, ny_{MLO}]$, the position of the centroid of the ROI, $[x_m, y_m]$ and the maximum extent of the ROI, $\Delta r$, in the CC view are required to locate the ROI in the MLO view.

## 4.4 Overall Results

An example of the overall result of pre-processing a mammogram is shown in **Figure 4.14** for a mammogram of a right breast. The final processed image contains significantly less breast tissue that has to be searched for the ROI than the original unprocessed test image.

## 4.5 Summary

A novel, simple method of finding the breast edge using areas enclosed by iso-intensity contours was presented that improves on traditional thresholding methods for segmentation, by incorporating spatial information into the segmentation. The method does not rely on models of the breast or background and borders. Results were evaluated by comparison to breast borders drawn by three radiologists in their normal working environment. The effect of various pre-processing methods on the accuracy of the automated borders was investigated. Results were generally good for those

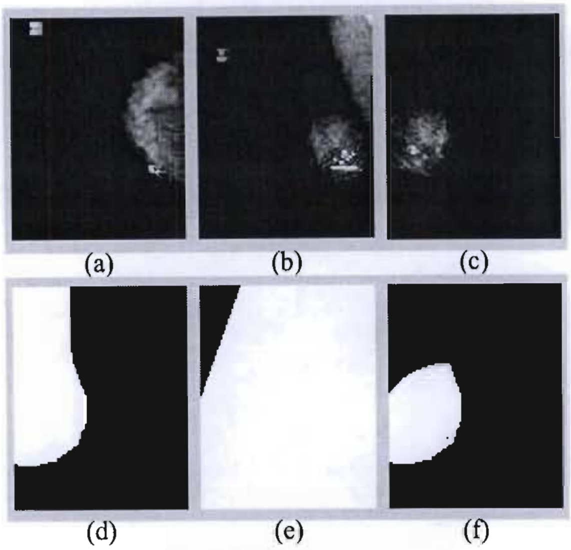


**Figure 4.14**: Example of the result of applying the pre-processing methods to a mammogram. **(a)** Reference image. **(b)** Test image. **(c)** Result of pre-processing test image. **(d)** Breast tissue mask excluding background. **(e)** Mask excluding pectoral muscle. **(f)** Mask of reduced search region. All processing is performed with the breast pointing to the right.

images containing clear breast edges. It was found that smoothing with a Lorentzian kernel as a pre-processing method, with the width automatically determined for each mammogram worked acceptably well for those with clear breast edges. The best results for mammograms with clear breast edges was 3.0 mm±0.3 mm.

The semi-automatic algorithm used to remove the pectoral muscle was based on the work of Karssemeijer [1998]. The arc method was used to define an annular, reduced search region, by using the position of the ROI in one standard mammographic view and the positions of the nipple in both views, following the work of Paquerault et al. [2002]. The overall result of the three pre-processing steps is a significantly reduced region in the test image, which is searched for a match.

# Chapter 5

# Quantification of Image Texture

The ability to quantify the characteristics of image texture forms the foundation of this project. The concept of texture and different methods of quantifying texture are discussed in this chapter.

## 5.1 What is *texture*?

An important characteristic of images is *texture* and while texture has no universally accepted definition, it is common to find words like *smooth*, *fine*, *grainy* and *coarse* used to describe it [Gonzalez & Wintz 1987, Sonka et al. 1999]. Image texture can also be described as the variation in grey-level from pixel to pixel or region to region. If the grey-level is interpreted as an elevation on a surface then texture is a measurement of the surface properties [Haralick 1979, Russ 1995, Sonka et al. 1999]. Some examples of different textures and the corresponding texture viewed as a surface are shown in **Figure 5.1**.

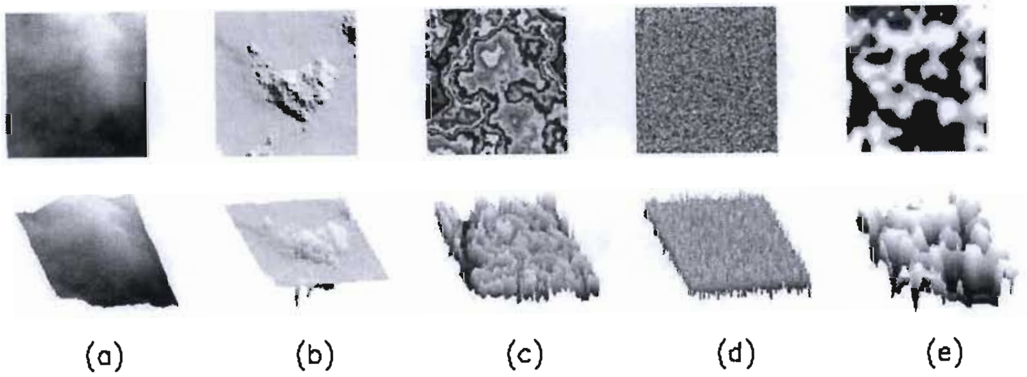


**Figure 5.1**: Examples of different types of image texture. Texture has no universal definition, but words like smooth, fine, grainy and coarse are used to describe it. In an image, if the grey-levels are interpreted as elevations on a surface then texture is a measurement of the surface properties. The texture images, shown in the top row, are represented as surfaces in the bottom row, to demonstrate the variation in surface properties with image texture

## 5.2 Texture Quantification Methods

In order to compare textures, there must be some method of quantifying textural characteristics, i.e. extracting how the grey-levels in an image are arranged, relative to each other. Many methods have been developed to quantify texture and some of these are briefly discussed. There are three categories of methods to quantify texture in an image [Gonzalez & Wintz 1987]:

1. statistical methods that only examine the distributions of the grey-levels without taking spatial information into account, e.g. grey-level histograms;
2. structural methods where locations of pixels and grey-levels are taken into account, e.g. grey-level co-occurrence matrices; and

3. spectral methods, e.g. the autocorrelation function.

The results of the above methods are often an estimate of a *probability density function*, which incorporates information about the frequency of the grey-levels, pixel location and scale information. These probability density functions generally require large matrices and are memory-intensive during computation. Therefore, statistical measures are calculated from the probability density functions to reduce computation and memory requirements. For example, if the probability density function is a histogram then the mean, variance, skewness and kurtosis extracted from the histogram are examples of the statistical measures that summarise the general shape of the histogram.

## 5.3 Probability Density Functions

Let $X = [x_0, \dots, x_{n-1}]$ be a discrete random variable with a finite number of states, $n$. $X$ is governed by a discrete probability distribution, that is an assignment of a probability, $p(x_i)$, to each state, $x_i$ (for $i = 0, 1, \dots, n-1$), denoted by:

$$P(X) = [p(x_0), p(x_1), p(x_2), \dots, p(x_{n-1})] \tag{5-1}$$

$P(X)$ is referred to as the *probability density function.*

## 5.4 Statistical Methods of Texture Quantification

Statistical approaches to quantifying texture are based on probability density functions of only the grey-levels in the image [Gonzalez & Wintz 1987].

### 5.4.1 Grey-level Histograms

The most common probability density function used for statistical textural description is the grey-level histogram. For a random variable $X = [x_0, x_1, \dots, x_{q-1}]$ with an origin, $x_0$, and a bin width, $h$, define the bins of the histogram to be the intervals [$x_0$+mh,$x_0$+(m+1)h)] for $m \in \mathbb{Z}$. If there are $n = \frac{\max(X)-\min(X)}{h}$ bins, then the histogram, $P_h(X)$ is defined as [Silverman 1986]:

$$P_h(X) = [p_h(x_0), p_h(x_1), \dots, p_h(x_{q-1})] \qquad \text{with} \tag{5-2}$$

$$p_h(x_i) = \frac{\text{number of entries in same bin as } x_i}{nh} \qquad \text{for each } x_i \text{ in } X,\ i = 0, \dots, q-1 \tag{5-3}$$

The bin width controls the amount of smoothing inherent in the histogram.

While histograms are computationally simple, they are not ideal for estimating probability densities because the resulting function is not continuous. The discontinuous nature of histograms causes problems when derivatives are required. There are other methods of estimating probability density functions, which do not suffer from the problems associated with histograms, but these methods are more computationally intensive than histograms. Some examples are [Silverman 1986]:

1. Naive estimators using Parzen windows
2. Kernel estimator
3. $k$-nearest neighbours
4. Adaptive kernel estimator

## 5.5 Structural Methods of Quantifying Texture

Statistical methods of quantifying texture are limited by the fact that spatial information about the locations of the pixels is not taken into account. Structural methods of texture quantification take spatial information into account.

### 5.5.1 Grey-Level Co-occurrence Matrices (GLCMs)

A popular structural method of quantifying texture is to use *grey-level co-occurrence matrices* (GLCMs), which has the advantage of including information about the relative positions of pixels in an image [Haralick et al. 1973, Gonzalez & Wintz 1987]. GLCMs are also known as spatial grey-level dependence (SGLD) matrices and incorporate information about the distributions of the locations of pixels and their grey-levels.

Consider an image, $I$, of bit-depth, $lbits$ and of dimensions $M \times N$. Let $l = 2^{lbits}$ represent the number of grey-levels. Also let $I[j,k] = m$ and $I[j - d\sin\theta, k + d\cos\theta] = n$, with $m,n \in [0,\ldots,l]$. Then the element of the GLCM, $\mathbf{G}_{\theta,d}[m,n]$, for two pixels of grey-levels $m$ and $n$, located a distance $d$ apart in the direction $\theta$, can be defined as:

$$\mathbf{G}_{\theta,d}[m,n] = \sum_{j=0}^{M-1}\sum_{k=0}^{N-1} \delta(I[j,k] = m, I[j - d\sin\theta, k + d\cos\theta] = n) \tag{5-4}$$

where

$$\delta(x,y) = \begin{cases} 1 & \text{if x=y} \\ 0 & \text{if } \mathrm{x} \neq \mathrm{y} \end{cases} \tag{5-5}$$

A schematic of the calculation of the GLCM is shown in **Figure 5.2**.

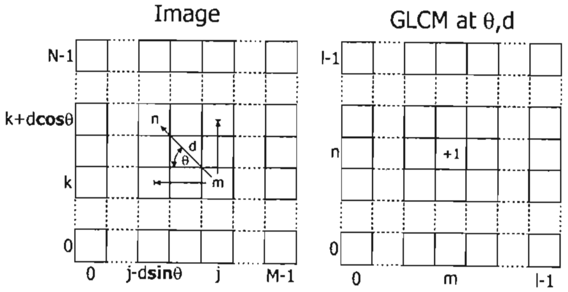


**Figure 5.2**: Schematic of GLCM ($\mathbf{G}_{\theta,d}$) calculation, at a distance $d$ in the direction $\theta$, for an $M \times N$ image, $I$, with $l$ grey-levels. $I[j,k] = m$ and $I[j - d\sin\theta, k + d\cos\theta] = n$, with $m,n \in [0,\ldots,l]$

$\mathbf{G}_{\theta,d}[m,n]$ is an estimate of the probability that a pair of pixels at an angle $\theta$ relative to each other and $d$ pixels apart, will have values $[m,n]$. $\mathbf{G}_{\theta,d}$ is therefore dependent on the location and grey-level of the pixels and texture measures based on $\mathbf{G}_{\theta,d}$ will have some information relating to the relative positions of the pixels [Gonzalez & Wintz 1987].

For digital images, define $\overline{\mathbf{G}}_d$ as the $\mathbf{G}_{\theta,d}$ over all angles, such that:

$$\overline{\mathbf{G}}_d = \frac{1}{8}(\mathbf{G}_{0^\circ,d} + \mathbf{G}_{45^\circ,d} + \mathbf{G}_{90^\circ,d} + \mathbf{G}_{135^\circ,d} + \mathbf{G}_{180^\circ,d} + \mathbf{G}_{225^\circ,d} + \mathbf{G}_{270^\circ,d} + \mathbf{G}_{315^\circ,d}) \quad (5\text{-}6)$$

However,

$$\begin{aligned}\mathbf{G}_{180^\circ,d} &= \mathbf{G}^T_{0^\circ,d}\\ \mathbf{G}_{225^\circ,d} &= \mathbf{G}^T_{90^\circ,d}\\ \mathbf{G}_{270^\circ,d} &= \mathbf{G}^T_{45^\circ,d}\\ \mathbf{G}_{315^\circ,d} &= \mathbf{G}^T_{135^\circ,d}\end{aligned}$$

Now, **Equation 5-6** can be simplified to:

$$\begin{aligned}\overline{\mathbf{G}}_d &= \frac{1}{8}(\mathbf{G}_{0^\circ,d} + \mathbf{G}^T_{0^\circ,d} + \mathbf{G}_{45^\circ,d} + \mathbf{G}^T_{45^\circ,d} + \mathbf{G}_{90^\circ,d} + \mathbf{G}^T_{90^\circ,d} + \mathbf{G}_{135^\circ,d} + \mathbf{G}^T_{135^\circ,d}) \quad (5\text{-}7)\\ &= \frac{1}{8}\left[(\mathbf{G}_{0^\circ,d} + \mathbf{G}_{45^\circ,d} + \mathbf{G}_{90^\circ,d} + \mathbf{G}_{135^\circ,d}) + (\mathbf{G}_{0^\circ,d} + \mathbf{G}_{45^\circ,d} + \mathbf{G}_{90^\circ,d} + \mathbf{G}_{135^\circ,d})^T\right]\end{aligned}$$

which shows that $\overline{\mathbf{G}}_d$ is a symmetric matrix of size $l \times l$ with non-negative elements.

$$\begin{aligned}\mathbf{G}_{\theta,d}[m,n] &= \mathbf{G}_{\theta,d}[n,m] \quad (5\text{-}8)\\ \mathbf{G}_{\theta,d}[m,n] &\geq 0\end{aligned}$$

#### 5.5.1.1 GLCM Algorithm

The calculation of the GLCM is based on **Equation 5-7** with $\theta \in [0, 45^\circ, 90^\circ, 135^\circ]$ and $d$ restricted to integer multiples of the pixel separation [Chan et al. 1995]. The algorithm is detailed in **Algorithm 5.1**. The algorithm uses two IDL functions (`SHIFT` and `HIST_2D` explained in **Appendix D)** and is explained schematically in **Figure 5.3**. Since the GLCM examines the relationship between pixels in an image, $A$, separated by a distance, $d$, at a relative angle, $\theta$, the `SHIFT` function is used to create an image, $B$, which is the image $A$ with the pixels shifted by $d$ and $\theta$ towards the reference pixels. The `HIST_2D` function is then applied to portions of images $A$ and $B$ resulting in the GLCM of $A$ at the distance, $d$, and relative angle, $\theta$. This algorithm is considerably quicker than traditional loop-based algorithms to calculate the GLCM and a comparison of computation times is shown in **Figure 5.4**.

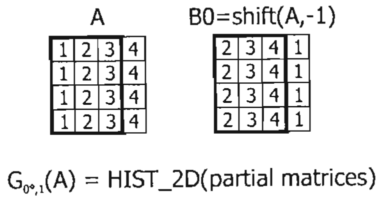


**Figure 5.3**: Schematic of the new IDL GLCM algorithm using IDL's `HIST_2D` and `SHIFT`. The algorithm is demonstrated on an image *A*. *B* is the image *A* with the columns shifted to the left by 1 pixel. The GLCM is computed by taking `HIST_2D` of the matrices bordered in thicker lines at for $\theta = 0°$ and $d$=1. `HIST_2D` and `SHIFT` are described in **Appendix D**.

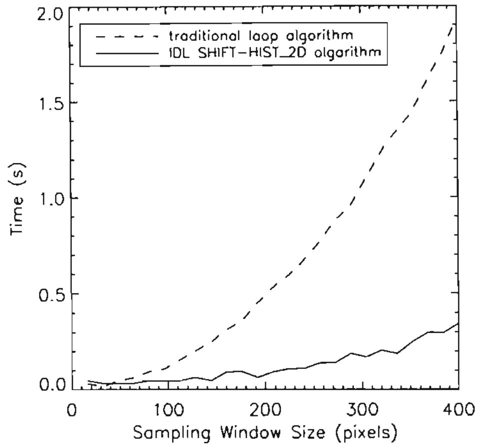


**Figure 5.4**: Comparison of computational time between the new IDL GLCM algorithm and a loop-based algorithm.

**Algorithm 5.1**: Calculation of average GLCM for θ=[0°,45°,90°,135°] based on **Equation 5-7**.

**Data**: Image **A**, distance in GLCM calculation, *d*, bit-depth of images, *nbits*, size of images, *sz*

**Result**: GLCM **G** of image **A** averaged over θ=[0°,45°,90°,135°].

```
    ; IDL's SHIFT and HIST_2D functions simplify the GLCM calculation
$max_l = 2^{nbits} - 1$
    ; 0° calculation
B0 = SHIFT(A,d,0)
c0 = HIST_2D(A[d:sz[0]-1,*], B0[d:sz[0]-1,*], max = $max_l$, min = 0)
    ; 45° calculation
B45 = SHIFT(A,d,d)
c45 = HIST_2D(A[d:sz[0]-1,d:sz[1]-1], B45[d:sz[0]-1,d:sz[1]-1], max = $max_l$, min = 0)
    ; 90° calculation
B90 = SHIFT(A,0,d)
c90 = HIST_2D(A[*,d:sz[1]-1], B90[*,d:sz[1]-1], max = $max_l$, min = 0)
    ; 135° calculation
B135 = SHIFT(A,d,-d)
c135 = HIST_2D(A[d:sz[0]-1,0:sz[1]-d-1], B135[d:sz[0]-1,0:sz[1]-d-1], max = $max_l$, min = 0)
c = c0+c45+c90+c135
G = (c+TRANSPOSE(c))/8
```

#### 5.5.1.2 GLCM Example

The GLCM, $\mathbf{G}_{\theta,d}$, with $\theta=0°$ and $d=1$, for the following $4 \times 4$ image, $I$, with 4 grey-levels is given by:

$$I = \begin{matrix} 0 & 1 & 0 & 1 \\ 1 & 2 & 1 & 2 \\ 2 & 3 & 2 & 3 \\ 3 & 0 & 3 & 0 \end{matrix} \quad \mathbf{G}_{0,1} = \begin{bmatrix} 0 & 3 & 0 & 3 \\ 3 & 0 & 3 & 0 \\ 0 & 3 & 0 & 3 \\ 3 & 0 & 3 & 0 \end{bmatrix}$$

### 5.5.2 GLCM-based Texture Measures

The dimensions of $\mathbf{G}_{\theta,d}$ depend only upon the number of grey-levels in the image [Gonzalez & Wintz 1987] which for an 8-bit image, means a $256 \times 256$ matrix. Working with these large matrices is made easier by the use of *texture measures*, which are extracted from the GLCMs and generally contain information image characteristics, like homogeneity, contrast, complexity, presence of organised structures and grey-level transitions within the image.

The point to note is that a specific texture measure cannot be uniquely related to a specific image characteristic [Chan et al. 1995]. So, while some texture measures describe certain physical textural characteristics, most cannot be directly related to such a characteristic. Robert Haralick's [Haralick et al. 1973, Haralick 1979] texture measures are detailed and used in this study.

Calculate a probability density function, $P$, by normalising $\mathbf{G}_{\theta,d}$ determined from an image with $l$ grey-levels:

$$P = \frac{\mathbf{G}_{\theta,d}}{\sum_{i=0}^{l-1}\sum_{j=0}^{l-1}\mathbf{G}_{\theta,d}(i,j)} \tag{5-9}$$

The following texture measures can then be defined [Haralick et al. 1973, Haralick 1979] (a description, where appropriate, is given of the physical textural characteristic that a texture measure describes):

1. **Maximum Probability in** $P$: gives the strongest response to the position operator defined by $d$ and $\theta$ that is used to construct the GLCM.

$$\max_{ij} p_{ij} \tag{5-10}$$

2. **Entropy,** $H$: The entropy of a random variable is a measure of the uncertainty associated with the random variable. It is a measure of the amount of information required (on average) to describe the random variable. As a result entropy can be taken to be a measure of non-uniformity in the image. Entropy is at its maximum when all elements of the GLCM are equal [Abramson 1963, Bradley et al. 1995, Mudigonda et al. 2000]. Entropy is discussed in greater detail in **Appendix C** (page 247).

$$H = -\sum_{i=0}^{l-1}\sum_{j=0}^{l-1} p_{ij}\log_2(p_{ij}) \tag{5-11}$$

3. **Second Angular Moment/Energy**: The second angular moment (or energy) is a measure of homogeneity. The second angular moment has its lowest value when all elements of

12. **Difference Entropy**:

$$\sum_{k=0}^{2n-2} p_{x-y}(k)\log_2(p_{x-y}(k)) \tag{5-21}$$

13. **Information Measure of Correlation 1**:

$$\frac{H-H_1}{max[H_x,H_y]} \tag{5-22}$$

where

$$H_1 = -\sum_{i=0}^{l-1}\sum_{j=0}^{l-1} p_{ij}\log_2(p_x(i)p_y(j))$$

$$H_x = -\sum_{i=0}^{l-1}\sum_{j=0}^{l-1} p_x(i)\log_2(p_x(i))$$

$$H_y = -\sum_{i=0}^{l-1}\sum_{j=0}^{l-1} p_y(j)\log_2(p_y(j))$$

14. **Information Measure of Correlation 2**:

$$\sqrt{1-\exp 2(H-H_2)} \tag{5-23}$$

where

$$H_2 = -\sum_{i=0}^{l-1}\sum_{j=0}^{l-1} p_x(i)p_y(j)\log_2(p_x(i)p_y(j))$$

In most medical imaging applications, texture measures are extracted from GLCMs, which are good at quantifying the spatial relationship between tonal pixels and are invariant to monotonic grey-level transformations. However, GLCM-based texture measures do not consider primitive shapes and are not appropriate if the texture consists of large primitives. Another limiting factor is the memory requirements [Sonka et al. 1999].

#### 5.5.2.1 Categorising of Texture Measures

Gotlieb & Kreyszig [1990] put Haralick's texture measures into 4 classes as follows:

1. Classifiers that express visual textural characteristics
   (a) second angular moment or homogeneity
   (b) contrast
   (c) correlation
2. Classifiers that are based on statistics
   (a) inverse difference moment
   (b) sum average
   (c) sum variance
   (d) difference variance
3. Classifiers that are based on information theory
   (a) entropy

    (b) sum entropy
    (c) difference entropy
4. Classifiers that are based on information measures of correlation
    (a) information measure of correlation 1
    (b) information measure of correlation 2
    (c) maximal correlation coefficients

Gotlieb & Kreyszig [1990] found that contrast, inverse difference moment and entropy appeared most often in the classification of various types of textures.

## 5.6 Spectral Textural Approaches

For statistical methods like histograms, the only information that is encoded into the probability density function is the number of times a grey-level appears in the image. For structural methods like GLCMs, information about neighbouring pixels and grey-levels are encoded, but no information about large-scale structures or periodicity of structures is taken into account.

Spectral approaches of quantifying texture examine the periodic and scale properties of the image.

### 5.6.1 The Autocorrelation Function

The autocorrelation function (ACF) of a function $f(x,y)$, denoted by $f \star f$, is the self-convolution of $f(x,y)$ and is defined as [Bracewell 1965]:

$$f \star f = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(u,v) f(x-u, y-v)\, du\, dv \tag{5-24}$$

If however, $f(x,y)$ is not periodic and has a finite length, $k_x, k_y$ then the autocorrelation function, $\gamma(u,v)$, of this function $f_{k_x,k_y}(x,y)$, which is zero outside of the finite length is:

$$\gamma(u,v) = \frac{\int_{-0.5k_x}^{0.5k_x} \int_{-0.5k_y}^{0.5k_y} f^{*}_{k_x,k_y}(u,v) f_{k_x,k_y}(x+u, y+v)\, du\, dv}{\int_{-0.5k_x}^{0.5k_x} \int_{-0.5k_x}^{0.5k_x} f^{*}_{k_x,k_y}(u,v) f_{k_x,k_y}(u,v)\, du\, dv} \tag{5-25}$$

The ACF finds linear spatial relationships between texture primitives. If texture primitives are large, then the ACF decreases slowly as distance increases and decreases quickly if the primitives are small. If the texture primitives are periodic, then the ACF changes periodically with distance [Singh & Singh 2002].

The ACF of an image can be used to determine scale sizes within which there is correlation in the image. If the sums of the components of autocorrelation function in the horizontal and vertical directions are examined, the minima indicate the scale lengths in each direction. These scale lengths can then be used as an indication of the characteristic lengths of the textures in the image. An example of the ACF of an image is shown in **Figure 5.5**.

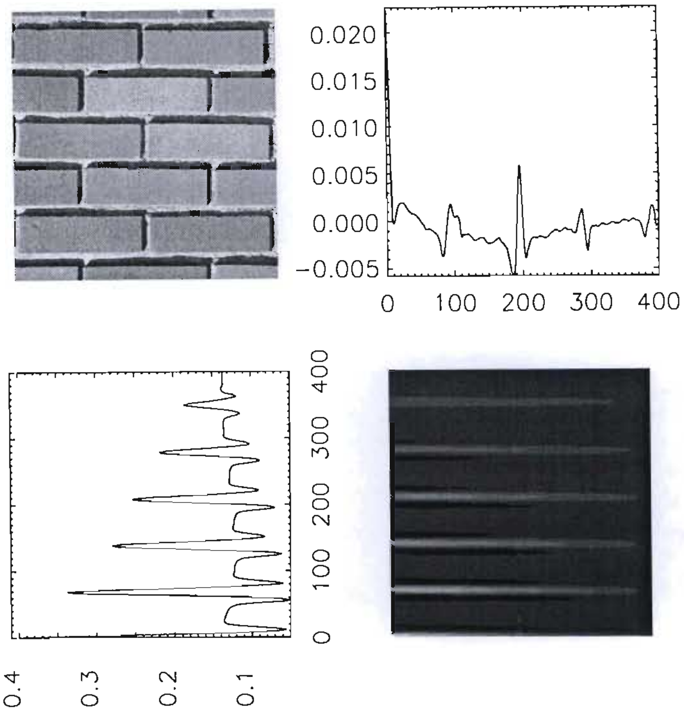


**Figure 5.5**: Example of the autocorrelation function of an image. **(top-left)** The original image. **(top-right)** The sum of the vertical components of the ACF. The minima indicate the scale sizes in the horizontal direction in the image and correspond to the widths of the bricks in the original image. **(bottom-left)** The sum of the horizontal components of the ACF. The minima indicate the scale sizes in the vertical direction in the image and correspond to the heights of the bricks in the original image. **(bottom-right)** The autocorrelation function. The periodic nature of the structures in the image are reflected in the autocorrelation function.

## 5.7 Summary

The analysis of texture forms an important part of image processing. Textural information must be quantified before images can be compared. The main quantification methods are statistical, structural and spectral. Statistical methods (like grey-level histograms) only incorporate the number of grey-levels in an image. Structural methods (like grey-level co-occurrence matrices) incorporate numbers of grey-levels and information about the location of the pixels in the image. Spectral methods (like the autocorrelation function) examine the periodic and scale properties of an image.

# Chapter 6

# Similarity Metrics

Similarity metrics quantify how similar two quantities are to each other and are a critical component of texture analysis algorithms where textures are being compared. Two types of similarity metrics are discussed: distance metrics and information theory metrics.

## 6.1 Distance Similarity Metrics

### 6.1.1 Introduction

The distance between two quantities is a simple indication of how similar the two quantities are, with similarity increasing as distance decreases. Common distance metrics are the Euclidean distance metric and the Mahalanobis distance metric.

### 6.1.2 Euclidean Distance Metric, $D_E$

The Euclidean distance metric, $D_E$, is the most commonly used metric to calculate distance. $D_E$ between two points, $x$ and $y$, in $n$ dimensions is defined as [PlanetMath 2006]:

$$D_E(x,y) = \sqrt{\sum_{i=0}^{n-1} (x_i - y_i)^2} \tag{6-1}$$

The strength of using $D_E$ as a similarity metric is that it is a simple calculation, but suffers from the weakness that a single value can dominate the result.

### 6.1.3 Euclidean Distance Metric with Standardised Variables, $D_{ES}$

The standardised Euclidean distance metric, $D_{ES}$, differs from the traditional Euclidean distance in that the inputs are standardised to a normal distribution with a mean of zero and a variance of one. This has the advantage over **Equation 6-1** in that the inputs are scaled to the same range and problems resulting from a single variable dominating the $D_E$ calculation do not arise. $D_{ES}$ between two points, $x$ and $y$, in $n$ dimensions is [De Maesschalck et al. 2000]:

$$D_{ES}(x,y) = \sqrt{\sum_{i=0}^{n-1} (\tilde{x}_i - \tilde{y}_i)^2} \tag{6-2}$$

where $\tilde{x}$ and $\tilde{y}$ represent $x$ and $y$ that have each been standardised to a normal distribution with a mean of zero and a variance of one.

### 6.1.4 Mahalanobis Distance Metric, $D_M$

A slight modification on the Euclidean distance metric is the Mahalanobis distance metric, $D_M$. In 1936, P. C. Mahalanobis derived the formula for the Mahalanobis distance, which considers correlations between the inputs. The Mahalanobis distance between the $i^{th}$ sample of two variables, $X$ and $Y$, in $n$ dimensions, with $m$ samples, i.e. $X = [X_0, X_1, X_2, \ldots, X_{m-1}]$ and $X_i = [x_{i,0}, x_{i,1}, x_{i,2}, \ldots, x_{i,n-1}]$ with covariance matrix $\mathbf{C}$ is [De Maesschalck et al. 2000]:

$$D_M(X_i, Y_i) = \sqrt{(X_i - Y_i)\mathbf{C}^{-1}(X_i - Y_i)}. \tag{6-3}$$

The covariance matrix $\mathbf{C}$ is an $n \times n$ matrix and contains the sample covariances between each variable. The sample covariances are calculated over all samples of a specific variable. For two observations, $x$ and $y$, with $m$ samples, $x = (x_0, x_1, x_2, \ldots, x_{m-1})$ and $y = (y_0, y_1, y_2, \ldots, y_{m-1})$, with means of $\bar{x}$ and $\bar{y}$ respectively, the covariance between $x$ and $y$ is defined as:

$$C_{xy} = \frac{\sum_{i=0}^{m-1}(x_i - \bar{x})(y_i - \bar{y})}{m-1} \tag{6-4}$$

### 6.1.5 Differences between Distance Metrics

A significant problem with $D_E$ is that a single value can dominate the final result. The inputs to the $D_E$-calculation are standardised, resulting in $D_{ES}$, to overcome this problem. However, the standardisation process does not consider any correlations between inputs, which is addressed by $D_M$.

### 6.1.6 The Use of Distance Metrics as Similarity Metrics

In medical imaging applications, distance metrics often appear in classifiers like the linear discriminant analysis classifier, where the distance metrics are used to compute the distance between classes. For example, Rangayyan et al. [1997*a*], used the Mahalanobis distance metric in a classifier to classify masses as malignant or benign, while Chan et al. [1995] used the Mahalanobis distance in a linear discriminant analysis classifier to discriminate between normal and abnormal tissue. The algorithm to compute the various distance maps using texture measures as inputs is detailed on page 75.

## 6.2 Information Theory Similarity Metrics

### 6.2.1 Introduction

In 1948 Claude E. Shannon published a paper entitled *A Mathematical Theory of Communication* where he described the problem of encoding a message such that the capacity of a channel to transmit the message at a given rate is a maximum [Shannon 1948]. Shannon introduced two quantities: *entropy*, a measure of the uncertainty associated with a received message and *mutual information*, a measure of the information shared between the transmitter and receiver at either end of the communication channel. Since then, both quantities have been applied to many problems outside of the field of communication, notably in image processing.

Mutual information is described as an information theory similarity metric.

**Algorithm 6.1**: Calculation of distance maps

```
Input: tm, ncols, nrows, loc
Output: $M_{DE}$, $M_{DES}$, $M_{DM}$
    ; tm - array of texture measures
    ; $M_{DE}$ - Euclidean distance map
    ; $M_{DES}$ - Standardised Euclidean distance map
    ; $M_{DM}$ - Mahalanobis distance map
DE = fltarr(ncols,nrows)
DES = fltarr(ncols,nrows)
DM = fltarr(ncols,nrows)
Ref = tm[*,nrows·ncols]    ; Extract reference texture measures
tm0 = STANDARDIZE(tm)    ; Standardise array
Ref0 = tm0[*,nrows·ncols]    ; Extract reference texture measures
C = COVARIANCE(tm)    ; Calculate covariance matrix
C1 = INVERT(C)    ; Invert covariance matrix
    ; Calculate distance maps
for k = 0 to nrows·ncols-1 do
    DE[k] =TOTAL(tm[*,k] -Ref[*])$^2$
    DES[k] =TOTAL(tm0[*,k] -Ref0[*])$^2$
    DM[k] =TOTAL((tm[*,k] -Ref[*])·C1· (tm[*,k]-Ref[*]))
endfor
$M_{DE}$ = SQRT(DE)
$M_{DES}$ = SQRT(DES)
$M_{DM}$ = SQRT(DM)
```

### 6.2.2 Mutual Information

Shannon [1948] introduced the concept of mutual information as a measure of the information content between the transmitter and receiver across an information channel. More generally mutual information can be interpreted as a measure of the information that two quantities have in common. Mutual information is defined as:

$$\begin{aligned} MI(X \cdot Y) &= H(X) - H(X|Y) && \text{(6-5)} \\ &= H(Y) - H(Y|X) && \text{(6-6)} \\ &= H(X) + H(Y) - H(X \cdot Y). && \text{(6-7)} \end{aligned}$$

where

$$H(X) = -\sum_{i=0}^{n-1} p(x_i) \log p(x_i). \tag{6-8}$$

is the entropy and

$$H(X|Y) = -\sum_{i=0}^{n-1} \sum_{j=0}^{n-1} p(x_i \cdot y_j) \log p(x_i|y_j) \tag{6-9}$$

is the conditional entropy. Entropy was introduced as a measure of the uncertainty associated with a set of probabilities and the conditional entropy $H(X|Y)$ is a measure of the uncertainty in $X$ given knowledge of $Y$ [Shannon 1948]. A detailed description of entropy is given in **Appendix C**.

**Equation 6-5** describes mutual information as the uncertainty of the source less the uncertainty of what was lost to noise given what was received, **Equation 6-6** describes mutual information as the uncertainty of what was received less the uncertainty of what was lost to noise given what was sent and **Equation 6-7** describes mutual information as the total uncertainty of what was sent and received less the joint uncertainty between what was sent and received. The latter is depicted in **Figure 6.1**.

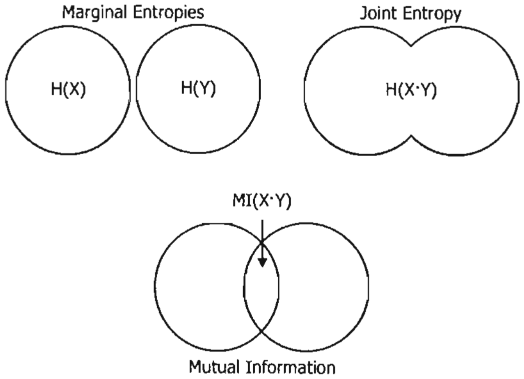


**Figure 6.1**: The mutual information between two quantities can be interpreted as the total uncertainty associated with each quantity less the joint uncertainty between each quantity, as is depicted. [Studholme et al. 1999]

Mutual information can also be defined as [Abramson 1963]:

$$MI(X\cdot Y)=\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}p(x_i\cdot y_j)\log\frac{p(x_i\cdot y_j)}{p(x_i)p(y_j)} \tag{6-10}$$

It can be shown that **Equation 6-10** is equivalent to **Equation 6-7**.

$$\begin{aligned}
MI(X\cdot Y) &= \sum_{i,j=0}^{n-1}p(x_i\cdot y_j)\log\frac{p(x_i\cdot y_j)}{p(x_i)p(y_j)}\\
&= \sum_{i,j=0}^{n-1}p(x_i\cdot y_j)\log p(x_i\cdot y_j)-\sum_{i,j=0}^{n-1}p(x_i\cdot y_j)\log p(x_i)-\sum_{i,j=0}^{n-1}p(x_i\cdot y_j)\log p(y_j)\\
&= -H(X\cdot Y)-\sum_{i=0}^{n-1}\left[\sum_{j=0}^{n-1}p(x_i\cdot y_j)\right]\log p(x_i)-\sum_{j=0}^{n-1}\left[\sum_{i=0}^{n-1}p(x_i\cdot y_j)\right]\log p(y_j)\\
&= -H(X\cdot Y)-\sum_{i=0}^{n-1}p(x_i)\log p(x_i)-\sum_{j=0}^{n-1}p(y_j)\log p(y_j)\\
&= -H(X\cdot Y)+H(X)+H(Y)
\end{aligned}$$

Similarly, by using **Equation 6-9**, it can be shown that **Equation 6-10** is equivalent to **Equations 6-5** and **6-6**.

The most obvious characteristic of mutual information is that it depends on *both* $X$ and $Y$. If $MI(X\cdot Y)$ is the information in $X$ about $Y$, then any information that $X$ contains about $Y$, $Y$ also contains about $X$. Therefore, $MI(X\cdot Y)=MI(Y\cdot X)$, which shows that mutual information is commutative [Abramson 1963].

The chain rule is used to determine the mutual information between more than two quantities:

$$MI(X_1 \cdot X_2 \cdot Y) = MI(X_1 \cdot Y) + MI(X_2 \cdot Y|X_1) \tag{6-11}$$

$$MI(X_1 \cdot X_2 \cdot X_3 \cdot Y) = MI(X_1 \cdot Y) + MI(X_2 \cdot Y|X_1) + MI(X_3 \cdot Y|X_1 \cdot X_2) \tag{6-12}$$

### 6.2.2.1 Mutual Information Algorithm

The algorithm for the calculation of mutual information is detailed in **Algorithm 6.2**.

**Algorithm 6.2**: Calculation of mutual information

```
Data: Probability density functions associated with random variables $X$ and $Y$, $P_X$ and $P_Y$ and the joint
    probability density function associated with random variables, $P_{X \cdot Y}$
Result: Mutual information between $X$ and $Y$, $MI$
    ; Check that dimensions of $P_X$, $P_Y$ and $P_{X \cdot Y}$ correspond
if dimensions of input matrices do not correspond then
    print, 'Dimensions of inputs are incompatible'
    break
endif
$MI = 0$
for $i=0$ to $l-1$ do
    for $j=0$ to $l-1$ do
        $MI = MI + P_{X \cdot Y}[i,j] \cdot \log\left(\frac{P_{X \cdot Y}[i,j]}{P_X[i] \cdot P_Y[j]}\right)$
    endfor
endfor
return $MI$
```

### 6.2.2.2 Properties of Mutual Information

The following properties of mutual information, *MI*, make it a reasonable similarity metric [Abramson 1963]:

1. $MI \geq 0$ unless the two quantities are completely independent then $MI=0$.
2. *MI* increases as dependency between two quantities increases.
3. *MI* is independent of the actual value of the probability.

Maes et al. [1997] summarised the following properties of mutual information:

$$\begin{array}{lll}
non-negativity: & MI(X \cdot Y) & \geq 0 \\
independence: & MI(X \cdot Y) & = 0 \Leftrightarrow P(X \cdot Y) = P(X) \cdot P(Y) \\
symmetry: & MI(X \cdot Y) & = MI(Y \cdot X) \\
self-information: & MI(X \cdot X) & = H(X) \\
boundedness: & MI(X \cdot Y) & \leq MI(H(X), H(Y)) \\
 & & \leq \frac{1}{2}(H(X) + H(Y)) \\
 & & \leq \max(H(X), H(Y)) \\
 & & \leq H(X \cdot Y) \\
 & & \leq H(X) + H(Y)
\end{array}$$

#### 6.2.2.3 Theoretical Maximum of Mutual Information

Consider two quantities $X$ and $Y$. The theoretical maximum of $MI(X \cdot Y)$ is given by $\ln n$ where $n$ is the number of states of $X$ and $Y$.

$$\begin{aligned} \max(MI(X \cdot Y)) &= \max(H(X)) + \max(H(Y)) - \max(H(X \cdot Y)) \\ &= \ln n + \ln n - \ln n \\ &= \ln n \end{aligned}$$

#### 6.2.2.4 Examples of Mutual Information

Consider three 2-bit images, $X$, $Y$ and $Z$.

$$X = \begin{matrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 2 & 2 & 2 \\ 2 & 2 & 3 & 3 \end{matrix} \quad Y = \begin{matrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 2 & 2 & 2 \\ 2 & 2 & 3 & 3 \end{matrix} \quad Z = \begin{matrix} 2 & 2 & 3 & 3 \\ 0 & 2 & 2 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{matrix}$$

Using grey-level histograms, the corresponding probability density functions are:

$$\begin{matrix} P(X) = \left[\frac{5}{16}, \frac{4}{16}, \frac{5}{16}, \frac{2}{16}\right] \\ P(Y) = \left[\frac{5}{16}, \frac{4}{16}, \frac{5}{16}, \frac{2}{16}\right] \\ P(Z) = \left[\frac{5}{16}, \frac{4}{16}, \frac{5}{16}, \frac{2}{16}\right] \end{matrix} \; P(X \cdot Y) = \begin{bmatrix} 0 & 0 & 0 & \frac{2}{16} \\ 0 & 0 & \frac{5}{16} & 0 \\ 0 & \frac{4}{16} & 0 & 0 \\ \frac{5}{16} & 0 & 0 & 0 \end{bmatrix} P(X \cdot Z) = \begin{bmatrix} 0 & \frac{2}{16} & 0 & 0 \\ \frac{3}{16} & \frac{2}{16} & 0 & 0 \\ 0 & 0 & \frac{2}{16} & \frac{2}{16} \\ \frac{2}{16} & 0 & \frac{3}{16} & 0 \end{bmatrix}$$

The mutual information between $X$ and $Y$ and $X$ and $Z$ is then:

$$MI(X \cdot Y) = 1.33$$

$$MI(X \cdot Z) = 0.740$$

It is clear that the mutual information for equal input probabilities (i.e. between $X$ and $Y$) is higher than the case when the input probabilities are not equal (i.e. between $X$ and $Z$). This example has a maximum mutual information of 1.33 when $X = Y$. The theoretical maximum of mutual information for a 2-bit system is $\ln(2^2) = 1.39$ which only applies when all states have equal probabilities, which is not the case for the example above.

### 6.2.3 Probability Density Estimation

The strength of using mutual information as a similarity metric lies in how the probability densities are estimated for the mutual information calculation. There are two methods of estimating a probability density function: parametric and non-parametric. Parametric methods are used when the general functional form of the density is known and the problem is one of determining the parameters that exactly describe the data at hand. For example, if the data has a normal distribution then all that is required to describe the distribution is to determine the amplitude, offset and standard deviation [Silverman 1986].

Nonparametric methods are used to estimate probability density functions when there is no known functional form. The following are examples of non-parametric methods of probability density function estimation [Silverman 1986]:

1. histograms
2. Naive estimators using Parzen windows
3. Kernel estimator
4. $k$-nearest neighbours
5. Adaptive kernel estimator

For images, there are no known functional forms so non-parametric methods of estimating probability density functions have to be used. Some examples of non-parametric image-based probability density estimates are histograms and grey-level co-occurrence matrices. For more information on the different methods of estimating probability functions see Silverman [1986].

### 6.2.4 The Use of Mutual Information as a Similarity Metric

In medical imaging, mutual information has been highly successful at registering images from different modalities and has also been used in the selection of features, stereo matching and model (template) matching.

#### 6.2.4.1 Registration of Images

In diagnostic medical imaging, each modality provides some information on the same patient's anatomy. Therefore there should be some information in common between the modalities. In general, the aim of any registration algorithm is to find a spatial transformation that results in two images being perfectly aligned. The registration of multi-modal medical images is an important aid to surgery. For example in microsurgery, the tumours are identified with MRI imaging while the stereotactic technology uses CT. It is then useful to transfer the co-ordinates of the tumour from the MRI image to the CT image [Wells et al. 1996].

Mutual information has been shown to be a robust similarity metric and has been successfully used to register multi-modal medical images [Wells et al. 1996, Maes et al. 1997, Wirth et al. 2002]. The mutual information for two images $X$ and $Y$ that have been aligned through some geometric transformation is maximal [Maes et al. 1997]. $X$ and $Y$ can describe the behaviour of any of the following properties related to images, but most of the methods used thus far involve grey-levels:

1. intensities or grey-levels
2. texture measures
3. scale
4. any shape signature

Hutton & Braun [2003] and Maintz & Viergever [1998*a*, 1998*b*] described some of the methods used to register images and Pluim et al. [2003] gave an overview of the use of mutual information to register medical images.

Viola & Wells [1995] and Gilles [1996] used Parzen windows with a Gaussian kernel to estimate the probability density function and mutual information to register medical images from different modalities. Gilles found that mutual information was more robust than normalised cross correlation.

Wells et al. [1996] used the maximisation of mutual information to register MRI, CT and PET images. Parzen windows were used to estimate the probability density functions. The mutual information was maximised by using a stochastic analogue of gradient descent. The technique described is quite general as it does not require segmentation or any assumptions about the nature of the signals used.

Wirth et al. [2002] used thin plate splines together with mutual information to register mammograms for bilateral and temporal comparison. The probability density functions were estimated from grey-level histograms. Results were described as ‘promising’, but the changeable nature of the breast under compression led to some problems.

#### 6.2.4.2 Feature Selection

Feature selection is the extraction of a reduced set of features to reduce the dimensionality of any classification problem.

Fisher & Principe [1998] and Tourassi et al. [2001] used mutual information to optimise the features extracted from medical images. The reduced set of features was then fed into a classification scheme (e.g. artificial neural networks).

#### 6.2.4.3 Segmentation

Tsai et al. [2004] used shape-based active contours coupled with mutual information to segment pelvic MRI images. No quantitative evaluation of results was presented and results were described as “performing very well in segmenting the anatomical regions of interest.” Parzen density estimation was used to estimate the probability density function.

#### 6.2.4.4 Template/Object Matching

Mutual information has been used as similarity metric for template matching [Sista et al. 1995, Egnal 2000, Shams et al. 2001]. Tourassi et al. [2003] used mutual information as a similarity metric for template matching in a knowledge-based mammographic CAD-system for discrimination of masses from normal tissue. Histograms were used to determine the probability density functions. An $A_z$=0.88±0.01 was achieved for discrimination between malignant and normal tissue and an $A_z$=0.86±0.01 was achieved for discrimination between benign and normal tissue.

Filev et al. [2005] compared the effectiveness of twelve similarity metrics in matching the correspondence between masses in temporal mammograms. The size of the search region was varied and the average template size was 17 mm×17 mm. The similarity metrics were:

1. Pearson's correlation coefficient
2. cosine coefficient
3. Goodman and Kruskal's gamma coefficient
4. Pearson's correlation coefficient standardised by median
5. mutual information (scaled) with histogram values linearly scaled between minimum and maximum of the grey-level
6. mutual information (unscaled) with histogram computed from original pixel values
7. ordinal measure
8. increment sign correlation coefficient
9. pattern intensity
10. rank transform
11. extended Jaccard transform
12. gradient difference

Pearson's correlation coefficient, the cosine coefficient and Goodman and Kruskal's gamma coefficient performed best and were the most robust. While mutual information was robust, it was $6^{th}$ best, and the scaled mutual information consistently performed better than the unscaled mutual information. Also, mutual information did not perform well for small template sizes.

## 6.3 Comparison of Similarity Metrics

Distance metrics are simple to compute while the probability density functions required by mutual information can be time-consuming to compute. One significant advantage of using distance similarity metrics and mutual information for classification is that no training is required unlike other methods of classification e.g. artificial neural networks or linear discriminant analysis.

## 6.4 Summary

Similarity metrics quantify how similar two quantities are to each other. The similarity metrics that are used in this research are: Euclidean distance, Mahalanobis distance and mutual information. The Euclidean distance similarity metric suffers from the problem that a single input can dominate the final distance value, if this input is considerably larger than the other inputs. Standardising the inputs to a normal distribution with a variance of 1 and a mean of 0 solves this problem. However, standardisation does not consider correlations between inputs. Hence, the Mahalanobis distance similarity metric is often preferred.

Mutual information has been shown to be a robust similarity metric in image registration problems, but has also been applied to template matching, feature selection and segmentation problems.

# Chapter 7

# Evaluation of Results

Consider a medical diagnostic test to determine the presence of a disease. The results of the diagnostic test are analysed by a physician who classifies the results and delivers a diagnosis as to whether the disease is present or not.

Mammography is a radiological diagnostic test and a radiologist interprets the resulting mammogram and delivers a decision that summarises the mammogram as normal or abnormal. This task is simplified when there is no uncertainty in the image characteristics of the mammogram and the radiologist can state with certainty that the mammogram is normal or abnormal. However, when there is some overlap between the image characteristics of normal and abnormal breasts, then the radiologist may shift his decision threshold to err on the side of caution and classify the mammogram as abnormal. This type of decision would most likely be followed by further diagnostic tests which would be used to deliver a more certain diagnosis.

A mammographic CAD-system analysing a mammogram fills the same role as the radiologist in classifying the mammogram, and the methods used to evaluate the accuracy of medical diagnostic tests can therefore be applied to evaluate the accuracy of CAD-algorithms.

For this study, the accuracy of the algorithms is evaluated on two levels. Firstly, the area of the matched region in the matching map is compared with the region marked by a radiologist in the original test image, by using the area under the receiver operating characteristic curve, $A_{ROC}$ [Bushberg et al. 2002]. Secondly, the matched region in the matching map is compared with surrounding regions, to determine contrast, $C_{fb}$.

## 7.1 Evaluation Terminology

The results of any diagnostic test must be compared to the truth to evaluate the accuracy of the test. The terminology used to describe how the diagnosis relates to the truth is summarised in **Table 7.1** and forms the basis of determining the accuracy of any diagnostic test [Bushberg et al. 2002].

**Table 7.1**: Terminology used during evaluation to describe how the diagnosis relates to the ground truth data.

| | Actually Abnormal | Actually Normal |
|---|---|---|
| Diagnosed as abnormal | True-Positive (*TP*) | False-Positive (*FP*) |
| Diagnosed as normal | False-Negative (*FN*) | True-Negative (*TN*) |

Two quantities are required to evaluate the performance of a diagnostic test, the true-positive fraction and the false-positive fraction [Bushberg et al. 2002].

The true-positive fraction, *TPF*, (also known as sensitivity) is the fraction of abnormal cases that is actually diagnosed as abnormal and is defined in **Equation 7-1**.

$$TPF = \frac{TP}{TP+FN} \tag{7-1}$$

The false-positive fraction, *FPF*, is the fraction of normal cases diagnosed as abnormal and is defined in **Equation 7-2**.

$$FPF = \frac{FP}{FP+TN} \tag{7-2}$$

The true-negative fraction (*TNF*) (also known as specificity) is the fraction of normal cases that is actually diagnosed as normal and is defined in **Equation 7-3**.

$$TNF = \frac{TN}{FP+TN} \tag{7-3}$$

The perfect diagnostic test has *TPF*=1, *FPF*=0 and *TNF*=1 for all decision thresholds. Diagnostic tests are often described in terms of their sensitivity and specificity, but this has the disadvantage of examining the decision space at a single decision threshold. The effect of decision threshold on sensitivity and specificity is illustrated in **Figure 7.1** where the histograms of the results of two hypothetical medical diagnostic tests, X and Y, are shown.

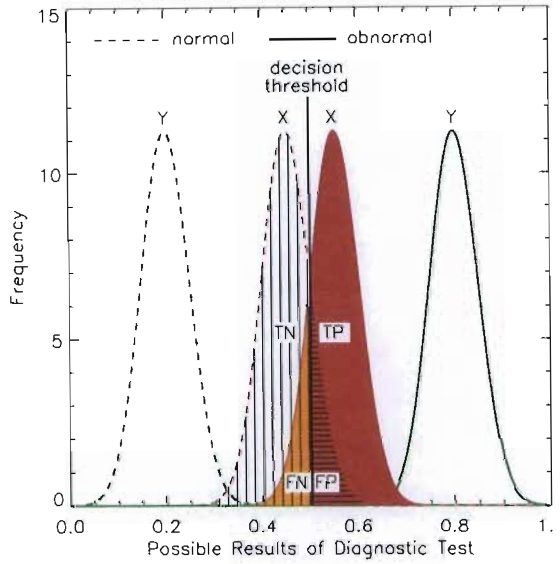


**Figure 7.1**: Effect of decision threshold on sensitivity and specificity. The histograms of the results of two hypothetical medical diagnostic tests, X and Y, to test for the presence of a disease, are shown. The solid line indicates positive results (i.e. that the disease is present) and the dashed line indicates negative results (i.e. that the disease is not present). (Adapted from Bushberg et al. [2002])

Test Y can be considered to be the perfect test as results for patients having the disease (solid line) and patients not having the disease (dashed line) are well separated. If the decision threshold is set at 0.5, then test Y will have a sensitivity of 100% and a specificity of 100%. Test X is more realistic, with the results for diseased and undiseased patients overlapping. In this case, any decision threshold will lead to some undiseased patients being classified as diseased and vice-versa. If a low decision threshold is selected, the number of false-negative results decreases (higher sensitivity), but the number of false-positive results increases (lower specificity). Selecting a high threshold increases the number of false-negative results (lower sensitivity) and decreases the number of false-positive results (higher specificity) [van Erkel & Pattynama 1998, Bushberg et al. 2002].

The dependence of sensitivity and specificity on the decision threshold makes it difficult to compare the discriminating power of diagnostic tests. Fortunately, there is a method of evaluating the accuracy of diagnostic tests that is independent of the decision threshold: receiver operating characteristic analysis.

## 7.2 Receiver Operating Characteristic (ROC) Analysis

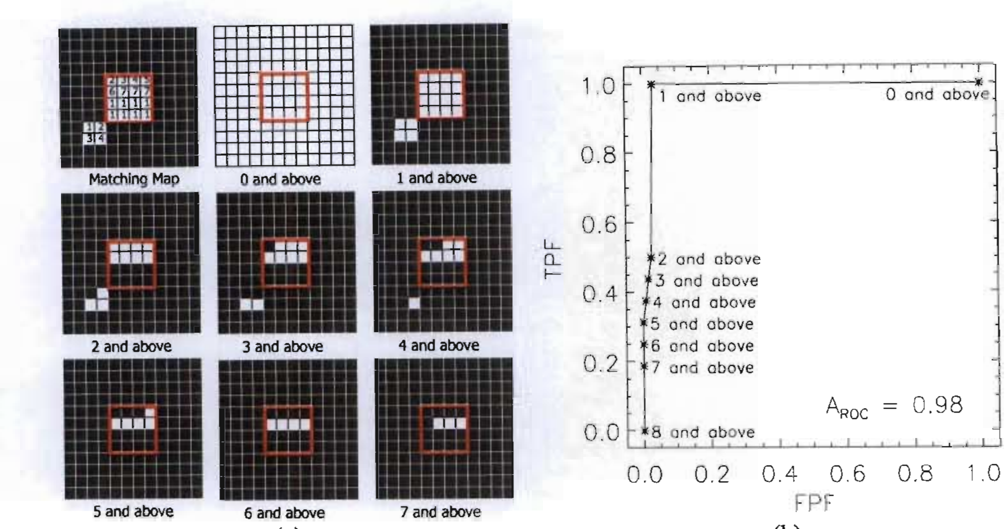
Receiver operating characteristic (ROC) analysis originated in signal theory as a model of how well a receiver is able to detect a signal in the presence of noise, by using the distinction between hit rate (or *TPF*) and false alarm rate (or *FPF*) as two separate performance measures. ROC analysis is now widely used in medical data analysis to study the effect of a decision threshold on the overall accuracy of a diagnostic test and has been widely accepted as the most effective method of quantifying the performance of a diagnostic test (e.g. a CAD algorithm). The keystone of ROC analysis is the ROC curve - a plot of *TPF vs. FPF* at various decision thresholds, which describes the balance between sensitivity and specificity [Bradley 1997, Liu et al. 2004]. ROC analysis further allows for different tests to be compared and ranked, using the area under the ROC curve as a measure of how discriminating a test is.

## 7.3 Example: Determination of the ROC Curve

A simple example to determine the ROC curve is detailed. Consider the matching map shown at the top left in **Figure 7.2**. The ground truth region is enclosed by the red square. The matching map is a 3-bit map (i.e. 8 grey-levels). **Table 7.2** shows details of the various steps of the determination of the ROC curve. The various steps are also shown graphically in **Figure 7.2(a)** with the ROC-curve shown in **Figure 7.2(b)**.

**Table 7.2**: Example: calculation of values for the ROC curve for matching map in **Figure 7.2(a)**.

| Notes | Grey-level | *TP* | *FP* | *TN* | *FN* | *TPF* | *FPF* |
|---|---|---|---|---|---|---|---|
| Start below minimum | 0 and above | 16 | 128 | 0 | 0 | 1 | 1 |
| | 1 and above | 16 | 4 | 124 | 0 | 1 | 4/128 |
| | 2 and above | 8 | 3 | 125 | 8 | 8/16 | 3/128 |
| | 3 and above | 7 | 2 | 126 | 9 | 7/16 | 2/128 |
| | 4 and above | 6 | 1 | 127 | 10 | 6/16 | 1/128 |
| | 5 and above | 5 | 0 | 128 | 11 | 5/16 | 0 |
| | 6 and above | 4 | 0 | 128 | 12 | 4/16 | 0 |
| | 7 and above | 3 | 0 | 128 | 13 | 3/16 | 0 |
| Stop above maximum | 8 and above | 0 | 0 | 128 | 16 | 0 | 0 |



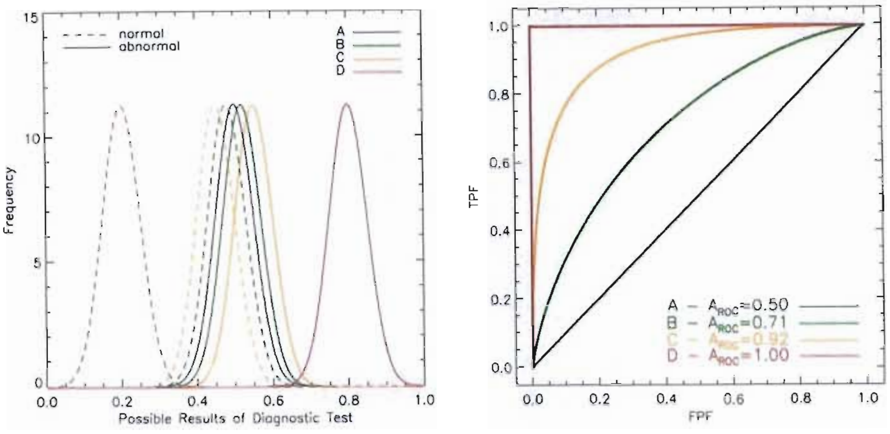
(a) (b)

**Figure 7.2**: Example demonstrating the calculation of values for the ROC curve. **(a)** Detailed matching maps, at each decision threshold, used to calculate *FP* and *TP* for the ROC curve. **(b)** ROC curve.

## 7.4 Area under the ROC Curve, $A_{ROC}$

An entire ROC curve is needed to describe the performance of a diagnostic test and the area under the ROC curve has been proven to be a powerful index for assessing the performance of a diagnostic test [Bradley 1997]. It provides a useful summary of the performance of a test and can be interpreted as the average value of sensitivity over all specificities or vice versa [Liu et al. 2004], with a larger $A_{ROC}$ indicating that a higher *TPF* (and correspondingly a smaller *FPF*) has been achieved. The area under the curve varies between 0.5 (corresponding to a random test) and 1.0 (corresponding to a perfect test) [van Erkel & Pattynama 1998, Bushberg et al. 2002, Horng et al. 2002, Van Schalkwyk 2003, Fawcett 2004, Flach 2004]. Values for the area between 0 and 0.5 correspond to a negative correlation between the test results and the truth. In terms of a medical diagnostic test, this means that if a truth value of $x$ corresponds to the disease being present, the test indicates that for a value of $x$ that the disease is not present.

Consider four medical diagnostic tests to test for the presence of a disease, to illustrate the use of ROC curves to compare diagnostic tests. The histogram of the results of the various tests are shown in **Figure 7.3** as A, B, C and D. The part of each distribution indicated by the solid line represents positive test results while the dashed line indicates negative test results. **Figure 7.3(a)** shows that each test has a varying degree of overlap between positive and negative results. The ROC curves for each test are shown in **Figure 7.3(b)**. Test A has an area under the ROC curve of 0.5, which is expected since there was no separation between the results of the test being negative and positive. For Test B, which has a slight separation between negative and positive results, the area under the ROC curve is slightly higher at 0.71. For Test C, the area under the ROC curve is 0.92 and for diagnostic test D, which has the best separation, $A_{ROC}$=1.0 [van Erkel & Pattynama 1998, Bushberg et al. 2002].



(a) (b)

**Figure 7.3**: Comparison of four diagnostic tests using ROC analysis. **(a)** Histogram of the results of four tests. The solid line indicates the positive results and the dashed line indicates the negative results. **(b)** The ROC curves are obtained by plotting *TPF* and *FPF* at various decision thresholds. The area under each curve is used as an indication of how discriminating each test is, with an area of 0.5 being no better than random. (Adapted from Bushberg et al. [2002])

Greiner et al. [2000] categorised ranges of $A_{ROC}$ as follows:

1. non-informative ($A_{ROC}$=0.5)
2. less accurate ($0.5 < A_{ROC} \leq 0.7$)
3. moderately accurate ($0.7 < A_{ROC} \leq 0.9$)
4. highly accurate ($0.9 < A_{ROC} < 1$)
5. perfect ($A_{ROC}$=1)

## 7.5 Calculating the Area under the ROC Curve

The area under the ROC curve can be calculated using two methods. The first is to use basic integral calculus methods to directly calculate the area and the second is to fit a model to the data and calculate the area under the fitted curve, analytically.

### 7.5.1 Trapezoidal Rule

If there are $n$ points on the ROC curve, $\alpha$=*TPF*and $\beta$=*TNF*, then 1-$\beta$=*FPF*, then $A_{ROC}$ can be calculated using the trapezoidal rule for integration.

$$A_{ROC} = \sum_{i=0}^{n} (1-\beta_i)\cdot\Delta\alpha + 0.5\cdot[\Delta(1-\beta)\cdot\Delta\alpha] \tag{7-4}$$

where $\Delta(1-\beta) = (1-\beta_i) - (1-\beta_{i-1})$ and $\Delta\alpha = \alpha_i - \alpha_{i-1}$.

### 7.5.2 Binormal Model of ROC Analysis

A successful model for fitting data to ROC curves is the binormal model, which assumes that the decision variable can be transformed to a pair of normal distributions with [Liu et al. 2004]:

$$f(x|negative) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{x^2}{2}) \tag{7-5}$$

for true-negative cases and

$$f(x|positive) = \frac{b}{\sqrt{2\pi}} \exp\left(-\frac{(bx-a)^2}{2}\right) \tag{7-6}$$

for true-positive cases. Following Liu et al. [2004], let $N(\mu, \sigma)$ represent a normal distribution with mean $\mu$ and standard deviation $\sigma$. Then **Equation 7-5** becomes $N(0,1)$ and **Equation 7-6** becomes $N(a/b, 1/b)$. Using the binormal model, the *TPF*, *FPF*, and $A_{ROC}$ can easily be computed. If $x_c$ is a critical value then a case is diagnosed as positive if and only if $x > x_c$. *TPF* is then given by

$$TPF(x_c) = \int_{x_c}^{\infty} \frac{b}{\sqrt{2\pi}} \exp\left(-\frac{(bx-a)^2}{2}\right) dx = \Phi(a - bx_c) \tag{7-7}$$

and *FPF* is given by:

$$FPF(x_c) = \int_{x_c}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = \Phi(-x_c) \tag{7-8}$$

where $\Phi(z)$ is the cumulative standard normal distribution function. As $x_c$ varies from $-\infty$ to $\infty$, $TPF(x_c)$ and $FPF(x_c)$ sweeps out the ROC curve. The area under this binormal ROC curve, denoted by $A_z$ (with $z$ indicating the use of the binormal model) can be expressed as:

$$A_z = \Phi\left(\frac{a}{\sqrt{1+b^2}}\right) \tag{7-9}$$

Liu et al. [2004] gives full details of the derivation of **Equation 7-9**.

## 7.6 ROC Analysis in Mammographic CAD

ROC analysis has been widely used in the evaluation of mammographic CAD algorithms, for example, Miller & Astley [1992], Chan et al. [1995], Wei et al. [1995], Rangayyan et al. [1997*b*], Wei et al. [1995], Chan et al. [1999], Huo et al. [1999], Chang et al. [2001], Mudigonda et al. [2001], Qian et al. [2001], Tourassi et al. [2001, 2003].

## 7.7 Contrast, $C_{fb}$

Contrast is the local change in brightness and is defined as the ratio of the average brightness of an object to the average brightness of the background (**Equation 7-10**) [Sonka et al. 1999].

$$C_{fb} = \frac{f-b}{f+b} \tag{7-10}$$

$f$ is the average grey-level of the foreground, $b$ is the average grey-level of the background and $-1 \le C_{fb} \le 1$. Negative values arise when the foreground is darker than the background and positive values arise when the foreground is brighter than the background. A contrast of 0 means that the object cannot be seen against the background.

For this study, $f$ and $b$ are determined from the two regions indicated in **Figure 7.4**. The centroid and maximum extent of the ROI ($d_{max}$) from the centroid are used to define three circles of radii $d_{max}$, $1.05d_{max}$ and $1.10d_{max}$. The average of the grey levels within the smallest circle defines $f$ and the average of the grey-levels in the region between the two outer circles defines $b$.

**Figure 7.5** shows some examples of contrast values. Generally, contrasts above 0.5 indicate that the foreground (white square) can be easily seen against the background.

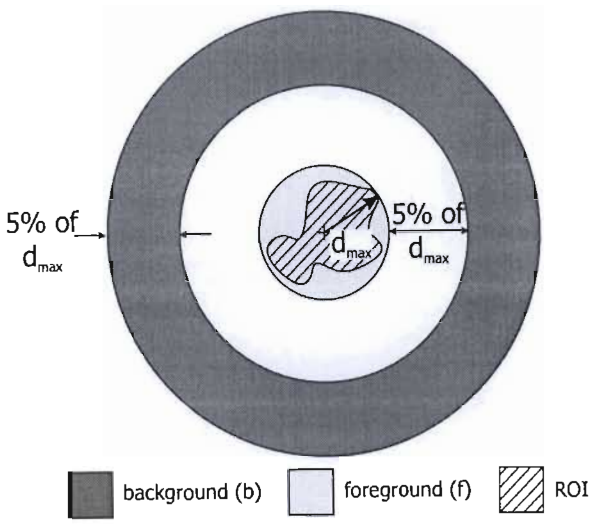


**Figure 7.4**: Schematic (not to scale) showing which regions are used to determine the foreground ($f$) and background ($b$) values for the contrast calculation. The grey-levels within the shaded regions are averaged to obtain values for $f$ and $b$. $d_{max}$ represents the maximum extent, from the centroid, of the ROI.

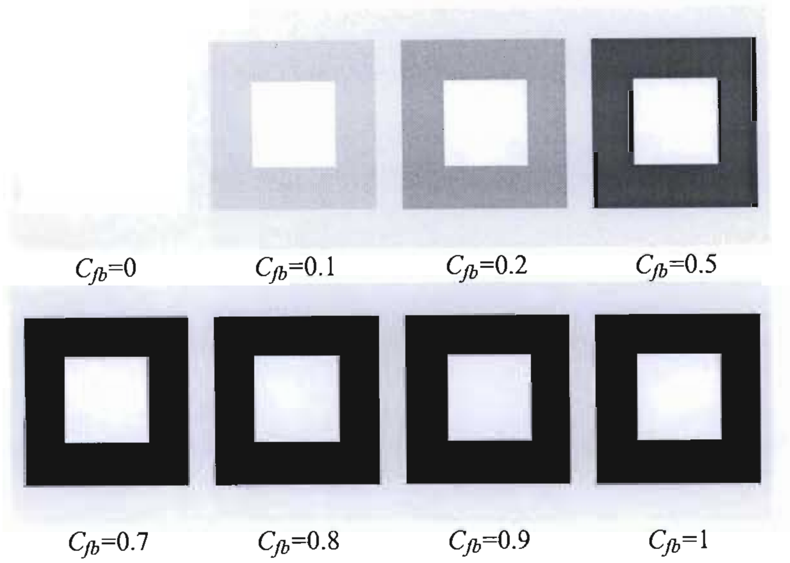


**Figure 7.5**: Images showing examples of contrast. Generally, contrasts above 0.5 indicate that the foreground (white square) can be easily seen against the background.

## 7.8 Example: Evaluation of Matching

**Figure 7.6** shows an example to illustrate the evaluation methodology. The reference ROI was matched to the test image by using a mutual information similarity metric with GLCMs. The matching map (**Figure 7.6(d)**), generated for a sampling window size of 80 pixels, at a bit-depth of 8 bits and for $d$=1, shows that the match is generally quite good. All three instances of the reference region are well matched, with high intensities in the matching map. The right-most reference region, which has been rotated by 90°, is not as well matched as the two, unrotated, reference regions, and shows that the matching algorithm is not invariant to rotation. There are also matches to other textures, which appear to be similar to the reference texture at the scale of the reference ROI of 80 pixels. The detected regions in the matching map have a contrast of $C_{fb}$=0.63.

The ROC extracted from the matching map (**Figure 7.6(e)**) has $A_{ROC}$=0.97. This is, therefore, an example of a good match with high $A_{ROC}$ and a reasonably high $C_{fb}$.

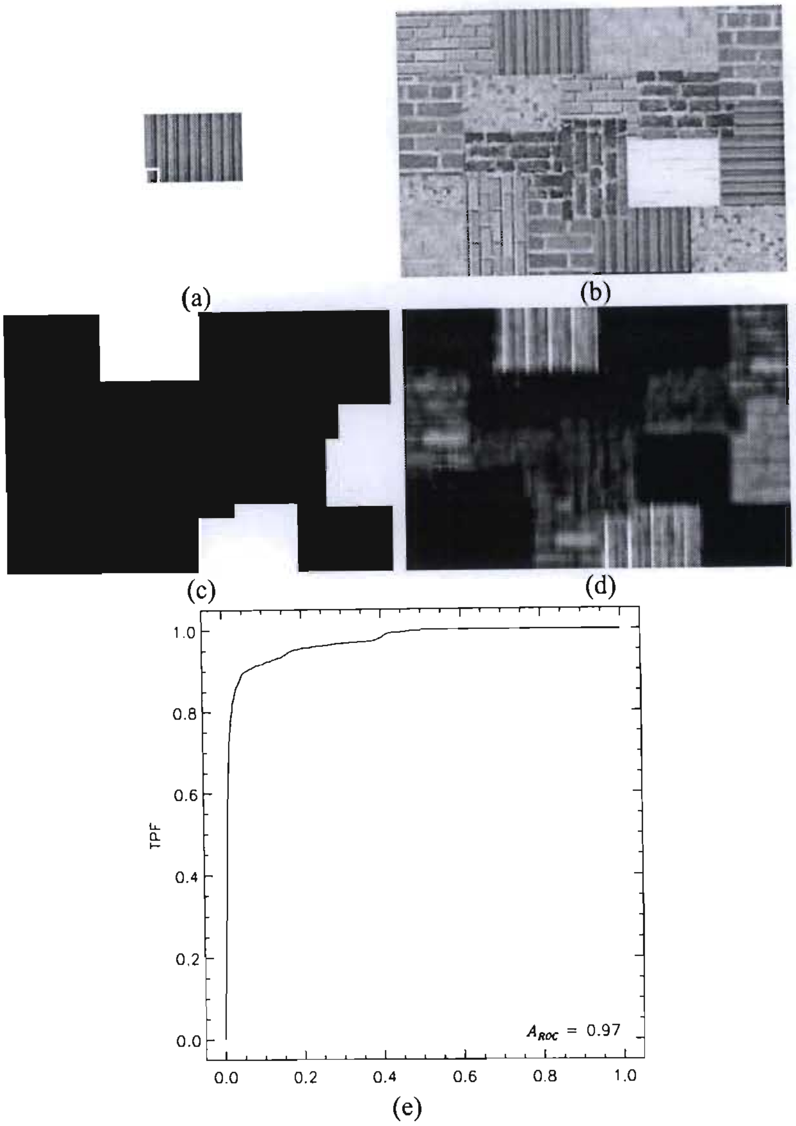


**Figure 7.6**: Example to illustrate the evaluation algorithm **(a)** Reference image with reference ROI of 80 pixels marked by white square in bottom left corner. **(b)** Test image. **(c)** Ground truth image. **(d)** Matching map at a sampling window size of 80 pixels, a step size of 20 pixels, a bit-depth of 8-bits and $d$=1. The contrast of the detected regions compared to the background of the matching map is 0.63. **(e)** ROC curve with $A_{ROC}$=0.97.

## 7.9 Summary

ROC analysis is a standard method of evaluating and ranking medical diagnostic tests. To perform any evaluation, the 'truth' must be known so that it can be compared with the output of the test. The evaluation and ranking of the CAD algorithms developed here is analogous to that of a standard medical diagnostic test and is therefore perfectly suited to the use of ROC analysis. For this study, the accuracy of the matching maps is evaluated by using ROC analysis and contrast. The area under the ROC curve gives a measure of how much of the ROI has been matched while contrast gives a measure of how well the matched ROI stands out in the matching map.

# Chapter 8

# Matching Methods

The aim of this study is to investigate whether using texture analysis methods with suitable similarity metrics allows a suspicious feature from one mammographic view to be matched with the same suspicious feature in other mammographic views of the same breast. Two texture analysis methods are investigated in this research: texture measure matching (hereafter referred to as **TM-matching**) and mutual information matching (hereafter referred to as **MI-matching**). The exact details of these methods are described in this chapter and are based on the theory described in **Chapters 4** to **7**.

The method, shown schematically in **Figure 3.5** (page 42), is as follows: The location of the reference ROI is used to define an annular search region in the test image. Textural characteristics of the reference ROI are compared to textural characteristics of equally sized sub-images from the annular region in the test image. The result of this comparison process is a map of similarity called a matching map, with the brighter regions on the matching map corresponding to a greater similarity between those regions in the test image and the reference ROI.

## 8.1 Effect of Algorithm Parameters on Matching Accuracy

The effect of various *matching parameters* (sampling window size, sampling window step size, *d* in the GLCM, bit-depth and number of bins in the grey-level histogram) is investigated for each of the matching methods. The matching parameters are depicted in **Figure 8.1**. The choice of the values of the parameters used in this study is based on the many mammographic CAD studies using texture analysis, some of which are listed in **Table 8.1**.

### 8.1.1 Sampling Window Size, $w$

A sampling window of size $M \times N$ placed at position $[x, y]$ in a large image extracts an $M \times N$ sub-image from that large image. For texture analysis, sampling windows are needed because texture is described as a variation of grey-level from pixel to pixel or region to region. Therefore individual pixels cannot be used for texture analysis as, individually, they do not contain non-local information. The sampling window must be large enough for the extracted sub-image to exhibit similar characteristics to those of the region in the image from where it was extracted. At the same time, the sampling window must be as small as possible to enable accurate detection of borders between neighbouring textural regions [Wang et al. 1996].

Sampling windows are defined by their size and position. In mammographic texture analysis, square sampling windows centred on a feature of interest, are often used with window

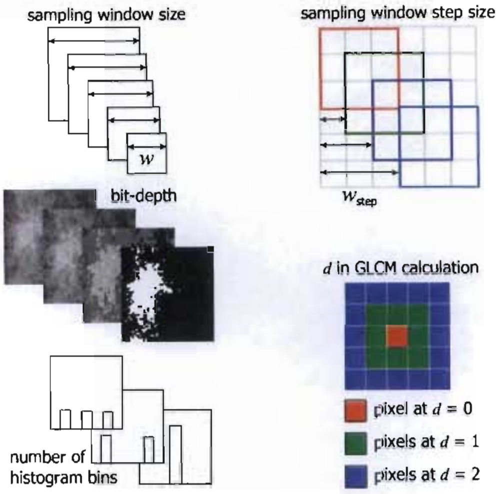


**Figure 8.1**: Parameters affecting matching accuracy. All examples (except sampling window step size) increase from bottom to top. Sampling window step size increases from top to bottom.

**Table 8.1**: Summary of a few studies and the matching parameter values that have been used for mammographic textural analysis.

| Reference | Spatial Resolution ($\mu$m/pixel) | Bit-depth (bits) | $d$ (pixels) | $w$ (pixels) |
|---|---|---|---|---|
| Yin et al. [1991] | 400 | 8 | – | – |
| Kegelmeyer et al. [1994] | 280 | 10 | – | – |
| Laine et al. [1994] | 200 | 10 | – | – |
| Suckling et al. [1995] | 50 | 8 | – | 16,64 |
| Tahoces et al. [1995] | 350 | 10 | – | – |
| Wei et al. [1995] | 100 | 8 | 1–16 | 256 |
| Petrick et al. [1998] | 100 | 12 | 1–48 | 256 |
| Rangayyan et al. [1997*a*] | 50 | 8 | – | – |
| Rangayyan et al. [1997*b*] | 124 | 10 | – | – |
| Karssemeijer [1998] | 200 | 12 | – | – |
| Karssemeijer [1998] | 50 | 8 | 1 | – |
| Sahiner et al. [1998*a*] | 100 | 12 | 1–16 | 256 |
| Chan et al. [1999] | 100 | 12 | 1–20 | – |
| Huo et al. [1999] | 100 | 10 | – | – |
| Bovis et al. [2000] | – | 8 | 1–9 | – |
| Mudigonda et al. [2000] | 50 | 8 | 1–10 | – |
| te Brake et al. [2000] | 200 | 12 | – | – |
| Chang et al. [2001] | 400 | 12 | – | 125 |
| Li et al. [2002] | 100 | 14 | – | – |
| Timp & Karssemeijer [2004] | 200 | 12 | – | – |
| Filev et al. [2005] | 800 | 12 | – | 21–61 with $w_{step}$=1 |

sizes ranging between 0.8 mm and 50 mm [Suckling et al. 1995, Petrick et al. 1998, Petrick et al. 1999, Chang et al. 2001].

For this study square sampling windows with sizes ranging from 4.06 mm (4 pixels) to 143.2 mm (400 pixels) were used. For the mammograms, the size of the suspicious feature determined the upper bound to the size of the sampling window.

### 8.1.2 Sampling Window Step Size, $w_{step}$

The sampling window step size defines the amount by which the sampling window is incrementally moved and determines how well boundaries between textures are resolved. If $w_{step}$ is too large then boundaries are not resolved, while if $w_{step}$ is too small then unnecessary time is spent on computation. In mammographic texture analysis most researchers use single sampling windows centred on an object of interest to compute textural features and in the single study where moving sampling windows were used, the sampling window step size was 0.8 mm corresponding to 1 pixel [Filev et al. 2005].

For this study, $w_{step}$ was varied from 4 pixels to 100 pixels for the mosaic images and was fixed at 1.02 mm (4 pixels) for the two sets of mammograms.

### 8.1.3 Distance, *d*, in the GLCM

The distance, *d*, used in the GLCM calculation defines the scale at which the texture in the image is analysed. If *d* is small relative to the texture coarseness, the GLCM values cluster around the diagonal while for large *d* the values are more spread out. In practice, small values of *d* yield the best results [Wang et al. 1996]. In mammographic texture analysis, *d*-values have varied from 0.05 mm [Mudigonda et al. 2000] to 2 mm [Chang et al. 1999].

For this study, *d* was varied from 1 pixel to 10 pixels for the mosaic images and between 0.258 mm (1 pixel) and 2.58 mm (10 pixels)for the mammograms and stereotactic biopsy mammograms.

### 8.1.4 Bit-depth of images, *nbits*

The bit-depth of images controls the number of shades of grey that are available to describe the information in an image (**Appendix B**). The more shades of grey available, the more detail can be depicted in the image. However, if the bit-depth of an image is reduced, then the image features are effectively smoothed since detail is lost.

In mammographic texture analysis, fixed values of either 8 bits, 10 bits, 12 bits or 14 bits have been used [Yin et al. 1991, Suckling et al. 1995, Rangayyan et al. 1997*a*, Karssemeijer 1998, Bovis et al. 2000, Mudigonda et al. 2000, Laine et al. 1994, Kegelmeyer et al. 1994, Tahoces et al. 1995, Rangayyan et al. 1997*b*, Huo et al. 1999, Chan et al. 1999, Petrick et al. 1998, Karssemeijer 1998, Petrick et al. 1999, te Brake et al. 2000, Chang et al. 2001, Timp & Karssemeijer 2004, Li et al. 2002].

For this study bit-depth was varied from 5 bits to 8 bits.

### 8.1.5 Number of Histogram Bins, *nbins*

The probability density function estimated by a grey-level histogram is dependent on the number of histogram bins. Like bit-depth, the number of histogram bins controls the effective smoothing in the resulting probability density function. If there are too many bins then each bin is sparsely populated and the histogram is not continuous. However, if there are too few bins then the histogram is too smooth and features are lost.

Tourassi et al [2001,2003] investigated the effect of the number of histogram bins on mammographic texture analysis. Histograms with 64 bins, 128 bins and 256 bins were used.

For this study, 16 bins, 32 bins, 64 bins and 128 bins were used.

### 8.1.6 Ranges of Values for Matching Parameters

The spatial resolutions for the mammograms and stereotactic biopsy mammograms listed in **Table 8.2** are used to convert the pixel-values of the matching parameters to physical sizes. The ranges of the values used for the matching parameters are summarised in **Table 8.3**. Distances and widths are given in pixels (as used in the calculations).

**Table 8.2**: Summary of spatial resolutions and bit-depths for the three sets of images used in this study. The spatial resolution of the mosaic images is not listed since the spatial resolution of the mosaic images is not linked to the physical sizes of the features in these images.

| Image Set | Spatial Resolution (mm per pixel) | | Bit-depth (bits) | |
|---|---|---|---|---|
| | Original | Analysis | Original | Analysis |
| Mosaics | – | – | 8 | 5 to 8 |
| Mammograms | 0.050 | 0.254 | 10 | 5 to 8 |
| Stereotactic Biopsy Mammograms | 0.048 | 0.254 | 12 | 5 to 8 |

**Table 8.3**: Ranges of values for the matching parameters used in this study. Physical sizes are obtained by using the spatial resolutions given in **Table 8.2**.

| Parameter | Minimum | Maximum | Interval |
|---|---|---|---|
| $w$ (pixels) | 16 | $\leq 400^{a}$ | 16 |
| $w_{step}$ (pixels) [b] | 4 | $\leq 400$ | 25%-100% of $w$ |
| *nbits* (bits) | 5 | 8 | 1 |
| $d$ (pixels) | - | - | [1,2,5,10] |
| *nbins* (pixels) | $2^4$ | $2^7$ | factor 2 |

[a] The maximum of 400 pixels refers to that for the mosaic images. The maximum for the mammograms and stereotactic biopsy mammograms is obtained from the maximum extent of the ROI.

[b] $w_{step}$ was varied by a fraction of $w$ for the mosaic images. For the mammogram images, $w_{step}$ was fixed with $w_{step} = 4$ pixels.

## 8.2 Formats of Results

**Figure 8.2** shows some examples of the formats used to present the evaluation results. **Figure 8.2(a)** is a scatter plot of the area under the ROC curve *vs.* contrast to examine the general behaviour of a particular matching method. Points with $A_{ROC}>0.75$ and $C_{fb}>0.75$ are preferred since these indicate good matches.

**Figure 8.2(b)** shows examples of two plots with $\kappa$ *vs.* $w$ for different values of *nbits* (left) and $d$ (right) to examine how matching accuracy varies across different values of $w$, for different values of *nbits* or *d*. These plots are generated for each image pair. Dependence on matching accuracy is indicated by a separation of the different coloured plots.

**Figure 8.2(c)** shows example plots of κ averaged over all images as a function of *d* (left) and *nbits* (right). These plots give insight to the overall behaviour of the matching parameters *nbits* and *d* as sampling window size is varied and are generated for each sampling window size. Dependence on matching accuracy is indicated by a separation of the different coloured plots [Chan et al. 1995].

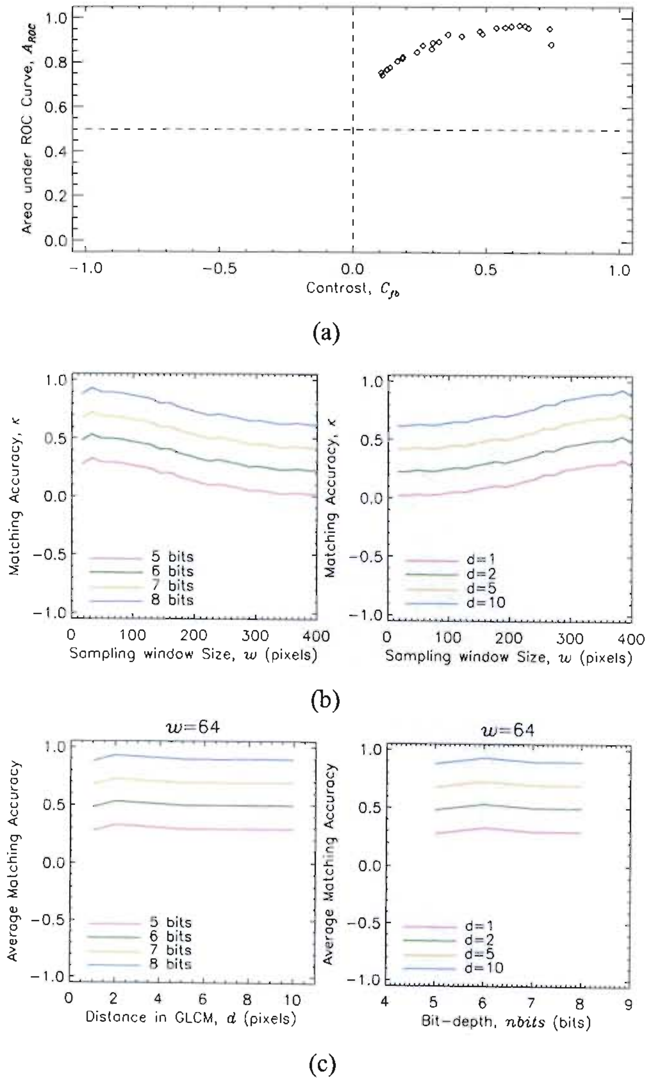


**Figure 8.2**: Example plots of the different formats used to present evaluation data. Each plot reveals a specific pattern of behaviour for each of the parameters being investigated. **(a)** Scatter plot of the area under the ROC curve *vs.* contrast to examine the general behaviour of a particular matching method. **(b)** Examples of κ *vs.* *w* for different values of *nbits* (left) and *d* (right) to examine how matching accuracy varies across different values of *w*. **(c)** Examples of κ averaged over all images as a function of *d* (left) and *nbits* (right). These plots give insight to the overall behaviour of the matching parameters *nbits* and *d* as *w* is varied.

## 8.3 Texture Measure Matching (TM-Matching)

In TM-matching, the image textural information is quantified by texture measures extracted from GLCMs (§5.5.2 on page 68). These texture measures are computed for the reference and test ROIs, and are compared using three distance similarity metrics: Euclidean distance, standardised Euclidean distance, Mahalanobis distance (§6.1 on page 73).

The GLCMs are computed at four angles (θ=[0°,45°,90°,135°]) and then averaged to remove any directional effects that may be introduced by the change in breast compression between mammographic views. Malignant masses are also usually homogeneous with regards to texture [Chan et al. 1995], so averaging the GLCMs is ideally suited to detecting homogeneous textures. The following thirteen texture measures are calculated from the averaged GLCMs at different values of *d*:

maximum probability
inertia
sum average
sum variance
information measure of correlation 1
entropy
inverse difference moment
sum entropy
difference average
energy
correlation
difference entropy
difference variance

The resulting 13-dimensional texture measure vectors are then used as the inputs to the distance similarity metrics, resulting in distance maps.

Various combinations of these thirteen GLCM-based texture measures have been used in mammographic texture analysis by, for example, Wei et al. [1995], Petrick et al. [1998], Christoyianni et al. [1999], Petrick et al. [1999], Bovis et al. [2000], Sahiner et al. [2001], Sahiner et al. [1998*a*] and Tourassi et al. [2001].

If texture measures are used to quantify texture for purposes of matching, then a good match is characterised by similar texture measure values for the reference and test ROIs.

The TM-matching algorithm is detailed in **Algorithm 8.1**.

## 8.4 Mutual Information Matching (MI-Matching)

MI-matching uses grey-level histograms and GLCMs to quantify the textural information content in the image and mutual information as the similarity metric.

This study uses grey-level histograms because most mutual information image registration algorithms use grey-level histograms to estimate the probability density functions [Wirth et al. 2002, Maes et al. 1997, Pluim et al. 1998, Tourassi et al. 2001, Maintz et al. 1998]. For the calculation of mutual information, a joint, two-dimensional histogram between the reference and test ROIs is required, as well as individual, one-dimensional histograms of the reference and test ROIs. The IDL function (`HIST_2D`) used to compute the two-dimensional histogram is described in **Appendix D**. Examples of two-dimensional histograms are shown in **Figure 8.3** for two sample images. The joint two-dimensional histograms of each image with itself (**Figure 8.3(c,g)**) are only populated on the diagonals, while the joint two-dimensional histograms between R1 and R7

**Algorithm 8.1**: TM-matching using GLCM-based texture measures and distance metrics

```
Data: Test image, $T$, Reference image, $R$, centroid of reference ROI, $[x_c, y_c]$, sampling window size, $w$,
    sampling window step size, $w_{step}$, required image bit-depth, $nbits$, original image bit-depth, $obits$,
    distance in GLCM, $d$
Result: Distance map, $D$
  1. Pre-processing
     (a) Change bit-depth of reference and test images using $nbits$ and $obits$
     (b) Apply appropriate masks if image is a mammogram
  2. Process reference image
     (a) $R_{sub} = R[x_c - ½w : x_c + ½w - 1, y_c - ½w : y_c + ½w - 1]$    ; Define sub-image
     (b) Calculate GLCM of $R_{sub}$    ; Algorithm 5.1
     (c) Calculate texture measures, $TM_R$, of $R_{sub}$    ; Equations 5-10 to 5-22.
  3. Define dimensions of texture measure array
     $ncols$ = height of $T$, $nrows$ = width of $T$
     $xdim = (ncols - w)/step + 1$
     $ydim = (nrows - w)/step + 1$
     $TMArray$ = FLOATARRAY$[xdim, ydim, 13]$
  4. Calculate texture measures for the test ROIs
     for $j = 0$ to $nrows-1$ in $step$ do
   for $i = 0$ to $ncols-1$ in $step$ do
      $id = (i/step)$
      $jd = (j/step)$
      if $j + w - 1 < nrows$ and $i + w - 1 < ncols$ then
              (a) $T_{sub} = T[x_c - ½w : x_c + ½w - 1, y_c - ½w : y_c + ½w - 1]$
              (b) Calculate GLCM of $T_{sub}$
              (c) Calculate texture measures, $TM_T$, of $T_{sub}$
              (d) $TMArray[id, jd, *] = TM_T[*]$
      endif
   endfor
     endfor
  5. Compare $TM_R$ to each entry in $TMArray$ to calculate distance maps, $D$, for the Euclidean,
     standardised Euclidean and Mahalanobis distance metrics.    ; Algorithm 6.1
```

**(Figure 8.3(i,j))** are not symmetrical.

If grey-level histograms are used to quantify texture for purposes of matching, then a good match is characterised by similar one-dimensional histograms for each of the reference and test ROIs and also a densely populated diagonal in the joint two-dimensional histogram.

However, grey-level histograms do not incorporate spatial information, so this study uses the full GLCM as an estimate of a probability density function that incorporates spatial information. Hseu et al. [1999] has used the full GLCM for image registration with mutual information, but

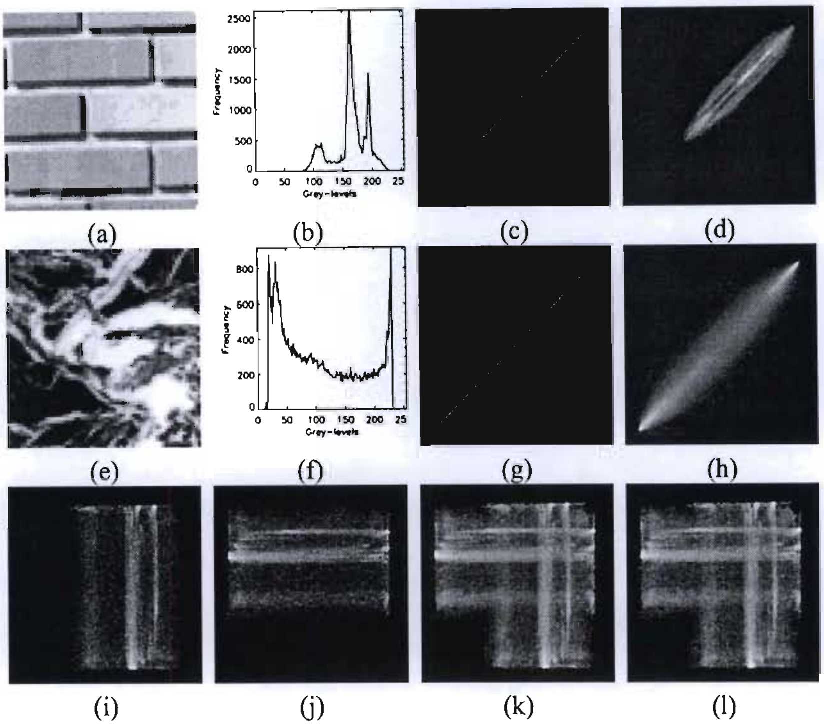


**Figure 8.3**: Examples of individual and joint grey-level histograms and GLCMs averaged over $\theta$=[0°,45°,90°,135°] at $d$=1. **(a)** Image R1. **(b)** Grey-level histogram of R1. **(c)** Joint histogram of R1 and R1. **(d)** GLCM of R1. **(e)** Image R7. **(f)** Grey-level histogram of R7. **(g)** Joint histogram of R7 and R7. **(h)** GLCM of R7. **(i)** Joint histogram of R1 and R7. **(j)** Joint histogram of R7 and R1. **(k)** Joint GLCM of R1 and R7. **(l)** Joint GLCM of R7 and R1. The joint GLCMs are symmetric while the joint histograms are not.

the full GLCM has not been applied to a template-matching problem or to any problem in mammographic CAD. For the calculation of mutual information, individual GLCMs of the reference and test images as well as a joint GLCM between the reference and test images is required. The individual GLCMs are calculated from **Equation 5-4** (page 65). The joint GLCM is based on **Equation 5-4** and for a reference image, $I_R$, and a test image, $I_T$, the joint GLCM between $I_R$ and $I_T$ is given by:

$$\mathbf{G}_{\theta,d}[m,n] = \sum_{j=0}^{M-1}\sum_{k=0}^{N-1} \delta(I_R[j,k] = m, I_T[j - d\sin\theta, k + d\cos\theta] = n) \qquad (8\text{-}1)$$

**Figure 8.3** shows examples of GLCMs for two sample images. The GLCMs of each image with itself (**Figure 8.3(d,h)**) are populated around the diagonals, while the joint GLCMs (**Figure 8.3(k,l)**) are symmetrical.

If GLCMs are used to quantify texture for purposes of matching, then a good match is characterised by reference, test and joint GLCMs that are similarly populated.

**Figure 8.3** shows that the joint histograms are not commutative, i.e. the joint histogram of R1 and R7 is not equal to the joint histogram of R7 and R1, while the joint GLCMs are commutative.

The MI-matching algorithm is detailed in **Algorithm 8.2**.

**Algorithm 8.2**: MI-matching using grey-level histograms, GLCMs and mutual information

```
Data: Test image, $T$, Reference image, $R$, centroid of reference ROI, $[x_c, y_c]$, sampling window size, $w$,
    sampling window step size, $w_{step}$, required image bit-depth, $nbits$, original image bit-depth, $obits$,
    distance in GLCM, $d$
Result: Mutual information map, $MIMAP$
  1. Pre-processing
     (a) Change bit-depth of reference and test images using $nbits$ and $obits$
     (b) Apply appropriate masks if image is a mammogram
  2. Process reference image
     (a) $R_{sub} = R[x_c - ½w : x_c + ½w - 1, y_c - ½w : y_c + ½w - 1]$     ; Define sub-image
  3. Define dimensions of mutual information maps
     $ncols$ = height of $T$, $nrows$ = width of $T$
     $xdim = (ncols - w)/w_{step} + 1$
     $ydim = (nrows - w)/w_{step} + 1$
     $MIMAP = \text{FLOATARRAY}[xdim, ydim]$
  4. Calculate mutual information map, $MIMAP$
     for $j = 0$ to $nrows\text{-}1$ in $w_{step}$ do
  for $i = 0$ to $ncols\text{-}1$ in $w_{step}$ do
     $id = (i/w_{step})$
     $jd = (j/w_{step})$
     if $j + w - 1 < nrows$ and $i + w - 1 < ncols$ then
        $T_{sub} = T[x_c - ½w : x_c + ½w - 1, y_c - ½w : y_c + ½w - 1]$
          ; For method ∈ [histogram, GLCM] compute individual ($G_R$, $G_T$) and
            joint ($G_{RT}$) probability density functions.
        $MIMap[id, jd] = \text{MI}(G_R, G_T, G_{RT})$     ; Algorithm 6.2
     endif
  endfor
     endfor
```

### 8.4.1 Details of Histogram Calculation

In order to reduce computation time, the individual and joint grey-level histograms were determined by taking the minimum grey-level, $f_{min} = \min[R_{sub}, T_{sub}]$, and maximum grey-level, $f_{max} = \max[R_{sub}, T_{sub}]$, for the pair of reference and test sub-images. The range of grey-levels given by $f_{max} - f_{min}$ was then divided into the required number of bins, *nbins*.

### 8.4.2 Details of GLCM Calculation

In order to reduce computation time, the GLCM was computed by taking the minimum grey-level, $f_{min} = \min[R_{sub}, T_{sub}]$, and maximum grey-level, $f_{max} = \max[R_{sub}, T_{sub}]$, for the pair of reference and test sub-images. The GLCM was determined between the range of grey-levels given by $f_{max} - f_{min}$.

## 8.5 Determination of the Matching Map

The maps of distance and mutual information that result from the TM- and MI-matching algorithms have to be converted to matching maps. A matching map is defined as having the optimal match indicated by maximum intensity. For MI-matching, the optimal match occurs at the maximum of the mutual information map and the matching map is identical to the mutual information map.

For TM-matching, the optimum match on a distance map is at a minimum of distance and so the distance maps must be transformed to matching maps. The matching map, $M_D$, corresponding to a distance map, $D$, of size $M \times N$, is given by:

$$M_D[i,j] = \frac{1}{D[i,j]} \quad \text{for } i = 0,\ldots,M-1 \text{ and } j = 0,\ldots,N-1. \tag{8-2}$$

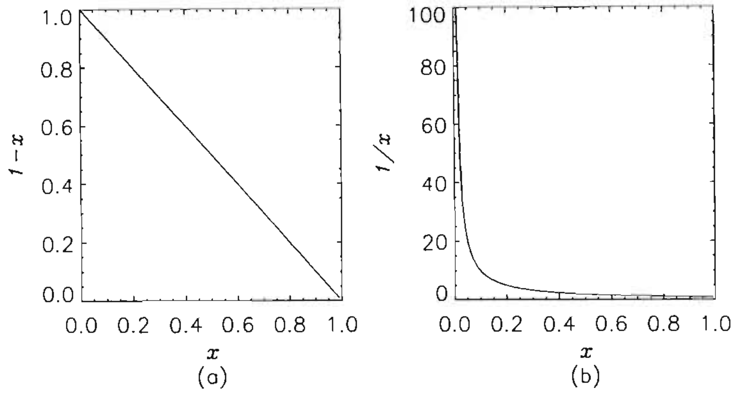


**Figure 8.4**: Functions for converting distance maps to matching maps. **(a)** $1-x$ is a linear function that has equal preference for all values of $x$. **(b)** $1/x$ is an inverse function that puts a greater preference on lower values or $x$ and is preferred for converting distance maps to matching maps.

The transformation in **Equation 8-2 (Figure 8.4(b))** was chosen because low distance values are enhanced in the matching map, while high distance values are suppressed. The enhancement and suppression is greater for this inverse transformation than for example, a linear transformation like that in **Figure 8.4(a)**, which would also achieve the required conversion from (optimum at minimum) distance map values to (optimum at maximum) matching map values. The greater enhancement of regions with good matches and suppression of regions with poor matches means that the contrast of the high intensity regions indicating a good match is also enhanced when compared to the low intensity background in the matching map.

## 8.6 Dimensions and Spatial Resolution of the Matching Map

The sampling window size and the sampling window step size affect the spatial resolution and dimensions of the matching map. The dimensions, $[m_x, m_y]$, and spatial resolution, $m_s$, of the matching map for a sampling window size of $w$ and a sampling window step size of $w_{step}$ is:

$$m_x = \frac{(n_x - w)}{w_{step}} + 1, \quad m_y = \frac{(n_y - w)}{w_{step}} + 1 \tag{8-3}$$

$$m_s = n_s \cdot w \tag{8-4}$$

where $n_x$, $n_y$ are the width and height of the test image with spatial resolution $n_s$.

The spatial resolution of the matching map is dependant on the sampling window size alone while the dimensions of the matching map are dependant on the sampling window size and the sampling window step size.

Examples of matching maps at different values of $w_{step}$, for the same value of $w$ are shown in **Figure 8.5**. The matching maps are shown to scale to demonstrate the differences in the dimensions of the matching map while the spatial resolution or pixel size is the same.

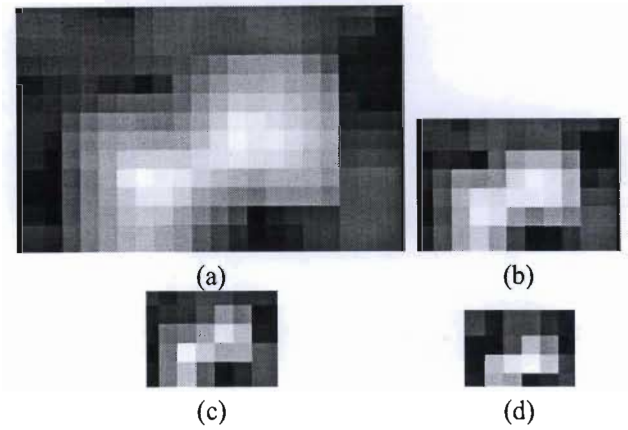


**Figure 8.5**: Effect of sampling window step size on dimensions of matching map demonstrated on mosaic reference image R7 and test image T2, for $w$=400 and **(a)** $w_{step}$=100, **(b)** $w_{step}$=200, **(c)** $w_{step}$=300, **(d)** $w_{step}$=400. Matching maps are shown to scale to demonstrate differences in map dimensions and the same spatial resolution.

For large values of $w$ and $w_{step}$, the spatial resolution of the matching map is very coarse while for small $w$ and small $w_{step}$, the spatial resolution of the matching map is very fine. If $w$ is too large then the sampled sub-image might contain information from multiple textural regions and if $w$ is too small then the sampled sub-image might not contain enough information. Similarly, if $w_{step}$ is too large, boundaries between textural regions may be missed, resulting in a poor match and if $w_{step}$ is too small, computation time may be unnecessarily increased.

## 8.7 Calculation of $A_{ROC}$ and $C_{fb}$

Matching maps are scaled between 0 and 255 (8 bits of information) to enable comparison with the ground truth data at discrete decision thresholds. *FPF* and *TPF* values are computed at each decision threshold and are used to generate a ROC curve for each map. The trapezoidal rule is used to compute the area under the ROC. The ground truth data is also used in the calculation of the contrast of the detected regions when compared to the background in the matching map. The algorithm for evaluation of the matching results is detailed in **Algorithm 8.3**.

**Algorithm 8.3**: Evaluation of results of matching methods

```
Data: Matching map, $M$, and ground-truth map, $T$. $M$ is scaled between 0 and 255 (corresponding to
    8 bits of information).
Result: Area under curve, $A_{ROC}$, and contrast, $C_{fb}$.
for $g = 0$ to 255 do
    for $i = 0$ to width-1 do
        for $j = 0$ to height-1 do
            ; B is the binary map formed by thresholding M at g
            if $M[i,j] \geq l$ then $B[i,j] = 1$else $B[i,j] = 0$
            if $B[i,j] = 1$ and $T[i,j] = 1$ then $TP = TP + 1$
            if $B[i,j] = 0$ and $T[i,j] = 0$ then $TN = TN + 1$
            if $B[i,j] = 1$ and $T[i,j] = 0$ then $FP = FP + 1$
            if $B[i,j] = 0$ and $T[i,j] = 1$ then $FN = FN + 1$
        endfor
    endfor
    $TPF[g] = (TP)/(TP + FN)$     ; eq.7-1
    $FPF[g] = (FP)/(FP + TN)$     ; eq.7-2
endfor
$A_{ROC} = \sum_{g=0}^{255} FPF[g] \cdot (TPF[g] - TPF[g-1]) + \frac{1}{2} \cdot (FPF[g] - FPF[g-1]) \cdot (TPF[g] - TPF[g-1])$
    ; Contrast Calculation
    1. Determine maximum extent of ROI, $d_{max}$.
    2. Use $d_{max}$ to define a foreground mask, $F$, and a background mask, $B$, as indicated in Figure 7.4.
    3. Use $F$ and $B$ to determine $f$ and $b$, the average grey-levels in the foreground and background
       respectively.
    4. Calculate Contrast $C_{fb} = (f - b)/(f + b)$     ; eq.7-10
```

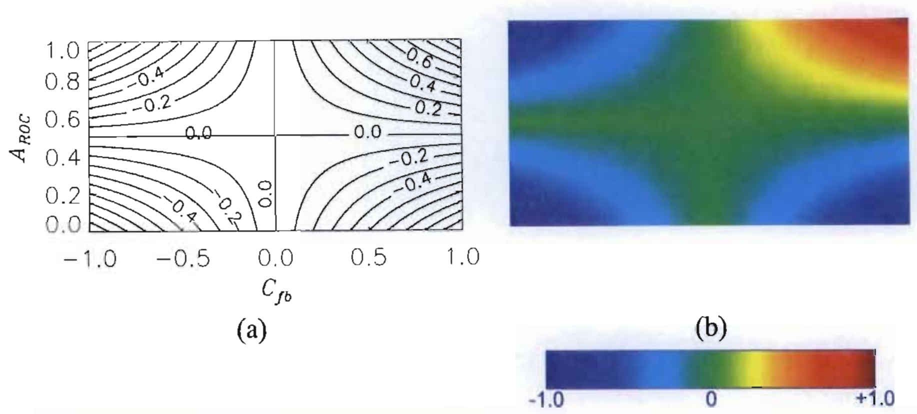
## 8.8 Quantification of Matching Accuracy, κ

The selection of the best combination of $A_{ROC}$ and $C_{fb}$-values is facilitated by the novel use of a *combined* $A_{ROC} \backslash C_{fb}$*-value* referred to as *matching accuracy*, κ, and calculated as follows:

$$\kappa = \begin{cases} -2 \cdot (A_{ROC} - 0.5) \cdot C_{fb} & \text{if } A_{ROC} \leq 0.5 \text{ and } C_{fb} \leq 0 \\ 2 \cdot (A_{ROC} - 0.5) \cdot C_{fb} & \text{otherwise} \end{cases} \tag{8-5}$$

with $-1.0 \leq \kappa \leq 1.0$ because $0 \leq A_{ROC} \leq 1$ and $-1 \leq C_{fb} \leq 1$.

The combination of $A_{ROC}$ and $C_{fb}$ described in **Equation 8-5** was selected because both $A_{ROC}$ and $C_{fb}$ are equally preferred. **Figure 8.6(a)**, a contour plot of the κ surface, shows that $\kappa \rightarrow 0$ as $A_{ROC} \rightarrow 0.5$ and as $C_{fb} \rightarrow 0$. Only points in the upper right quadrant are favoured as this is the only region where $\kappa > 0$. A good match has $A_{ROC} \approx 1$ and $C_{fb} \approx 1$. Acceptable matches should have $A_{ROC} > 0.5$ as this indicates that the matching is better than random and $C_{fb}$ should be positive as this indicates that the matched region is at least brighter than the background in the matching map.



(a) (b)

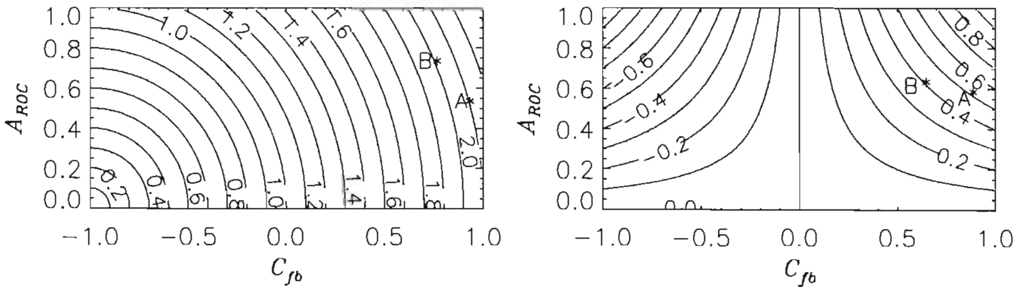
**Figure 8.6**: Quantification of matching accuracy, κ, to rank combinations of $A_{ROC}$ and $C_{fb}$. **(a)** Contour plot of κ surface. **(b)** κ surface as a pseudo-colour image. κ tends to 0 as $A_{ROC}$ approaches 0.5 and as $C_{fb}$ approaches 0 and 255 also only favours $A_{ROC}$- and $C_{fb}$-values in the upper right quadrant as this is the only region where κ>0. $A_{ROC}$ and $C_{fb}$ are equally weighted.

Two other ranking schemes (shown in **Figure 8.7**) were also investigated:

$$\kappa_{alt1} = \sqrt{A_{ROC}^2 + (C_{fb}+1)^2} \tag{8-6}$$

$$\kappa_{alt2} = A_{ROC} \cdot C_{fb} \tag{8-7}$$

These schemes were not used since both place a higher preference on contrast, which results in a point with a very high $C_{fb}$-value, but a very low $A_{ROC}$-value being ranked higher than a point with a slightly lower $C_{fb}$-value and a slightly higher $A_{ROC}$-value. Higher values of $A_{ROC}$ are preferred to higher values of contrast since the former indicates that a larger region has been correctly matched.



(a) (b)

**Figure 8.7**: Alternate methods of combining $A_{ROC}$ and $C_{fb}$ to rank matching results that were also investigated. **(a)** Ranking using **Equation 8-6** yields a better match at point A which has $A_{ROC}$=0.5 and $C_{fb}$=0.94 compared to point B which has $A_{ROC}$=0.70 and $C_{fb}$=0.77. **(b)** Ranking using **Equation 8-7** yields a better match at point A which has $A_{ROC}$=0.55 and $C_{fb}$=0.89 than at point B which has $A_{ROC}$=0.60 and $C_{fb}$=0.65. Both these ranking schemes gave $C_{fb}$ a higher preference than $A_{ROC}$, which was not desired.

## 8.9 Sampling Window Sizes and the Autocorrelation Function (ACF)

Texture analysis depends strongly on the size of the sampling window. A sampling window needs to be large enough to contain sufficient information for the texture analysis, but at the same time needs to be small enough to ensure that single textures are sampled. In this study, the dependence of matching accuracy on sampling window size is investigated for a range of values, but it is desirable to have an independent method of determining the sampling window size which results in an optimal matching accuracy, for any image. The autocorrelation function (ACF) described in §5.6.1 (page 71) was investigated as a method of determining an optimal sampling window size.

The ACF was used to determine the characteristic scale widths and scale heights of textures in the reference images. The ACF analysis was performed at different bit-depths as the scale sizes of textures are expected to change with bit-depth. These results were compared to the sampling window sizes that result in an optimal matching accuracy (from the investigation of the matching accuracy of the matching algorithms) using a linear correlation analysis. If there is a correlation, then the ACF can be used to select an optimal sampling window size for any image. The algorithm to compute the ACF scale sizes is described in **Algorithm 8.4**.

**Algorithm 8.4**: Use of autocorrelation function to determine sampling window sizes

```
Data: Reference image, $R$, & size, $sz$ of square sampling window that completely encloses the ROI in $R$.
Result: ACF-width, $w_{ACF}$, and ACF-height, $h_{ACF}$.
  1. Extract sub-image, $R_{sub}$ from $R$, of size $sz$ centred on centroid of ROI.
  2. Compute autocorrelation function, ACF, of $R_{sub}$, as detailed in Equation 5-25.
  3. Compute sum of ACF columns, $\mathrm{ACF}_{cols}$. $w_{ACF} = \min(\mathrm{ACF}_{cols})$.
  4. Compute sum of ACF rows, $\mathrm{ACF}_{rows}$. $h_{ACF} = \min(\mathrm{ACF}_{rows})$.
```

## 8.10 Significance and Correlation Analysis

### 8.10.1 The *t*-Test

A paired *t*-test analysis is used to assess whether the mean results of two data sets are statistically different from each other. The *t*-test produces two results: a *t*-value and a *p*-value. The *t*-value is a ratio between the difference of the means and of the standard error of the difference of the means. If the *t*-value is negative, then the mean of the first data set is less than the mean of the second data set. The *p*-value is determined by using the *t*-value, the number of data points in each data set (or degrees of freedom) and the significance level (or alpha value). The alpha value refers to the odds that the observed result is due to chance. If a significance level of 0.05 is assumed and if the data sets have $p < 0.05$, then the data sets being compared can be described as statistically different. If $p > 0.05$ then the data sets being compared are not statistically different [Trochim 2005].

The *t*-test analysis is performed using IDL's `TM_TEST` function, and is used to compare the best results for each matching method to determine how similar the matching performance of a method is to the matching performances of the rest.

### 8.10.2 The Pearson Linear Correlation Coefficient

The Pearson linear correlation coefficient gives an indication of the degree to which two data sets are linearly related. The linear Pearson correlation coefficient, $\rho$, ranges from -1.0 to 1.0, with a strong linear correlation for $\rho=\pm 1$ and a random, non-linear relationship between the data sets for $\rho \approx 0$ [StatSoft, Inc. 2003].

The Pearson linear correlation coefficient is calculated using IDL's `CORRELATE` function.

In this study, the Pearson linear correlation coefficient is primarily used to determine whether the results of the autocorrelation function analysis are correlated with the optimal sampling window sizes from the matching analysis.

## 8.11 Summary

TM-matching uses GLCM-based texture measures to quantify texture and distance similarity metrics to compare textures between images to determine similarity. MI-matching uses grey-level histograms and GLCMs to quantify texture and mutual information as a similarity metric to compare textures between images to determine similarity. Results are evaluated by computation of the area under the ROC curve, $A_{ROC}$, and contrast, $C_{fb}$, of the matched region in the matching map. Matching accuracy, $\kappa$, is defined as a combination of $A_{ROC}$ and $C_{fb}$. Matching parameters like sampling window size, sampling window step size, bit-depth, distance in the GLCM and the number of histogram bins are varied to investigate their effect on matching accuracy. The autocorrelation function is investigated as a possible method of independently extracting an optimal sampling window size.

The image processing components used in this study are summarised in **Table 8.4**, while the abbreviations used for the different methods are summarised in **Table 8.5**.

**Table 8.4**: Summary of texture-based image processing methods used in this study

| | TM-matching | MI-matching |
|---|---|---|
| Texture Quantification | 1. GLCM-based texture measures | 1. grey-level histograms<br>2. GLCMs |
| Similarity Metric | 1. Euclidean distance ($D_E$)<br>2. Standardised Euclidean distance ($D_{ES}$)<br>3. Mahalanobis distance ($D_M$) | 1. mutual information |

**Table 8.5**: Abbreviations of the different matching methods used in this study

| Method | Abbreviation |
|---|---|
| GLCM-based texture measures with $D_E$ | TM-$D_E$ |
| GLCM-based texture measures with $D_{ES}$ | TM-$D_{ES}$ |
| GLCM-based texture measures with $D_M$ | TM-$D_M$ |
| Histograms with mutual information | MI-histograms |
| GLCMs with mutual information | MI-GLCMs |

# Chapter 9

# Matching Results: Mosaic Images

The TM-matching and MI-matching algorithms were applied to test images made up of a mosaic of single texture reference images to test matching performance under the ideal condition of knowing how the reference images are transformed in the test image. The matching results were evaluated and these results are presented and discussed in this chapter.

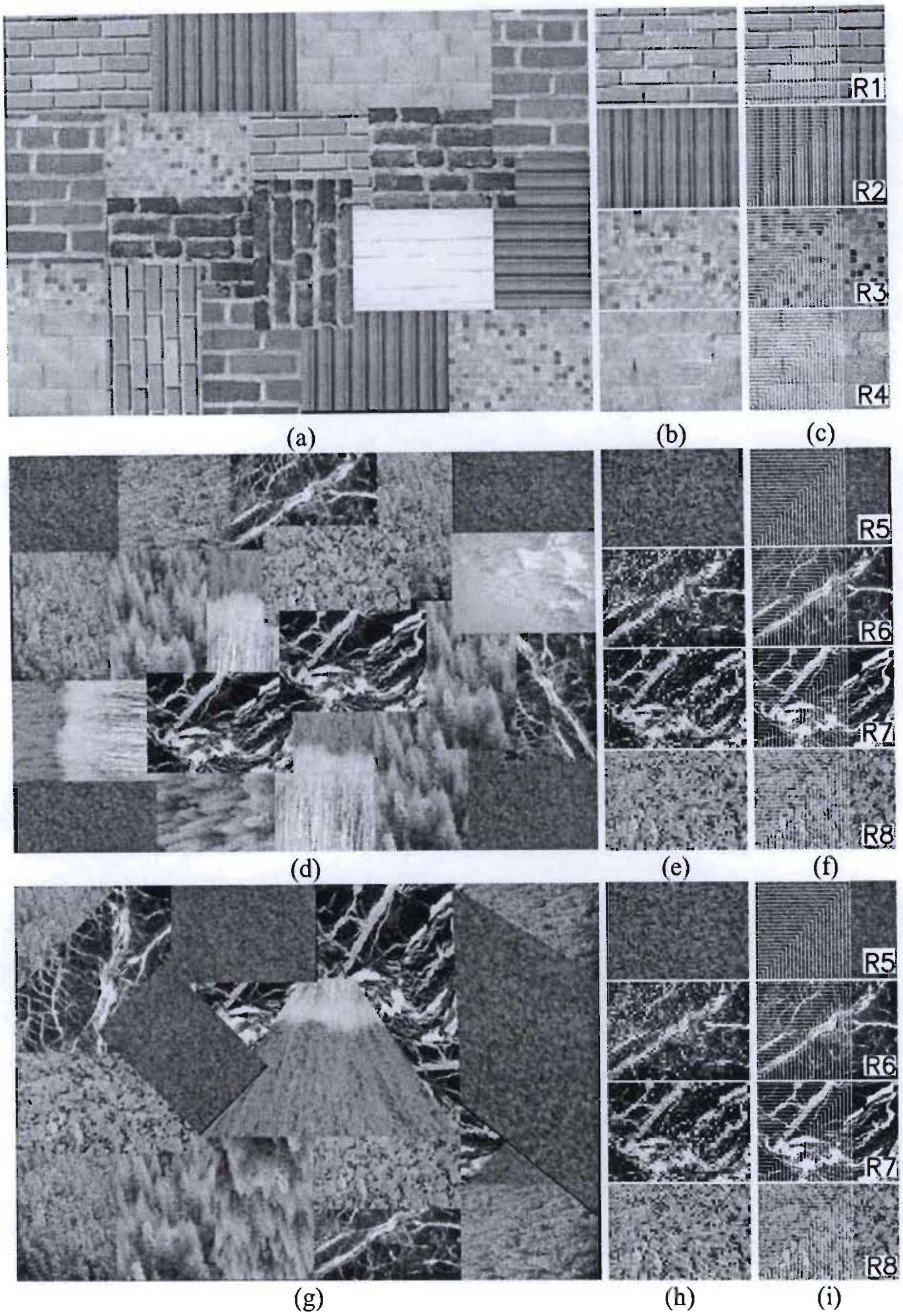
The purpose of this chapter is two-fold. Firstly, while not all textures included in the mosaic images are similar to the mammographic textures, testing the performance of the matching algorithms on test images with clear borders between single textures and where there are known transformations of the single textures will give insight into whether the matching algorithms have any potential for identifying similar textures in mammograms.

Secondly, the investigation using mosaic images will also give some insight into how the matching accuracy is affected by the various matching parameters. This information will be used to guide the selection of a set of optimal matching parameters, when the matching algorithms are applied to the pairs of mammograms and the stereotactic biopsy mammograms.

## 9.1 Details of the Mosaic Images used in this Study

Three mosaic images were constructed from single texture images (at 8 bits) [Softkey International, Inc. 1996] and were used to test the performance of the matching algorithms under the conditions of knowing exactly how the reference regions were transformed in the test image. For two of the mosaic images (test image T1 and test image T2), the reference images were identical to those in the mosaic images, with some rotated by 90°or 180°. In test image T3 some reference images were scaled up or distorted or rotated by 45°, 90°, 135° or 180°, to simulate the possible effects of changing the compression between mammographic views. Ground-truth data was manually extracted from the test images and was used to evaluate matching accuracy.

**Figure 9.1** shows the mosaic test images, their corresponding reference images and the placement of sampling windows of different sizes, within each reference image. All sampling windows are placed with their bottom left corner at the image origin (i.e. bottom left corner as defined in **Appendix B**). All test images were 2 400 pixels × 1 600 pixels, and the reference images were 600 pixels × 400 pixels. Each mosaic image pair is referred to by the reference and test image labels, for example, the image pair consisting of reference image R7 and test image T2, is R7T2.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

**Figure 9.1**: Mosaic images. Each mosaic image pair is referred to by the reference and test image labels, for example, the image pair consisting of reference image R7 and test image T2, is R7T2 **(a)** Test image T1 with images of bricks and tiles combined at various orientations (0°, 90°, 180°). **(b)** Reference images for T1, (from top) R1, R2, R3 and R4. **(c)** Positions of sampling windows of different sizes (white lines) overlaid on reference images. **(d)** Test image T2 with images of trees and granite combined at various orientations (0°, 90°, 180°). **(e)** Reference images for T2, (from top) R5, R6, R7 and R8. **(f)** Positions of sampling windows of different sizes (white lines) overlaid on reference images. **(g)** Test image T3 with images of trees and granite combined at various orientations (0°, 45°, 90°, 135°, 180°) and at various scaling factors and distortions (to simulate magnification and compression possible effects). **(h)** Reference images for T3, (from top) R5, R6, R7 and R8. **(i)** Positions of sampling windows of different sizes (white lines) overlaid on reference images.

**Table 9.1**: Details of transformations of reference textures in mosaic images listing the number of times each reference texture appears in the test image and the number of times the reference texture is either rotated, scaled (up or down) or rotated.

| Image Pairs | # of occurrences | # rotated | # scaled up or down | # distorted |
|---|---|---|---|---|
| R1T1 | 3 | 1 | 0 | 0 |
| R2T1 | 3 | 1 | 0 | 0 |
| R3T1 | 3 | 0 | 0 | 0 |
| R4T1 | 2 | 0 | 0 | 0 |
| R5T2 | 4 | 0 | 0 | 0 |
| R6T2 | 2 | 1 | 0 | 0 |
| R7T2 | 2 | 0 | 0 | 0 |
| R8T2 | 1 | 0 | 0 | 0 |
| R5T3 | 4 | 2 | 1 | 1 |
| R6T3 | 2 | 1 | 1 | 0 |
| R7T3 | 4 | 0 | 4 | 1 |
| R8T3 | 2 | 0 | 1 | 0 |

## 9.2 Effect of Sampling Window Step Size on Matching Accuracy

The template matching algorithms were applied to the mosaic image pairs, generating matching maps for each combination of matching parameters. The results of varying the sampling window step size, $w_{step}$, are presented. For the mosaic images $w_{step}$ was varied as a fraction, $w_{sf}$, of the sampling window size, $w$, with:

$$w_{step} = w \cdot w_{sf} \tag{9-1}$$

Typical plots of matching accuracy, κ, *vs.* $w$, are shown for 4 values of $w_{sf}$ in **Figure 9.2** for the matching of image pair R7T2, at a bit-depth of 8 bits and $d$=1 pixel (for the GLCMs) or *nbins*=128 bins (for the histograms). No method shows a clear dependence of matching accuracy on $w_{sf}$. Apart from TM-$D_E$ (for $w$<260 pixels) matching accuracy plots are smoothest for $w_{sf}$=0.25 and are most erratic for $w_{sf}$-values of 0.75 and 1.00.

As described in §8.6 (page 99), the dimensions and spatial resolution of the matching map are dependent on $w$ and $w_{step}$. When $w$ and $w_{step}$ are very large, the resulting matching map is very small (**Figure 8.5(d)**), with a coarse spatial resolution. The coarse resolution means that when the map is scaled between 0 and 255 for the evaluation, a single high-intensity pixel can cover a very large area in the matching map and can indicate a 'good' match leading to a high κ-value, at large values of $w$. This would explain why the matching accuracy plots are so erratic for the larger values of $w_{sf}$ and also why there are some κ-values for $w_{sf}$=0.75 and $w_{sf}$=1.0 (for the larger sampling windows) that are higher than the κ-values for $w_{sf}$=0.25 and $w_{sf}$=0.50 at the same sampling window sizes.

Unfortunately, as a consequence of choosing the sampling window step size as a fraction of the sampling window size, the results of this investigation have two varying parameters, viz. $w$ and $w_{step}$ because each value of $w$ had a different value of $w_{step}$. It is therefore impossible to determine which parameter ($w$ or $w_{step}$) is affecting the matching accuracy.

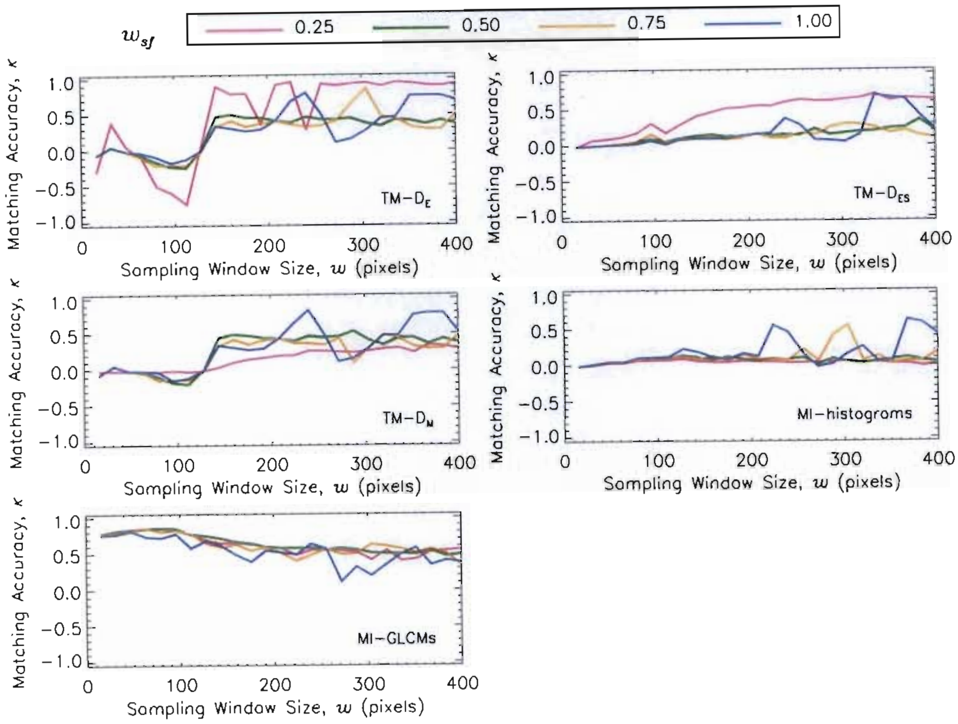


**Figure 9.2**: Typical example of the effect of sampling window step size on matching accuracy demonstrated on mosaic image pair R7T2, with matching accuracy plotted as a function of sampling window size, for the 4 values of $w_{sf}$, indicated by the different colours. Results with $w_{sf}$=0.25 are generally smoothest.

This investigation of the dependence of sampling window step size on matching accuracy for the mosaic images led to the conclusion that the sampling window step size:

1. should be independent of the sampling window size, and
2. should be fixed.

For the mosaic images, matching accuracy curves were smoothest for $w_{sf}$=0.25, so only results for $w_{sf}$=0.25 are presented and discussed in the following sections.

## 9.3 Examples of Matching Maps

Examples of typical matching maps for TM- and MI-matching are shown in **Figure 9.3**. These maps were obtained by selecting the combination of matching parameters that yielded the maximum and minimum values of κ for each method. Details of the exact matching parameters for these maps are given in **Table 9.2**.

The matching maps for the minimum κ-values are shown in the left column in **Figure 9.3**. Three methods (TM-$D_E$, MI-histograms, MI-GLCMs) have negative contrasts because the background is brighter than the intensity of region to be matched. The remaining two methods (TM-$D_{ES}$, TM-$D_M$) have $A_{ROC}$<0.5 indicating that the match is no better than random.

The matching maps for the maximum κ-values are shown in the right column in **Figure 9.3**. All methods have very high values for $A_{ROC}$, but TM-$D_M$ and MI-histograms have low values for

$C_{fb}$, which leads to an overall low value for $\kappa$. The effect of a positive, but low $C_{fb}$-value can be seen in the maps for the latter two methods. It can be clearly seen that matched region is not as easily discerned from the background as, for example, for the best matching map for MI-GLCMs.

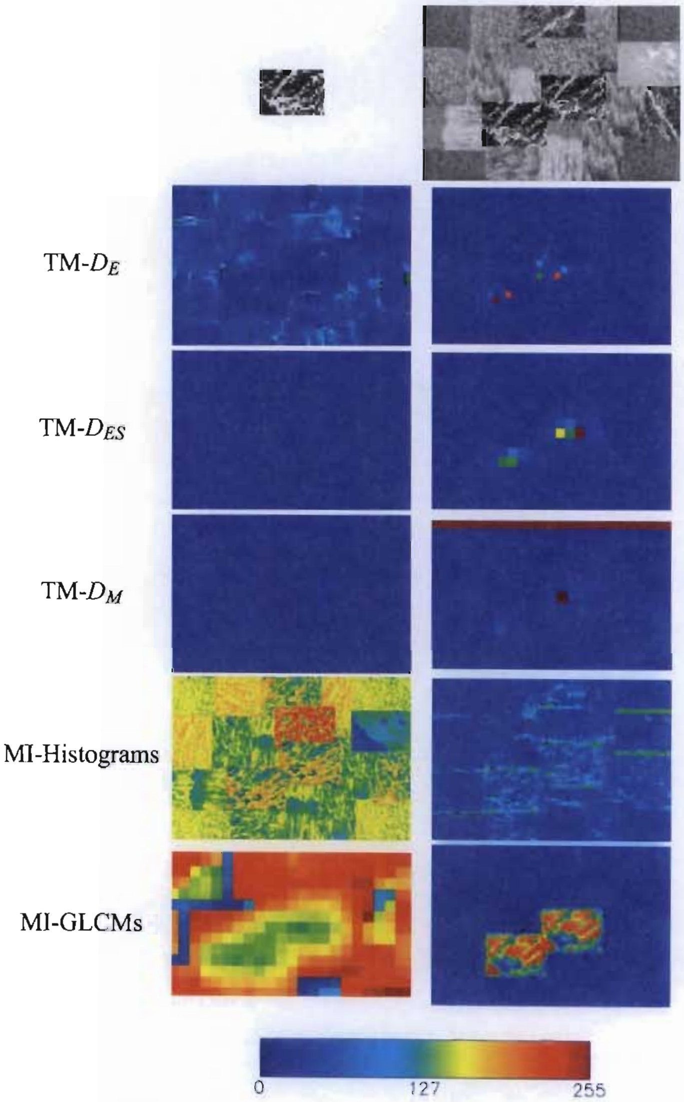


**Figure 9.3**: Examples of matching maps (not to scale) for image pair R7T2. The reference image R7 is shown at the top, left and the test image T2 is shown at the top, right. The maps with the minimum and maximum $\kappa$-values, for each method, are shown in the left and right columns, respectively, with the colour bar at the bottom indicating the intensity scale for the maps.

**Table 9.2**: Summary of matching parameters for the example matching maps of mosaic for image pair R7T2, listed at the maximum and minimum $\kappa$-values. Negative contrasts indicate that the background is brighter than the region to be matched.

| Method | Minimum $\kappa$ | | | | | | Maximum $\kappa$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *w* | *nbits* | *d*/*nbins* | $A_{ROC}$ | $C_{fb}$ | $\kappa$ | *w* | *nbits* | *d*/*nbins* | $A_{ROC}$ | $C_{fb}$ | $\kappa$ |
| TM-$D_E$ | 112 | 8 | 2 | 0.06 | -0.85 | -0.75 | 224 | 8 | 1 | 0.99 | 0.99 | 0.97 |
| TM-$D_{ES}$ | 96 | 5 | 10 | 0.43 | 0.15 | -0.02 | 336 | 5 | 10 | 0.98 | 0.82 | 0.80 |
| TM-$D_M$ | 96 | 6 | 10 | 0.39 | 0.09 | -0.02 | 368 | 8 | 1 | 0.90 | 0.46 | 0.36 |
| MI-histograms | 16 | 8 | 128 | 0.52 | -0.01 | -0.00 | 80 | 6 | 16 | 0.83 | 0.25 | 0.16 |
| MI-GLCMs | 384 | 5 | 10 | 0.16 | -0.12 | -0.08 | 48 | 8 | 10 | 1.00 | 0.94 | 0.93 |

## 9.4 TM-Matching Evaluation Results

The TM-matching algorithm (**Algorithm 8.1** on page 96) was applied to the three test images on page 106. The results of the dependence of matching accuracy on the matching parameters are discussed for each of the three distance similarity metrics.

Examples of the median, average and standard deviation for each of the texture measures is shown in **Table 9.3** for reference image R7 to demonstrate that there is a clear difference in the ranges of the values used as inputs to the distance similarity metric. The values listed were calculated over the full range of matching parameters. The sum variance values are considerably higher than the values for the other texture measures. The sum variance texture measure most likely dominates the Euclidean distance calculation.

**Table 9.3**: Median, average and standard deviation of texture measures.

| Texture Measures for R7 | Median | Average | Standard Deviation |
|---|---|---|---|
| Entropy | 9.06 | 7.71 | 4.13 |
| Energy | 0.004 | 0.008 | 0.012 |
| Inertia | 162 | 810 | 1399 |
| Inverse Difference Moment | 0.183 | 0.184 | 0.130 |
| Correlation | 0.574 | 0.498 | 0.329 |
| Sum Average | 44.1 | 63.3 | 65.9 |
| Sum Entropy | 6.15 | 5.23 | 2.82 |
| Difference Entropy | 4.46 | 3.93 | 2.20 |
| Sum Variance | 1132 | 4263 | 6866 |
| Difference Average | 8.11 | 13.1 | 13.8 |
| Difference Variance | 94.9 | 450 | 748 |
| Information Measure of Correlation 1 | 1.46 | 1.18 | 0.593 |
| Maximum Probability | 0.035 | 0.046 | 0.046 |

### 9.4.1 Computation Times

Computation times for a single map were dependent on the values used for *nbits* and $w$ and ranged from a few minutes to an hour. The calculation of the texture measures was the most time-consuming portion of the calculation, when compared to the time required for the GLCM calculation and the distance map calculation.

### 9.4.2 Evaluation of TM-Matching with $D_E$

The evaluation results of applying the TM-matching algorithm with the Euclidean distance similarity metric, $D_E$, to mosaic images are presented and discussed. The dependence of matching accuracy on *nbits*, $d$ and $w$ is examined. Results are presented in the formats described in §8.2 (page 93). Dependence on matching accuracy is indicated by separation of the different coloured plots.

### Effect of *nbits* on Matching Accuracy, κ

Typical examples of κ *vs.* *w* for different values of *nbits* are shown in **Figure 9.4(a)** to demonstrate the dependence of matching accuracy on *nbits* and *w*. There is some dependence of matching accuracy on *nbits* at the smaller sampling windows for R2T1 and R5T2, but the results for R7T2 and R7T3 overlap almost exactly indicating that matching accuracy is independent of *nbits* for these two examples.

The results of the autocorrelation function analysis are plotted as vertical lines for each value of *nbits*, but there is very little difference in the results at the different bit-depths. The ACF-widths are significantly different from the ACF-heights. There appears to be poor correlation between the ACF results and the maximum values of κ. The ACF results are discussed in §9.7.

**Figure 9.4(b)** shows average matching accuracy (over all image pairs) plotted as a function of *d*, to highlight the general dependence of matching accuracy on *nbits*. The dependence of matching accuracy on *nbits* is influenced by the sampling window size. There is some separation of the plots, for $w \leq 80$ pixels, showing a slight increase in matching accuracy as *nbits* decreases. Matching accuracy is independent of *nbits* for the larger sampling windows since all these plots overlap completely.

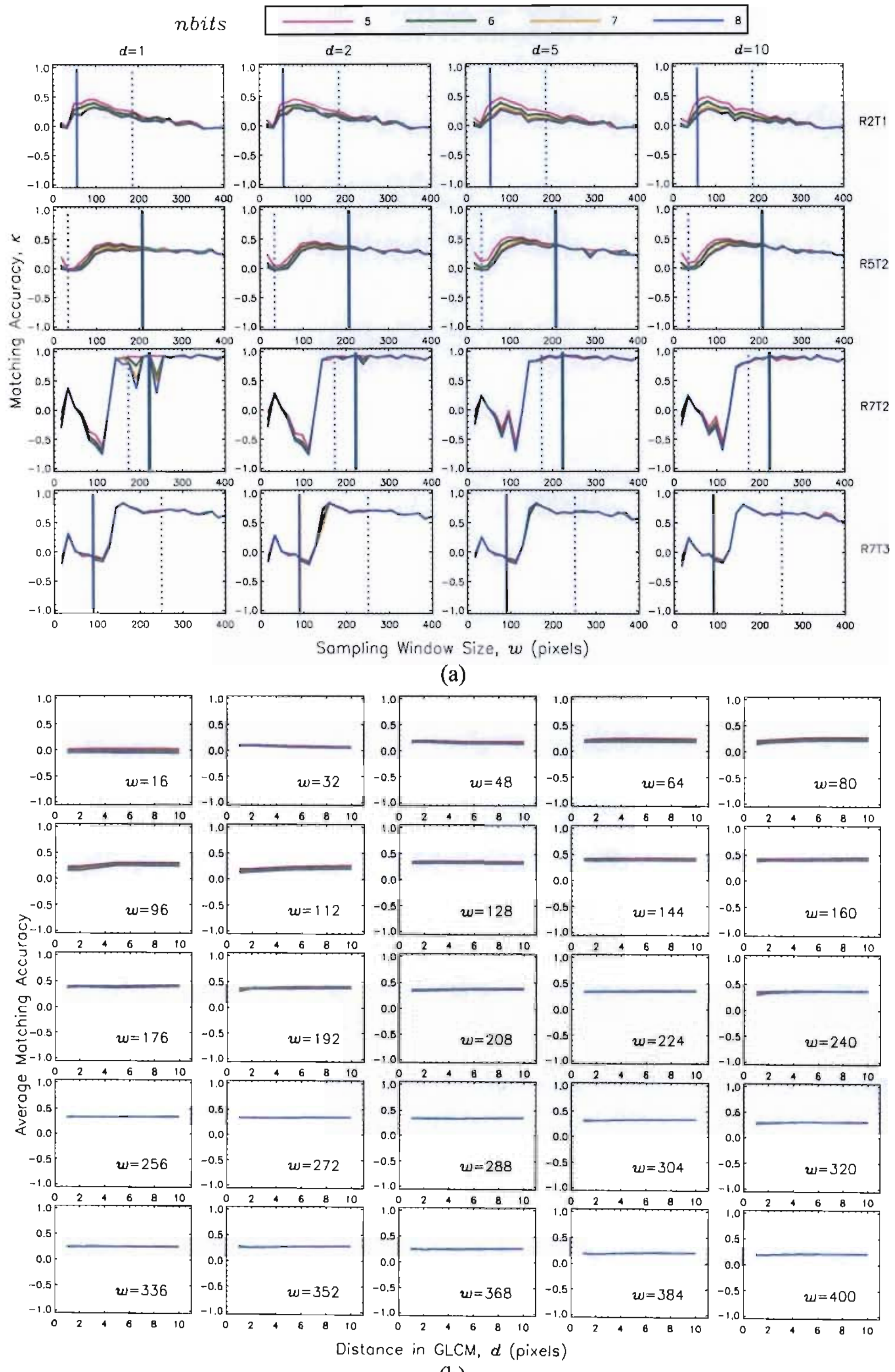
### Effect of *d* on Matching Accuracy

Some examples of κ *vs.* *w* are shown in **Figure 9.5(a)** to demonstrate the dependence of matching accuracy on *d*. There is some separation of the plots, but there is no consistent pattern to the dependence of matching accuracy on *d*. There are a few window sizes in R7T2 where the 5-bit plot deviates from the rest and this is most likely a consequence of the different step size for each sampling window size.

**Figure 9.5(b)** shows average matching accuracy (over all image pairs) plotted as a function of *nbits*, to highlight the general dependence of matching accuracy on *d*. The dependence of matching accuracy on *d* is influenced by the sampling window size. Most sampling windows show no separation of the *d*-plots, indicating an independence of matching accuracy on *d*. Some sampling windows show a slight separation between the plots. These show that matching accuracy improves as *d* increases.

### Effect of *w* on Matching Accuracy

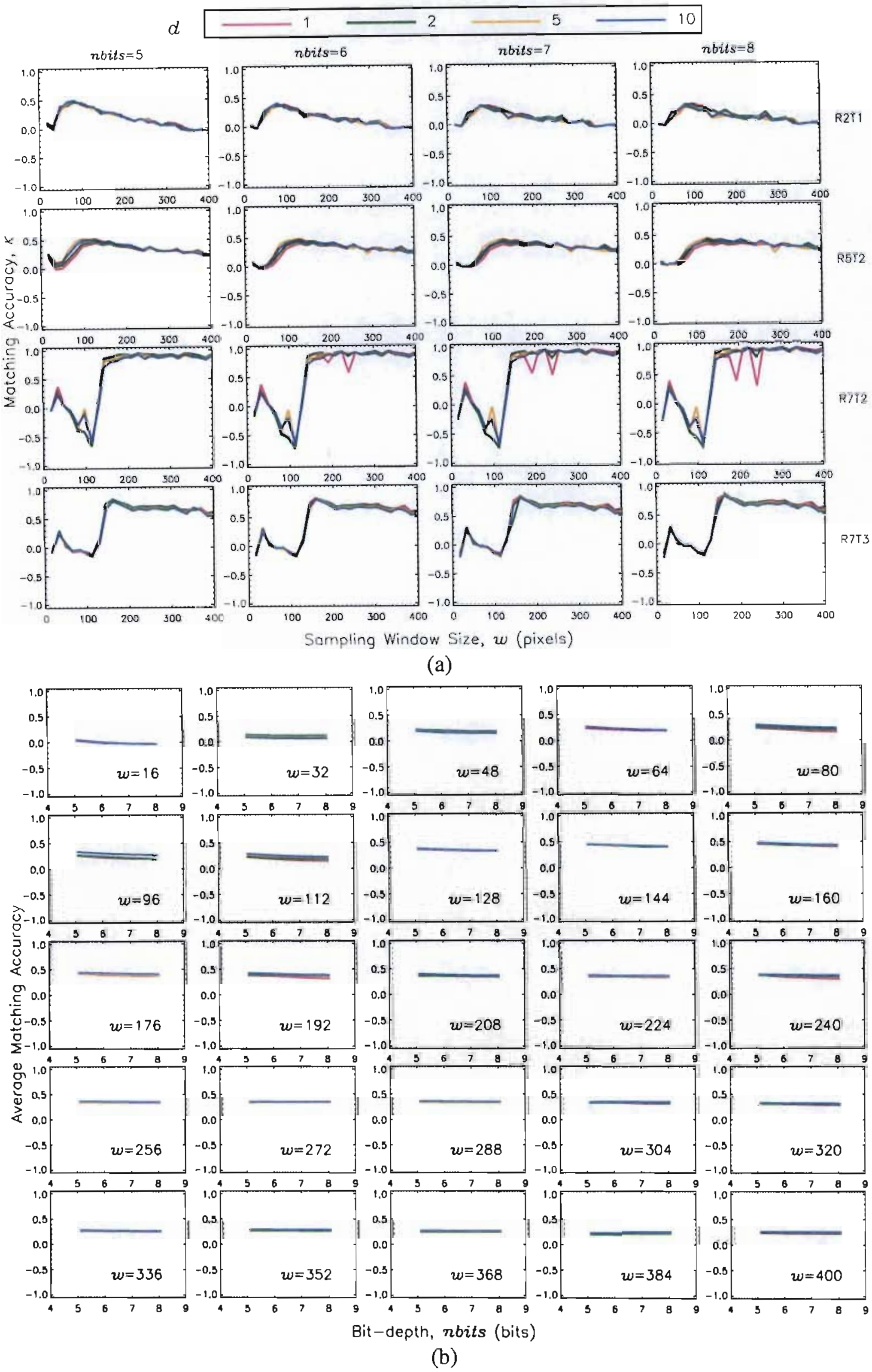
**Figures 9.4(a)** and **9.5(a)** show that matching accuracy varies with *w*, but that the variation is different for each reference image (R2, R5 and R7) and is similar for the same reference image (R7T2 and R7T3). It is, therefore, likely that the dependence of matching accuracy on *w* is influenced by the scale sizes of structures in the reference image.



(a)

(b)

**Figure 9.4**: Effect of *nbits* on matching accuracy for TM-$D_E$ applied to mosaic image pairs. **(a)** Typical examples of κ *vs.* *w*. The colours indicate different values of *nbits*. Each row contains the information for a single image pair. The vertical lines indicate the results of the ACF analysis. The solid line represents the ACF-width and the dotted line represents the ACF-height, at different bit-depths. **(b)** Average matching accuracy as a function of *d*, at different values of *nbits* to highlight the general dependence of matching accuracy on *nbits* for the mosaic images.



(a)

(b)

**Figure 9.5**: Effect of $d$ on matching accuracy for TM-$D_E$ applied to mosaic image pairs. **(a)** Typical examples of κ *vs.* $w$. The colours indicate different values of $d$. Each row contains the information for a single image pair. **(b)** Average matching accuracy as a function of bit-depth, at different values of $d$ to highlight the general dependence of matching accuracy on $d$ for the mosaic images.

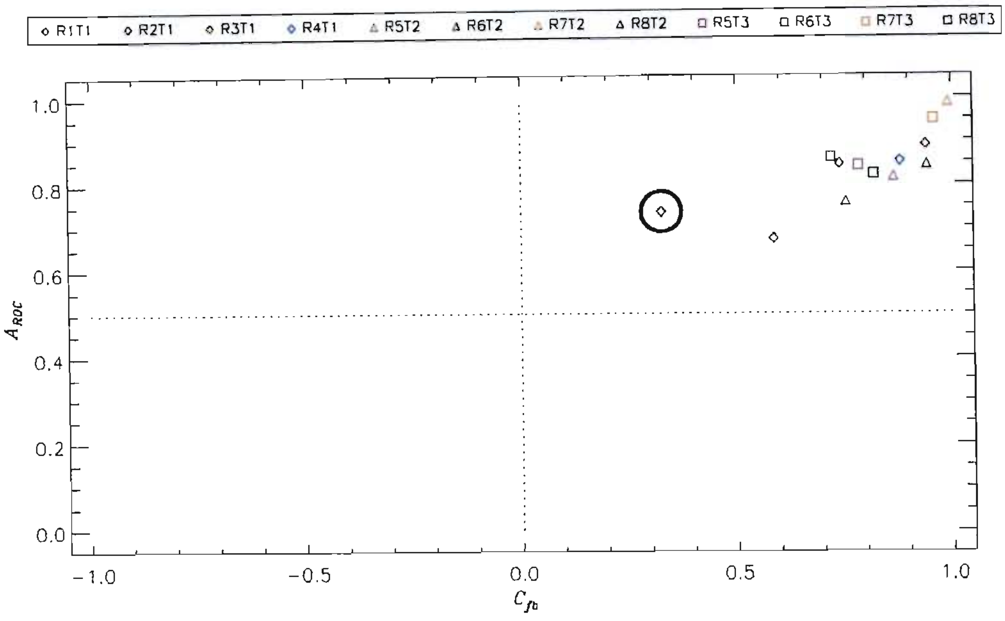


**Figure 9.6**: Scatter plot of the best results for TM-$D_E$ applied to the mosaic images. The circled point refers to a match for image pair R3T1 with the origin of the reference sampling window placed away from the origin of the reference image, to demonstrate that the position of the reference sampling window affects matching accuracy.

**Summary of Best Results**

The $A_{ROC}$- and $C_{fb}$-values corresponding to the best matches for each image pair are plotted in **Figure 9.6**. Matching is generally good with $0.65 < A_{ROC} < 1.0$ (average $0.84 \pm 0.08$), $0.55 < C_{fb} < 1.0$ (average $0.83 \pm 0.12$) and matching accuracy, $0.20 < \kappa < 1.0$ (average $0.59 \pm 0.20$). The high $A_{ROC}$-values indicate that most of the reference region was found in the test image, and the high $C_{fb}$-values indicate that the matched regions were easily discerned on the matching map.

The worst match occurred for image pair R3T1 with $A_{ROC}$=0.67 and $C_{fb}$=0.58. This poor match is most likely because reference image R3 is made up of many small tiles, each a slightly different shade of grey. The matching accuracy would then depend on the position of the reference sampling window. This dependence was confirmed by moving the reference sampling window away from the origin of the reference image by 100 pixels along the image width. The full TM-matching algorithm was applied to image pair R3T1 for the new reference sampling position and the result, circled in **Figure 9.6**, has a completely different matching result to that for the original reference sampling position, highlighting the sensitivity of the method to the position of the reference sampling window. The new (circled) result is however slightly worse than the original result.

### 9.4.3 Evaluation of TM-Matching with $D_{ES}$

The evaluation results of applying the TM-matching algorithm with the standardised Euclidean distance similarity metric, $D_{ES}$, to mosaic images are presented and discussed. The dependence of matching accuracy on *nbits*, *d* and *w* is examined. Results are presented in the formats described

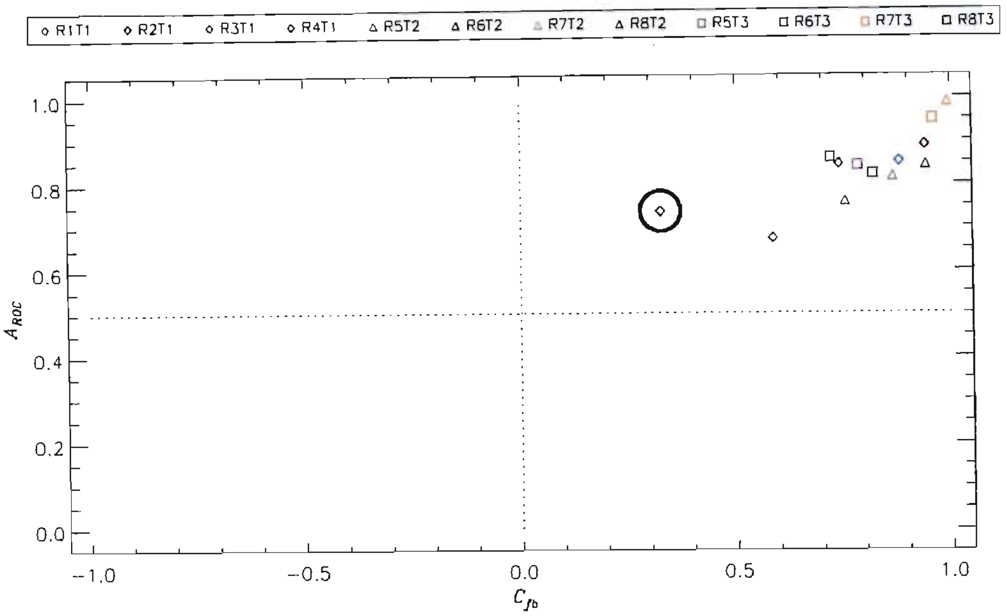


**Figure 9.6**: Scatter plot of the best results for TM-$D_E$ applied to the mosaic images. The circled point refers to a match for image pair R3T1 with the origin of the reference sampling window placed away from the origin of the reference image, to demonstrate that the position of the reference sampling window affects matching accuracy.

### Summary of Best Results

The $A_{ROC}$- and $C_{fb}$-values corresponding to the best matches for each image pair are plotted in **Figure 9.6**. Matching is generally good with $0.65 < A_{ROC} < 1.0$ (average $0.84 \pm 0.08$), $0.55 < C_{fb} < 1.0$ (average $0.83 \pm 0.12$) and matching accuracy, $0.20 < \kappa < 1.0$ (average $0.59 \pm 0.20$). The high $A_{ROC}$-values indicate that most of the reference region was found in the test image, and the high $C_{fb}$-values indicate that the matched regions were easily discerned on the matching map.

The worst match occurred for image pair R3T1 with $A_{ROC}$=0.67 and $C_{fb}$=0.58. This poor match is most likely because reference image R3 is made up of many small tiles, each a slightly different shade of grey. The matching accuracy would then depend on the position of the reference sampling window. This dependence was confirmed by moving the reference sampling window away from the origin of the reference image by 100 pixels along the image width. The full TM-matching algorithm was applied to image pair R3T1 for the new reference sampling position and the result, circled in **Figure 9.6**, has a completely different matching result to that for the original reference sampling position, highlighting the sensitivity of the method to the position of the reference sampling window. The new (circled) result is however slightly worse than the original result.

### 9.4.3 Evaluation of TM-Matching with $D_{ES}$

The evaluation results of applying the TM-matching algorithm with the standardised Euclidean distance similarity metric, $D_{ES}$, to mosaic images are presented and discussed. The dependence of matching accuracy on *nbits*, *d* and *w* is examined. Results are presented in the formats described in §8.2 (page 93). Dependence on matching accuracy is indicated by separation of the different coloured plots.

### Effect of *nbits* on Matching Accuracy

Typical examples of $\kappa$ *vs.* $w$ to demonstrate the dependence of matching accuracy on *nbits* are shown in **Figure 9.7(a)**. All examples, except R7T3, show no dependence of matching accuracy on bit-depth. The results for R7T3 are separated for most values of sampling window size, and show that matching accuracy improves as bit-depth decreases.

The results of the autocorrelation function analysis are plotted as vertical lines for each value of *nbits*, but there is very little difference in the results at the different bit-depths. The ACF-widths are significantly different from the ACF-heights. There appears to be poor correlation between the ACF results and the maximum values of $\kappa$. The ACF results are discussed in §9.7.

**Figure 9.7(b)** shows the average matching accuracy (over all image pairs) plotted as a function of $d$, to highlight the general dependence of matching accuracy on *nbits*. The dependence of matching accuracy on bit-depth is influenced by the size of the sampling window. There is no separation for $w \leq 64$ pixels, and the degree of separation increases slightly as sampling window size increases. It appears as if matching accuracy improves slightly as bit-depth decreases.

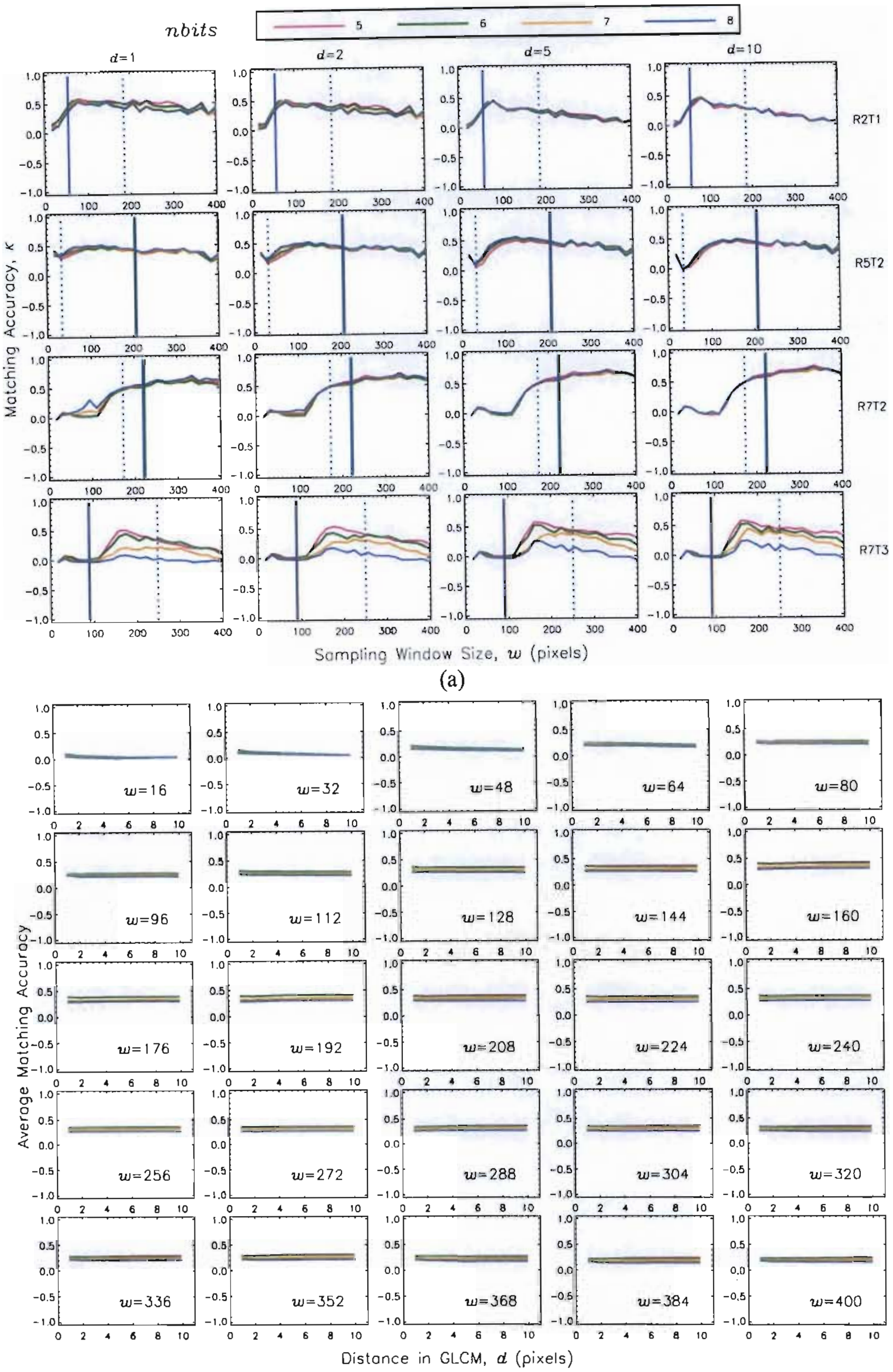
### Effect of $d$ on Matching Accuracy

Some examples of $\kappa$ *vs.* $w$ are shown in **Figure 9.8(a)** to demonstrate the dependence of matching accuracy on $d$. All plots show some separation indicating that there is a dependence of matching accuracy on $d$, but matching accuracy improves with decreasing $d$ for R2T1, R5T2 and R7T2 and improves with increasing $d$ for R7T3.

**Figure 9.8(b)** shows the average matching accuracy (over all image pairs) plotted as a function of *nbits*, to highlight the general dependence of matching accuracy on $d$. There is some separation of the different $d$-plots for $w \leq 80$ pixels. The separation is not significant, but it appears that matching accuracy improves slightly as $d$ decreases.

### Effect of $w$ on Matching Accuracy

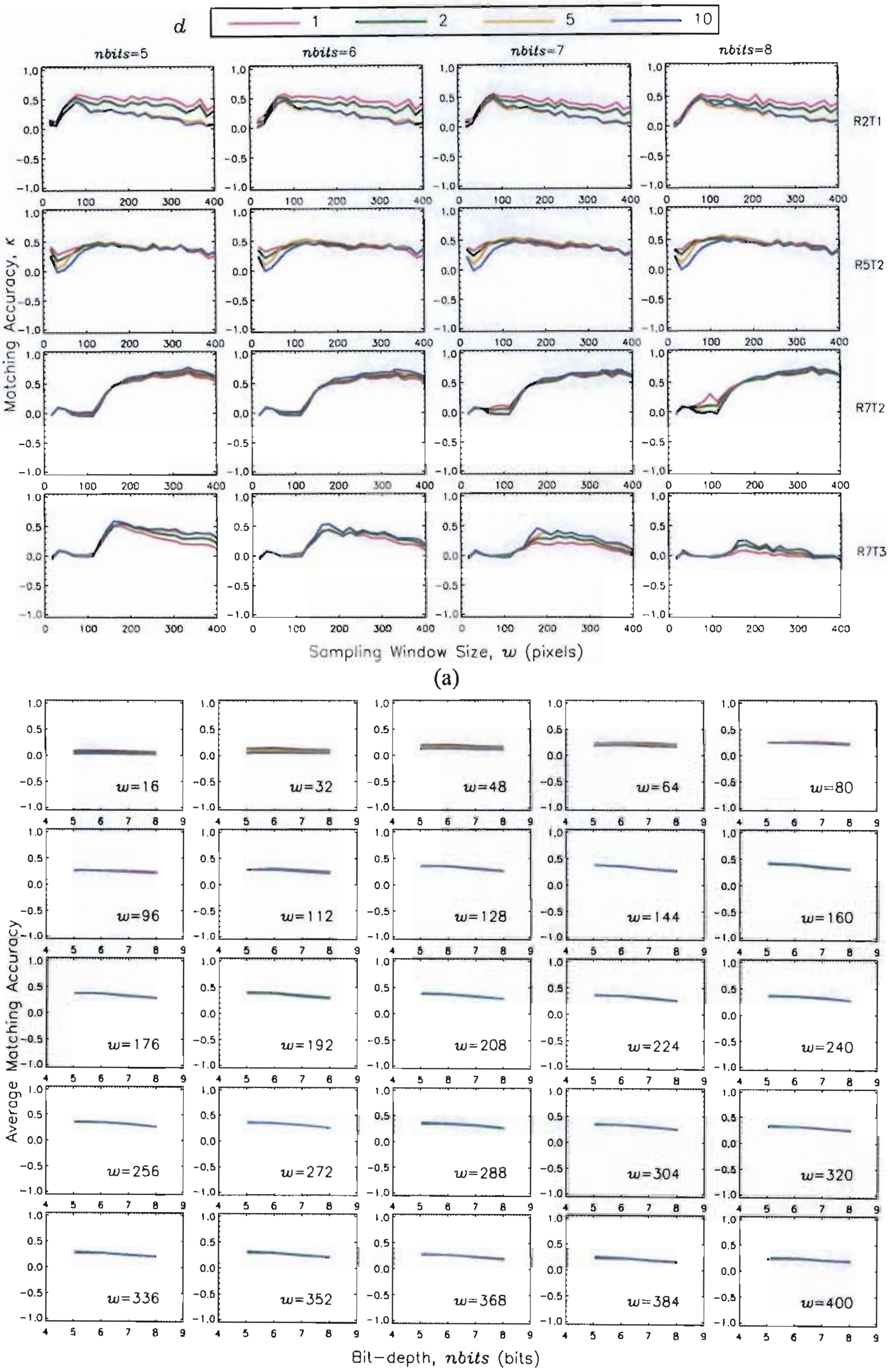
**Figures 9.7(a)** and **9.8(a)** show that matching accuracy is dependent on sampling window size. Since the variation of matching accuracy is different for the different reference images (R2, R5 and R7), but is similar for the R7T2 and R7T3, it is most likely that the dependence of matching accuracy on sampling window size, is influenced by the scale sizes of structures in the reference image.



(a)

(b)

**Figure 9.7**: Effect of *nbits* on matching accuracy for TM-$D_{ES}$ applied to mosaic image pairs. **(a)** Typical examples of κ *vs.* *w*. The colours indicate different values of *nbits*. Each row contains the information for a single image pair. The vertical lines indicate the results of the ACF analysis. The solid line represents the ACF-width and the dotted line represents the ACF-height, at different bit-depths. **(b)** Average matching accuracy as a function of *d*, at different bit-depths to highlight the general dependence of matching accuracy on bit-depth for the mosaic images.



(a)

(b)

**Figure 9.8**: Effect of $d$ on matching accuracy for TM-$D_{ES}$ applied to mosaic image pairs. **(a)** Typical examples of $\kappa$ *vs.* $w$. The colours indicate different values of $d$. Each row contains the information for a single image pair. **(b)** Average matching accuracy as a function of bit-depth, at different values of $d$ to highlight the general dependence of matching accuracy on $d$ for the mosaic images.

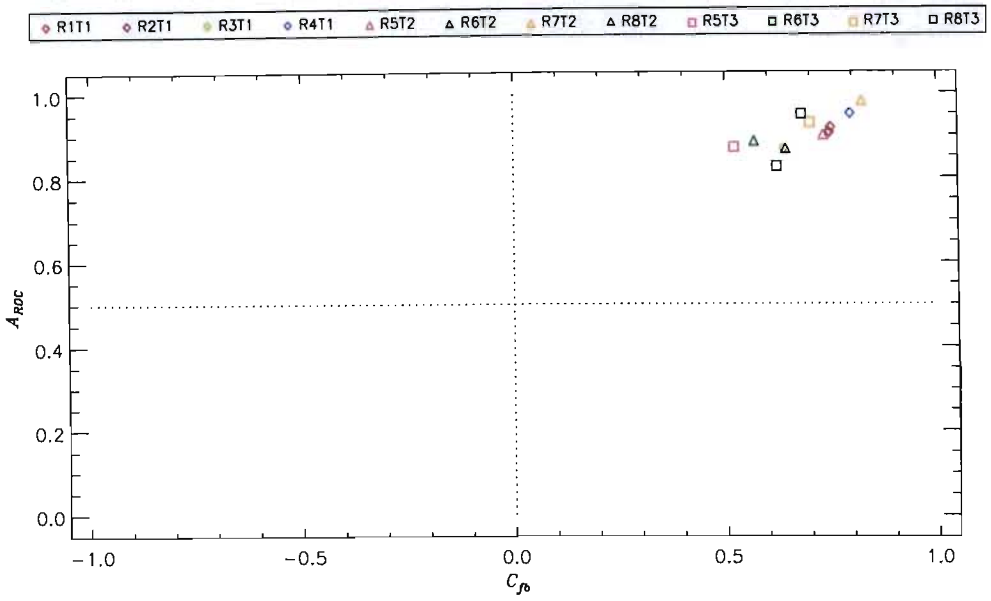


**Figure 9.9**: Scatter plot of the best results for TM-$D_{ES}$ applied to the mosaic images.

**Summary of Best Results**

The values of $A_{ROC}$ and $C_{fb}$ corresponding to the best matches for each image pair are plotted in **Figure 9.9**. All image pairs have $0.80 < A_{ROC} < 1.0$ (average $0.91 \pm 0.04$) and $0.50 < C_{fb} < 0.85$ (average $0.68 \pm 0.09$), with $0.40 < \kappa < 0.80$ (average $0.56 \pm 0.13$). The high $A_{ROC}$-values indicate that most of the reference ROI was matched to the corresponding region in the test image and the high $C_{fb}$-values indicate that the matched regions can be discerned on the matching map. The results are more clustered than those for $D_E$, with no outliers as was the case for image pair R3T1 with $D_E$.

Results of plotting $\kappa$ as a function of $w$ are generally smoother than the equivalent plots for TM-$D_E$, showing that standardisation has 'evened' out the results. It is also interesting that for TM-$D_E$, the separation of the different *nbits*-plots was only clearly seen for the small sampling windows, while for TM-$D_{ES}$, the separation of the different *nbits*-plots are clearly seen for most sampling window sizes, with almost no separation of the plots for the small sampling windows. This is also most likely another consequence of the standardisation, because results from the larger and smaller window sizes have been normalised.

### 9.4.4 Evaluation of TM-Matching with $D_M$

The evaluation of matching results from applying the TM-matching algorithm with the Mahalanobis distance similarity metric, $D_M$, to mosaic images are presented and discussed. The dependence of matching accuracy on *nbits*, *d* and *w* is examined. Results are presented in the formats described in §8.2 (page 93). Dependence on matching accuracy is indicated by separation of the different coloured plots.

### Effect of *nbits* on Matching Accuracy

Typical examples of κ *vs.* *w* to demonstrate the dependence of matching accuracy on *nbits* are shown in **Figure 9.10(a)**. There is some dependence of matching accuracy on bit-depth at the smaller values of *d* for R5T2 and for the larger sampling windows for R7T2. There is no separation of *nbits*-plots for R2T1 and R7T3. R5T2 shows matching accuracy increasing as bit-depth decreases and R7T2 shows matching accuracy increasing as bit-depth increases.

The results of the autocorrelation function analysis are plotted as vertical lines for each value of *nbits*, but there is very little difference in the results at the different bit-depths. The ACF-widths are significantly different from the ACF-heights. There appears to be poor correlation between the ACF results and the maximum values of κ. The ACF results are discussed in §9.7.

**Figure 9.10(b)** shows the average matching accuracy (over all image pairs) plotted as a function of *d*, to highlight the general dependence of matching accuracy on *nbits*. Most plots overlap almost exactly, but there is a slight separation of the *nbits*-plots at small values of *d* for the larger sampling windows. The separation is insufficient to determine a general pattern to dependence of matching accuracy on bit-depth.

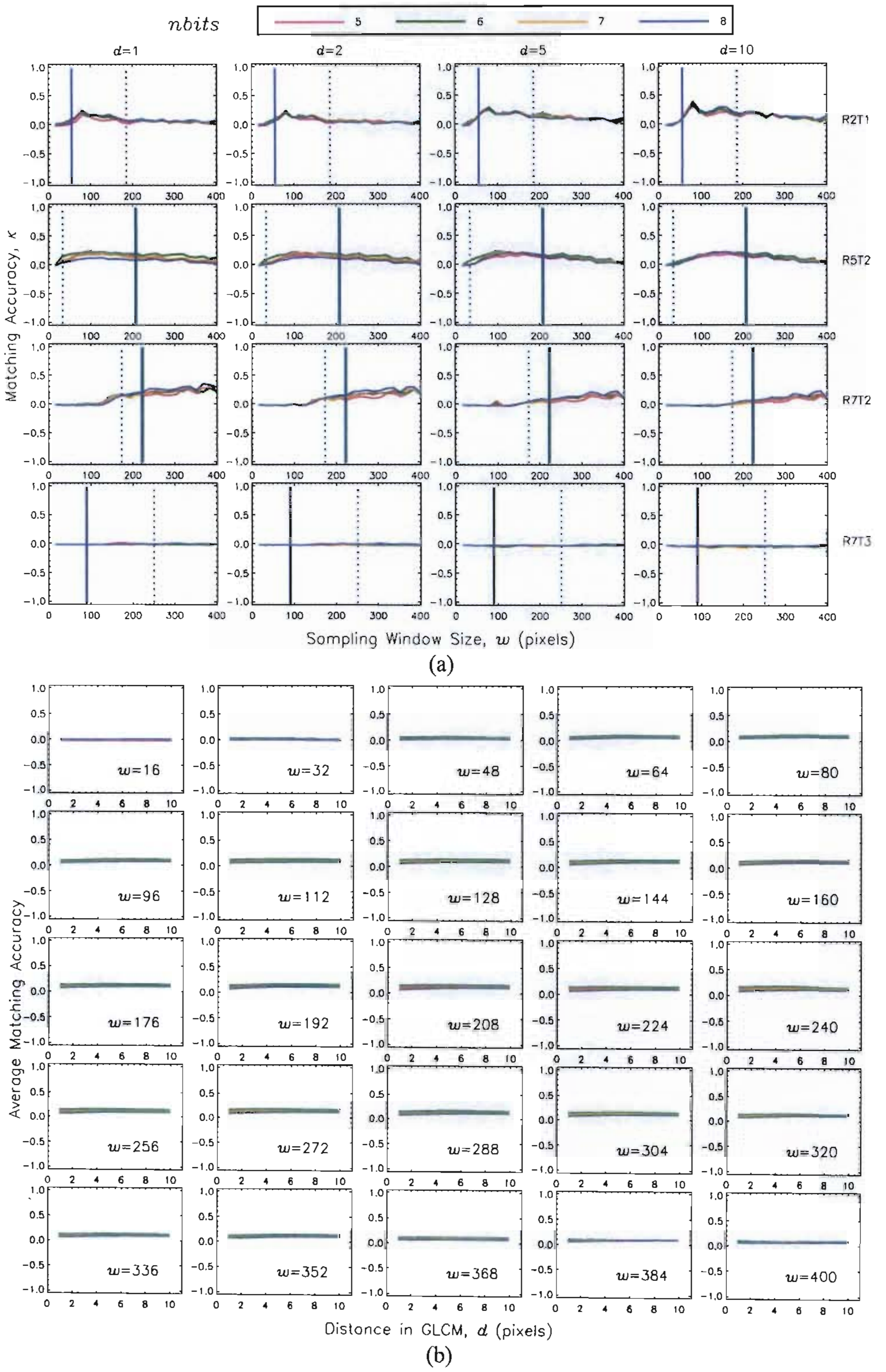
### Effect of *d* on Matching Accuracy

Some examples of κ *vs.* *w* are shown in **Figure 9.11(a)** to demonstrate the dependence of matching accuracy on *d*. Plots for R2T1, R5T2 and R7T2 are separated, indicating a dependence of matching accuracy on *d*. For R2T1 and R5T2 matching accuracy improves as *d* increases and for R7T2, matching accuracy improves as *d* decreases. All plots for R7T3 overlap completely and matching accuracy is independent of *d* for R7T3. The match for R7T3 is very poor as the *nbits*-plots have values of κ close to 0.

**Figure 9.11(b)** shows the average of the matching accuracy (over all image pairs) as a function of *nbits*, to highlight the general dependence of matching accuracy on *d*. There is no dependence of matching accuracy on *d* as all plots overlap almost exactly for all sampling window sizes.

### Effect of *w* on Matching Accuracy

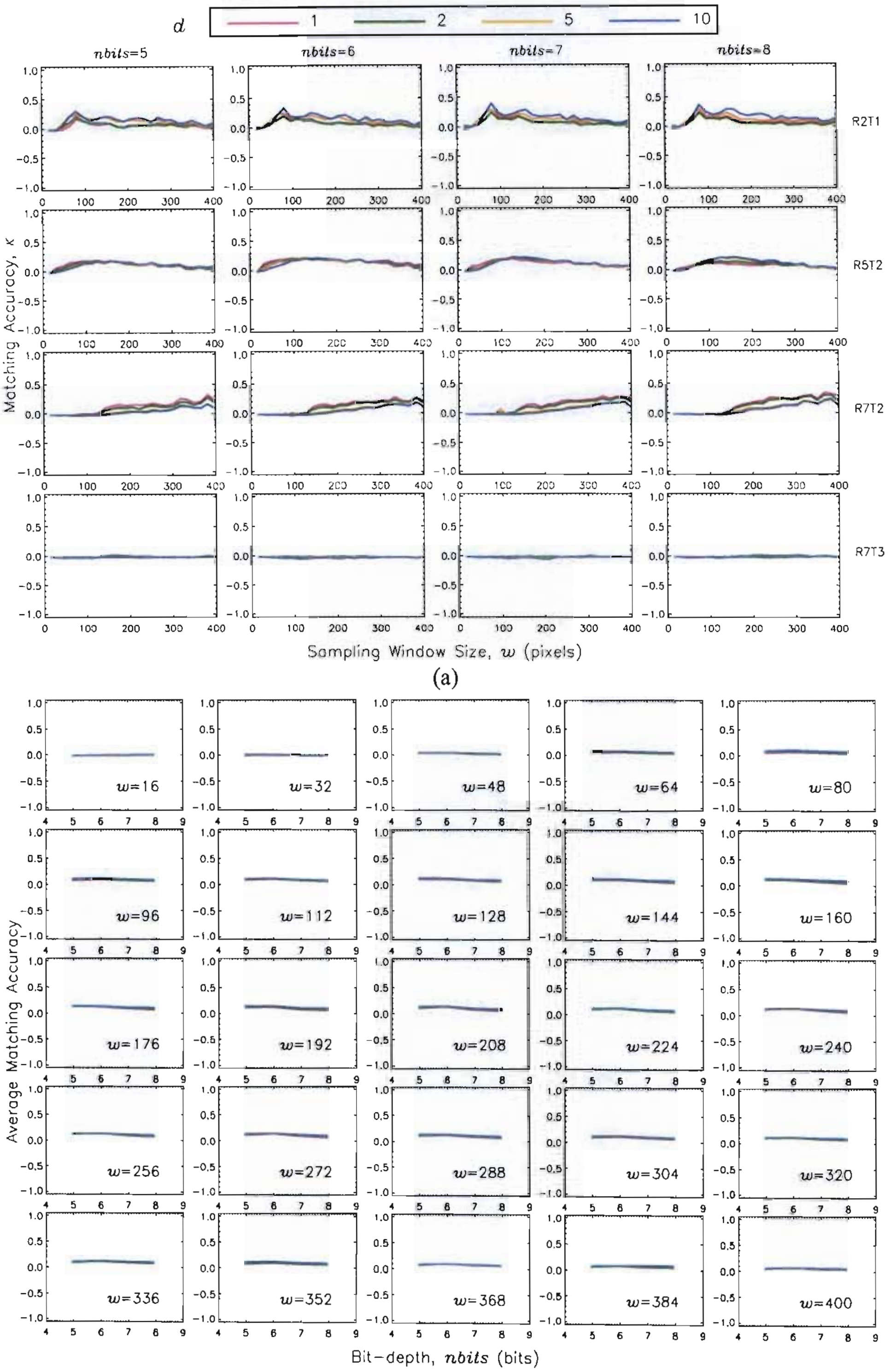
**Figures 9.10(a)** and **9.11(a)** show that matching accuracy is dependent on sampling window size. Since the variation of matching accuracy is different for the different reference images (R2, R5 and R7), it is most likely that the dependence of matching accuracy on sampling window size, is influenced by the scale sizes of structures in the reference image.



(a)

(b)

**Figure 9.10**: Effect of *nbits* on matching accuracy for TM-$D_M$ applied to mosaic image pairs. **(a)** Typical examples of $\kappa$ *vs.* $w$. The colours indicate different values of *nbits*. Each row contains the information for a single image pair. The vertical lines indicate the results of the ACF analysis. The solid line represents the ACF-width and the dotted line represents the ACF-height, at different bit-depths. **(b)** Average matching accuracy as a function of $d$, at different bit-depths to highlight the general dependence of matching accuracy on bit-depth for the mosaic images.



(a)

(b)

**Figure 9.11**: Effect of *d* on matching accuracy for TM-$D_M$ applied to mosaic image pairs. **(a)** Typical examples of κ *vs.* *w*. The colours indicate different values of *d*. Each row contains the information for a single image pair. **(b)** Average matching accuracy as a function of bit-depth, at different values of *d* to highlight the general dependence of matching accuracy on *d* for the mosaic images.

### Summary of Best Results

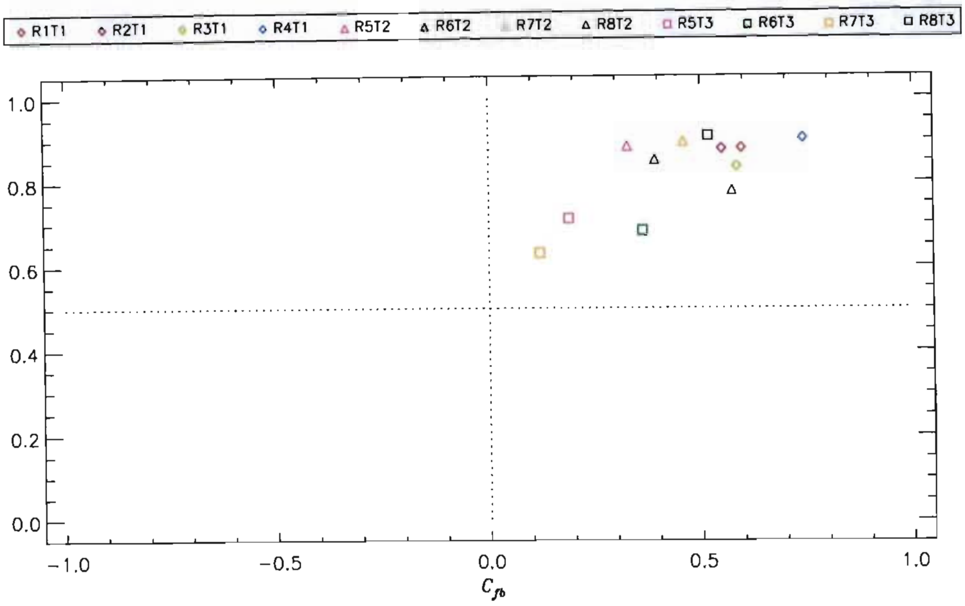


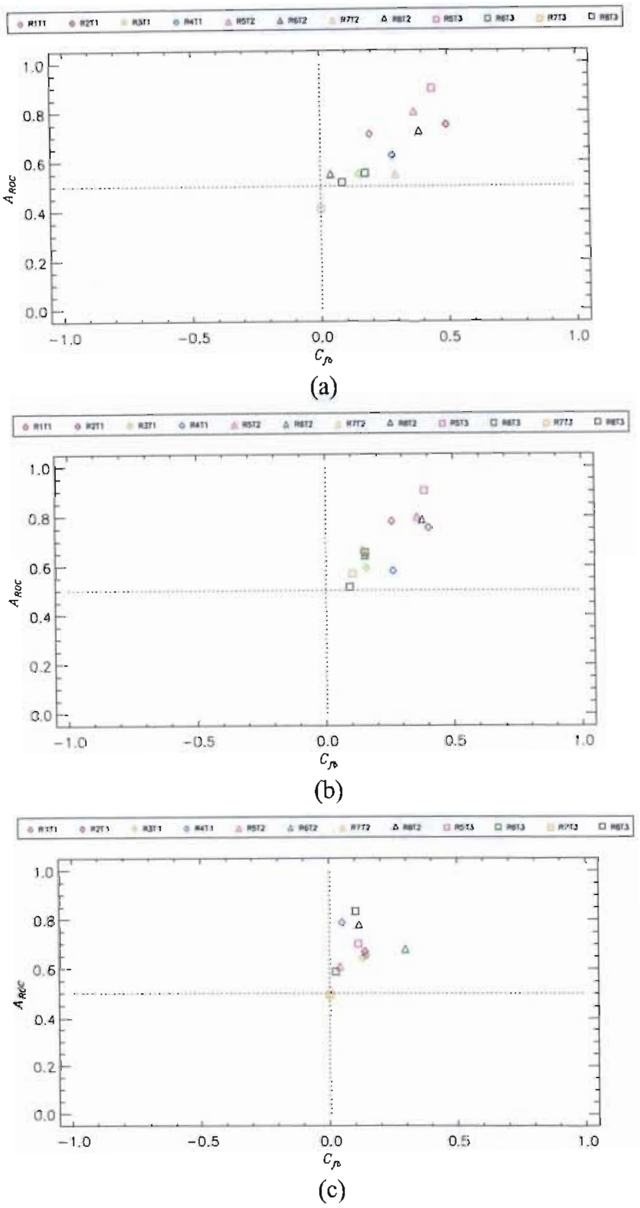
**Figure 9.12**: Scatter plot of the best results for TM-$D_M$ applied to the mosaic images.

The values of $A_{ROC}$ and $C_{fb}$ corresponding to the best matches for each image pair are plotted in **Figure 9.12**. All images have $0.6 < A_{ROC} < 0.95$ (average $0.82 \pm 0.10$) and $0.10 < C_{fb} < 0.75$ (average $0.45 \pm 0.18$), with $0.10 < \kappa < 0.75$ (average $0.31 \pm 0.17$). Results are quite scattered and matching is very poor for 3 image pairs (R5T3, R6T3 and R7T3). The high $A_{ROC}$-values indicate that most of the reference ROI was matched to the corresponding region in the test image, but the low $C_{fb}$-values indicate that the matched regions are difficult to discern on the matching map.

The average $A_{ROC}$- and $C_{fb}$-values are lower than those for $D_E$ and $D_{ES}$ and the poor matching accuracies might be an indication that there is poor correlation between the texture measures to warrant using $D_M$.

### 9.4.5 Effect of a Reduced Set of Texture Measures on Matching Accuracy

TM-matching was performed using a reduced set of texture measures to test whether the reduced set of texture measures might improve accuracy. The texture measures used were: correlation, inverse difference moment, entropy, sum entropy, difference entropy and the first information measure of correlation, and were selected according to the criteria laid out in Gotlieb & Kreyszig [1990] (page 70). The results are summarised in **Figure 9.13** and are worse than the results obtained by using 13 texture measures, when compared to the results in **Figure 9.20**. This might be a consequence of using the same reduced set of texture measures for all images since each image might have its own set of texture measures that best describe the textural information contained in the image.



(a)

(b)

(c)

**Figure 9.13**: Scatter plot of the best matching results for TM-matching with a reduced set of texture measures. Results are worse than those obtained from using 13 texture measures, and both $A_{ROC}$ and $C_{fb}$ have lower values. **(a)** TM-$D_E$. **(b)** TM-$D_{ES}$. **(c)** TM-$D_M$.

## 9.5 MI-Matching Evaluation Results

The MI-matching algorithm (**Algorithm 8.2** on page 98) was applied to the three test images on page 106. The results of the dependence of matching accuracy on the matching parameters are discussed for grey-level histograms and GLCMs.

### 9.5.1 Computation Times

Computation times were dependent on the values used for *nbits* and *w* and ranged from a few minutes to an hour, for a single map. Computation times were however faster than for TM-matching, since mutual information is computed faster than the texture measures.

### 9.5.2 Evaluation of MI-Matching with Histograms

The evaluation results of applying the MI-matching algorithm with histograms to mosaic images are presented and discussed. The dependence of matching accuracy on *nbits*, *nbins* and *w* is examined. Results are presented in the formats described in §8.2 (page 93). Dependence on matching accuracy is indicated by separation of the different coloured plots.

**Effect of *nbits* on Matching Accuracy**

Typical examples of $\kappa$ *vs.* $w$ to demonstrate the dependence of matching accuracy on *nbits* are shown in **Figure 9.14(a)**. There is some separation of the *nbits*-plots for some values of $d$, for each image pair. R2T1 shows matching accuracy improving as bit-depth decreases. R5T2 and R7T2 show some dependence of matching accuracy on bit-depth for the smaller sampling windows, but the separation is insufficient to extract a general pattern of behaviour. R7T3 shows some separation of the *nbits*-plots for $80 < w < 250$, with matching accuracy improving as bit-depth increases.

The results of the autocorrelation function analysis are plotted as vertical lines for each value of *nbits*, but there is very little difference in the results at the different bit-depths. The ACF-widths are significantly different from the ACF-heights. There appears to be poor correlation between the ACF results and the maximum values of $\kappa$. The ACF results are discussed in §9.7.

**Figure 9.14(b)** shows the average matching accuracy (over all image pairs) as a function of *nbins*, to highlight the general dependence of matching accuracy on *nbits*. There is some separation of the *nbits*-plots for $w < 128$ pixels, but the separation is insufficient to determine a general pattern to the dependency of matching accuracy on *nbits*. Matching accuracy is independent of *nbits* for the larger sampling windows, since all these plots overlap completely.

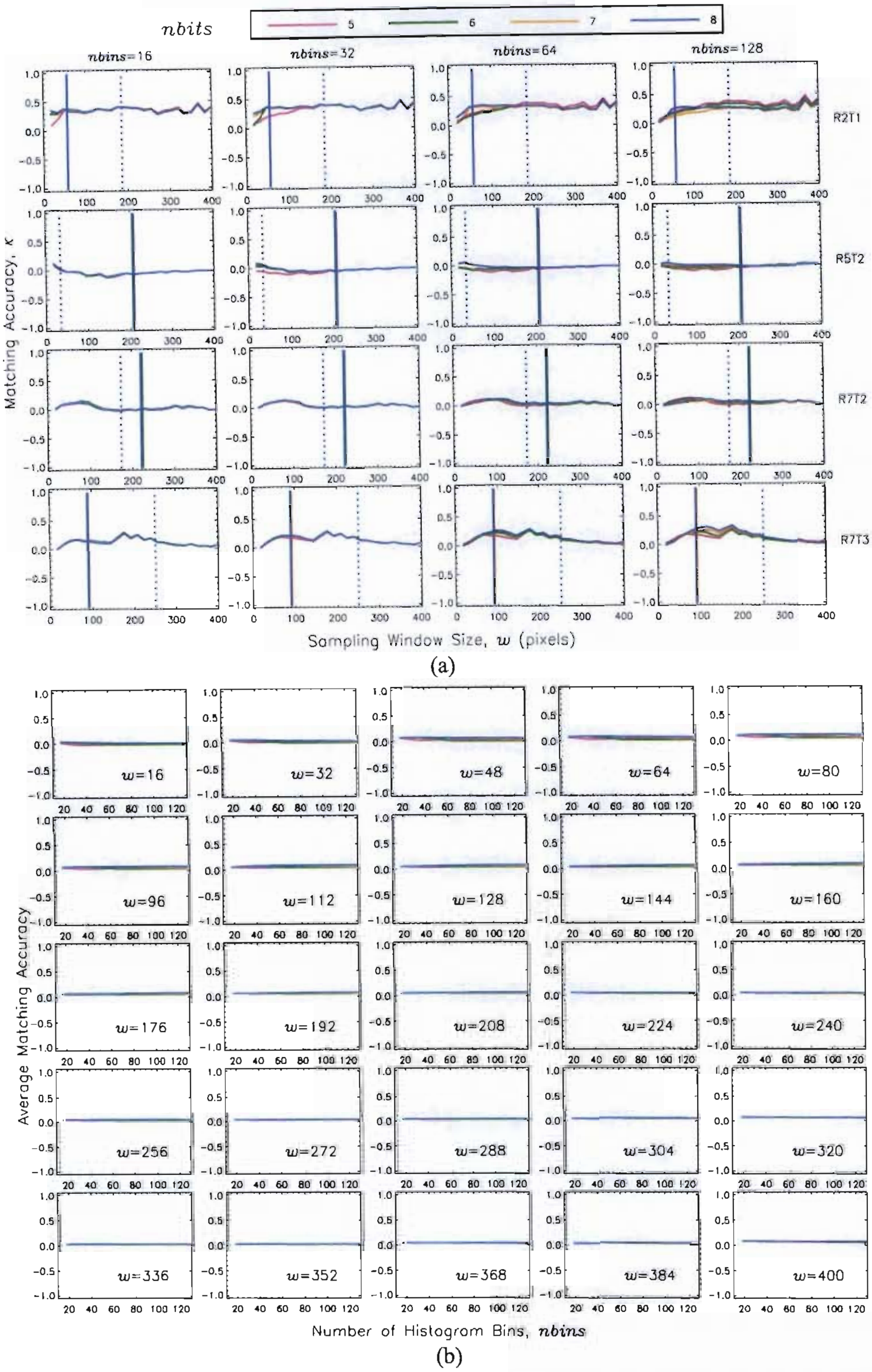


**Figure 9.14**: Effect of *nbits* on matching accuracy for MI-histograms applied to mosaic image pairs. **(a)** Typical examples of $\kappa$ *vs.* $w$. The colours indicate different values of *nbits*. Each row contains the information for a single image pair. The vertical lines indicate the results of the ACF analysis. The solid line represents the ACF-width and the dotted line represents the ACF-height, at different bit-depths. **(b)** Average matching accuracy as a function of $d$, at different bit-depths to highlight the general dependence of matching accuracy on bit-depth for the mosaic images.

### Effect of *nbins* on Matching Accuracy

Some examples of $\kappa$ *vs.* $w$ are shown in **Figure 9.15(a)** to demonstrate the dependence of matching accuracy on *nbins*. There is some separation for portions of all plots. R2T1 has the greatest degree of separation between *nbins*-plots at the higher bit-depths and shows that matching accuracy improve as *nbins* decreases for the higher bit-depths. R5T2 and R7T2 show a slight separation of the *nbins*-plots at the smaller sampling windows, with results showing that matching accuracy increases as *nbins* decreases. R7T3 shows separation of the *nbins*-plots for $80 < w < 250$, with matching accuracy improving as *nbins* increases, over this region.

**Figure 9.15(b)** shows the average matching accuracy (over all image pairs) plotted as a function of *nbits*, to highlight the general dependence of matching accuracy on *nbins*. There is slight separation of the plots for $w \leq 128$ pixels, and results show that matching accuracy improves as *nbins* decreases. For the larger sampling windows, there no dependence of matching accuracy on *nbins* as all plots overlap almost exactly.

### Effect of $w$ on Matching Accuracy

**Figures 9.14(a)** and **9.15(a)** show that matching accuracy varies with sampling window size, but the variation is not as clear as it was for the distance similarity metrics. The plots of *nbits* and *nbins* are generally much flatter than the results for the distance similarity metrics. Although, since both **Figures 9.14(b)** and **9.15(b)** show the greatest separation for the smaller sampling windows, it may be concluded that matching accuracy is optimal for the smaller sampling windows.

### Summary of Best Results

The values of $A_{ROC}$ and $C_{fb}$ corresponding to the best matches for each image pair are plotted in **Figure 9.16**. All images have $0.65 < A_{ROC} < 1.0$ (average $0.77 \pm 0.10$) and $0.08 < C_{fb} < 0.70$ (average $0.29 \pm 0.18$), with $0.0 < \kappa < 0.40$ (average $0.16 \pm 0.11$). There is a wider spread of results compared to the results for the three distance similarity metrics. The generally low $C_{fb}$-values indicate that the matched regions are difficult to discern on the matching map. Specific images (e.g. R2T1) seem to respond better to histograms than others, but the matching accuracy of $\kappa$=0.39 is still lower than the results for the distance similarity metrics.

In general, matching accuracies are much lower than for the distance similarity metrics.

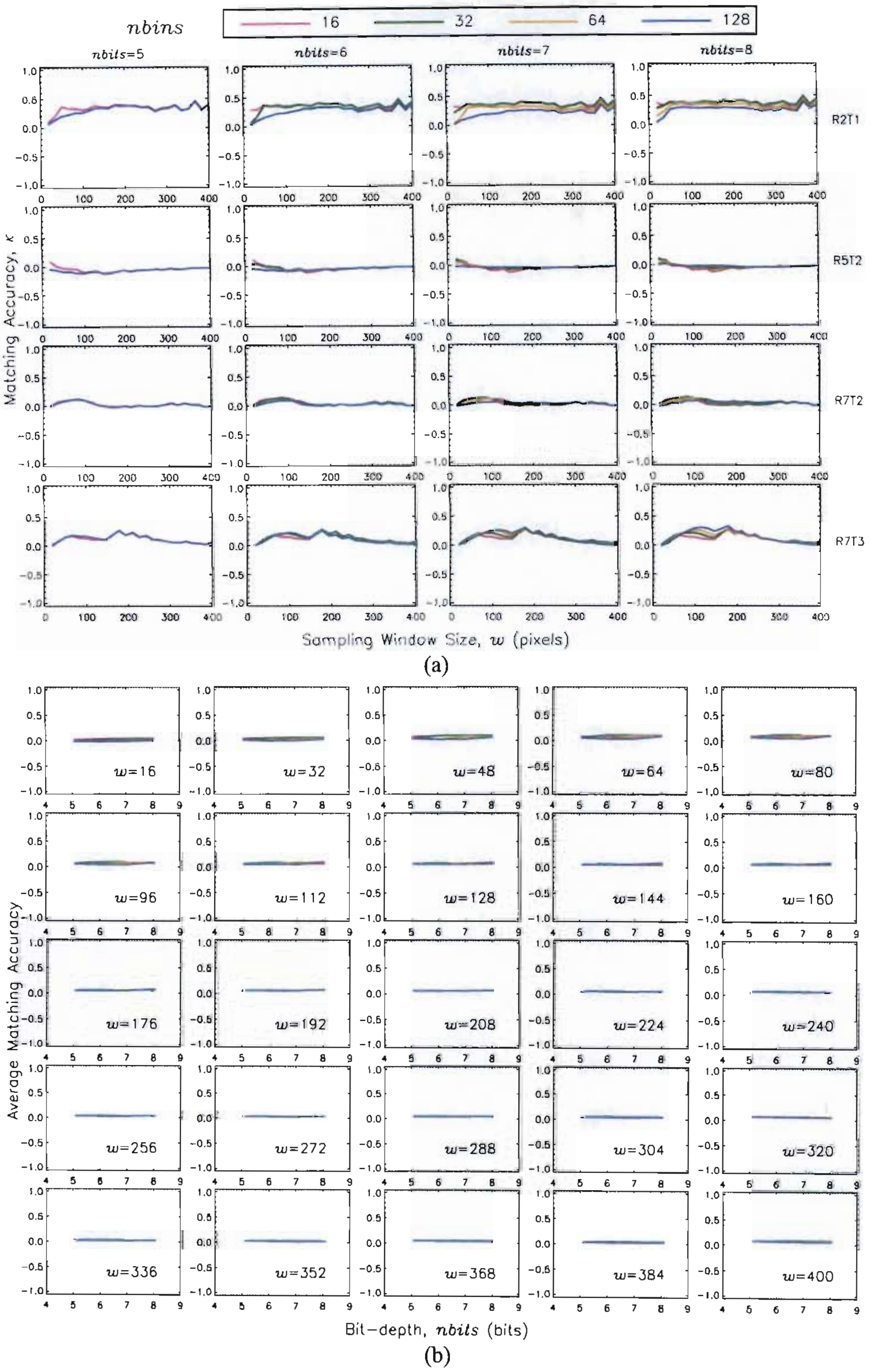


**Figure 9.15**: Effect of *nbins* on matching accuracy for MI-histograms applied to mosaic image pairs. **(a)** Typical examples of κ *vs.* *w*. The colours indicate different values of *nbins*. Each row contains the information for a single image pair. **(b)** Average matching accuracy as a function of bit-depth, at different values of *nbins* to highlight the general dependence of matching accuracy on *nbins* for the mosaic images.

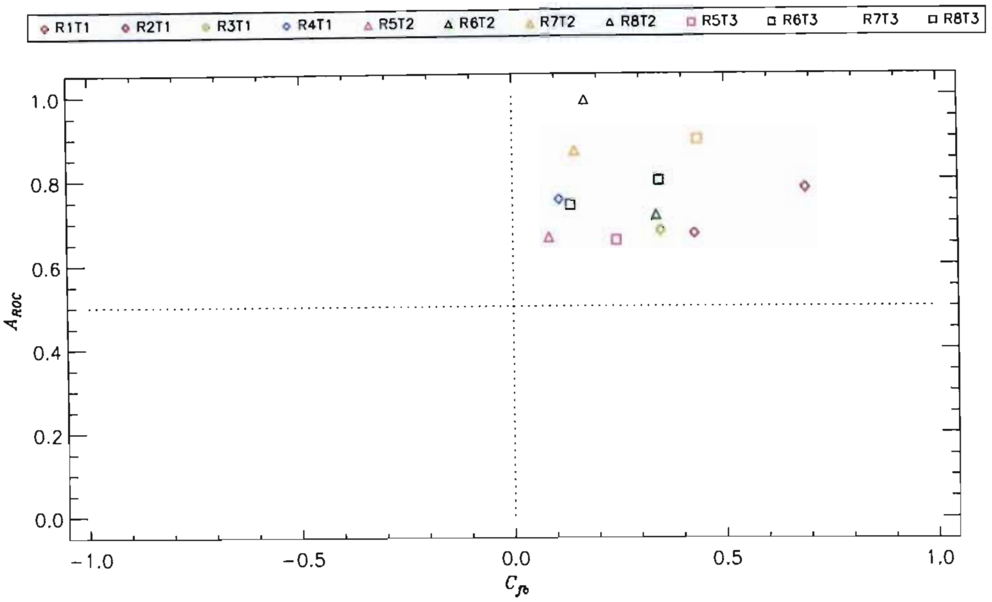


**Figure 9.16**: Scatter plot of the best results for MI-histograms applied to the mosaic images.

### 9.5.3 Evaluation of MI-Matching with GLCMs

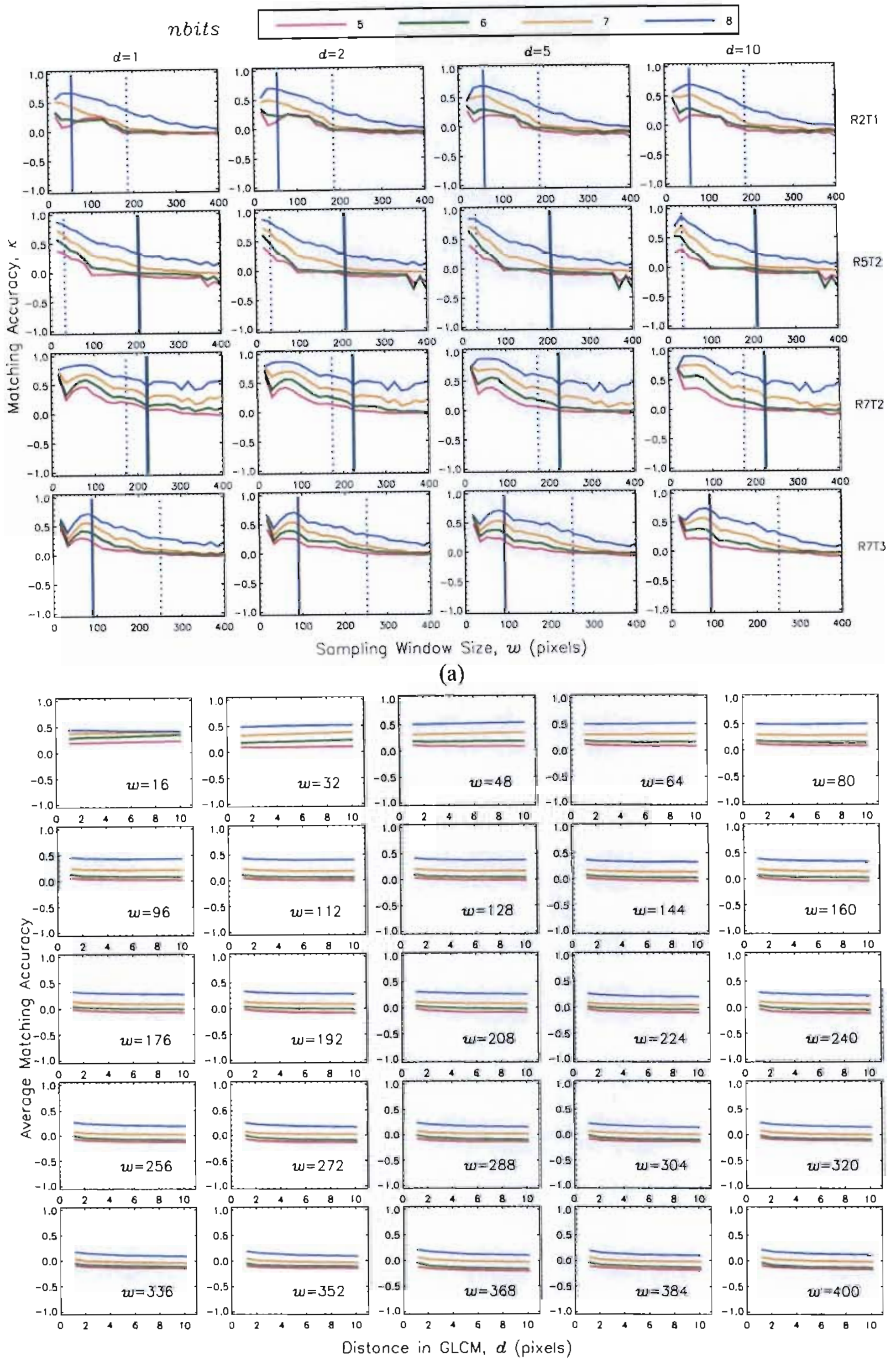
The evaluation results of applying the MI-matching algorithm with GLCMs to mosaic images are presented and discussed. The dependence of matching accuracy on *nbits*, *d* and *w* is examined. Results are presented in the formats described in §8.2 (page 93). Dependence on matching accuracy is indicated by separation of the different coloured plots.

#### Effect of *nbits* on Matching Accuracy

Typical examples of $\kappa$ *vs.* $w$ to demonstrate the dependence of matching accuracy on *nbits* are shown in **Figure 9.17(a)**. There is clear separation of the *nbits*-plots, with all results showing that matching accuracy improves as bit-depth increases.

The results of the autocorrelation function analysis are plotted as vertical lines for each value of *nbits*, but there is very little difference in the results at the different bit-depths. The ACF-widths are significantly different from the ACF-heights. There appears to be some correlation between the ACF results and the maximum values of $\kappa$, for R2T1, R5T2 and R7T3. The ACF results are discussed in §9.7.

**Figure 9.17(b)** shows the average matching accuracy (over all image pairs) plotted as a function of $d$, to highlight the general dependence of matching accuracy on *nbits*. There is a clear separation of the *nbits*-plots for all sampling window sizes. Matching accuracy improves as bit-depth increases, irrespective of the sampling window size.



(a)

(b)

**Figure 9.17**: Effect of *nbits* on matching accuracy for MI-GLCMs applied to mosaic image pairs. **(a)** Typical examples of κ *vs*. *w*. The colours indicate different values of *nbits*. Each row contains the information for a single image pair. The vertical lines indicate the results of the ACF analysis. The solid line represents the ACF-width and the dotted line represents the ACF-height, at different bit-depths. **(b)** Average matching accuracy as a function of *d*, at different bit-depths to highlight the general dependence of matching accuracy on bit-depth for the mosaic images.

### Effect of $d$ on Matching Accuracy

Some examples of $\kappa$ *vs.* $w$ are shown in **Figure 9.18(a)** to demonstrate the dependence of matching accuracy on $d$ in the GLCM calculation. Generally, all plots show some separation, with the more clear separations showing matching accuracy increasing for decreasing $d$.

**Figure 9.18(b)** shows the average of the matching accuracy (over all image pairs) as a function of *nbits*, to highlight the general dependence of matching accuracy on $d$. The degree of separation increases as the sampling window size increases, but matching accuracy improves as $d$ decreases for all sampling window sizes.

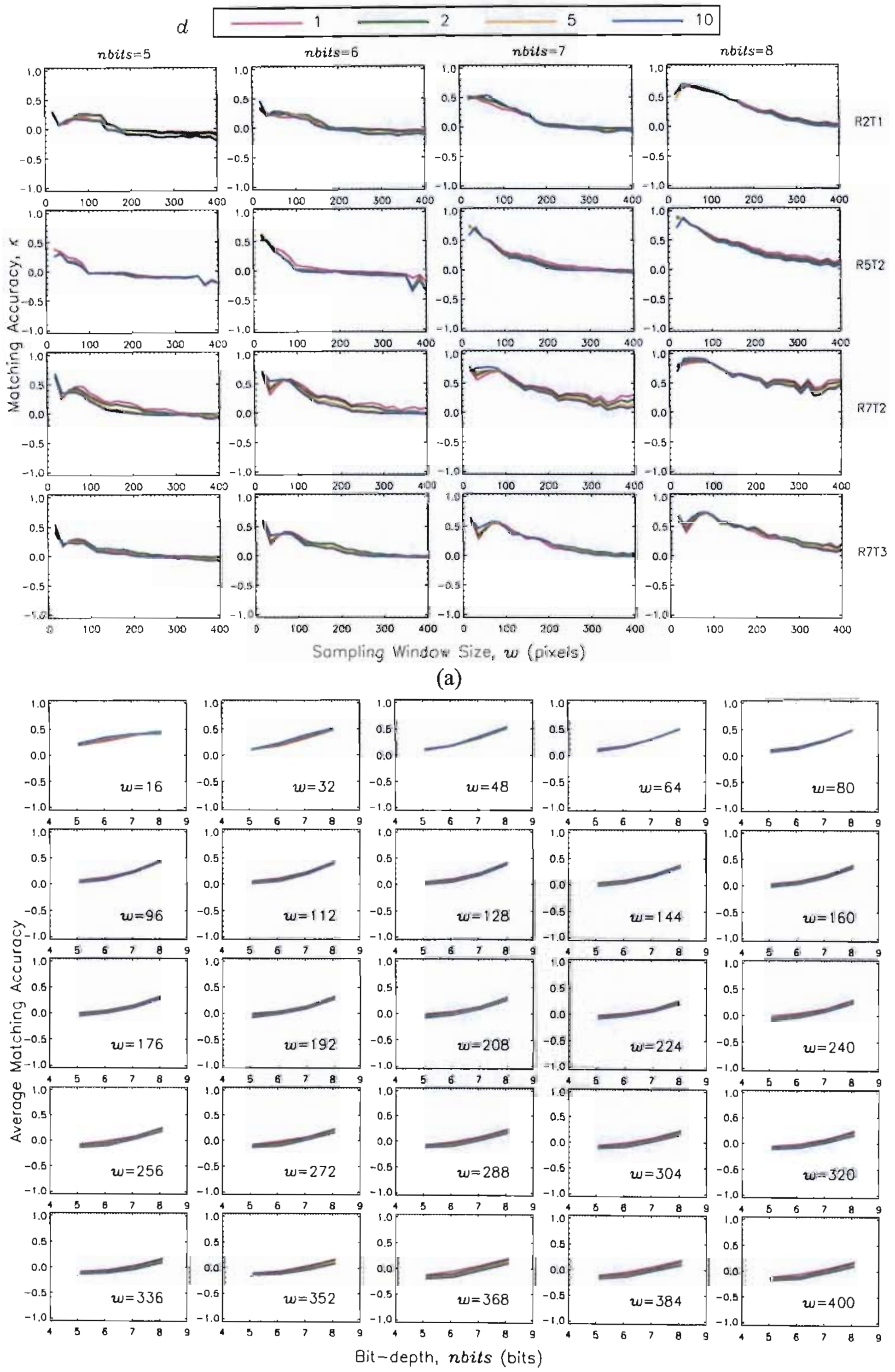
### Effect of $w$ on Matching Accuracy

**Figures 9.17(a)** and **9.18(a)** show that there is some dependence of matching accuracy on sampling window size, and that matching accuracy appears to be optimal for the smaller sampling windows. The position of the peak in **Figures 9.17(a)** and **9.18(a)** varies with each image, and so there is possibly a relationship to the scale sizes of the textures in the images.

### Summary of Best Results

The values of $A_{ROC}$ and $C_{fb}$ corresponding to the best matches for each image pair are plotted in **Figure 9.19**. All image pairs have $0.85 < A_{ROC} < 1.0$ (average $0.94 \pm 0.06$) and $0.40 < C_{fb} < 0.95$ (average $0.70 \pm 0.16$), with $0.25 < \kappa < 0.95$ (average $0.63 \pm 0.19$). Results are clustered in the high $A_{ROC}$ high $C_{fb}$ region of the scatter plot and all but one image (R6T3) have $A_{ROC} \geq 0.8$ and $C_{fb} \geq 0.5$, indicating that most of the region to be matched was identified and that the matched region is clearly discernible from the background.

Matching accuracy improves as bit-depth increases and appears to improve as $d$ decreases. Matching accuracy is optimal for the smaller sampling windows. Results for the dependence of matching accuracy on bit-depth was independent of sampling window size, unlike for the distance similarity metrics and MI-histograms.



(a)

(b)

**Figure 9.18**: Effect of *d* on matching accuracy for MI-GLCMs applied to mosaic image pairs. **(a)** Typical examples of κ *vs. w*. The colours indicate different values of *d*. Each row contains the information for a single image pair. **(b)** Average matching accuracy as a function of bit-depth, at different values of *d* to highlight the general dependence of matching accuracy on *d* for the mosaic images.

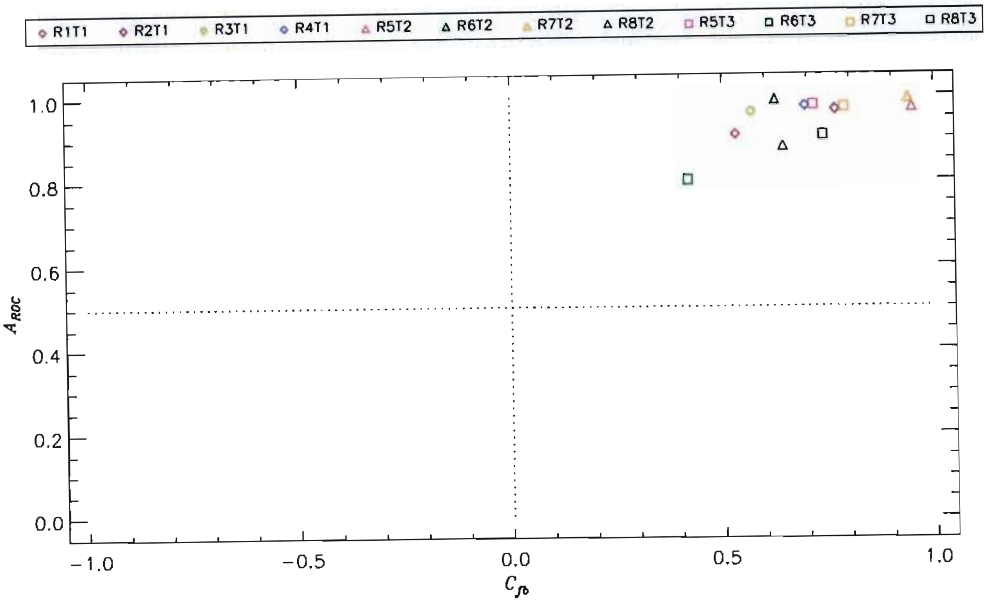


**Figure 9.19**: Scatter plot of the best results for MI-GLCMs applied to the mosaic images.

## 9.6 Overall Matching Results

### 9.6.1 Results

**Figure 9.20** shows the best matching results for each of the twelve mosaic image pairs, for the five matching methods. The $A_{ROC}$- and $C_{fb}$-values are clustered for TM-$D_E$, TM-$D_{ES}$ and MI-GLCMs, but are quite scattered for TM-$D_M$ and MI-histograms. All the results have $A_{ROC}>0.5$ and $C_{fb}>0.0$, indicating that the match is better than random and that the matched region is brighter than the background. Both $A_{ROC}$- and $C_{fb}$-values are quite high for TM-$D_E$, TM-$D_{ES}$ and MI-GLCMs, while TM-$D_M$ and MI-histograms both have low $C_{fb}$-values. This is confirmed by examining the averages of best matching accuracies that are summarised in **Table 9.4** for each method.

**Table 9.4**: Average of the best matching accuracies for each matching method applied to mosaic images.

| Method | Average $A_{ROC}$ | Average $C_{fb}$ | Average $\kappa$ |
|---|---|---|---|
| TM-$D_E$ | 0.84±0.08 | 0.79±0.18 | 0.55±0.23 |
| TM-$D_{ES}$ | 0.91±0.04 | 0.68±0.09 | 0.56±0.13 |
| TM-$D_M$ | 0.82±0.09 | 0.45±0.18 | 0.31±0.17 |
| MI-histograms | 0.77±0.10 | 0.29±0.18 | 0.15±0.11 |
| MI-GLCMs | 0.94±0.06 | 0.70±0.16 | 0.63±0.19 |

The combination of matching parameters that yielded the maximum $\kappa$-value is listed for each image pair in **Tables 9.5** and **9.6** for TM- and MI-matching, respectively. It appears that TM-matching has optimal matches for large sampling windows and low bit-depths, while MI-matching has optimal matches for small sampling windows and high bit-depths. These results are consistent with the dependence of matching accuracy on bit-depths and sampling window sizes for MI-GLCMs.

It was hoped that the untransformed reference textures (R3T1, R4T1, R5T2, R7T3 and R8T2) would have near perfect matching results, for all methods. However, apart from R7T2 (which has

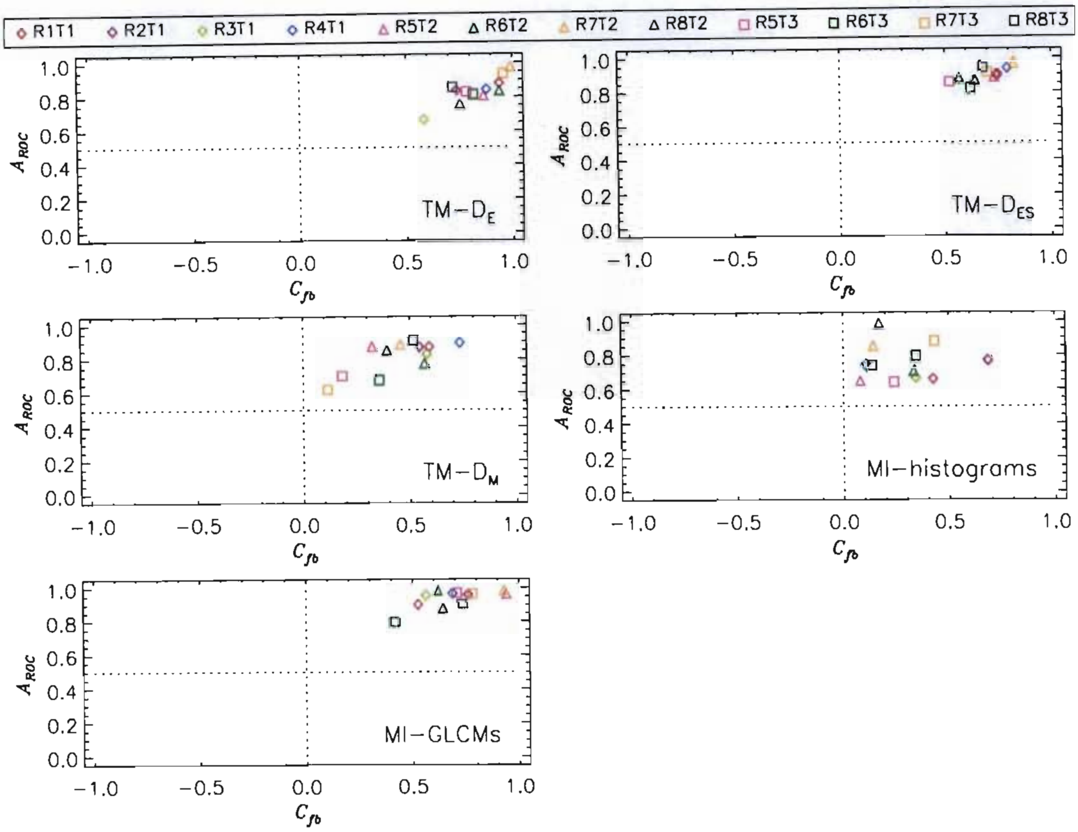


**Figure 9.20**: Scatter plot of $A_{ROC}$ *vs.* $C_{fb}$ with the best matching accuracies for TM- and MI-matching applied to the mosaic images.

κ>0.80 for all methods) and R5T2 (which has κ=0.90 for MI-GLCMs), matching results are very poor for the rest. Also, R7T3, which has R7 rotated and scaled in T3 has κ=0.86 for TM-$D_E$. Upon examination of **Figure 9.20**, it can be seen that image pair with the best and worst matching accuracy, is different for each method. There is therefore no clear pattern to the behaviour of the matching algorithms with the matching performances for each image pair.

**Table 9.5**: Matching parameters and evaluation results for the best matches of TM-matching applied to mosaic images, listed for each image pair.

| Images | TM-$D_E$ | | | | | | TM-$D_{ES}$ | | | | | | TM-$D_M$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *w* | *d* | *nbits* | $A_{ROC}$ | $C_{fb}$ | κ | *w* | *d* | *nbits* | $A_{ROC}$ | $C_{fb}$ | κ | *w* | *d* | *nbits* | $A_{ROC}$ | $C_{fb}$ | κ |
| R1T1 | 176 | 10 | 6 | 0.89 | 0.94 | 0.73 | 320 | 10 | 8 | 0.92 | 0.75 | 0.63 | 400 | 2 | 6 | 0.88 | 0.60 | 0.45 |
| R2T1 | 80 | 10 | 5 | 0.85 | 0.74 | 0.51 | 80 | 1 | 5 | 0.91 | 0.75 | 0.61 | 80 | 10 | 7 | 0.88 | 0.55 | 0.41 |
| R3T1 | 160 | 1 | 5 | 0.67 | 0.58 | 0.20 | 160 | 1 | 8 | 0.87 | 0.64 | 0.47 | 176 | 1 | 6 | 0.83 | 0.59 | 0.39 |
| R4T1 | 80 | 5 | 7 | 0.85 | 0.88 | 0.62 | 320 | 1 | 5 | 0.95 | 0.80 | 0.72 | 160 | 10 | 5 | 0.90 | 0.74 | 0.60 |
| R5T2 | 128 | 5 | 5 | 0.82 | 0.86 | 0.55 | 128 | 5 | 8 | 0.90 | 0.73 | 0.59 | 128 | 5 | 7 | 0.88 | 0.33 | 0.25 |
| R6T2 | 80 | 1 | 6 | 0.84 | 0.94 | 0.65 | 160 | 10 | 7 | 0.89 | 0.57 | 0.44 | 368 | 2 | 6 | 0.78 | 0.57 | 0.32 |
| R7T2 | 224 | 1 | 8 | 0.99 | 0.99 | 0.97 | 336 | 10 | 5 | 0.98 | 0.82 | 0.80 | 368 | 1 | 8 | 0.90 | 0.46 | 0.36 |
| R8T2 | 80 | 10 | 5 | 0.76 | 0.75 | 0.39 | 128 | 2 | 5 | 0.87 | 0.64 | 0.47 | 112 | 10 | 6 | 0.85 | 0.39 | 0.28 |
| R5T3 | 128 | 10 | 5 | 0.84 | 0.78 | 0.54 | 128 | 10 | 5 | 0.88 | 0.52 | 0.39 | 240 | 1 | 7 | 0.71 | 0.19 | 0.08 |
| R6T3 | 288 | 1 | 7 | 0.82 | 0.82 | 0.53 | 288 | 10 | 7 | 0.83 | 0.62 | 0.41 | 288 | 10 | 7 | 0.68 | 0.36 | 0.13 |
| R7T3 | 160 | 2 | 8 | 0.95 | 0.96 | 0.86 | 160 | 5 | 5 | 0.93 | 0.70 | 0.60 | 176 | 1 | 5 | 0.63 | 0.12 | 0.03 |
| R8T3 | 176 | 10 | 7 | 0.86 | 0.72 | 0.52 | 304 | 10 | 5 | 0.95 | 0.68 | 0.61 | 288 | 1 | 5 | 0.91 | 0.52 | 0.42 |

**Table 9.6**: Matching parameters and evaluation results for the best matches of MI-matching applied to mosaic images, listed for each image pair.

| Images | MI-Histograms | | | | | | MI-GLCMs | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $w$ | $nbins$ | $nbits$ | $A_{ROC}$ | $C_{fb}$ | $\kappa$ | $w$ | $d$ | $nbits$ | $A_{ROC}$ | $C_{fb}$ | $\kappa$ |
| R1T1 | 400 | 16 | 8 | 0.67 | 0.43 | 0.15 | 32 | 10 | 8 | 0.91 | 0.53 | 0.44 |
| R2T1 | 368 | 32 | 8 | 0.78 | 0.69 | 0.39 | 48 | 5 | 8 | 0.97 | 0.77 | 0.72 |
| R3T1 | 400 | 16 | 6 | 0.68 | 0.35 | 0.13 | 48 | 10 | 8 | 0.96 | 0.57 | 0.53 |
| R4T1 | 16 | 32 | 8 | 0.75 | 0.11 | 0.06 | 16 | 1 | 8 | 0.98 | 0.70 | 0.66 |
| R5T2 | 16 | 16 | 8 | 0.67 | 0.08 | 0.03 | 16 | 2 | 8 | 0.97 | 0.95 | 0.90 |
| R6T2 | 96 | 32 | 5 | 0.72 | 0.34 | 0.15 | 208 | 1 | 8 | 0.99 | 0.63 | 0.62 |
| R7T2 | 80 | 16 | 6 | 0.87 | 0.15 | 0.11 | 48 | 10 | 8 | 1.00 | 0.94 | 0.93 |
| R8T2 | 64 | 128 | 8 | 0.99 | 0.17 | 0.17 | 16 | 2 | 7 | 0.88 | 0.64 | 0.49 |
| R5T3 | 384 | 16 | 5 | 0.66 | 0.24 | 0.08 | 32 | 10 | 8 | 0.98 | 0.71 | 0.69 |
| R6T3 | 96 | 32 | 5 | 0.80 | 0.35 | 0.21 | 112 | 2 | 8 | 0.80 | 0.42 | 0.25 |
| R7T3 | 176 | 128 | 8 | 0.90 | 0.43 | 0.34 | 80 | 10 | 8 | 0.98 | 0.79 | 0.75 |
| R8T3 | 64 | 64 | 7 | 0.74 | 0.14 | 0.07 | 16 | 5 | 8 | 0.91 | 0.74 | 0.60 |

### 9.6.2 Statistical Significance Analysis

The results of a paired $t$-test analysis (§8.10 on page 103) of the best $\kappa$-values for each method is presented in **Table 9.7**. For a significance level of 0.05, the average values of $\kappa$ for TM-$D_E$, TM-$D_{ES}$ and MI-GLCMs are not statistically different ($p > 0.05$) from each other, but are statistically different from the average results for TM-$D_M$ and MI-histograms ($p < 0.05$).

**Table 9.7**: Results of the $t$-test analysis for the best matching accuracies from the mosaic images for the various matching methods.

| Method | TM-$D_{ES}$ | | TM-$D_M$ | | MI-histograms | | MI-GLCMs | |
|---|---|---|---|---|---|---|---|---|
| | t-value | p-value | t-value | p-value | t-value | p-value | t-value | p-value |
| TM-$D_E$ | 0.37 | 0.72 | 3.64 | 0.00 | 6.45 | 0.00 | -0.55 | 0.59 |
| TM-$D_{ES}$ | | | 4.20 | 0.00 | 8.48 | 0.00 | -1.06 | 0.30 |
| TM-$D_M$ | | | | | 2.70 | 0.01 | -4.38 | 0.00 |
| MI-histograms | | | | | | | -7.43 | 0.00 |

### 9.6.3 TM-Matching

The use of TM-matching with distance similarity metrics generally yielded results with very good matching accuracies for $D_E$ and $D_{ES}$, but poor matching accuracies for $D_M$. The difference in matching results for TM-$D_E$ and TM-$D_{ES}$ were not statistically different, but the results of both these methods differed significantly from those of TM-$D_M$. The poor results of TM-$D_M$ might indicate a poor correlation between the texture measures and the use of $D_M$ as a similarity metric is inappropriate for these images.

Also, while the best $A_{ROC}$- and $C_{fb}$-values for TM-$D_{ES}$ were more clustered than for TM-$D_E$, standardisation did not significantly improve results over using $D_E$ as a similarity metric, since the average matching accuracies for these methods were statistically similar.

Matching accuracy was worse for a reduced set of texture measures, but this was probably because each image pair has a unique set of texture measures that best describes the textural characteristics of the images. A more rigorous feature selection algorithm would have to be used to investigate this hypothesis.

One factor that might have affected the matching accuracy for TM-matching concerns the averaging of the GLCMs at the four angles, which could cause the loss of some angle-dependent information, which might otherwise improve the matching accuracy.

### 9.6.4 MI-Matching

MI-histograms performed very poorly as a matching scheme. Average $A_{ROC}$ and $C_{fb}$ values were low, indicating that only some of the region to be matched was identified and that the matched region was not easily discernable from the background in the matching map. The poor results for MI-histograms were most likely due to two reasons.

The first reason stems from the method used to compute the histograms (§8.4.1). The histogram was computed between the minimum and maximum grey-levels of both the reference and test sub-images. This range was then divided into *nbins* bins. This implementation has two disadvantages, firstly the width of the histogram bins varies as the test image is sampled, and secondly, the histograms are not invariant to shifts in grey-level. This method of determining the histogram was chosen so that the histograms for the reference and test regions would have equal bin sizes, but it was not realised that the bin sizes would vary for each sampled window across the image.

The second reason might be because histograms do not incorporate any spatial information, and using histograms to describe texture, which explicitly relies on variations of grey-level between pixels at different positions, is not completely appropriate.

MI-GLCMs performed very well as a matching scheme. Average $A_{ROC}$ and $C_{fb}$ values were high, indicating that most of the region to be matched was identified and that the matched region was easily discernable from the background in the matching map. Average $A_{ROC}$ and $C_{fb}$ values were considerably higher than those for MI-histograms.

One factor that might have affected the matching accuracy for MI-GLCMs concerns the averaging of the GLCMs at the four angles, which could cause the loss of some angle-dependent information, which might otherwise improve the matching accuracy.

## 9.7 Results of ACF Analysis

The autocorrelation function was used to determine the characteristic scale width, $w_{ACF}$, and scale height, $h_{ACF}$, of the textural features in each reference image. The ACF results were generated at different bit-depths because the characteristic scales of textural features were expected to change as bit-depth changed. However, there was no difference in the ACF results at each bit-depth. The ACF results are displayed as vertical lines in **Figures 9.4, 9.7, 9.10, 9.14** and **9.17**, but only **Figure 9.17** for MI-GLCMs showed some correlation between the ACF results and the maximum $\kappa$-value.

The results of the linear Pearson correlation analysis (§8.10 on page 103) in **Table 9.8** were computed between the optimal sampling window sizes and $w_{ACF}$ and $h_{ACF}$ (**Tables 9.9** and **9.10**). The correlation coefficients indicate that there is a poor correlation between the optimal sampling window sizes and the results of the ACF analysis.

According to the ACF results, the characteristic scale sizes for each reference texture is often rectangular, so the poor correlation might be a factor of the square sampling windows used in the matching algorithm. The optimal sampling window sizes that have emerged from the matching analysis are therefore not appropriate for comparison to the rectangular scale sizes.

**Table 9.8**: Correlation coefficients between results of ACF analysis and optimal sampling window sizes for the mosaic images. The linear Pearson correlation coefficient was computed between the optimal sampling window sizes and $w_{ACF}$ ($\rho_w$) and $h_{ACF}$ ($\rho_h$). Results indicate that there is a poor correlation between the optimal window sizes and the ACF results.

| Method | 5-bits | | 6-bits | | 7-bits | | 8-bits | |
|---|---|---|---|---|---|---|---|---|
| | $\rho_w$ | $\rho_h$ | $\rho_w$ | $\rho_h$ | $\rho_w$ | $\rho_h$ | $\rho_w$ | $\rho_h$ |
| TM-$D_E$ | 0.12 | 0.03 | 0.17 | -0.02 | 0.17 | -0.05 | 0.17 | -0.04 |
| TM-$D_{ES}$ | -0.15 | -0.08 | -0.33 | 0.06 | -0.05 | -0.07 | -0.24 | 0.14 |
| TM-$D_M$ | -0.10 | 0.07 | -0.34 | 0.15 | 0.01 | -0.04 | -0.14 | 0.11 |
| MI-histograms | -0.17 | -0.05 | -0.18 | -0.05 | -0.18 | -0.06 | -0.28 | -0.05 |
| MI-GLCMs | -0.07 | 0.05 | -0.22 | 0.18 | -0.10 | -0.03 | 0.07 | -0.04 |

**Table 9.9**: Optimal sampling window sizes from TM-matching and results of ACF analysis for mosaic images. Although the ACF analysis was performed at different bit-depths, there was no difference between the results at the different bit-depths.

| Image Pair | 5-bits | | | 6-bits | | | 7-bits | | | 8-bits | | | ACF | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $D_E$ | $D_{ES}$ | $D_M$ | $D_E$ | $D_{ES}$ | $D_M$ | $D_E$ | $D_{ES}$ | $D_M$ | $D_E$ | $D_{ES}$ | $D_M$ | $ACF_h$ | $ACF_w$ |
| R1T1 | 176 | 288 | 352 | 176 | 320 | 400 | 176 | 320 | 400 | 176 | 320 | 400 | 56 | 186 |
| R2T1 | 80 | 80 | 80 | 80 | 80 | 80 | 96 | 80 | 80 | 96 | 80 | 80 | 209 | 34 |
| R3T1 | 160 | 160 | 208 | 176 | 160 | 176 | 176 | 160 | 192 | 176 | 160 | 192 | 222 | 174 |
| R4T1 | 80 | 320 | 160 | 80 | 320 | 240 | 80 | 160 | 80 | 80 | 160 | 192 | 90 | 251 |
| R5T2 | 128 | 128 | 128 | 128 | 128 | 128 | 128 | 128 | 128 | 128 | 128 | 160 | 141 | 364 |
| R6T2 | 80 | 160 | 368 | 80 | 160 | 368 | 80 | 160 | 368 | 128 | 160 | 368 | 168 | 204 |
| R7T2 | 224 | 336 | 384 | 224 | 336 | 336 | 224 | 368 | 336 | 224 | 336 | 368 | 268 | 135 |
| R8T2 | 80 | 128 | 128 | 80 | 128 | 112 | 80 | 80 | 112 | 80 | 16 | 80 | 236 | 97 |
| R5T3 | 128 | 128 | 256 | 128 | 128 | 240 | 128 | 128 | 240 | 128 | 128 | 256 | 141 | 364 |
| R6T3 | 288 | 192 | 336 | 288 | 192 | 272 | 288 | 288 | 288 | 288 | 192 | 384 | 168 | 204 |
| R7T3 | 160 | 160 | 176 | 160 | 176 | 224 | 160 | 176 | 304 | 160 | 176 | 304 | 268 | 135 |
| R8T3 | 144 | 304 | 288 | 176 | 144 | 224 | 176 | 224 | 304 | 176 | 48 | 304 | 236 | 97 |

**Table 9.10**: Optimal sampling window sizes from MI-matching with grey-level histograms (GLH) and GLCMs and results of ACF analysis for mosaic images. There was no variation of the results of the ACF analysis at the different bit-depths.

| Image Pair | 5-bits | | 6-bits | | 7-bits | | 8-bits | | ACF | |
|---|---|---|---|---|---|---|---|---|---|---|
| | GLH | GLCMs | GLH | GLCMs | GLH | GLCMs | GLH | GLCMs | $ACF_h$ | $ACF_w$ |
| R1T1 | 400 | 16 | 400 | 32 | 400 | 32 | 400 | 32 | 56 | 186 |
| R2T1 | 368 | 16 | 368 | 16 | 368 | 48 | 368 | 48 | 209 | 34 |
| R3T1 | 400 | 16 | 400 | 32 | 400 | 48 | 112 | 48 | 222 | 174 |
| R4T1 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 90 | 251 |
| R5T2 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 141 | 364 |
| R6T2 | 96 | 80 | 96 | 16 | 96 | 16 | 96 | 208 | 168 | 204 |
| R7T2 | 80 | 16 | 80 | 16 | 64 | 16 | 80 | 48 | 268 | 135 |
| R8T2 | 48 | 16 | 48 | 16 | 64 | 16 | 64 | 16 | 236 | 97 |
| R5T3 | 384 | 16 | 384 | 32 | 384 | 32 | 384 | 32 | 141 | 364 |
| R6T3 | 96 | 16 | 96 | 80 | 96 | 112 | 96 | 112 | 168 | 204 |
| R7T3 | 176 | 16 | 176 | 16 | 176 | 16 | 176 | 80 | 268 | 135 |
| R8T3 | 48 | 16 | 48 | 16 | 64 | 16 | 64 | 16 | 236 | 97 |

## 9.8 Summary of the Effects of Matching Parameters

### 9.8.1 Effect of $w_{step}$ on Matching Accuracy

Unfortunately, as a consequence of varying the sampling window step size, $w_{step}$, as a fraction of the window size, the results had two varying parameters, viz. $w$ and $w_{step}$, and a study of the effect of $w_{step}$ on matching accuracy could not be carried out. The results did indicate that the plots of $\kappa$ *vs.* $w$ were smoother for smaller values of $w_{step}$.

### 9.8.2 Effect of $w$ on Matching Accuracy

Sampling window size had the most significant effect on matching accuracy. For TM-matching and MI-histograms, it is quite clear from the results that the choice of the optimal sampling window size is dependent on the specific reference image, while MI-GLCMs showed that matching accuracy increased as sampling window size decreased.

Unfortunately, the analysis using the autocorrelation function to determine sampling window sizes showed that there was a poor correlation between the optimal sampling window sizes of the matching algorithms and the ACF analysis. The poor correlation results might be a consequence of the changing window step size for each window size or the use of square sampling windows, since the ACF results indicated that the scale widths and heights of the mosaic textures were not equal. Therefore, rectangular sampling windows should probably be used to improve matching accuracy for the mosaic images.

### 9.8.3 Effect of *nbits* on Matching Accuracy

Matching accuracy increased with decreasing *nbits* for TM-$D_E$ (for small sampling windows) and for TM-$D_{ES}$ (for all, but the very small sampling windows). Matching accuracy was independent of *nbits* for TM-$D_M$ and for the other sampling window sizes for TM-$D_E$ and TM-$D_{ES}$.

As bit-depth decreases, images with large regions made up of similar grey-levels get the slight variations evened out. So the image at the lower bit-depth has large areas of the same grey-level. This physical change in the image could therefore mean that the chances of a match are improved, at the smaller sampling windows, since there are now fewer differences in grey-level (or texture) between the reference and test images. At the larger window sizes, the structural detail in the image starts appearing and matching is not as good as at the smaller sampling windows. This explains why TM-$D_E$ shows a dependence of matching accuracy on *nbits* for small sampling windows. The improvement of matching accuracy with decreasing *nbits* for large sampling windows is probably an artefact of the standardisation process.

MI-histograms showed no significant dependence of matching accuracy on bit-depth.

MI-GLCMs showed a clear dependence of matching accuracy on bit-depth, with matching accuracy improving as bit-depth increased, for all sampling windows.

This can be understood from the point of view that the higher the bit-depth, the more information is contained in the image. At high bit-depths, only textures similar to the reference texture will be matched with a high value of mutual information. As bit-depth decreases, fewer grey-levels represent the information in the image, and more textures will be similar to the reference texture and will also have high values of mutual information. However, at the lower bit-depths, there will be many false-positive detections and the overall matching accuracy will decrease as bit-depth decreases. The GLCMs used to estimate the probability density functions for the mutual information calculation work directly with the grey-levels in the image.

The results for TM-$D_E$ and TM-$D_{ES}$ are contrary to the results obtained by Chan et al. [1995] who varied bit-depth between 4 bits and 9 bits and found the optimal bit-depth to be either 7 bits or 8 bits, while the results for MI-GLCMs are consistent with those of Chan et al..

### 9.8.4 Effect of *d* on Matching Accuracy

For both TM-matching and MI-matching there was no significant dependence of matching accuracy on *d*. TM-$D_{ES}$, MI-histograms and MI-GLCMs showed a slight improvement of matching accuracy as *d* was decreased and TM-$D_E$ showed a slight improvement of matching accuracy as *d* was increased.

The lack of dependence of matching accuracy on *d* is most likely coincidentally due to the maximum value of *d* being not very large and textures being similar within this range.

### 9.8.5 Effect of *nbins* on Matching Accuracy

Matching accuracy improves slightly as the number of histogram bins decreases.

Results are consistent with those of Tourassi et al. [2003] who used mutual information as a similarity metric for template matching in a knowledge-based mammographic CAD-system for discrimination of masses from normal tissue and found that results were most accurate for the fewest number of histogram bins (64 bins).

### 9.8.6 Sensitivity of Matching Methods to Choice of Parameter Values

**Figure 9.21** shows all pairs of $A_{ROC}$- and $C_{fb}$-values, across all parameters, for each matching method. An examination of this data will give an idea of how sensitive each method is to the choice of parameter values. If all the points are clustered in one particular region, then the method is not sensitive to the choice of values of the matching parameters, but if the points are very scattered then the choice of values for the matching parameters is critical to ensure an optimal match. Results for TM-$D_E$ are clustered around $A_{ROC}$=0.75 and $C_{fb}$=0.50, but there are many points that are widely scattered. Results for TM-$D_{ES}$, TM-$D_M$ and MI-histograms are generally more clustered. Unfortunately the results for the latter two methods are clustered about the [0.5,0] point. Results for MI-GLCMs are very scattered.

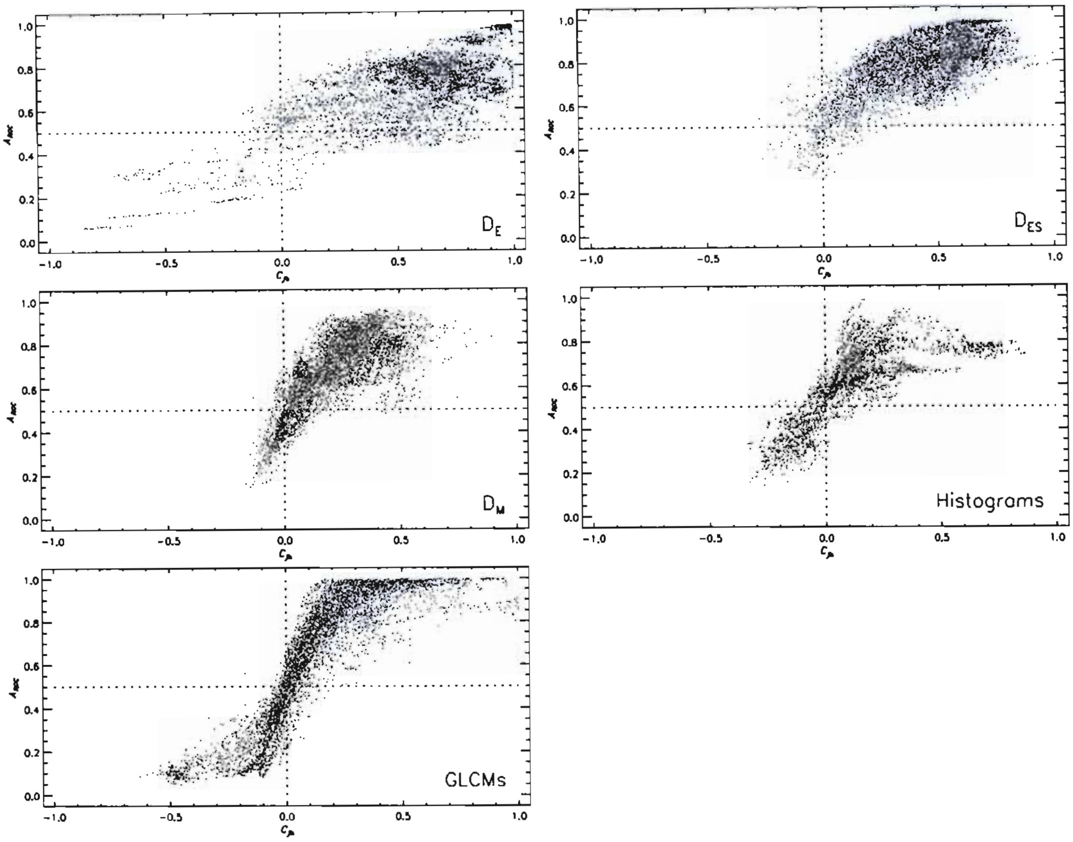


**Figure 9.21**: Scatter plot of $A_{ROC}$ *vs.* $C_{fb}$ for all combinations of matching parameters to demonstrate sensitivity of choice of matching parameter values for each method applied to the mosaic images.

These results indicate that MI-GLCMs is more sensitive to the choice of matching parameter values than TM-$D_E$, for this set of images.

In order to quantify the sensitivity of choice of parameter values, the number of pairs of $A_{ROC}$- and $C_{fb}$-values that fell into different zones was evaluated (as detailed in **Figure 9.22**). Zone 1 is the preferred zone, as these points have very high $A_{ROC}$- and very high $C_{fb}$-values. Zone 2 is next preferred as this contains points with very high $A_{ROC}$-values but lower contrast values than Zone 1. It is more preferable to have a good match, i.e. as few false-positives and a high true-positive fraction than to have a good contrast. Zone 3 has lower values of $A_{ROC}$, but similar values of contrast than Zone 1. Zone 4 has low values of $A_{ROC}$ and $C_{fb}$ but the values are still valid (i.e. $A_{ROC}>0.5$, $C_{fb}>0$). Zone 5 defines the regions with $A_{ROC}$ and $C_{fb}$-values that are not preferred (i.e. for $A_{ROC}< 0.5$ and $C_{fb}< 0$).

The results of this analysis is summarised in **Table 9.11** and shows that TM-$D_E$ is most accurate with most points falling into zone 1.

## 9.9 Comparison of TM- and MI-Matching for Mosaic Images

Overall, TM-$D_E$, TM-$D_{ES}$ and MI-GLCMs have the best average matching accuracies and these results are statistically similar. The matching accuracies for TM-$D_M$ and MI-histograms are very poor.

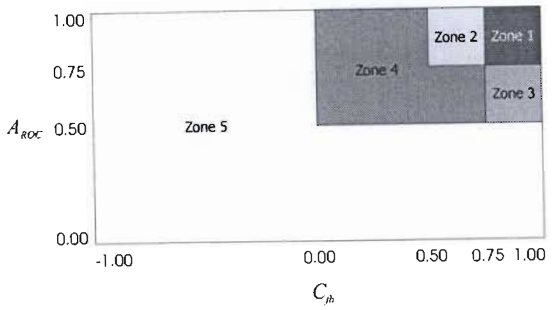


**Figure 9.22**: Schematic showing regions of scatter-plot used for zone analysis.

**Table 9.11**: Results of zone analysis for mosaic images to demonstrate sensitivity of choice of matching parameters for each method. An ideal method would have a high percentage of points in Zone1 and none in Zone 5.

| Method | Zone 1 (%) | Zone 2 (%) | Zone 3 (%) | Zone 4 (%) | Zone 5 (%) | Total (Zones 1,2 & 3)(%) |
|---|---|---|---|---|---|---|
| TM-$D_E$ | 32.5 | 34.7 | 19.6 | 12.4 | 0.820 | 86.8 |
| TM-$D_{ES}$ | 6.66 | 70.0 | 0.700 | 22.3 | 0.410 | 77.3 |
| TM-$D_M$ | 0.620 | 14.9 | 2.91 | 77.8 | 3.820 | 18.4 |
| MI-histograms | 3.63 | 23.2 | 1.07 | 65.6 | 6.51 | 27.9 |
| MI-GLCMs | 8.37 | 32.2 | 0.00 | 52.9 | 6.530 | 40.6 |

TM-$D_E$ and TM-$D_{ES}$ do not show clear dependencies of matching accuracy on the various matching parameters and it is difficult to recommend a set of optimal matching parameters that will give a maximal matching accuracy. TM-$D_{ES}$ is, however, less sensitive than TM-$D_E$ to the choice of matching parameter values.

MI-GLCMs shows clear dependencies of matching accuracy on the various matching parameters so it is easier to recommend a set of optimal matching parameters for this algorithm. MI-GLCMs is faster to compute than TM-matching with distance similarity metrics, but MI-GLCMs is quite sensitive to the choice of values of the matching parameters.

Overall, where MI-GLCMs has a weakness, TM-$D_E$ and TM-$D_{ES}$ have strengths and vice-versa. In summary, there are no significant differences between TM-$D_E$, TM-$D_{ES}$ and MI-GLCMs, and a possible hybrid-matching scheme using the results of all three methods might yield good matching results.

## 9.10 Summary

The TM-matching and MI-matching algorithms were applied to test images made up of a mosaic of single texture reference images to test matching performance under the ideal condition of knowing how the reference images are transformed in the test image.

The purpose of this chapter was two-fold: (1) to test the performance of the matching algorithms on images with clear borders between the textures, accurate ground truth data and where there are known transformations of the textures to give insight into whether the matching algorithms have any potential for identifying similar textures in mammograms; and (2) to investigate how the matching accuracy is affected by the various matching parameters.

TM-matching and MI-matching show some potential for use as matching schemes. Results in this chapter *do* provide evidence to support the hypothesis that using texture based image-processing methods allows a textural region to be matched with a reference texture. However, it has been shown that the choice of matching parameter values can significantly affect matching accuracy, so the following parameters will still be varied for the pairs of mammograms and the stereotactic biopsy mammograms: *nbits*, $d$, $w$. As a consequence of the effect of sampling window step size, $w_{step}$, on matching accuracy for the mosaic images, $w_{step}$ are fixed for the analysis of the pairs of mammograms and the stereotactic biopsy mammograms.

The top three matching methods, based on averages of $\kappa$ for the best matches, are: MI-GLCMs ($\kappa$=0.63±0.20), TM-$D_E$ ($\kappa$=0.56±0.13) and TM-$D_{ES}$ ($\kappa$=0.55±0.23). However, these three results are not statistically different ($p > 0.05$). TM-$D_M$ and MI-histograms perform poorly as matching schemes.

The recommended values of the matching parameters for implementation of the matching schemes on similar images are given in **Table 9.12**. The values for TM-$D_E$ and TM-$D_{ES}$ are obtained from the few cases where there was some dependence of matching accuracy on the matching parameters.

**Table 9.12**: Recommended values for matching parameters for application of TM-$D_E$, TM-$D_{ES}$ and MI-GLCMs to images similar to the mosaics

| Matching Parameter | TM-$D_E$ | TM-$D_{ES}$ | MI-GLCMs |
|---|---|---|---|
| $w_{step}$ | 4 pixels | 4 pixels | 4 pixels |
| $w$ | rectangular from ACF | rectangular from ACF | rectangular from ACF |
| *nbits* | 5 bits | 5 bits | 8 bits |
| $d$ | 10 pixels | 1 pixel | 1 pixel |

# Chapter 10

# Matching Results: Mammograms

The results of applying the matching algorithms to the mosaic images have demonstrated that the algorithms have the potential to match a known reference texture to the same texture in another image. The algorithms were applied to pairs of mammograms to confirm whether a reference region of interest (ROI) identified by a radiologist in one standard mammographic view can be matched to the corresponding region in another standard mammographic view, to determine whether the matching algorithms can be utilised in a CAD-system. The results are presented and discussed in this chapter.

## 10.1 Details of the Mammograms used in this Study

### 10.1.1 Selection of Mammograms

The 34 pairs of CC and MLO mammograms (from 26 patients) used in this study were arbitrarily selected from the patient archives at the Inkosi Albert Luthuli Central Hospital (Durban, South Africa), to represent a range of breast densities, mass sizes and patient ages. In 8 cases, mammograms from left and right breasts were used, accounting for the difference between the number of patients and the pairs of mammograms used in the study. The computed radiography (CR) images were exported in a DICOM format from the data archives at the hospital, at a bit-depth of 10 bits and 0.05 mm per pixel. For processing, images were resampled to 0.254 mm per pixel (100 dpi). The images were acquired on a Siemens Mammomat 3000 Nova mammography unit, with a focal spot size of 0.3 mm, a molybdenum anode and a 30 $\mu$m molybdenum filter. Two Fujifilm Fuji IP Cassettes with image plate sizes detailed in **Table 10.1** were used. The image reader is a Digiscan M (Fuji Photo Film Co Ltd).

**Table 10.1**: Specifications of computed radiography cassettes

| | 24 mm × 30 mm | 18 mm × 24 mm |
|---|---|---|
| Field-of-View | 296.4 mm × 236.4 mm | 237 mm × 177 mm |
| Image Matrix | 5928 pixels × 4728 pixels | 4740 pixels × 3540 pixels |

Since the matching results are independent of which view is used as a reference, each CC and MLO view was used separately as a reference image and a test image, resulting in the matching algorithms being applied to 68 pairs of mammograms. The 68 mammograms were divided into four categories based on the pathology of the suspicious ROI or overall diagnosis of the mammogram: 28 benign, 18 malignant, 10 normal and 12 indeterminate. Most patients get referred to the Inkosi Albert Luthuli Central Hospital for diagnostic tests so not all the mammograms had a full pathological history, since not all referring physicians recommended a biopsy. In these cases, the radiologist's report was used as a basis for the diagnosis of the mammogram.

The 'benign', 'indeterminate' and 'malignant' diagnoses refer to masses, while the 'normal' diagnosis refers to suspicious-looking regions in a normal mammogram. Masses were categorised as 'indeterminate' if the biopsy was inconclusive or the radiologist was unable to render a diagnosis based on the mammographic appearance.

The reference mammogram label and test mammogram label are used in combination to label the mammogram pair. For example, the mammogram pair consisting of reference image M1 and test image M0, has the mammogram pair label M1M0. The label for each mammogram is given in **Table 10.3**.

### 10.1.2 Characteristics of Mammograms

A radiologist identified the borders of the suspicious ROIs and classified the breast types as dense, mixed or fatty. The borders of the suspicious ROIs were marked in *MagicView*[1]. The borders were saved as DICOM images and were automatically extracted in IDL, for use as ground truth data. This eliminated the need to register the ground truth data with the mammograms. The areas of the regions enclosed by the radiologist-drawn borders that are automatically computed in *MagicView* were also saved.

Histograms of the areas of the suspicious ROIs along with their visibilities (compared to surrounding tissue) are shown in **Figure 10.1**. Visibility was automatically determined from the original mammograms (at 0.254 mm per pixel) before pre-processing. Visibility was defined to be the contrast of the ROI compared to the surrounding tissue and was computed from **Equation 7-10** and **Figure 7.4** (page 87). Visibility ranges between 0 for a very subtle ROI and 1 for a very obvious ROI. The histograms demonstrate that there is a significant range of sizes of ROIs used in this study and some of these ROIs are very subtle while others are more visible. A summary of the distributions of breast type is given in **Table 10.2**. Full details of the sizes and visibilities of the ROIs are shown in **Table 10.3**.

**Table 10.2**: Number of mammograms of a particular breast type as a function of diagnosis.

| | Dense | Mixed | Fatty |
|---|---|---|---|
| Benign | 4 | 10 | 14 |
| Indeterminate | 0 | 10 | 2 |
| Malignant | 0 | 8 | 10 |
| Normal | 2 | 4 | 4 |

### 10.1.3 Reduced Images

Reduced images of the mammograms are shown in **Figures 10.2** and **10.3** in grey-scale and pseudo-colour, respectively. Magnified versions of the ground truth ROIs are shown in **Figure 10.4**.

---

[1] The software interface used to view DICOM medical images. Search `http://www.medical.siemens.com/` for MagicView.

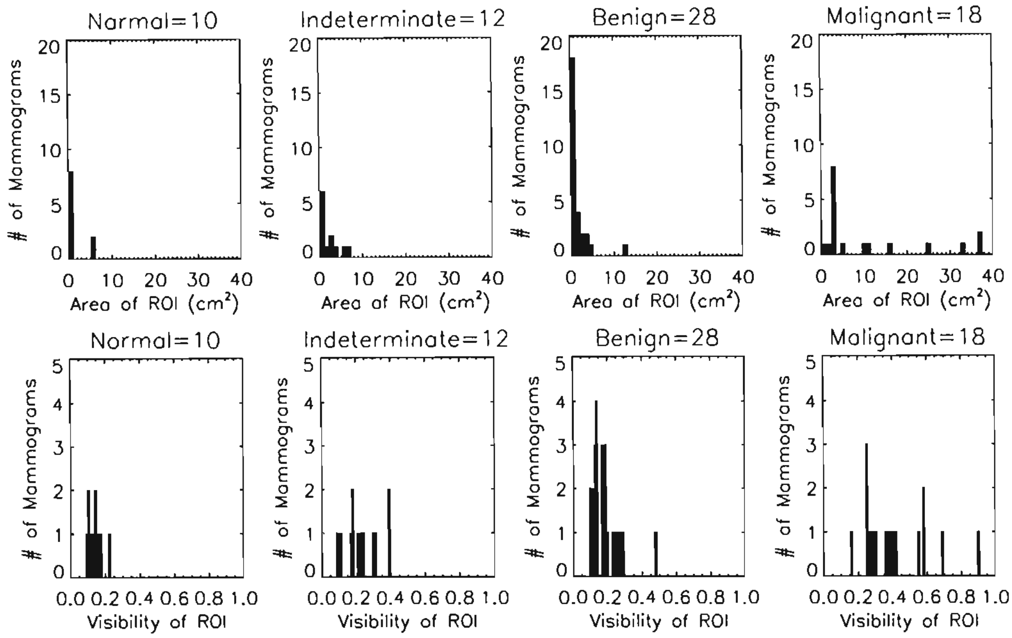


**Figure 10.1**: Histograms of areas and visibilities of the suspicious ROIs for the mammograms used in this study. There is a significant range in the sizes and visibilities.

**Table 10.3**: Characteristics of the suspicious ROIs in the mammograms used in this study.

| Label | Diagnosis | Breast Type | Area ($cm^2$) | Visibility | Label | Diagnosis | Breast Type | Area ($cm^2$) | Visibility |
|---|---|---|---|---|---|---|---|---|---|
| M0 | M | mixed | 3.14 | 0.25 | M34 | B | mixed | 1.99 | 0.19 |
| M1 | M | mixed | 4.53 | 0.25 | M35 | B | mixed | 2.84 | 0.24 |
| M2 | B | dense | 4.27 | 0.28 | M36 | M | fatty | 2.86 | 0.29 |
| M3 | B | dense | 1.21 | 0.29 | M37 | M | fatty | 2.71 | 0.27 |
| M4 | B | dense | 2.45 | 0.23 | M38 | I | mixed | 5.61 | 0.39 |
| M5 | B | dense | 3.4 | 0.48 | M39 | I | mixed | 6.32 | 0.40 |
| M6 | B | mixed | 0.91 | 0.19 | M40 | B | fatty | 0.58 | 0.12 |
| M7 | B | mixed | 0.62 | 0.14 | M41 | B | fatty | 0.66 | 0.14 |
| M8 | B | mixed | 0.25 | 0.17 | M42 | I | mixed | 2.07 | 0.17 |
| M9 | B | mixed | 0.64 | 0.19 | M43 | I | mixed | 3.84 | 0.31 |
| M10 | N | mixed | 0.33 | 0.1 | M44 | M | fatty | 15.76 | 0.58 |
| M11 | N | mixed | 0.81 | 0.14 | M45 | M | fatty | 0.41 | 0.37 |
| M12 | B | fatty | 0.18 | 0.26 | M46 | B | fatty | 12.69 | 0.12 |
| M13 | B | fatty | 0.14 | 0.17 | M47 | B | fatty | 0.39 | 0.18 |
| M14 | M | fatty | 2.45 | 0.36 | M48 | I | fatty | 0.22 | 0.09 |
| M15 | M | fatty | 2.38 | 0.39 | M49 | I | fatty | 0.53 | 0.30 |
| M16 | B | fatty | 0.73 | 0.14 | M50 | B | fatty | 0.41 | 0.13 |
| M17 | B | fatty | 0.23 | 0.25 | M51 | B | fatty | 0.25 | 0.14 |
| M18 | B | fatty | 1.6 | 0.09 | M52 | N | fatty | 0.38 | 0.09 |
| M19 | B | fatty | 0.21 | 0.17 | M53 | N | fatty | 0.22 | 0.15 |
| M20 | M | mixed | 2.68 | 0.25 | M54 | M | mixed | 36.86 | 0.69 |
| M21 | M | mixed | 2.95 | 0.26 | M55 | M | mixed | 24.6 | 0.90 |
| M22 | M | fatty | 3.29 | 0.16 | M56 | N | fatty | 0.27 | 0.11 |
| M23 | M | fatty | 3.29 | 0.3 | M57 | N | fatty | 0.16 | 0.18 |
| M24 | B | mixed | 0.22 | 0.19 | M58 | B | fatty | 0.27 | 0.10 |
| M25 | B | mixed | 0.25 | 0.13 | M59 | B | fatty | 0.19 | 0.11 |
| M26 | M | mixed | 37.02 | 0.58 | M60 | I | mixed | 0.92 | 0.24 |
| M27 | M | mixed | 32.88 | 0.55 | M61 | I | mixed | 0.61 | 0.11 |
| M28 | N | dense | 5.91 | 0.12 | M62 | I | mixed | 1.17 | 0.23 |
| M29 | N | dense | 5.54 | 0.15 | M63 | I | mixed | 0.56 | 0.18 |
| M30 | M | fatty | 10.67 | 0.4 | M64 | N | mixed | 0.51 | 0.12 |
| M31 | M | fatty | 10.26 | 0.42 | M65 | N | mixed | 0.48 | 0.22 |
| M32 | B | mixed | 1.46 | 0.18 | M66 | I | mixed | 2.35 | 0.18 |
| M33 | B | mixed | 4.11 | 0.13 | M67 | I | mixed | 2.56 | 0.21 |

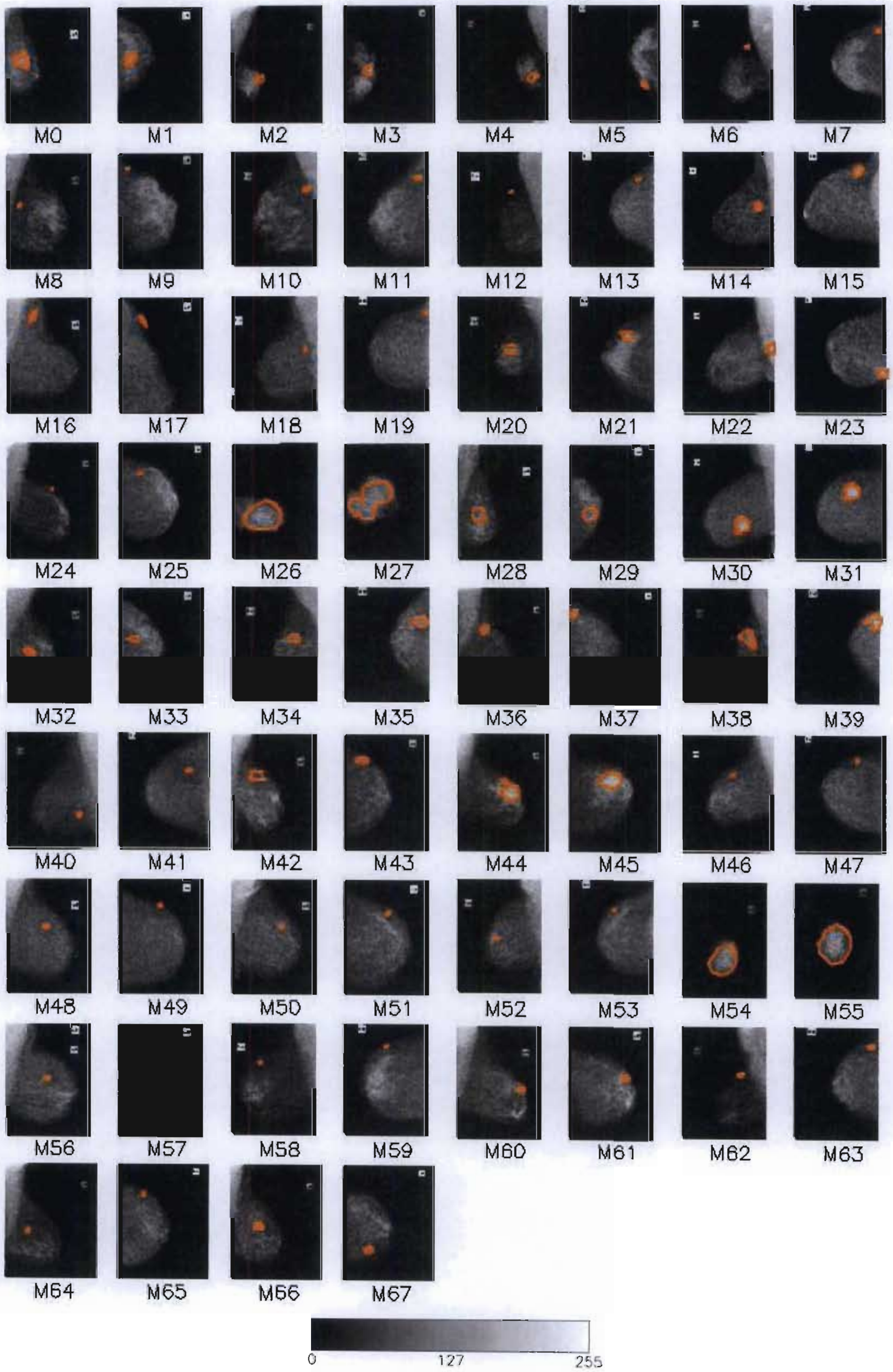


**Figure 10.2**: Reduced images of the mammograms used in this study in the original grey-scale. Images are paired from left to right, with four pairs per row. The ROI borders are shown in red. The bright rectangular feature in the upper right corner for left breast mammograms and in the upper left corner for right breast mammograms is the mammogram label.

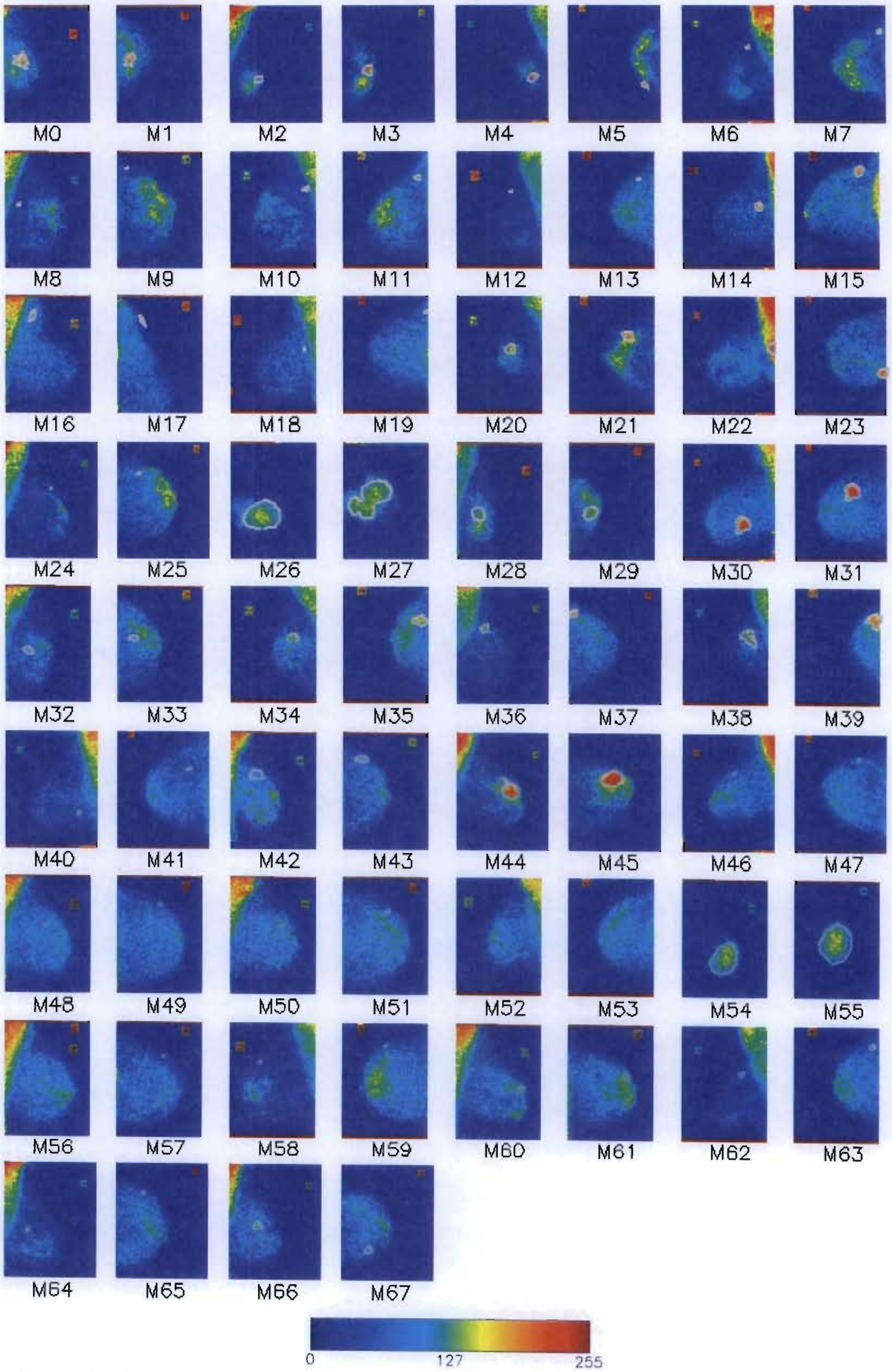


**Figure 10.3**: Reduced images of the mammograms used in this study in pseudo-colour to highlight the regions that cannot be easily seen in **Figure 10.2**. Images are paired from left to right, with four pairs per row. The ROI borders are shown in white, however at this scale, this is only obvious for the large, high contrast ROIs. The bright rectangular feature in the upper right corner for left breast mammograms and in the upper left corner for right breast mammograms is the mammogram label.

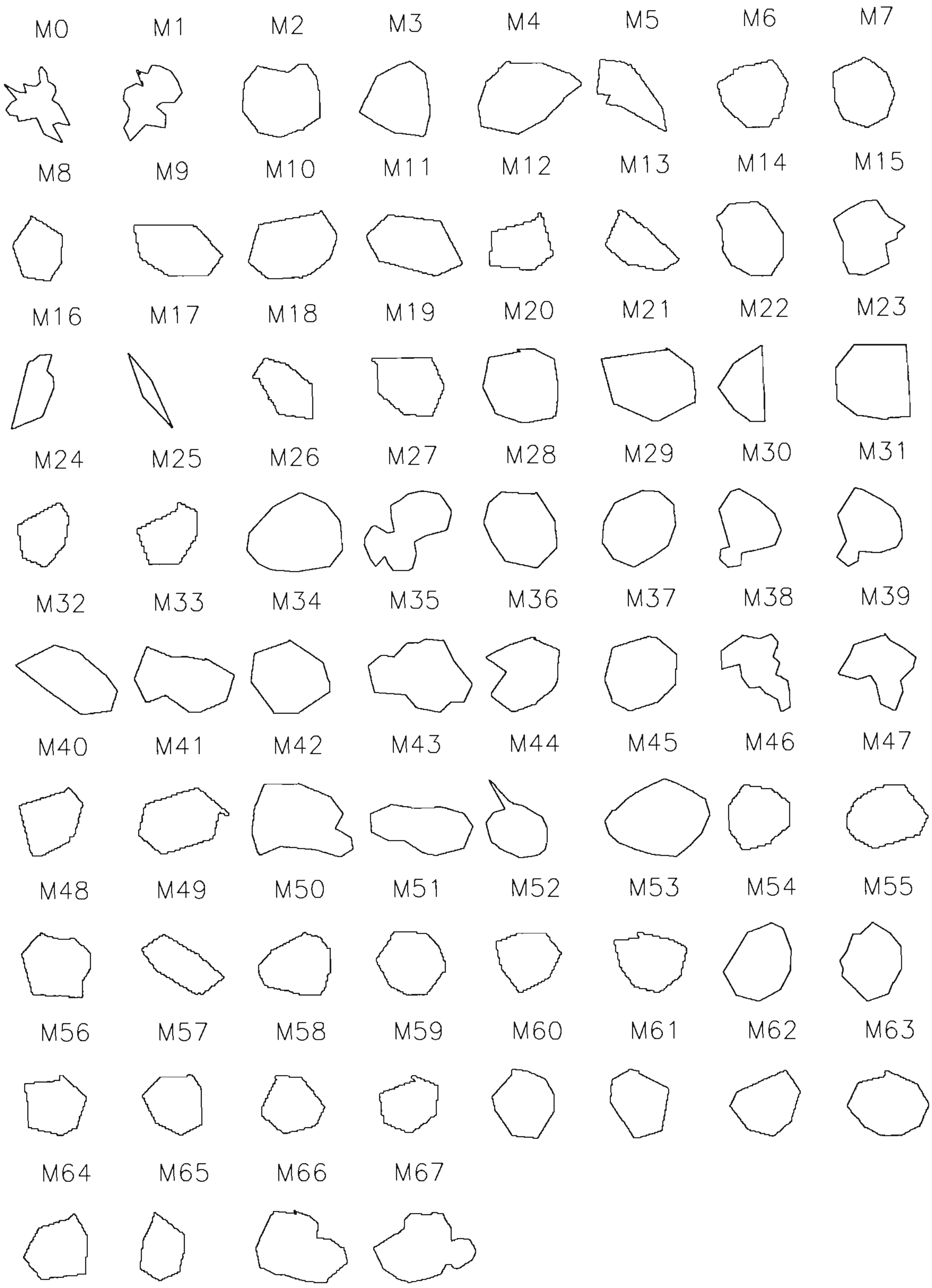


**Figure 10.4**: Magnified ROIs (not to scale) extracted from mammograms to demonstrate that some ROIs are spiculated (have a radiating pattern), while others are lobular (have lobes).

### 10.1.4 Sampling of Reference Images

The mammogram reference images are sampled differently to how the mosaic reference images were sampled because of the irregular shapes of the ROIs. Sampling windows were centred on the centroid of the ROI, and were then increased in size (in increments of 16 pixels) around the centroid until the ROI was completely enclosed. The size of the sampling window that completely enclosed the ROI was selected as the maximum window for that specific ROI. **Figure 10.5** shows sampling windows of different sizes placed around the centroid of a sample ROI.

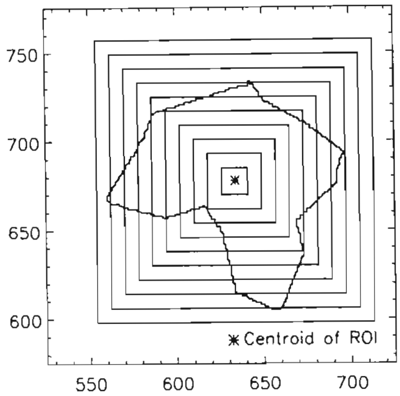


**Figure 10.5**: Placement of sampling windows in ROIs of mammograms. Sampling windows are centred on the centroid of the ROI. The maximum window is that which completely encloses the ROI.

## 10.2 Results of Pre-processing

The overall results of applying the masks obtained by the breast border, pectoral muscle and arc method algorithms are shown in **Figure 10.6**. Images are rotated so that the breast is oriented towards the right, for processing. The reduction of the search regions is greater for smaller ROIs than for the larger ROIs. The background mask has removed the mammogram labels.

## 10.3 Selection of Matching Parameters

The aim of the investigation using the mosaic images was to evaluate the matching algorithms under ideal matching conditions and to gain an understanding of the dependence of the various matching parameters on matching accuracy. The results from that investigation indicated that the matching parameters (*w*, *nbits* and *d*) depended on the individual images and were varied again for the mammograms. For TM-matching, the following texture measures were used: maximum probability, entropy, energy, inertia, inverse difference moment, correlation, sum average, sum entropy, difference entropy, sum variance, difference average, difference variance, information measure of correlation.

The sampling window step size was the only parameter that was fixed, at 4 pixels, for all mammograms and all combinations of matching parameters.

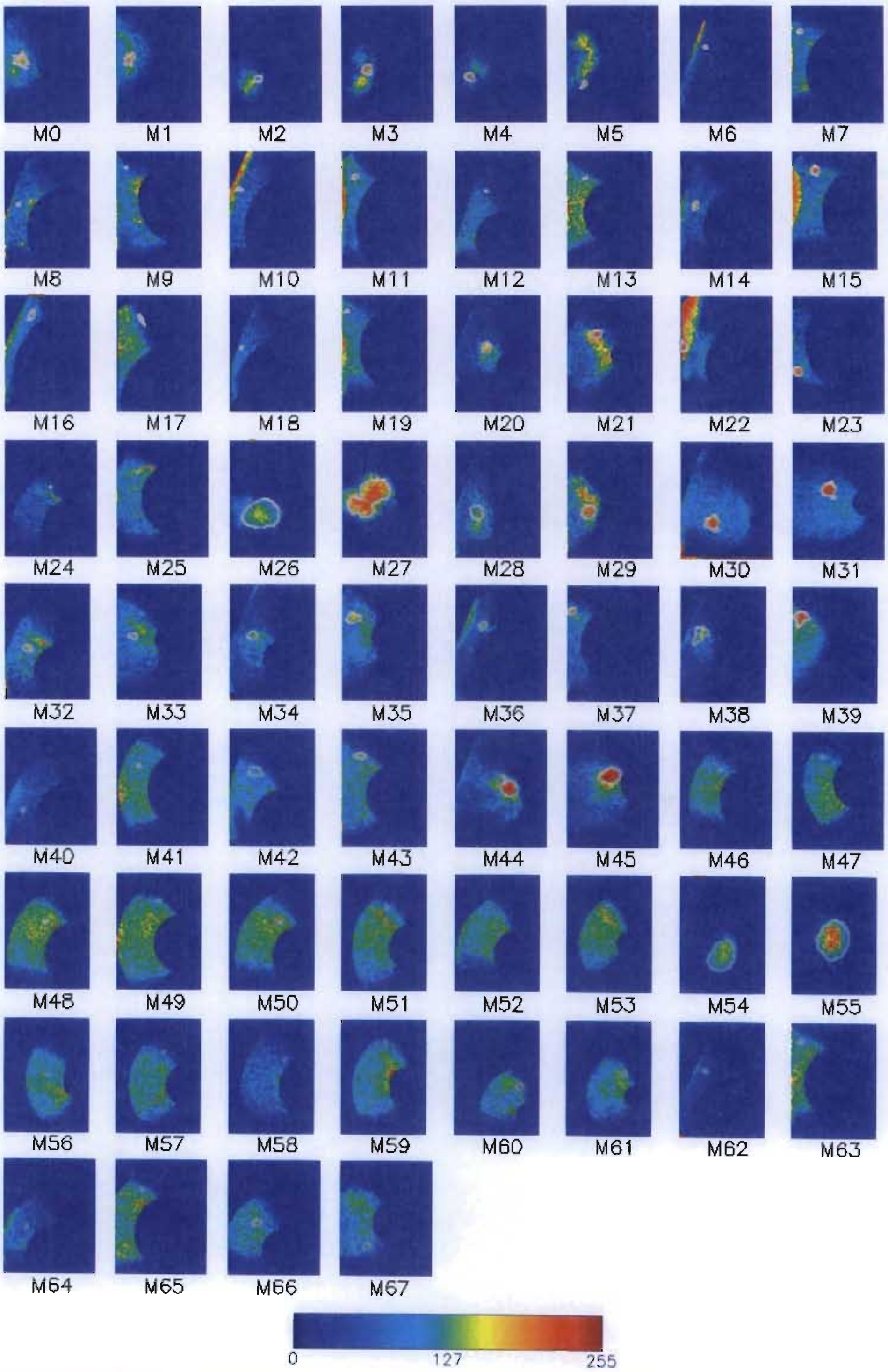


**Figure 10.6**: Results of pre-processing mammograms. Images show the regions of the breast that remain after application of the various segmentation methods. Images are rotated so that the breast is oriented towards the right, for processing. The ROI borders are indicated in white, however this is only obvious in some of the images. The reduction of the search regions is greater for smaller ROIs than for the larger ROIs. The background mask has removed the mammogram labels.

## 10.4 Examples of Matching Maps

Examples of matching maps for TM-matching and MI-matching are shown in **Figure 10.7**. These maps were obtained by selecting the maximum and minimum $\kappa$-values, over all matching parameters. Values of the actual matching parameters used for these maps are given in **Table 10.4**. Maps with a negative contrast imply that the region to be matched is darker than the background.

The example maps for the minimum $\kappa$-values appear in the left column of **Figure 10.4**. Three methods (TM-$D_E$, TM-$D_{ES}$, TM-$D_M$) have negative contrasts and $A_{ROC} < 0.5$ indicating very poor matches. The result for MI-histograms has a very low contrast and there was no match for MI-GLCMs.

The example maps for the maximum $\kappa$-values appear in the right column of **Figure 10.4**. All the maps show the ROI as the brightest feature, but all maps have other features in that have also been matched. These false-positive detections lower the $A_{ROC}$-value, lower the $C_{fb}$-value and therefore lower the overall matching accuracy. The matching accuracies for the remaining methods (TM-$D_E$, TM-$D_{ES}$, TM-$D_M$, MI-histograms) are similar with high values for $A_{ROC}$, but low values for $C_{fb}$. It is interesting to note that the optimal sampling window size for all distance similarity metrics are 80 pixels.

The maximum-$\kappa$ map for TM-$D_E$, has a bright band around the interface between the breast and the segmented background, while the maximum-$\kappa$ maps for TM-$D_{ES}$ and TM-$D_M$ show a bright region around the pectoral muscle. These bright regions lower the overall accuracy of the match. The MI-maps do not have any interface effects.

The matching maps for TM-$D_E$, TM-$D_{ES}$ and TM-$D_M$ also have non-zero backgrounds, while the matching maps for MI-histograms and MI-GLCMs have backgrounds that are very close to zero.

**Table 10.4**: Summary of matching parameters for the example matching maps of mammograms, listed at the maximum and minimum $\kappa$-values for image pair M0M1. Negative contrasts indicate that the background is brighter than the region to be matched.

| Method | Minimum $\kappa$ | | | | | | Maximum $\kappa$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $w$ | *nbits* | *d*/*nbins* | $A_{ROC}$ | $C_{fb}$ | $\kappa$ | $w$ | *nbits* | *d*/*nbins* | $A_{ROC}$ | $C_{fb}$ | $\kappa$ |
| TM-$D_E$ | 112 | 5 | 2 | 0.39 | -0.28 | -0.06 | 80 | 8 | 10 | 0.94 | 0.48 | 0.42 |
| TM-$D_{ES}$ | 16 | 5 | 1 | 0.49 | -0.04 | -0.00 | 80 | 5 | 10 | 0.98 | 0.32 | 0.30 |
| TM-$D_M$ | 16 | 5 | 10 | 0.35 | -0.02 | -0.00 | 80 | 7 | 1 | 0.87 | 0.32 | 0.24 |
| MI-histograms | 128 | 6 | 64 | 0.72 | 0.09 | 0.04 | 16 | 8 | 16 | 0.97 | 0.36 | 0.34 |
| MI-GLCMs | 32 | 5 | 1 | 0.50 | 0.00 | 0.00 | 32 | 6 | 2 | 0.96 | 0.99 | 0.91 |

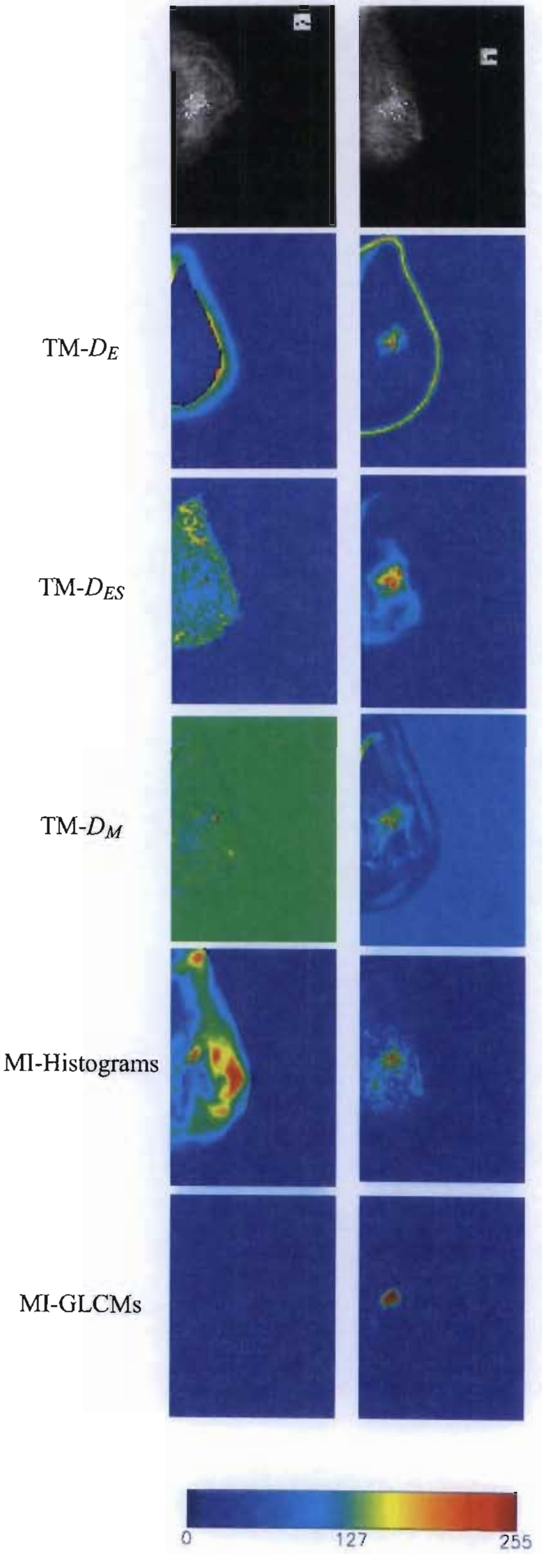


**Figure 10.7**: Example of matching maps (not to scale) for reference mammogram M0 (top, left) and test mammogram M1 (top, right). The maps for the minimum and maximum κ-values are shown in the left and right columns, respectively, with the colour bar at the bottom indicating the intensity scale for the maps.

## 10.5 TM-Matching Evaluation Results

The TM-matching algorithm (**Algorithm 8.1** on page 96) was applied to the mammogram images on page 145. The results of the dependence of matching accuracy on the matching parameters are discussed for each of the three distance similarity metrics. The results are colour-coded according to the breast tissue type of the reference ROI, benign (B), indeterminate (I), malignant (M) and normal (N).

Examples of the median, average and standard deviation for each of the texture measures is shown in **Table 10.5** for image M0 to demonstrate that there is a significant difference in the ranges of the values used as inputs to the distance similarity metric. The values listed were calculated over the full range of matching parameters. Sum variance and inertia have the highest values and these texture measures probably dominate the Euclidean distance calculation. The values listed for each texture measure differ from those values for the example mosaic image in **Table 9.3**. Therefore it is likely that each image has its own unique set of texture measures that best describe the image.

**Table 10.5**: Median, average and standard deviation of texture measures for mammogram M0.

| Texture Measures for M0 | Median | Average | Standard Deviation |
|---|---|---|---|
| Entropy | 9.15 | 9.23 | 2.31 |
| Energy | 0.002 | 0.006 | 0.010 |
| Inertia | 1 490 | 2 130 | 2 570 |
| Inverse Difference Moment | 0.024 | 0.037 | 0.032 |
| Correlation | 0.770 | 2.94 | 4.56 |
| Sum Average | 120 | 121 | 88.0 |
| Sum Entropy | 6.78 | 6.54 | 1.25 |
| Difference Entropy | 6.05 | 5.73 | 1.12 |
| Sum Variance | 1 360 | 2 040 | 2 420 |
| Difference Average | 33.3 | 32.2 | 22.7 |
| Difference Variance | 356 | 583 | 718 |
| Information Measure of Correlation 1 | 1.43 | 1.39 | 0.203 |
| Maximum Probability | 0.006 | 0.015 | 0.020 |

### 10.5.1 Computation Times

Computation times were dependent on the values used for *nbits* and $w$ and ranged from a few minutes to an hour, for a single map. As for the mosaic images, the calculation of the texture measures was the most time-consuming portion of the calculation.

### 10.5.2 Evaluation of TM-Matching with $D_E$

The results from using TM-matching with the Euclidean distance similarity metric, $D_E$, applied to pairs of mammograms, are presented and discussed. The dependence of matching accuracy on *nbits*, $d$ and $w$ is examined. Results are presented in the formats described in §8.2 (page 93). Dependence on matching accuracy is indicated by separation of the different coloured plots.

### Effect of *nbits* on Matching Accuracy

Typical examples of $\kappa$ *vs.* $w$ are shown in **Figure 10.8(a)** to demonstrate the effect of *nbits* and $w$ on matching accuracy. M0M1 and M2M3 show that matching accuracy is independent of bit-depth, for all values of $nbits>5$ bits, since these plots are exactly overlaid. The matching accuracy for $nbits=5$ bits is lower than the matching accuracy for higher bit-depths. For M39M38 and M28M29, the plots are noisy and the value of $w$ appears to influence the dependence of matching accuracy on *nbits*.

The results of the autocorrelation function analysis are plotted as vertical lines for each value of *nbits*, but there is very little difference in the results at the different bit-depths. The ACF-widths and ACF-heights are similar for three of the examples, but there appears to be poor correlation between the ACF results and the maximum values of $\kappa$. The ACF results are discussed in §10.8.

**Figure 10.8(b)** shows the average matching accuracy (over all image pairs) as a function of $d$, to highlight the general dependence of matching accuracy on *nbits*. The different degrees of separation of the *nbits*-plots at each sampling window size indicate that the dependence of matching accuracy on *nbits* is influenced by the size of the sampling window. This is clear because matching accuracy improves as bit-depth decreases for $w<112$ pixels and matching accuracy improves as bit-depth increases for $112\leq w\leq 160$ pixels. For $w>160$ pixels, there is no clear pattern of behaviour.

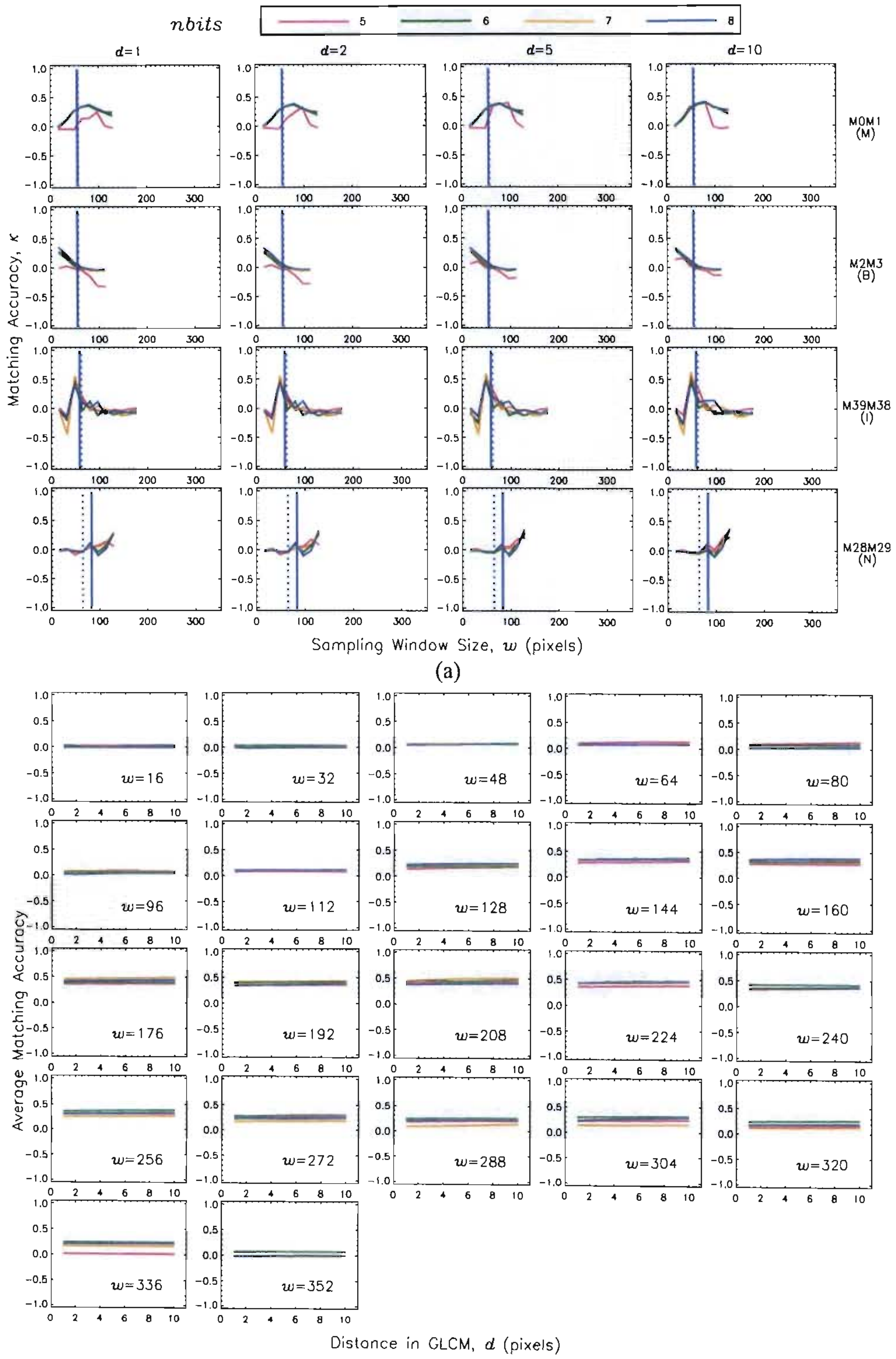
### Effect of *d* on Matching Accuracy

Some examples of $\kappa$ *vs.* $w$ are shown in **Figure 10.9(a)** to demonstrate how matching accuracy varies with $d$ and $w$. All plots for the different values of $d$ generally overlap almost completely, indicating that matching accuracy is independent of $d$ for these examples. There is some separation of the 5-bit plots for M0M1 and M2M3, but there is no consistent behaviour of matching accuracy with $d$ for these 5-bit plots.

**Figure 10.9(b)** shows the average matching accuracy (over all image pairs) as a function of *nbits*, to highlight the general dependence of matching accuracy on $d$. Plots at the different values of $d$ overlap almost completely for all sampling window sizes, indicating that matching accuracy is independent of $d$.

### Effect of *w* on Matching Accuracy

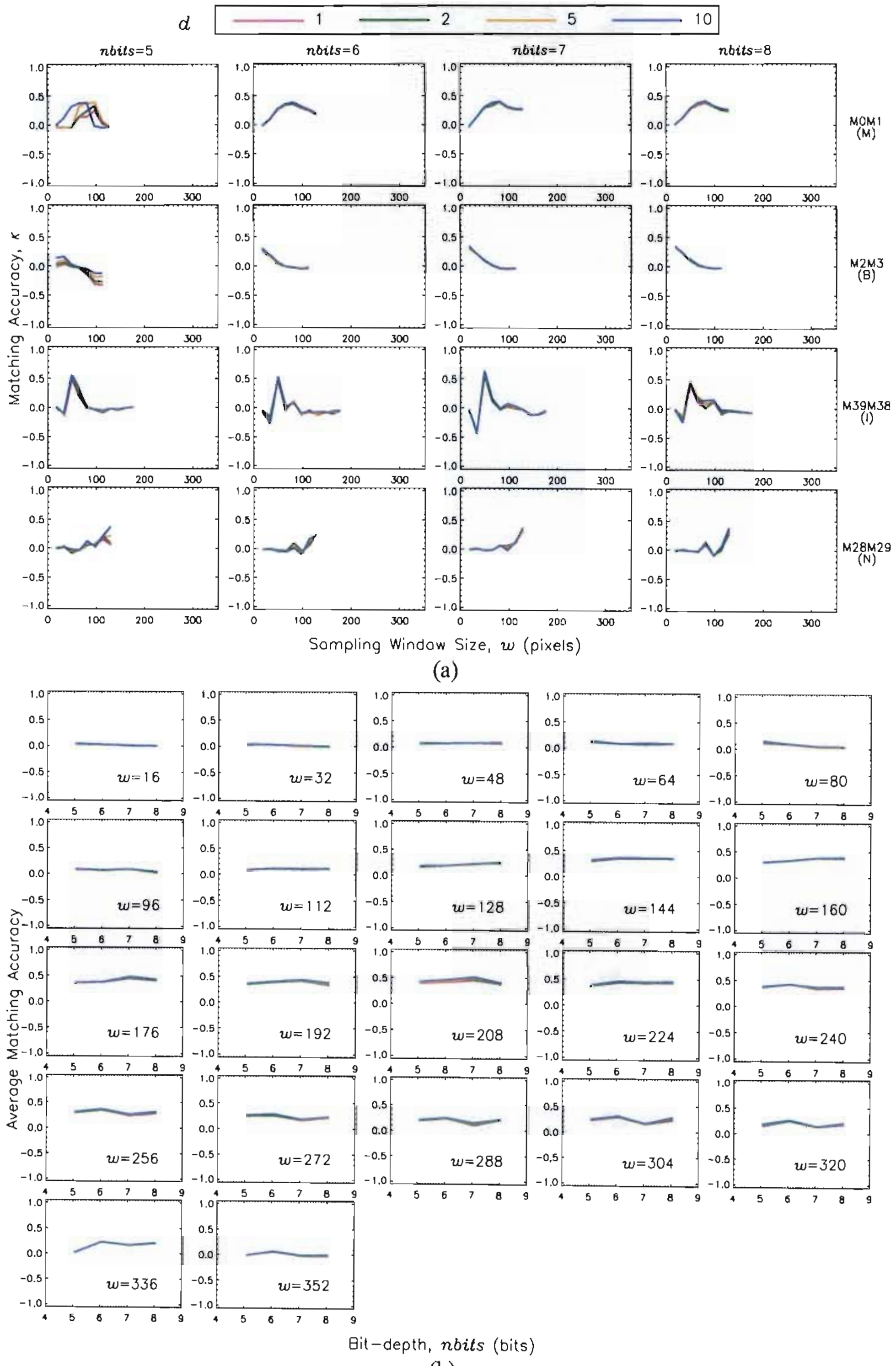
The example plots of $\kappa$ *vs.* $w$ in **Figures 10.8(a)** and **10.9(a)** show that matching accuracy varies as sampling window size varies. The variation is however unique to each image and the optimal sampling window size is most likely determined by the scale sizes of the textures in each image.



(a)

(b)

**Figure 10.8**: Effect of *nbits* on matching accuracy for TM-$D_E$ applied to pairs of mammograms. (**a**) Typical examples of $\kappa$ *vs.* *w*. Each row contains the information for a single image pair. The vertical lines indicate the results of the ACF analysis. The solid line represents the ACF-width and the dotted line represents the ACF-height, at different bit-depths. (**b**) Average matching accuracy as a function of *d*, at different bit-depths to highlight the general dependence of matching accuracy on bit-depth for pairs of mammograms.



(a)

(b)

**Figure 10.9**: Effect of $d$ on matching accuracy for TM-$D_E$ applied to pairs of mammograms. **(a)** Typical examples of $\kappa$ *vs.* $w$. Each row contains the information for a single image pair. **(b)** Average matching accuracy as a function of bit-depth, at different values of $d$ to highlight the general dependence of matching accuracy on $d$ for the pairs of mammograms.

### Summary of Best Results

The best matching results for each mammogram pair are plotted in **Figure 10.10** as functions of the ROI area and ROI visibility. This plot shows that $\kappa$ for the larger ROIs is relatively high and decreases as the ROIs get smaller. There is a wide range of matching accuracies for the very small ROIs. There is a poor correlation between matching accuracy and ROI area ($\rho$=0.54), but a good correlation between matching accuracy and ROI visibility ($\rho$=0.71).

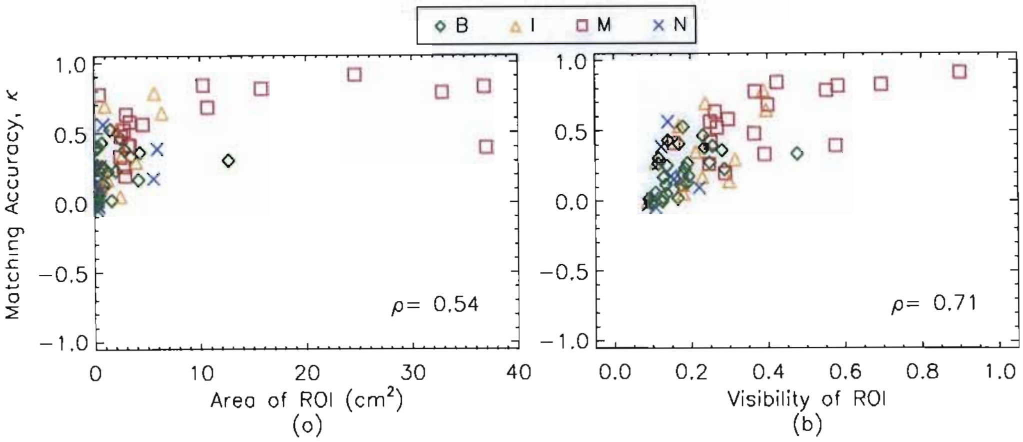


**Figure 10.10**: Best matching accuracy for TM-$D_E$ as functions of **(a)** ROI area and **(b)** ROI visibility for each mammogram pair.

**Figure 10.11** shows the best matching results for each mammogram pair, based on the diagnosis of the ROI. These results are summarised in **Table 10.6**. The results for the malignant masses are clustered towards the high $A_{ROC}$ values, but the $C_{fb}$-values range between 0.3 and 1.0. The average matching accuracy for the malignant masses is significantly higher than the matching accuracies for the other ROI-types($p < 0.02$) and all malignant masses were matched.

The results for the remaining ROI-types are quite scattered, and five ROIs (2 benign, 1 indeterminate, 2 normal) have not been matched. These ROIs have $A_{ROC} \leq 0.5$ and $C_{fb} \leq 0$.

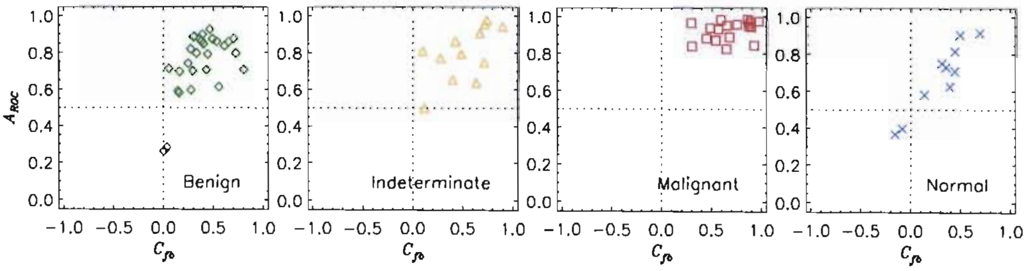


**Figure 10.11**: Scatter plot of the best results for each mammogram pair for TM-$D_E$, based on diagnosis.

**Table 10.6**: Average of the best matching results based on diagnosis for TM-$D_E$ applied to pairs of mammograms.

| Diagnosis | Average $A_{ROC}$ | Average $C_{fb}$ | Average $\kappa$ |
|---|---|---|---|
| Benign | 0.75±0.17 | 0.37±0.21 | 0.22±0.16 |
| Indeterminate | 0.80±0.15 | 0.50±0.26 | 0.34±0.27 |
| Malignant | 0.93±0.05 | 0.67±0.21 | 0.58±0.21 |
| Normal | 0.68±0.19 | 0.29±0.26 | 0.18±0.19 |
| All | 0.80±0.17 | 0.46±0.26 | 0.33±0.25 |

### 10.5.3 Evaluation of TM-Matching with $D_{ES}$

The results of applying TM-matching with a standardised Euclidean distance metric, $D_{ES}$, to pairs of mammograms, are presented and discussed. The effects of *nbits*, *d* and *w* on matching accuracy are examined. Results are presented in the formats described in §8.2 (page 93), with dependence on matching accuracy indicated by separation of the different coloured plots.

**Effect of *nbits* on Matching Accuracy**

Typical examples of $\kappa$ *vs.* $w$ are shown in **Figure 10.12(a)** to demonstrate variation of matching accuracy with *nbits* and *w*. All plots for *nbits*>5-bits overlap completely, for all examples. The 5-bit plots for M0M1, M39M38 and M28M29 deviate from the higher *nbits*-plots, but there is no consistent behaviour of matching accuracy for these plots.

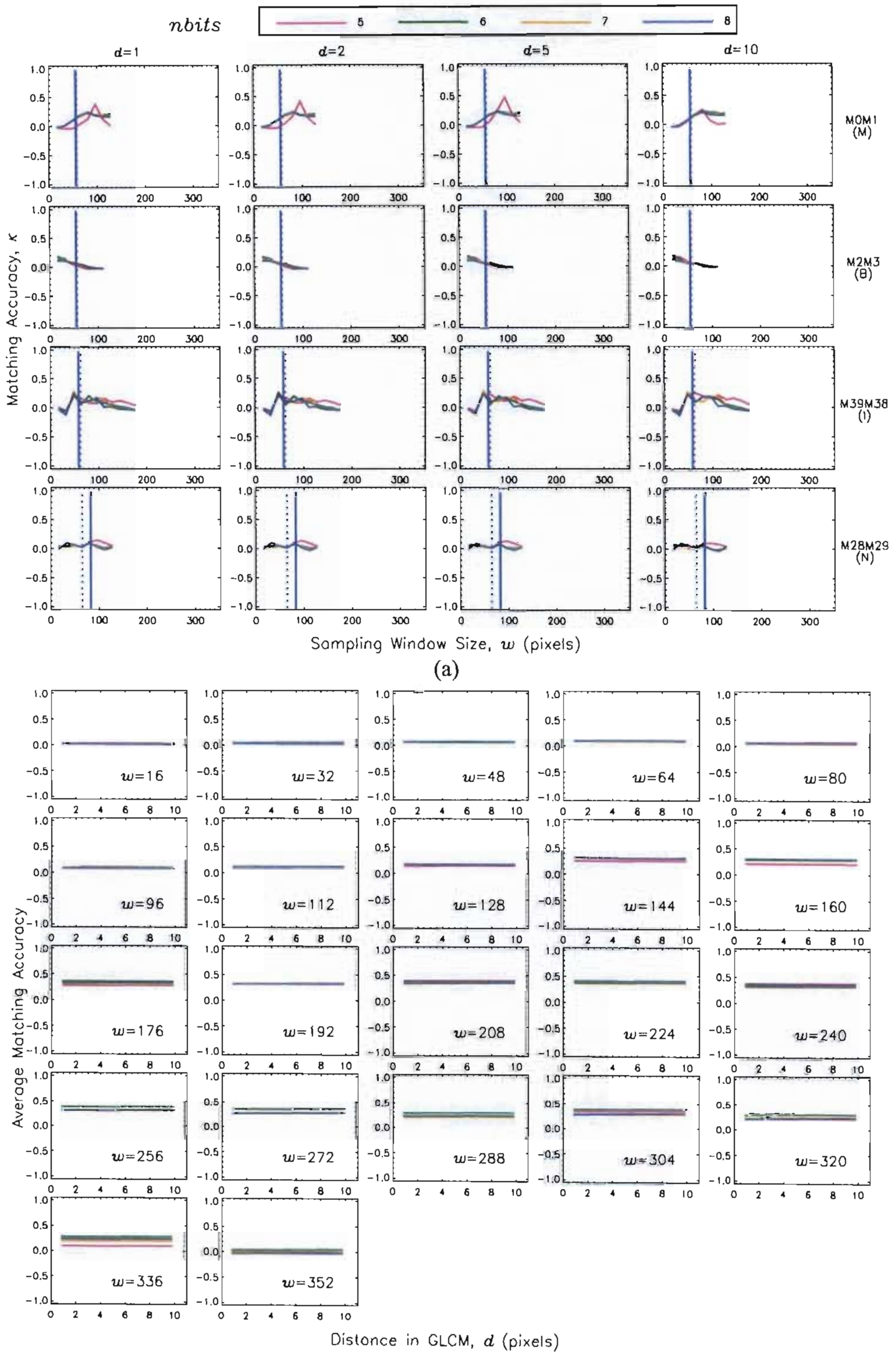
The results of the ACF analysis are plotted as vertical lines at each bit-depth, but there is very little difference between results at the different bit-depths. Also, the ACF-widths and ACF-heights are similar for three of the examples, but there appears to be poor correlation between the ACF results and the maximum values of $\kappa$. The ACF results are discussed in §10.8.

**Figure 10.12(b)** shows the average matching accuracy (over all image pairs) as a function of *d*, to highlight the general dependence of matching accuracy on *nbits*. The different degrees of separation for each sampling window size indicate that the dependence of matching accuracy on *nbits* is influenced by the size of the sampling window. Matching accuracy is dependent bit-depth for $w \leq 112$ pixels, since all these plots overlap almost completely. Matching accuracy increases as bit-depth increases for $w \geq 128$ pixels. For $w > 176$ pixels, there is no clear pattern of behaviour.

**Effect of *d* on Matching Accuracy**

Some examples of $\kappa$ *vs.* $w$ are shown in **Figure 10.13(a)** to demonstrate how matching accuracy varies with *d* and *w*. All plots generally overlap almost completely, indicating that matching accuracy is independent of *d* for these examples. There is some separation of plots for *nbits*=5 bits, but there is no consistent pattern of dependence of matching accuracy on *d*.

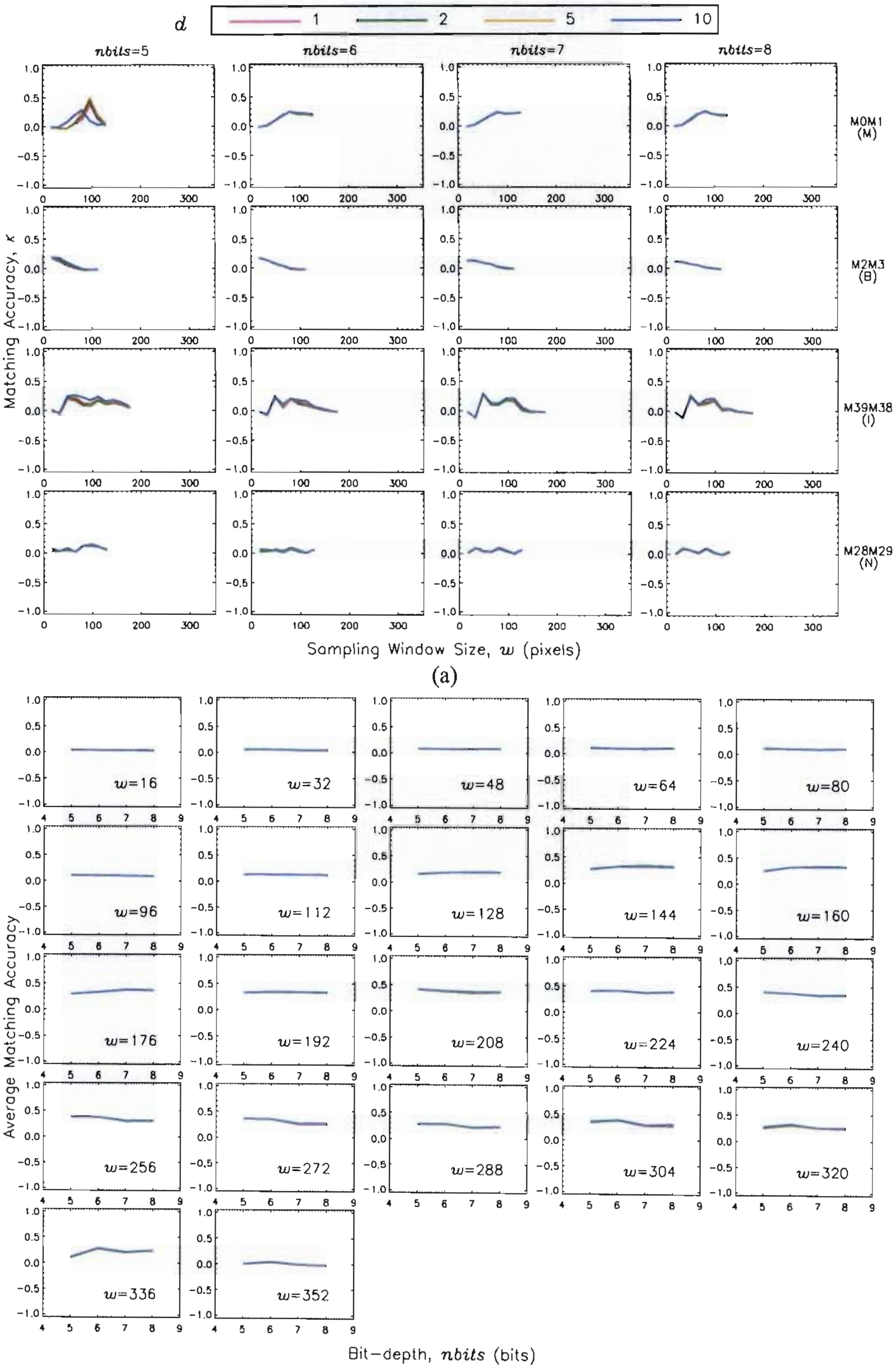
**Figure 10.13(b)** shows the average matching accuracy (over all image pairs) as a function of *nbits*, to highlight the general dependence of matching accuracy on *d*. Plots at the different values of *d* overlap almost completely for all sampling window sizes, indicating that matching accuracy is independent of *d*.



(a)

(b)

**Figure 10.12**: Effect of *nbits* on matching accuracy for TM-$D_{ES}$ applied to pairs of mammograms. **(a)** Typical examples of κ *vs. w*. Each row contains the information for a single image pair. The vertical lines indicate the results of the ACF analysis. The solid line represents the ACF-width and the dotted line represents the ACF-height, at different bit-depths. **(b)** Average matching accuracy as a function of *d*, at different bit-depths to highlight the general dependence of matching accuracy on bit-depth for pairs of mammograms.



(a)

(b)

**Figure 10.13**: Effect of $d$ on matching accuracy for TM-$D_{ES}$ applied to pairs of mammograms. **(a)** Typical examples of κ *vs.* $w$. Each row contains the information for a single image pair. **(b)** Average matching accuracy as a function of bit-depth, at different values of $d$ to highlight the general dependence of matching accuracy on $d$ for the pairs of mammograms.

### Effect of $w$ on Matching Accuracy

The example plots of $\kappa$ *vs.* $w$ in **Figures 10.12(a)** and **10.13(a)** show that matching accuracy varies as sampling window size varies. The variation is however unique to each image and the optimal sampling window size is most likely determined by the scale sizes of the textures in each image.

### Summary of Best Results

**Figure 10.14** shows the best matching accuracy as functions of ROI area and ROI visibility. Matching accuracy is generally higher for the larger, more visible ROIs. There is a correlation between matching accuracy and ROI area ($\rho=0.60$) and between matching accuracy and visibility ($\rho=0.74$).

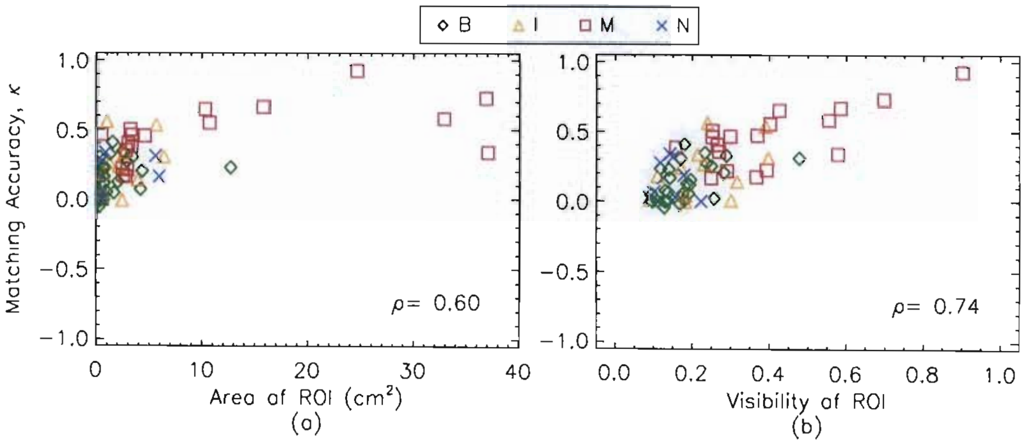


**Figure 10.14**: Best matching accuracy for TM-$D_{ES}$ as functions of **(a)** ROI area and **(b)** ROI visibility for each mammogram pair.

The best matching results for TM-$D_{ES}$ are shown separately for each ROI-type in **Figure 10.15**, with the results summarised in **Table 10.7**. Results appear similar to the results for TM-$D_E$. The results for the malignant masses are clustered at the high $A_{ROC}$-values for a wide range of $C_{fb}$-values. All malignant ROIs were matched, and had the highest matching accuracy. Results were statistically different ($p < 0.006$) to the results for the remaining ROI-types. The overall matching accuracies are lower than those for $D_E$.

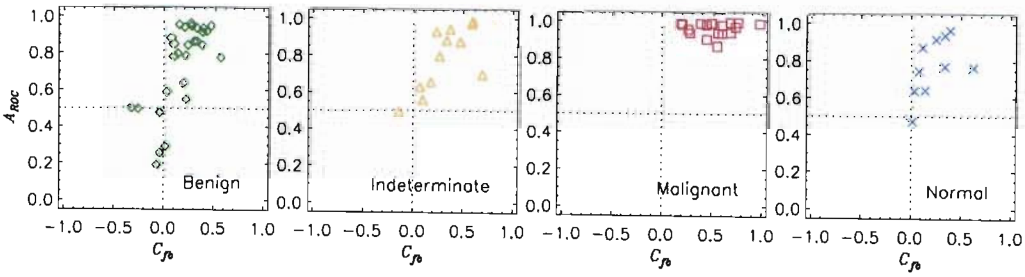


**Figure 10.15**: Scatter plot of the best results for each mammogram pair for TM-$D_{ES}$, based on diagnosis.

The results for the remaining ROI-types are scattered over a range of $A_{ROC}$- and $C_{fb}$-values. There were 10 ROIs (7 benign, 1 indeterminate, 2 normal) that were not matched.

**Table 10.7**: Average of the best matching results based on diagnosis for TM-$D_{ES}$ for pairs of mammograms.

| Diagnosis | Average $A_{ROC}$ | Average $C_{fb}$ | Average $\kappa$ |
|---|---|---|---|
| Benign | $0.74 \pm 0.23$ | $0.17 \pm 0.21$ | $0.14 \pm 0.13$ |
| Indeterminate | $0.79 \pm 0.17$ | $0.29 \pm 0.25$ | $0.22 \pm 0.20$ |
| Malignant | $0.96 \pm 0.04$ | $0.48 \pm 0.21$ | $0.45 \pm 0.20$ |
| Normal | $0.77 \pm 0.16$ | $0.21 \pm 0.19$ | $0.14 \pm 0.13$ |
| All | $0.81 \pm 0.20$ | $0.28 \pm 0.25$ | $0.24 \pm 0.21$ |

### 10.5.4 Evaluation of TM-Matching with $D_M$

The results of applying TM-matching with the Mahalanobis distance similarity metric, $D_M$, to pairs of mammograms, are presented and discussed. The effect of *nbits*, $d$ and $w$ on matching accuracy is examined. Results are presented in the formats described in §8.2 (page 93). Dependence on matching accuracy is indicated by separation of the different coloured plots.

#### Effect of *nbits* on Matching Accuracy

Typical examples of $\kappa$ *vs.* $w$ are shown in **Figure 10.16(a)** to demonstrate the effect of *nbits* and $w$ on matching accuracy. Matching accuracy is generally independent of $d$ for these examples, since most plots overlap completely.

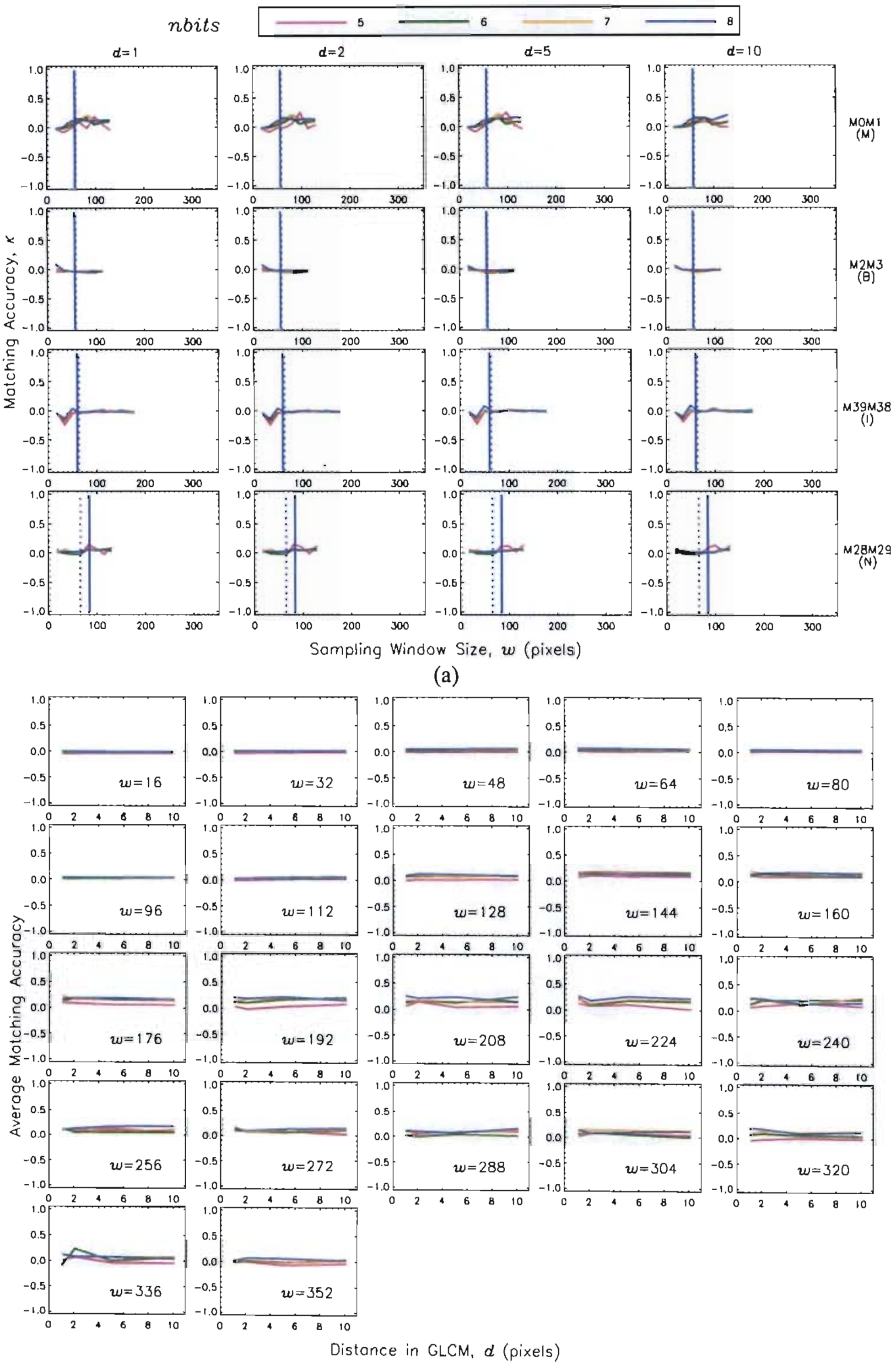
The results of the ACF analysis are plotted as vertical lines for each bit-depth, but there is very little difference in the results at the different bit-depths. The ACF-widths and ACF-heights are similar for three of the examples, but there appears to be poor correlation between the ACF results and the maximum values of $\kappa$. The ACF results are discussed in §10.8.

**Figure 10.16(b)** shows the average matching accuracy (over all image pairs) as a function of $d$, to highlight the general dependence of matching accuracy on *nbits*. The degrees of separation of the *nbits*-plots vary with sampling window size, but it appears that matching accuracy generally improves as bit-depth increases.

#### Effect of $d$ on Matching Accuracy

Some examples of $\kappa$ *vs.* $w$ are shown in **Figure 10.17(a)** to demonstrate how matching accuracy varies with $d$ and $w$. All $d$-plots generally overlap almost completely, indicating that matching accuracy is independent of $d$ for these examples. The 5-bit plots for M0M1 are separated with matching accuracy improving as $d$ decreases.

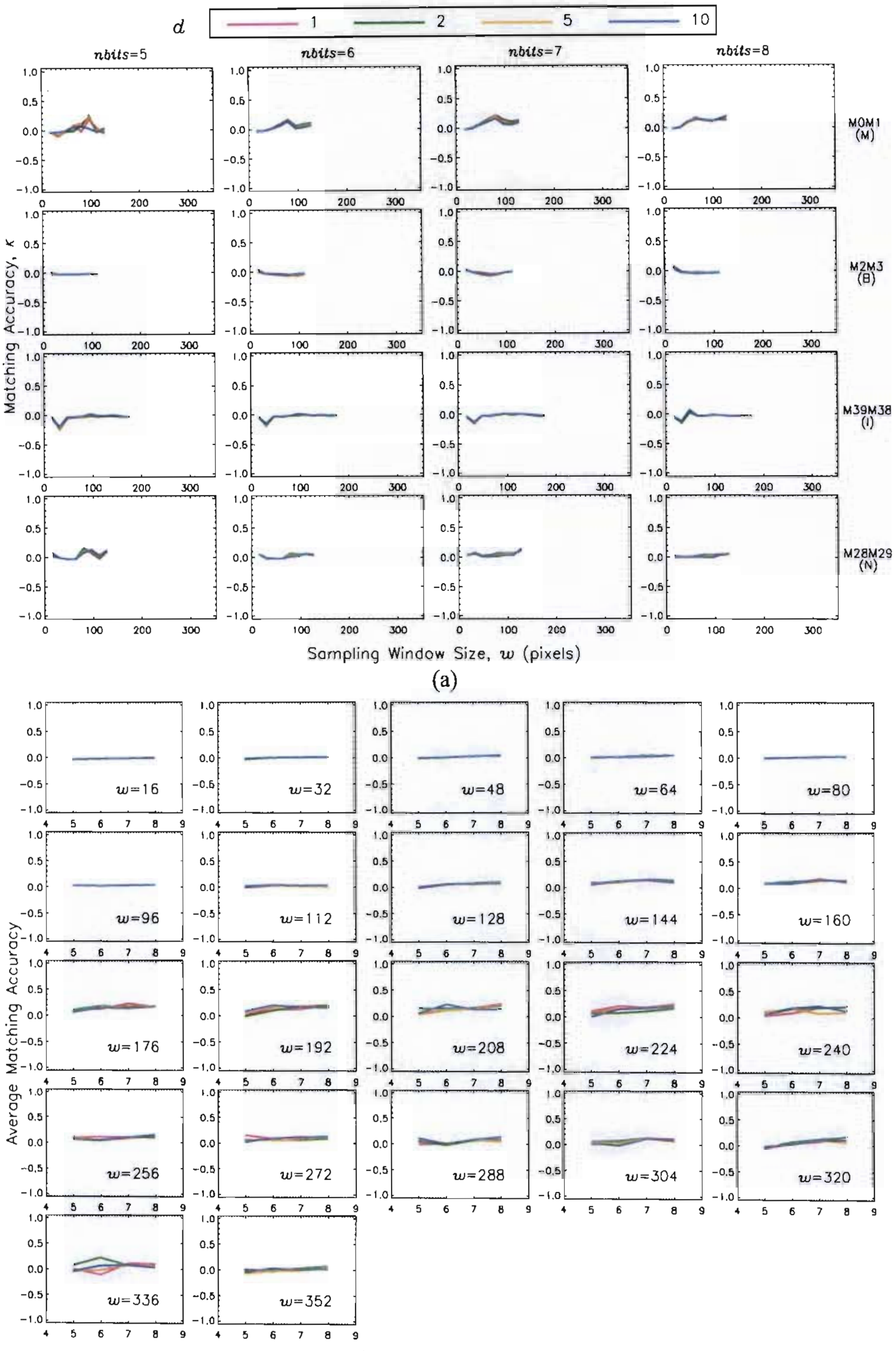
**Figure 10.17(b)** shows the average matching accuracy (over all image pairs) as a function of *nbits*, to highlight the general dependence of matching accuracy on $d$. Matching accuracy is independent of $d$ for $w \leq 160$ pixels, since these plots overlap almost completely. For $w > 160$ pixels, the dependence of matching accuracy on $d$ seems to be influenced by *nbits*, but there is no consistent pattern of behaviour.



(a)

(b)

**Figure 10.16**: Effect of *nbits* on matching accuracy for TM-$D_M$ applied to pairs of mammograms. **(a)** Typical examples of κ *vs.* *w*. Each row contains the information for a single image pair. The vertical lines indicate the results of the ACF analysis. The solid line represents the ACF-width and the dotted line represents the ACF-height, at different bit-depths. **(b)** Average matching accuracy as a function of *d*, at different bit-depths to highlight the general dependence of matching accuracy on bit-depth for pairs of mammograms.



(a)

(b)

**Figure 10.17**: Effect of $d$ on matching accuracy for TM-$D_M$ applied to pairs of mammograms. **(a)** Typical examples of $\kappa$ *vs.* $w$. Each row contains the information for a single image pair. **(b)** Average matching accuracy as a function of bit-depth, at different values of $d$ to highlight the general dependence of matching accuracy on $d$ for the pairs of mammograms.

### Effect of *w* on Matching Accuracy

The example plots of κ *vs.* *w* in **Figures 10.16(a)** and **10.17(a)** show that matching accuracy varies as sampling window size varies. The variation is however unique to each image and the optimal sampling window size is most likely determined by the scale sizes of the textures in each image.

### Summary of Best Results

**Figure 10.18** shows the best matching accuracy as functions of ROI area and ROI visibility. Matching accuracy is generally higher for the larger, more visible ROIs than for the smaller, less visible ROIs. There is a correlation between matching accuracy and ROI area (ρ=0.66) and between matching accuracy and ROI visibility (ρ=0.74).

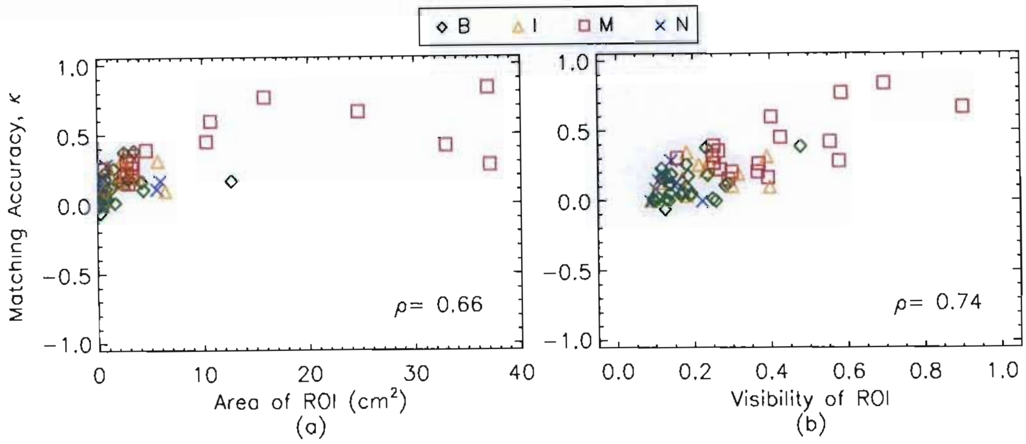


**Figure 10.18**: Best matching accuracy as functions of **(a)** ROI area and **(b)** ROI visibility for each mammogram pair.

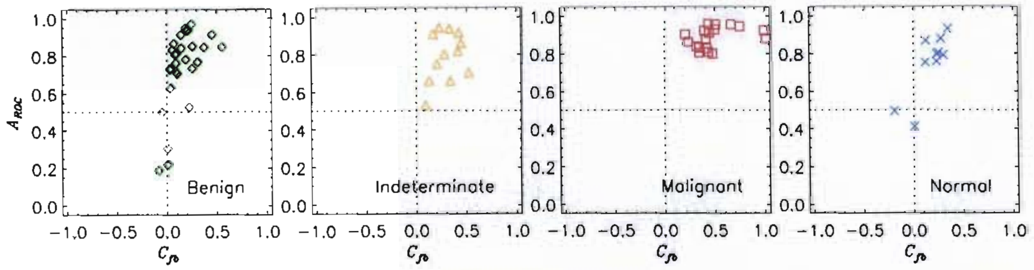


**Figure 10.19**: Scatter plot of the best results for each mammogram pair for TM-$D_M$, based on diagnosis.

**Figure 10.19** shows the best matching accuracies separated by ROI-type. These results are summarised in **Table 10.8**. Results for the malignant masses are clustered around high $A_{ROC}$-values but are spread over a range of $C_{fb}$-values. The malignant masses are matched with the highest accuracy and the average is statistically different ($p < 0.004$) to the averages for the rest of the ROI-types. Results for the remaining ROI-types are clustered around the lower $C_{fb}$-values and spread over a range of $A_{ROC}$-values. There are 6 ROIs (4 benign and 2 normal) that were not matched. All malignant and all indeterminate masses were matched. The overall matching accuracies are lower than those for $D_E$ and $D_{ES}$.

**Table 10.8**: Average of the best matching results based on diagnosis for TM-$D_M$ for pairs of mammograms.

| Diagnosis | Average $A_{ROC}$ | Average $C_{fb}$ | Average $\kappa$ |
|---|---|---|---|
| Benign | $0.74 \pm 0.21$ | $0.15 \pm 0.15$ | $0.10 \pm 0.11$ |
| Indeterminate | $0.79 \pm 0.13$ | $0.29 \pm 0.14$ | $0.18 \pm 0.11$ |
| Malignant | $0.89 \pm 0.06$ | $0.44 \pm 0.20$ | $0.36 \pm 0.19$ |
| Normal | $0.75 \pm 0.17$ | $0.16 \pm 0.16$ | $0.12 \pm 0.09$ |
| All | $0.79 \pm 0.17$ | $0.25 \pm 0.20$ | $0.19 \pm 0.17$ |

## 10.6 MI-Matching Evaluation Results

The MI-matching algorithm, using grey-level histograms and GLCMs to estimate the probability density functions (**Algorithm 8.2** on page 98) was applied to the mammogram images on page 145. The results are colour-coded according to the breast tissue type of the reference ROI, benign (B), indeterminate (I), malignant (M) and normal (N).

### 10.6.1 Evaluation of MI-Matching with Histograms

The results of applying MI-matching with histograms to pairs of mammograms, are presented and discussed. The dependence of matching accuracy on *nbits*, *nbins* and $w$ is examined. Results are presented in the formats described in §8.2 (page 93). Dependence on matching accuracy is indicated by separation of the different coloured plots.

#### Effect of *nbits* on Matching Accuracy

Typical examples of $\kappa$ *vs.* $w$ are shown in **Figure 10.20(a)** to demonstrate the effect of *nbits* and $w$ on matching accuracy. There is some separation of the plots for the smaller sampling windows, otherwise all plots overlap almost completely, for all examples. There is no clear dependence of matching accuracy on bit-depth.

The results of the autocorrelation function analysis are plotted as vertical lines for each value of *nbits*, but there is very little difference in the results at the different bit-depths. The ACF-widths and ACF-heights are similar for three of the examples, but there appears to be poor correlation between the ACF results and the maximum values of $\kappa$. The ACF results are discussed in §10.8.

**Figure 10.20(b)** shows the average matching accuracy (over all image pairs) as a function of *nbins*, to highlight the general dependence of matching accuracy on *nbits*. The slight separation of the *nbits*-plots for $w>160$ pixels indicates that matching accuracy improves as *nbits* decreases. Matching accuracy is independent of *nbits* for $w \leq 160$ pixels.

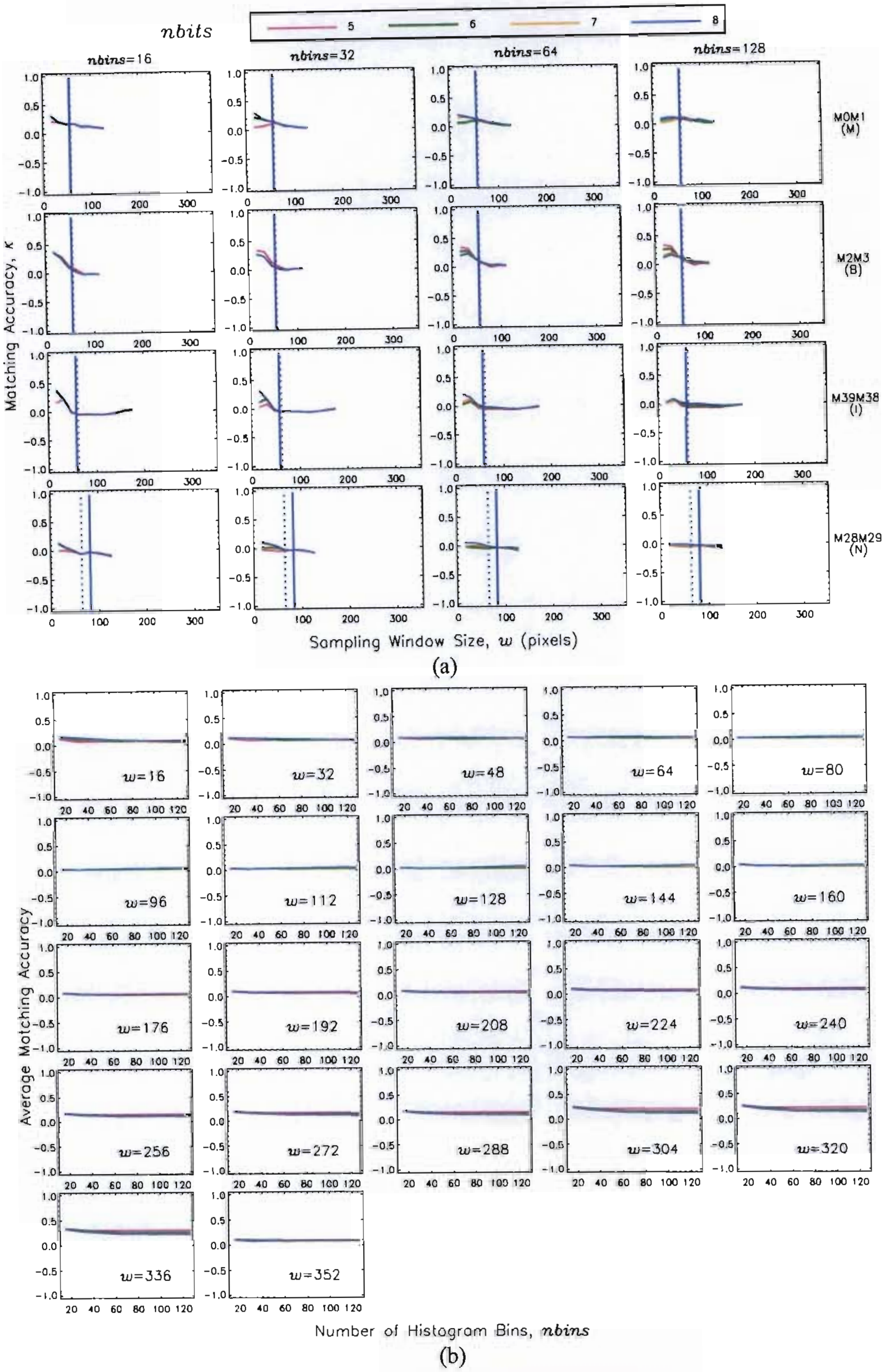


**Figure 10.20**: Effect of *nbits* on matching accuracy for MI-histograms applied to pairs of mammograms. **(a)** Typical examples of κ *vs.* *w*. Each row contains the information for a single image pair. The vertical lines indicate the results of the ACF analysis. The solid line represents the ACF-width and the dotted line represents the ACF-height, at different bit-depths. **(b)** Average matching accuracy as a function of *nbins*, at different bit-depths to highlight the general dependence of matching accuracy on bit-depth for pairs of mammograms.

### Effect of *nbins* on Matching Accuracy

Some examples of κ *vs.* *w* are shown in **Figure 10.21(a)** to demonstrate the effect *nbins* and *w* on matching accuracy. Three examples (M2M3, M39M38, M28M29) show some separation of the plots at the smaller window sizes, but the plots overlap almost completely, for the larger sampling windows. M0M1 shows that matching accuracy improves as *nbins* decreases.

**Figure 10.21(b)** shows the average matching accuracy (over all image pairs) as a function of *nbits*, to highlight the general dependence of matching accuracy on *nbins*. Separation of the *nbins*-plots increases as bit-depth increases and the degree of separation is influenced by the sampling window size, but matching accuracy, generally, improves as *nbins* is decreased.

### Effect of *w* on Matching Accuracy

The example plots of κ *vs.* *w* in **Figures 10.20(a)** and **10.21(a)** show that matching accuracy varies as sampling window size varies. The κ *vs.* *w* plots peak around a sampling window size of 16 pixels for all the examples.

### Summary of Best Results

**Figure 10.22** shows the best matching accuracy as functions of ROI area and ROI visibility. Matching accuracy is generally very low, irrespective of ROI area and there are only five very large malignant ROIs that have high matching accuracies. There is some dependence of matching accuracy on visibility, since matching accuracy appears to improve slightly as visibility increases. This is confirmed by the correlation coefficient of $\rho$=0.67 between matching accuracy and ROI visibility, while matching accuracy and ROI area are poorly correlated.

**Figure 10.23** shows the best matching accuracies separated by ROI-type. These results are summarised in **Table 10.9**. Results are generally clustered around high $A_{ROC}$-values and low $C_{fb}$-values for all ROI-types, and there are 3 image pairs (1 benign, 1 malignant, 1 normal) that were not matched. All indeterminate masses were matched. The average matching accuracies for each ROI-type are quite similar to each other, but are much lower than the average matching accuracies for the distance similarity metrics.

**Table 10.9**: Average of the best matching results based on diagnosis for MI-histograms for pairs of mammograms.

| Diagnosis | Average $A_{ROC}$ | Average $C_{fb}$ | Average κ |
|---|---|---|---|
| Benign | 0.84 ± 0.13 | 0.16 ± 0.15 | 0.12 ± 0.13 |
| Indeterminate | 0.85 ± 0.10 | 0.19 ± 0.10 | 0.13 ± 0.08 |
| Malignant | 0.88 ± 0.12 | 0.16 ± 0.20 | 0.15 ± 0.18 |
| Normal | 0.84 ± 0.06 | 0.12 ± 0.08 | 0.08 ± 0.06 |
| All | 0.85±0.11 | 0.16±0.15 | 0.12±0.13 |

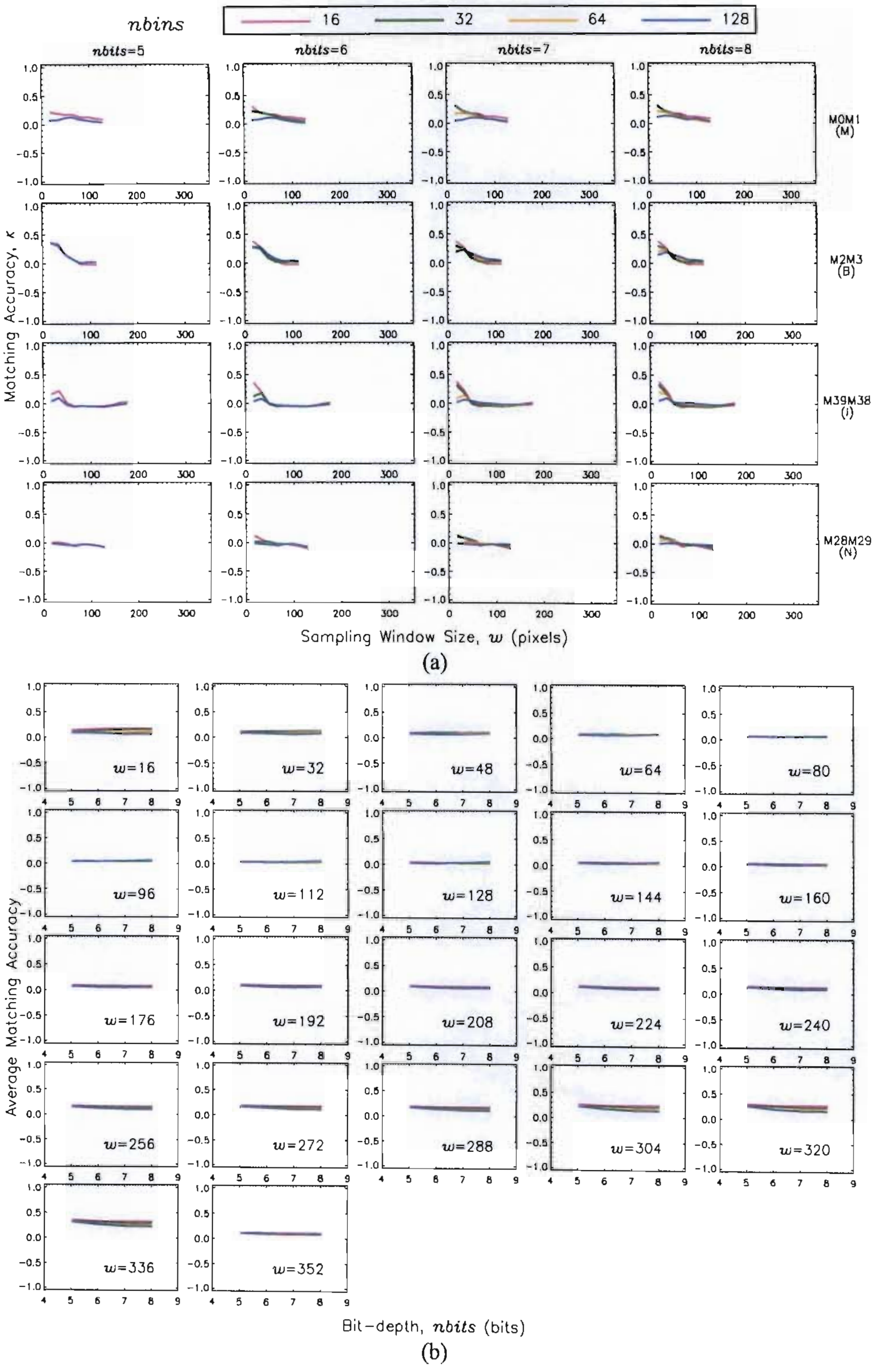


**Figure 10.21**: Effect of *nbins* on matching accuracy for MI-histograms applied to pairs of mammograms. **(a)** Typical examples of κ *vs.* *w*. Each row contains the information for a single image pair. **(b)** Average matching accuracy as a function of bit-depth, at different values of *nbins* to highlight the general dependence of matching accuracy on *nbins* for the pairs of mammograms.

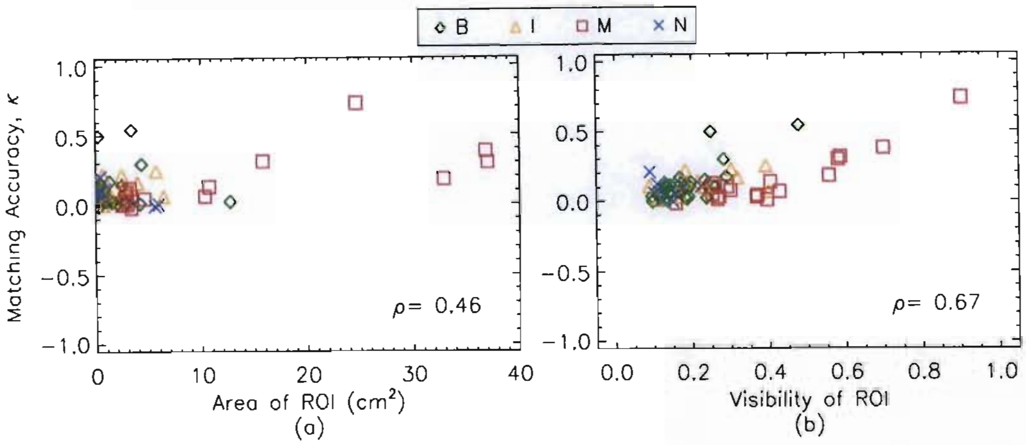


**Figure 10.22**: Best matching accuracy as functions of **(a)** ROI area and **(b)** ROI visibility for MI-histograms applied to pairs of mammograms.

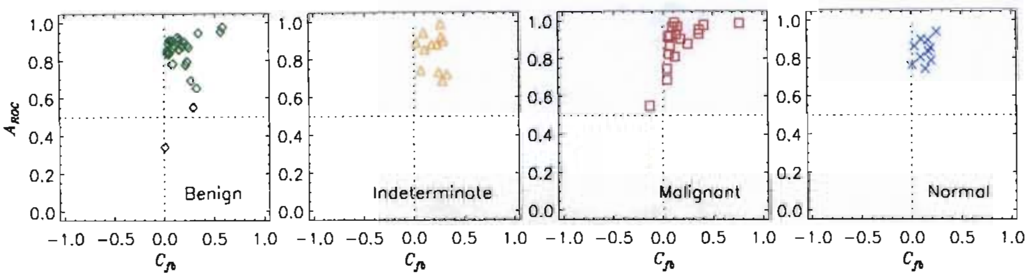


**Figure 10.23**: Scatter plot of the best results for each pair of mammograms for MI-histograms, based on diagnosis.

### 10.6.2 Evaluation of MI-Matching with GLCMs

The results of applying MI-matching with GLCMs to pairs of mammograms, are presented and discussed. The effect of *nbits*, *d* and *w* on matching accuracy is examined. Results are presented in the formats described in §8.2 (page 93). Dependence on matching accuracy is indicated by separation of the different coloured plots.

**Effect of *nbits* on Matching Accuracy**

Typical examples of κ *vs.* *w* are shown in **Figure 10.24(a)** to demonstrate the effect of *nbits* and *w* on matching accuracy. The examples show a very slight separation of the plots at different values of *nbits*, with matching accuracy generally improving as *nbits* increases.

The ACF results are plotted as vertical lines for each bit-depth, but there is no variation between the results at the different bit-depths. The ACF-widths and ACF-heights are similar for three of the examples, but the ACF results are poorly correlated with the maximum values of κ. The ACF results are discussed in §10.8.

**Figure 10.24(b)** shows the average matching accuracy (over all image pairs) as a function of *d*, to highlight the general dependence of matching accuracy on *nbits*. The different degrees of separation of the different *nbits*-plots, indicates that the dependence of matching accuracy on *nbits* is influenced by the size of the sampling window and by *d*. This is clear because the amount of separation of the plots at different values of *nbits* increases as sampling window size and *d* increase. However, there is a consistent improvement of matching accuracy as bit-depth increases.

### Effect of *d* on Matching Accuracy

Some examples of $\kappa$ *vs.* *w* are shown in **Figure 10.25(a)** to demonstrate how matching accuracy varies with *d* and *w*. All plots for the different values of *d* generally overlap almost completely, indicating that there is no dependence of matching accuracy on *d* for these examples.

**Figure 10.25(b)** shows the average matching accuracy (over all image pairs) as a function of *nbits*, to highlight the general dependence of matching accuracy on *d*. The dependence of matching accuracy on *d* is influenced by *nbits*, since the separation of the *d*-plots is greater at the smaller bit-depths and decreases as bit-depth increases. There are also a few sampling window sizes ($32 \leq w \leq 96$) with no separation, indicating an independence of matching accuracy on *d*. However, results indicate that matching accuracy, generally, improves as *d* decreases.

### Effect of *w* on Matching Accuracy

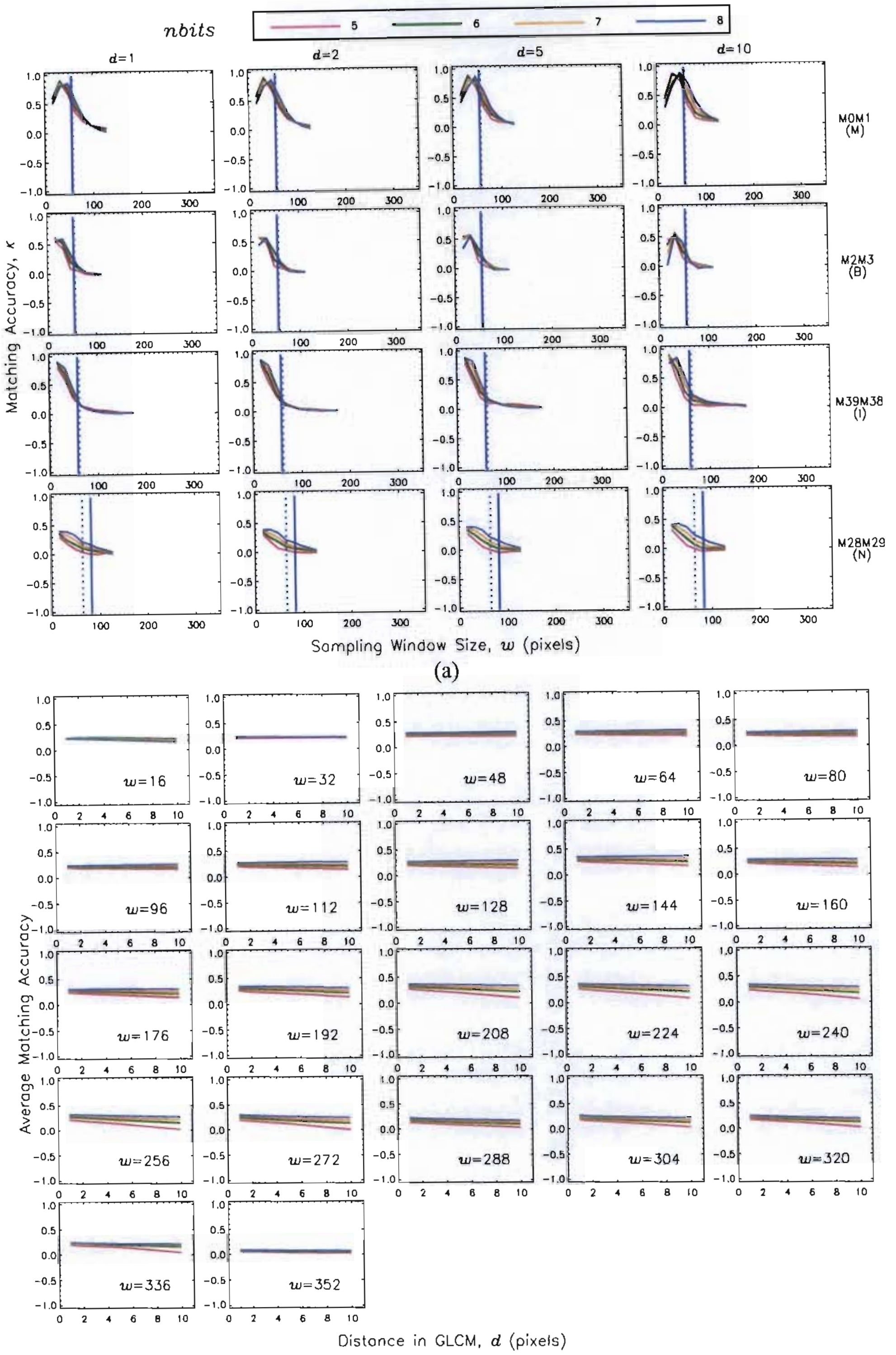
The example plots of $\kappa$ *vs.* *w* in **Figures 10.24(a)** and **10.25(a)** show that matching accuracy varies as sampling window size varies. Matching accuracy seems to peak at a small sampling window size and then decreases as sampling window size increases. The peak is most likely related to the scale sizes of the textures, and will therefore be unique to each image.

### Summary of Best Results

**Figure 10.26** shows the best matching accuracy as functions of ROI area and ROI visibility. The matching accuracy is spread over a wide range for the small, low visibility ROIs, while the large, high visibility ROIs generally have high matching accuracies. Matching accuracy is poorly correlated with ROI area ($\rho$=0.51) and there is a stronger correlation with ROI visibility ($\rho$=0.62).

**Figure 10.27** shows the best matching accuracies based on ROI-type. These results are summarised in **Table 10.10**. The results show that all, but three of the malignant masses are clustered around $A_{ROC}$=1 and $C_{fb}$=1. All malignant masses were matched. The malignant masses are matched with the highest accuracy and this is statistically different ($p < 0.002$) to the results for the other ROI-types.

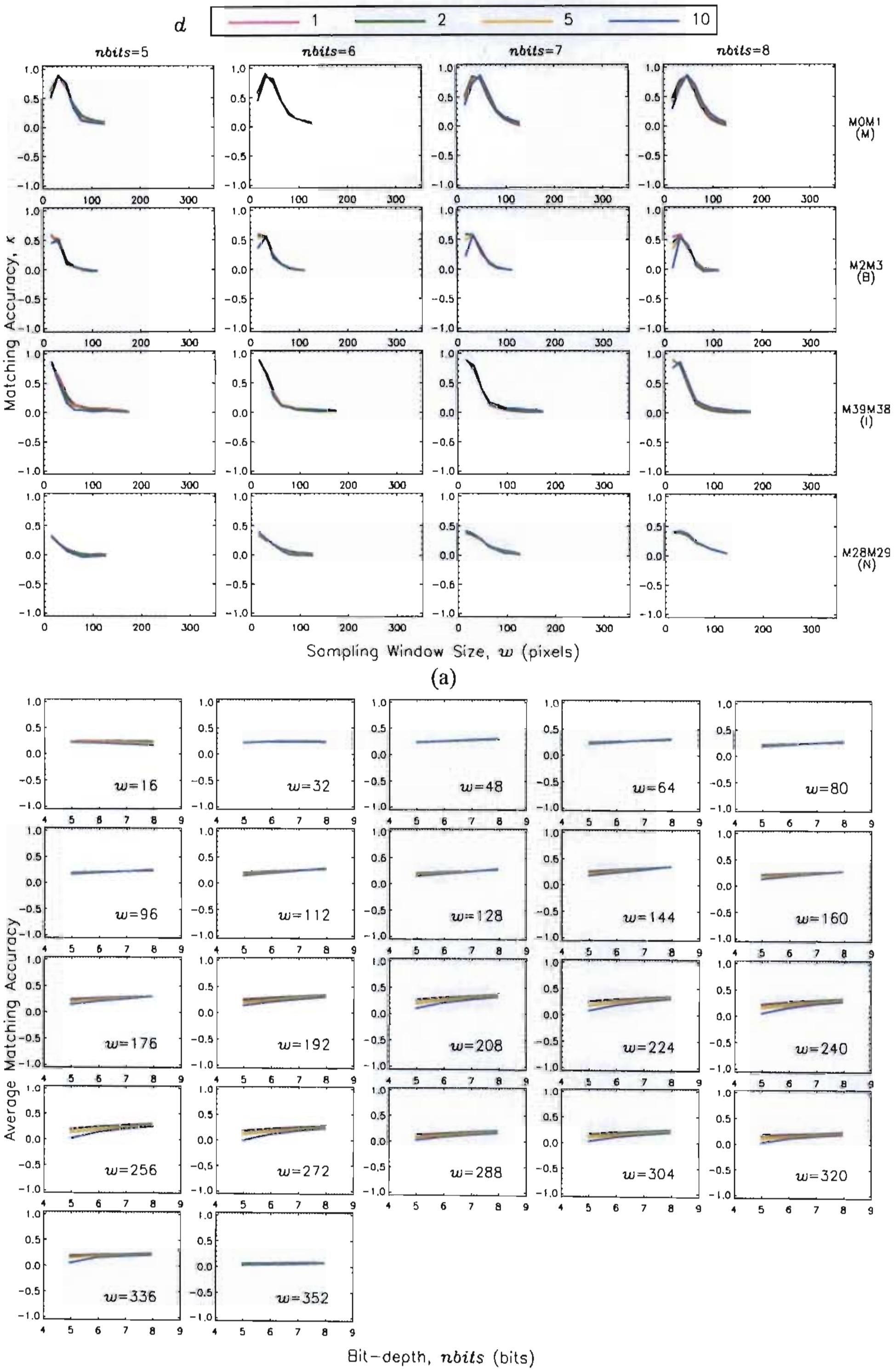
There is a significant spread of matching accuracies for the other ROI-types. There were 12 ROIs (8 benign, 2 indeterminate, 2 normal) that were not matched.



(a)

(b)

**Figure 10.24**: Effect of *nbits* on matching accuracy for MI-GLCMs applied to pairs of mammograms. **(a)** Typical examples of κ *vs.* *w*. Each row contains the information for a single image pair. The vertical lines indicate the results of the ACF analysis. The solid line represents the ACF-width and the dotted line represents the ACF-height, at different bit-depths. **(b)** Average matching accuracy as a function of *d*, at different bit-depths to highlight the general dependence of matching accuracy on bit-depth for pairs of mammograms.



(a)

(b)

**Figure 10.25**: Effect of *d* on matching accuracy for MI-GLCMs applied to pairs of mammograms. **(a)** Typical examples of κ *vs.* *w*. Each row contains the information for a single image pair. **(b)** Average matching accuracy as a function of bit-depth, at different values of *d* to highlight the general dependence of matching accuracy on *d* for the pairs of mammograms.

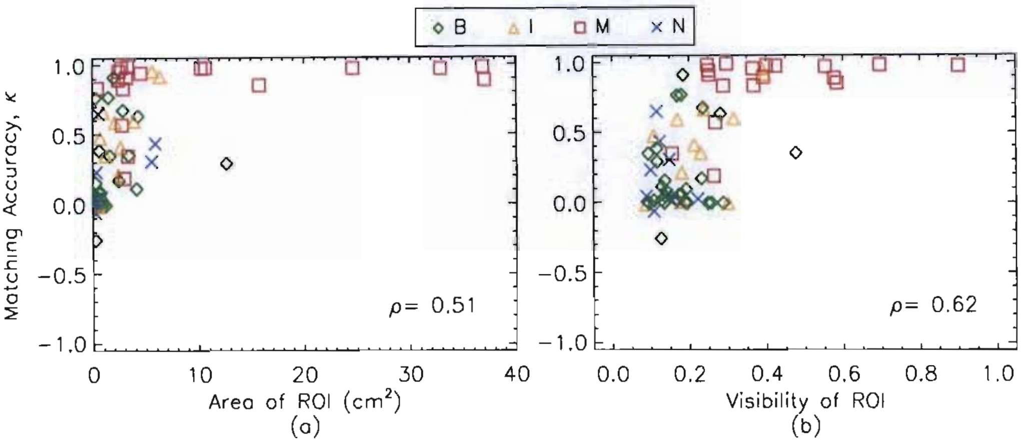


**Figure 10.26**: Best matching accuracy as functions of **(a)** ROI area and **(b)** ROI visibility for MI-GLCMs applied to pairs of mammograms.

**Table 10.10**: Average of the best matching results based on diagnosis for MI-GLCMs for pairs of mammograms.

| Diagnosis | Average $A_{ROC}$ | Average $C_{fb}$ | Average κ |
|---|---|---|---|
| Benign | 0.68 ± 0.26 | 0.29 ± 0.36 | 0.21 ± 0.30 |
| Indeterminate | 0.75 ± 0.28 | 0.59 ± 0.40 | 0.44 ± 0.34 |
| Malignant | 0.96 ± 0.05 | 0.90 ± 0.21 | 0.84 ± 0.23 |
| Normal | 0.71 ± 0.24 | 0.26 ± 0.29 | 0.17 ± 0.23 |
| All | 0.77±0.25 | 0.50±0.42 | 0.41±0.39 |

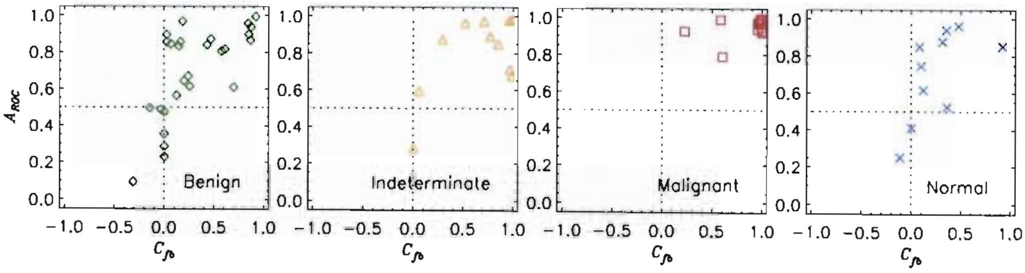


**Figure 10.27**: Scatter plot of the best results for each pair of mammograms for MI-GLCMs, based on diagnosis.

## 10.7 Overall Matching Results

The matching results for each of the five matching methods are compared and discussed.

### 10.7.1 Results

**Figure 10.28** shows the best matching results for each of the 68 pairs of mammograms, for each matching method. The $A_{ROC}$- and $C_{fb}$-values are only clustered for MI-histograms and are generally quite scattered for the remaining four methods (TM-$D_E$, TM-$D_{ES}$, TM-$D_M$, MI-GLCMs). Some results have $A_{ROC}<0.5$ and $C_{fb}<0.0$, indicating that the match was unsuccessful. Matching accuracies are quite high for TM-$D_E$, TM-$D_{ES}$ and MI-GLCMs, while TM-$D_M$ and MI-histograms both have low $C_{fb}$-values. This is confirmed by examining the average of the best matching accuracies that are summarised in **Table 10.11**. Results are generally lower than the equivalent values for the mosaic images.

This is most likely due to the quality of the ground truth data. Since there were clear borders between the textures in the mosaic images, the ground truth data was accurate. However, for the pairs of mammograms, only one radiologist marked the borders of the ROIs in each mammogram and there is no method of confirming the accuracy of the identified borders.

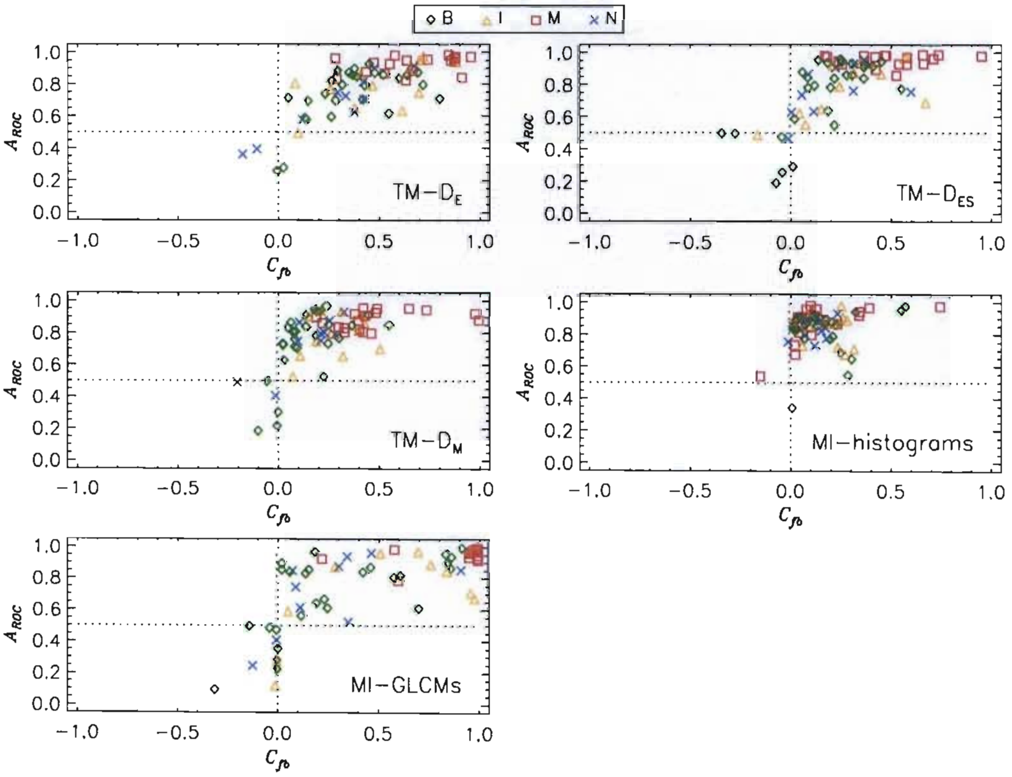


**Figure 10.28**: Scatter plot of $A_{ROC}$ *vs.* $C_{fb}$ with the best matching accuracies for all matching methods and for each mammogram pair.

**Table 10.11**: Average of the best matching accuracies for each matching method for pairs of mammograms.

| Method | average $A_{ROC}$ | average $C_{fb}$ | average $\kappa$ |
|---|---|---|---|
| TM-$D_E$ | 0.80±0.17 | 0.46±0.26 | 0.33±0.25 |
| TM-$D_{ES}$ | 0.81±0.20 | 0.28±0.25 | 0.24±0.21 |
| TM-$D_M$ | 0.79±0.17 | 0.25±0.20 | 0.19±0.17 |
| MI-histograms | 0.85±0.11 | 0.16±0.15 | 0.12±0.13 |
| MI-GLCMs | 0.77±0.25 | 0.50±0.42 | 0.41±0.39 |

### 10.7.2 Statistical Significance Analysis

The results of performing a paired *t*-test analysis (§8.10 on page 103) on the distribution of the best $\kappa$-values for each method is presented in **Table 10.12**. For a significance level of 0.05, the average values of $\kappa$ for TM-$D_E$ compared to MI-GLCMs and for TM-$D_{ES}$ compared to TM-$D_M$ are not statistically different.

**Table 10.12**: Results of significance (*t*-test) analysis for best matching accuracies from pairs of mammograms for the various matching methods. For a significance level of 0.05, the average values for the best matching accuracies of TM-$D_E$ compared to MI-GLCMs and TM-$D_{ES}$ compared to TM-$D_M$ are not statistically different.

| Method | TM-$D_{ES}$ | | TM-$D_M$ | | MI-histograms | | MI-GLCMs | |
|---|---|---|---|---|---|---|---|---|
| | t-value | p-value | t-value | p-value | t-value | p-value | t-value | p-value |
| TM-$D_E$ | 2.43 | 0.02 | 3.99 | 0.00 | 6.07 | 0.00 | -1.40 | 0.16 |
| TM-$D_{ES}$ | | | 1.56 | 0.12 | 3.77 | 0.00 | -3.27 | 0.00 |
| TM-$D_M$ | | | | | 2.36 | 0.02 | -4.38 | 0.00 |
| MI-histograms | | | | | | | -5.77 | 0.00 |

### 10.7.3 TM-Matching

The combination of matching parameters that yielded the best matches for TM-matching is listed in **Table A.3** for each image pair.

Most of the reference ROIs were well matched to the corresponding region in the test image, for all distance similarity metrics. $D_E$ performed best as a similarity metric with fewest non-matches and the highest average $\kappa$.

All malignant masses were matched for all distance similarity metrics, and had the highest matching accuracies. Results were generally clustered around high $A_{ROC}$-values, but were spread over a range of contrast values. The malignant masses probably have higher matching accuracies because these are generally texturally homogeneous and the method used of averaging the GLCM over four angles is ideally suited to detecting homogeneous textures.

Results for the benign and indeterminate masses and the normal ROIs were scattered across a range of $A_{ROC}$- and $C_{fb}$-values and there were a total of 5, 8 and 6 ROIs not matched for TM-$D_E$, TM-$D_{ES}$ and TM-$D_M$, respectively. Poor results for these ROIs could be because benign and normal regions in the breast are generally not homogeneous and directional information, which could improve the matching accuracy, might be lost by averaging the GLCMs over four angles.

Another factor that contributed to the poor matching results for the non-malignant ROIs was the ROI area and the ROI visibility. Most of the benign, indeterminate and normal ROIs were either very small ($<0.5\ cm^2$) or had a low visibility, two factors which generally resulted in a poor matching accuracy. There was no general correlation between matching accuracy and ROI area, but the larger ROIs were more accurately matched than the smaller ROIs. There was however a spread of matching accuracies for the very small ROIs. Matching accuracy was generally correlated with ROI visibility, and the ROIs with low visibilities had very poor matching accuracies.

The mammogram pairs that were not matched for each method are listed in **Table 10.13**.

**Table 10.13**: Pairs of mammograms that were not matched with TM-matching

| TM-$D_E$ | TM-$D_{ES}$ | | TM-$D_M$ |
|---|---|---|---|
| M56M57 | M7M6 | M47M46 | M7M6 |
| M52M53 | M13M12 | M56M57 | M12M13 |
| M25M24 | M25M24 | M65M64 | M25M24 |
| M48M49 | M50M51 | M66M67 | M52M53 |
| M58M59 | M59M58 | M58M59 | M58M59 |
| | | | M65M64 |

Of the total of 19 mammogram pairs that were not matched, there are only 12 unique pairs, with M25M24 and M58M59 appearing for all distance similarity metrics. M24 (area=0.22 cm$^2$ and visibility=0.19) and M25 (area=0.25 cm$^2$ and visibility=0.13) are benign masses in a mixed breast and are small with low visibilities. M58 (area=0.27 cm$^2$ and visibility=0.10) and M59 (area=0.27 cm$^2$ and visibility=0.11) are benign masses in a fatty breast and are also small with low visibilities. The remaining ROIs have varying areas, but visibility<0.30 for all ROIs (**Table 10.3**).

The generally low visibilities of the mammogram pairs that were not matched indicate that ROI visibility is an important factor in determining matching performance.

Overall matching accuracy decreased from TM-$D_E$ to TM-$D_{ES}$ to TM-$D_M$, which leads to the suspicion that standardisation ($D_{ES}$) and the use of the covariance matrix ($D_M$) are not suited to the analysis of the texture measures resulting from this set of mammograms.

The dependence of matching accuracy on sampling window size was unique to the image pair and the dependence of matching accuracy on bit-depth was influenced by the sampling window size. This is most likely because the scale sizes of features in the mammogram change as bit-depth is changed. This leads to matching at a different optimal sampling window size for each bit-depth. Matching accuracy was generally independent of $d$, for all distance similarity metrics.

Overall, the effects of the various matching parameters on matching accuracy are similar to the effects noted for the mosaic images. However, the average matching accuracies for each method were considerably lower than for the mosaic images.

### 10.7.4 MI-Matching

The combination of matching parameters that yielded the best matches for MI-matching is listed in **Table A.4** for each image pair.

The matching accuracy of MI-GLCMs was considerably higher than the matching accuracy of MI-histograms (at a significance level of 0.00 from Table ??).

The poor matching results for MI-histograms are most likely due to two reasons. The first reason stems from the method used to compute the histograms (§8.4.1). The histogram was computed between the minimum and maximum grey-levels of both the reference and test sub-images. This range was then divided into *nbins* bins. This implementation has two disadvantages, firstly the width of the histogram bins is not equal for the comparison over an image, and secondly, the histograms are not invariant to shifts in grey-level.

The second reason might be because histograms do not incorporate any spatial information. Therefore using histograms to estimate probability density functions for the purpose of matching textures does not completely describe the information contained in the image.

For GLCMs, the malignant masses had the highest matching accuracies, but the results for the remaining ROI-types were much lower. All malignant masses were well matched. The good results for malignant masses and the poor results for non-malignant ROIs are most likely because the averaging of the GLCMs over four angles is biased towards detecting the texturally homogeneous malignant masses, but is not suited to detecting the non-homogeneous textures associated with benign and normal breast tissue.

Matching accuracy was correlated with ROI visibility for MI-histograms and MI-GLCMs, but was poorly correlated with ROI area, for both methods.

MI-histograms had optimal matching accuracy at the smallest window sizes. MI-GLCMs also had optimal matching accuracy for the smaller sampling window sizes. Each image showed matching accuracy increasing to a peak at a small sampling window size and then decreased as the sampling window size increased further. The position of the peak is most likely determined by the scale sizes of textures and is unique to each image.

For MI-histograms matching accuracy was optimal at the lowest bit-depths and for the fewest number of histogram bins. For MI-GLCMs, matching accuracy was optimal at the highest bit-depth and the smallest *d*.

There were 3 image pairs that were not matched for MI-histograms and 12 that were not matched for MI-GLCMs. These are listed in **Table 10.14**. Of the 15 pairs not matched, only M58M59 was common to both MI-histograms and MI-GLCMs. For histograms, there was a range of ROI areas from 0.27 cm$^2$ to 5.91 cm$^2$. For GLCMs, ROI areas ranged from 0.16 cm$^2$ to 1.21 cm$^2$ and were generally smaller than those not matched for histograms. The generally low visibilities of the mammogram pairs that were not matched indicate that ROI visibility is an important factor in determining matching performance.

**Table 10.14**: Pairs of mammograms that were not matched with MI-matching

| MI-histograms | MI-GLCM | | | |
|---|---|---|---|---|
| M22M23 | M3M2 | M25M24 | M12M13 | M56M57 |
| M29M28 | M6M7 | M48M49 | M17M16 | M57M56 |
| M58M59 | M7M6 | M49M48 | M24M25 | M58M59 |

**Table 10.15**: Optimal sampling window sizes for mammograms for $D_E$, $D_{ES}$, $D_M$, histograms (GLH) and GLCMs (G).

| Image | 5-bits | | | | | 6-bits | | | | | 7-bits | | | | | 8-bits | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $D_E$ | $D_{ES}$ | $D_M$ | GLH | G | $D_E$ | $D_{ES}$ | $D_M$ | GLH | G | $D_E$ | $D_{ES}$ | $D_M$ | H | G | $D_E$ | $D_{ES}$ | $D_M$ | H | G |
| M0 | 80 | 80 | 80 | 16 | 32 | 80 | 80 | 80 | 16 | 32 | 80 | 80 | 80 | 16 | 48 | 80 | 80 | 128 | 16 | 48 |
| M1 | 80 | 80 | 64 | 32 | 16 | 80 | 80 | 80 | 16 | 16 | 80 | 80 | 64 | 16 | 16 | 80 | 80 | 64 | 16 | 32 |
| M2 | 32 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 32 |
| M3 | 32 | 48 | 48 | 32 | 80 | 32 | 48 | 32 | 32 | 80 | 48 | 48 | 32 | 32 | 64 | 48 | 48 | 48 | 32 | 80 |
| M4 | 16 | 16 | 48 | 16 | 48 | 16 | 16 | 32 | 16 | 64 | 48 | 32 | 48 | 16 | 16 | 64 | 32 | 32 | 16 | 16 |
| M5 | 16 | 16 | 16 | 32 | 32 | 16 | 16 | 16 | 32 | 32 | 16 | 16 | 32 | 32 | 32 | 16 | 48 | 48 | 32 | 32 |
| M6 | 32 | 16 | 16 | 48 | 16 | 16 | 16 | 16 | 16 | 16 | 48 | 16 | 32 | 32 | 16 | 16 | 16 | 48 | 16 | 16 |
| M7 | 48 | 48 | 16 | 32 | 48 | 16 | 48 | 48 | 16 | 48 | 48 | 48 | 48 | 16 | 32 | 48 | 48 | 16 | 16 | 48 |
| M8 | 16 | 16 | 32 | 16 | 16 | 32 | 16 | 32 | 16 | 16 | 32 | 16 | 32 | 16 | 16 | 32 | 32 | 16 | 16 | 16 |
| M9 | 48 | 16 | 48 | 32 | 16 | 32 | 48 | 16 | 32 | 16 | 16 | 16 | 16 | 32 | 16 | 16 | 16 | 16 | 32 | 32 |
| M10 | 48 | 48 | 16 | 48 | 32 | 16 | 64 | 48 | 48 | 32 | 16 | 48 | 48 | 48 | 32 | 32 | 32 | 16 | 48 | 48 |
| M11 | 16 | 16 | 16 | 32 | 32 | 32 | 32 | 32 | 32 | 48 | 64 | 16 | 32 | 32 | 48 | 32 | 32 | 32 | 32 | 64 |
| M12 | 16 | 16 | 32 | 16 | 16 | 16 | 16 | 32 | 16 | 16 | 32 | 32 | 32 | 16 | 16 | 32 | 32 | 32 | 16 | 16 |
| M13 | 16 | 32 | 32 | 16 | 16 | 32 | 32 | 32 | 16 | 32 | 32 | 32 | 16 | 16 | 32 | 32 | 32 | 16 | 16 | 16 |
| M14 | 48 | 96 | 96 | 16 | 32 | 48 | 64 | 96 | 16 | 32 | 48 | 64 | 48 | 16 | 32 | 48 | 64 | 48 | 16 | 32 |
| M15 | 64 | 64 | 64 | 32 | 16 | 48 | 64 | 48 | 16 | 16 | 48 | 64 | 48 | 16 | 16 | 48 | 64 | 48 | 16 | 32 |
| M16 | 112 | 112 | 64 | 16 | 96 | 112 | 112 | 112 | 16 | 112 | 16 | 16 | 112 | 32 | 112 | 32 | 32 | 48 | 32 | 112 |
| M17 | 16 | 32 | 64 | 32 | 96 | 16 | 32 | 32 | 48 | 96 | 96 | 32 | 16 | 48 | 16 | 48 | 32 | 16 | 48 | 16 |
| M18 | 16 | 16 | 32 | 16 | 16 | 16 | 32 | 32 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 |
| M19 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 |
| M20 | 80 | 80 | 80 | 16 | 32 | 80 | 80 | 64 | 16 | 32 | 80 | 80 | 48 | 16 | 32 | 64 | 80 | 48 | 16 | 32 |
| M21 | 64 | 64 | 48 | 16 | 48 | 64 | 64 | 64 | 16 | 32 | 64 | 64 | 64 | 16 | 48 | 64 | 64 | 64 | 16 | 48 |
| M22 | 80 | 96 | 96 | 48 | 16 | 96 | 96 | 96 | 32 | 16 | 96 | 96 | 96 | 16 | 16 | 96 | 96 | 96 | 16 | 16 |
| M23 | 96 | 112 | 96 | 32 | 32 | 96 | 112 | 128 | 16 | 32 | 96 | 112 | 112 | 16 | 32 | 112 | 112 | 128 | 16 | 32 |
| M24 | 32 | 16 | 16 | 16 | 16 | 32 | 32 | 16 | 16 | 16 | 16 | 16 | 32 | 16 | 16 | 32 | 16 | 32 | 16 | 16 |
| M25 | 32 | 32 | 32 | 32 | 32 | 32 | 16 | 32 | 16 | 32 | 32 | 16 | 16 | 16 | 32 | 32 | 32 | 16 | 16 | 32 |
| M26 | 16 | 16 | 256 | 16 | 32 | 256 | 32 | 32 | 16 | 32 | 80 | 16 | 32 | 16 | 48 | 256 | 16 | 32 | 16 | 48 |
| M27 | 256 | 256 | 256 | 16 | 32 | 256 | 256 | 32 | 16 | 32 | 256 | 256 | 32 | 16 | 48 | 256 | 256 | 32 | 16 | 48 |
| M28 | 128 | 96 | 80 | 32 | 16 | 128 | 80 | 80 | 16 | 16 | 128 | 80 | 128 | 16 | 16 | 128 | 80 | 128 | 16 | 32 |
| M29 | 32 | 32 | 96 | 64 | 16 | 48 | 48 | 32 | 16 | 16 | 128 | 48 | 32 | 32 | 16 | 96 | 48 | 80 | 16 | 32 |
| M30 | 176 | 144 | 144 | 80 | 64 | 112 | 144 | 144 | 32 | 80 | 112 | 144 | 176 | 16 | 80 | 128 | 144 | 144 | 16 | 80 |
| M31 | 96 | 112 | 176 | 48 | 64 | 112 | 112 | 144 | 32 | 64 | 160 | 144 | 144 | 16 | 64 | 112 | 176 | 128 | 16 | 64 |
| M32 | 80 | 32 | 64 | 48 | 16 | 32 | 32 | 32 | 16 | 16 | 80 | 48 | 48 | 16 | 16 | 48 | 48 | 48 | 16 | 16 |
| M33 | 32 | 16 | 112 | 32 | 16 | 32 | 112 | 80 | 16 | 16 | 32 | 80 | 80 | 16 | 16 | 16 | 80 | 80 | 16 | 16 |
| M34 | 48 | 80 | 80 | 128 | 16 | 64 | 80 | 64 | 16 | 32 | 64 | 80 | 80 | 16 | 32 | 64 | 80 | 80 | 16 | 32 |
| M35 | 64 | 64 | 96 | 32 | 16 | 64 | 64 | 80 | 16 | 16 | 80 | 64 | 80 | 16 | 16 | 64 | 64 | 64 | 16 | 16 |
| M36 | 16 | 80 | 80 | 16 | 32 | 32 | 80 | 80 | 16 | 32 | 32 | 80 | 80 | 16 | 48 | 32 | 80 | 80 | 16 | 48 |
| M37 | 32 | 64 | 64 | 64 | 16 | 80 | 80 | 96 | 16 | 16 | 80 | 96 | 64 | 16 | 16 | 80 | 80 | 64 | 16 | 16 |
| M38 | 64 | 64 | 144 | 16 | 48 | 64 | 64 | 64 | 16 | 64 | 64 | 64 | 64 | 16 | 64 | 64 | 64 | 64 | 16 | 64 |
| M39 | 48 | 64 | 96 | 32 | 16 | 48 | 48 | 96 | 16 | 16 | 48 | 48 | 112 | 16 | 16 | 48 | 48 | 48 | 16 | 16 |
| M40 | 48 | 48 | 64 | 16 | 16 | 64 | 64 | 64 | 16 | 16 | 64 | 64 | 48 | 16 | 16 | 48 | 48 | 64 | 16 | 16 |
| M41 | 48 | 48 | 48 | 48 | 16 | 48 | 48 | 48 | 32 | 16 | 32 | 48 | 16 | 32 | 16 | 16 | 48 | 48 | 32 | 16 |
| M42 | 64 | 80 | 112 | 32 | 16 | 64 | 32 | 32 | 16 | 16 | 64 | 48 | 48 | 16 | 16 | 64 | 48 | 48 | 16 | 16 |
| M43 | 32 | 16 | 96 | 16 | 16 | 16 | 112 | 112 | 32 | 16 | 16 | 16 | 112 | 32 | 16 | 80 | 16 | 112 | 16 | 16 |
| M44 | 144 | 176 | 160 | 16 | 96 | 176 | 176 | 176 | 16 | 112 | 176 | 176 | 176 | 16 | 112 | 160 | 176 | 176 | 16 | 128 |
| M45 | 128 | 112 | 112 | 32 | 32 | 128 | 128 | 128 | 16 | 32 | 128 | 128 | 128 | 16 | 32 | 112 | 128 | 96 | 16 | 48 |
| M46 | 16 | 32 | 32 | 16 | 16 | 32 | 32 | 32 | 16 | 16 | 32 | 32 | 32 | 16 | 16 | 48 | 32 | 32 | 16 | 16 |
| M47 | 16 | 32 | 32 | 16 | 16 | 16 | 32 | 32 | 16 | 16 | 32 | 32 | 32 | 16 | 16 | 32 | 16 | 32 | 32 | 16 |
| M48 | 32 | 16 | 32 | 32 | 32 | 48 | 16 | 16 | 32 | 48 | 48 | 48 | 32 | 32 | 48 | 48 | 48 | 32 | 32 | 48 |
| M49 | 16 | 32 | 48 | 16 | 48 | 16 | 48 | 16 | 16 | 48 | 16 | 48 | 32 | 16 | 48 | 48 | 48 | 32 | 16 | 16 |
| M50 | 32 | 16 | 32 | 32 | 32 | 16 | 16 | 32 | 32 | 32 | 16 | 32 | 32 | 32 | 32 | 16 | 16 | 16 | 32 | 16 |
| M51 | 16 | 32 | 32 | 16 | 16 | 16 | 32 | 32 | 16 | 16 | 16 | 32 | 32 | 16 | 16 | 32 | 32 | 32 | 16 | 16 |
| M52 | 32 | 16 | 32 | 32 | 16 | 16 | 16 | 16 | 16 | 16 | 32 | 16 | 16 | 16 | 32 | 32 | 16 | 32 | 16 | 16 |
| M53 | 16 | 16 | 16 | 16 | 16 | 16 | 32 | 32 | 16 | 32 | 32 | 32 | 32 | 16 | 32 | 32 | 32 | 32 | 16 | 32 |
| M54 | 176 | 16 | 288 | 16 | 32 | 176 | 176 | 176 | 16 | 48 | 176 | 176 | 176 | 16 | 48 | 176 | 176 | 160 | 16 | 48 |
| M55 | 32 | 32 | 272 | 16 | 48 | 128 | 32 | 128 | 16 | 64 | 128 | 16 | 144 | 16 | 64 | 144 | 16 | 144 | 16 | 80 |
| M56 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 16 | 16 | 16 | 32 | 16 |
| M57 | 16 | 32 | 32 | 16 | 32 | 16 | 32 | 32 | 16 | 32 | 16 | 32 | 16 | 16 | 32 | 32 | 32 | 32 | 32 | 16 |
| M58 | 16 | 32 | 16 | 16 | 32 | 32 | 32 | 16 | 16 | 32 | 16 | 32 | 16 | 16 | 32 | 32 | 32 | 32 | 16 | 32 |
| M59 | 32 | 32 | 32 | 32 | 32 | 16 | 32 | 32 | 16 | 32 | 32 | 16 | 16 | 16 | 32 | 16 | 32 | 16 | 16 | 32 |
| M60 | 64 | 64 | 64 | 16 | 16 | 64 | 64 | 64 | 16 | 16 | 64 | 64 | 64 | 16 | 16 | 64 | 64 | 64 | 16 | 16 |
| M61 | 48 | 32 | 32 | 32 | 16 | 32 | 48 | 32 | 16 | 16 | 32 | 48 | 32 | 16 | 16 | 48 | 32 | 16 | 16 | 16 |
| M62 | 16 | 16 | 32 | 16 | 16 | 48 | 32 | 32 | 16 | 16 | 48 | 32 | 32 | 16 | 16 | 48 | 32 | 32 | 16 | 32 |
| M63 | 48 | 48 | 48 | 16 | 16 | 48 | 48 | 48 | 16 | 48 | 32 | 48 | 48 | 16 | 48 | 32 | 48 | 48 | 16 | 48 |
| M64 | 16 | 16 | 32 | 16 | 16 | 48 | 32 | 16 | 16 | 16 | 48 | 32 | 16 | 16 | 16 | 32 | 32 | 48 | 16 | 16 |
| M65 | 32 | 32 | 32 | 48 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 32 | 32 | 48 | 16 | 16 | 32 | 32 | 48 |
| M66 | 48 | 112 | 64 | 48 | 16 | 32 | 32 | 112 | 112 | 16 | 16 | 64 | 16 | 112 | 16 | 16 | 16 | 112 | 112 | 16 |
| M67 | 80 | 48 | 48 | 48 | 32 | 48 | 48 | 64 | 32 | 32 | 48 | 48 | 80 | 32 | 32 | 48 | 48 | 80 | 32 | 32 |

**Table 10.16**: Correlation coefficients between results of ACF analysis and optimal sampling window sizes for mammograms. The linear Pearson correlation coefficient was computed between the optimal sampling window sizes and $w_{ACF}$ ($\rho_w$) and $h_{ACF}$ ($\rho_h$). Results indicate that the ACF results are correlated with the optimal window sizes.

| Method | 5-bits | | 6-bits | | 7-bits | | 8-bits | |
|---|---|---|---|---|---|---|---|---|
| | $\rho_w$ | $\rho_h$ | $\rho_w$ | $\rho_h$ | $\rho_w$ | $\rho_h$ | $\rho_w$ | $\rho_h$ |
| TM-$D_E$ | 0.54 | 0.60 | 0.85 | 0.86 | 0.74 | 0.73 | 0.88 | 0.86 |
| TM-$D_{ES}$ | 0.42 | 0.52 | 0.54 | 0.62 | 0.51 | 0.58 | 0.52 | 0.59 |
| TM-$D_M$ | 0.89 | 0.87 | 0.53 | 0.60 | 0.55 | 0.61 | 0.55 | 0.61 |
| MI-histograms | -0.01 | 0.12 | -0.04 | -0.03 | -0.07 | -0.05 | -0.11 | -0.09 |
| MI-GLCMs | 0.29 | 0.31 | 0.31 | 0.33 | 0.38 | 0.46 | 0.47 | 0.54 |

**Table 10.17**: Results of ACF analysis applied to mammograms.

| Image | 5-bits | | 6-bits | | 7-bits | | 8-bits | | Image | 5-bits | | 6-bits | | 7-bits | | 8-bits | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $w_{ACF}$ | $h_{ACF}$ | $w_{ACF}$ | $h_{ACF}$ | $w_{ACF}$ | $h_{ACF}$ | $w_{ACF}$ | $h_{ACF}$ | | $w_{ACF}$ | $h_{ACF}$ | $w_{ACF}$ | $h_{ACF}$ | $w_{ACF}$ | $h_{ACF}$ | $w_{ACF}$ | $h_{ACF}$ |
| M0 | 56 | 58 | 56 | 58 | 56 | 58 | 56 | 58 | M34 | 49 | 96 | 49 | 96 | 49 | 96 | 49 | 96 |
| M1 | 79 | 98 | 79 | 99 | 79 | 99 | 79 | 99 | M35 | 36 | 48 | 36 | 48 | 36 | 48 | 36 | 48 |
| M2 | 55 | 57 | 55 | 57 | 55 | 57 | 55 | 57 | M36 | 38 | 40 | 38 | 40 | 38 | 39 | 38 | 40 |
| M3 | 34 | 52 | 34 | 52 | 34 | 52 | 34 | 52 | M37 | 30 | 44 | 30 | 44 | 30 | 44 | 30 | 44 |
| M4 | 41 | 18 | 41 | 18 | 41 | 18 | 41 | 18 | M38 | 74 | 90 | 74 | 90 | 74 | 90 | 74 | 90 |
| M5 | 40 | 56 | 40 | 56 | 40 | 56 | 40 | 56 | M39 | 59 | 63 | 59 | 63 | 59 | 63 | 59 | 63 |
| M6 | 25 | 23 | 25 | 23 | 25 | 23 | 27 | 23 | M40 | 25 | 26 | 25 | 28 | 25 | 28 | 25 | 28 |
| M7 | 21 | 21 | 21 | 21 | 21 | 21 | 21 | 21 | M41 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 |
| M8 | 13 | 11 | 13 | 11 | 13 | 11 | 13 | 11 | M42 | 40 | 51 | 40 | 51 | 40 | 51 | 40 | 51 |
| M9 | 22 | 18 | 22 | 16 | 21 | 18 | 22 | 18 | M43 | 56 | 62 | 56 | 62 | 56 | 62 | 56 | 62 |
| M10 | 25 | 33 | 25 | 33 | 25 | 33 | 25 | 33 | M44 | 109 | 130 | 109 | 130 | 109 | 130 | 109 | 130 |
| M11 | 34 | 47 | 34 | 47 | 34 | 47 | 34 | 47 | M45 | 110 | 118 | 110 | 118 | 110 | 118 | 110 | 118 |
| M12 | 13 | 13 | 13 | 12 | 13 | 12 | 13 | 12 | M46 | 20 | 42 | 20 | 42 | 20 | 42 | 20 | 42 |
| M13 | 13 | 12 | 12 | 12 | 12 | 11 | 12 | 11 | M47 | 22 | 15 | 21 | 15 | 21 | 15 | 21 | 15 |
| M14 | 52 | 44 | 52 | 45 | 52 | 44 | 52 | 44 | M48 | 28 | 16 | 28 | 15 | 28 | 15 | 28 | 15 |
| M15 | 39 | 37 | 39 | 37 | 39 | 37 | 39 | 37 | M49 | 18 | 20 | 18 | 20 | 22 | 20 | 18 | 20 |
| M16 | 48 | 72 | 49 | 74 | 49 | 74 | 49 | 74 | M50 | 14 | 13 | 14 | 13 | 14 | 13 | 14 | 13 |
| M17 | 56 | 41 | 56 | 41 | 56 | 41 | 56 | 41 | M51 | 40 | 23 | 40 | 23 | 40 | 23 | 41 | 23 |
| M18 | 13 | 11 | 13 | 11 | 13 | 11 | 13 | 11 | M52 | 13 | 21 | 13 | 21 | 12 | 21 | 12 | 21 |
| M19 | 12 | 11 | 12 | 11 | 12 | 11 | 12 | 11 | M53 | 13 | 13 | 13 | 14 | 13 | 14 | 13 | 13 |
| M20 | 80 | 46 | 80 | 46 | 80 | 46 | 80 | 46 | M54 | 150 | 140 | 150 | 140 | 150 | 140 | 150 | 140 |
| M21 | 44 | 45 | 44 | 45 | 44 | 45 | 44 | 45 | M55 | 171 | 139 | 171 | 139 | 171 | 139 | 171 | 139 |
| M22 | 40 | 47 | 40 | 47 | 40 | 47 | 40 | 47 | M56 | 12 | 11 | 12 | 11 | 12 | 11 | 12 | 11 |
| M23 | 54 | 89 | 54 | 89 | 54 | 89 | 54 | 89 | M57 | 14 | 21 | 14 | 21 | 14 | 21 | 14 | 21 |
| M24 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | M58 | 27 | 17 | 25 | 13 | 24 | 17 | 24 | 17 |
| M25 | 16 | 17 | 16 | 17 | 16 | 17 | 16 | 17 | M59 | 11 | 8 | 11 | 6 | 11 | 6 | 11 | 6 |
| M26 | 206 | 193 | 206 | 193 | 206 | 193 | 206 | 193 | M60 | 30 | 53 | 30 | 53 | 30 | 53 | 30 | 53 |
| M27 | 143 | 142 | 143 | 142 | 143 | 142 | 143 | 142 | M61 | 26 | 28 | 26 | 28 | 26 | 28 | 26 | 28 |
| M28 | 83 | 65 | 83 | 65 | 83 | 65 | 83 | 65 | M62 | 21 | 22 | 21 | 22 | 21 | 22 | 21 | 22 |
| M29 | 101 | 81 | 101 | 81 | 101 | 81 | 101 | 81 | M63 | 17 | 25 | 17 | 25 | 17 | 25 | 17 | 25 |
| M30 | 81 | 80 | 81 | 80 | 81 | 80 | 81 | 80 | M64 | 22 | 18 | 16 | 19 | 16 | 19 | 16 | 19 |
| M31 | 80 | 100 | 80 | 100 | 80 | 101 | 79 | 101 | M65 | 20 | 44 | 20 | 44 | 20 | 44 | 20 | 44 |
| M32 | 33 | 52 | 34 | 52 | 34 | 52 | 34 | 52 | M66 | 48 | 57 | 48 | 57 | 48 | 57 | 48 | 57 |
| M33 | 29 | 36 | 29 | 36 | 29 | 36 | 29 | 36 | M67 | 37 | 41 | 37 | 38 | 37 | 38 | 37 | 38 |

## 10.9 Multiple Reference Regions

Since the reference ROIs of most of the non-malignant ROIs are most likely not texturally homogeneous, the position of the reference sampling window within the reference ROI would affect matching accuracy. In order to improve matching accuracy for this situation, each single reference sampling window was divided into four smaller reference sampling window, as depicted in **Figure 10.29**. The resulting matching map from each small reference sampling window was then averaged, in an attempt to highlight the different textures within the non-homogeneous ROIs.

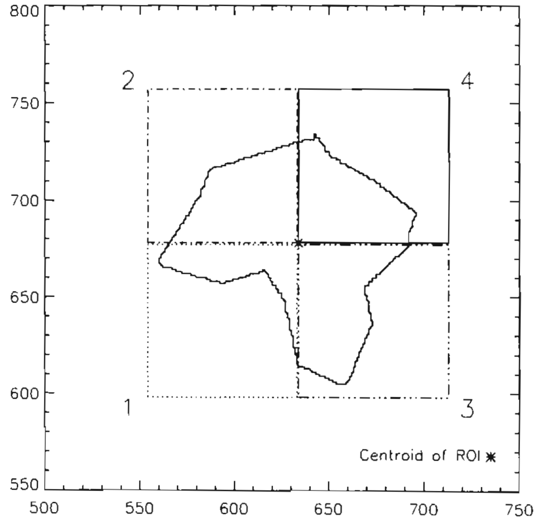


**Figure 10.29**: Placement of sampling windows for multiple (neighbouring) reference regions. Matching maps are generated for each of the four marked reference regions (indicated by the different line styles). The four matching maps are then averaged to highlight any differences in matching that might exist as a result of textural differences between the regions.

The best matching results of applying the matching algorithms with multiple reference regions to pairs of mammograms are summarised in **Figure 10.30**, for all methods. The results, for all methods except MI-histograms, are generally more scattered than for the results with a single reference region. For MI-histograms, the results are very clustered, with only 2 non-matches. TM-$D_E$ has 10 non-matches, TM-$D_{ES}$ has 16 non-matches, TM-$D_M$ has 9 non-matches and MI-GLCMs has 14 non-matches. Once again the only mammogram that is not matched for all methods is M58M59.

**Table 10.18**: Average of the best matching accuracies for single & multiple reference regions for all matching methods applied to pairs of mammograms.

| Method | Single Reference Region | | | Multiple Reference Regions | | |
|---|---|---|---|---|---|---|
| | $A_{ROC}$ | $C_{fb}$ | $\kappa$ | $A_{ROC}$ | $C_{fb}$ | $\kappa$ |
| TM-$D_E$ | 0.80±0.17 | 0.46±0.26 | 0.33±0.25 | 0.75±0.18 | 0.23±0.28 | 0.18±0.23 |
| TM-$D_{ES}$ | 0.81±0.2 | 0.28±0.25 | 0.24±0.21 | 0.73±0.22 | 0.19±0.23 | 0.15±0.19 |
| TM-$D_M$ | 0.79±0.17 | 0.25±0.20 | 0.19±0.17 | 0.73±0.19 | 0.18±0.16 | 0.12±0.14 |
| MI-histograms | 0.85±0.11 | 0.16±0.15 | 0.12±0.13 | 0.87±0.1 | 0.17±0.14 | 0.14±0.13 |
| MI-GLCMs | 0.77±0.25 | 0.50±0.42 | 0.41±0.39 | 0.76±0.26 | 0.44±0.44 | 0.36±0.4 |

Results indicate averaging results from multiple, neighbouring reference regions does not improve matching accuracy. While average results for single and multiple reference regions agree within errors, there are many more non-matches for the multiple reference region algorithms.

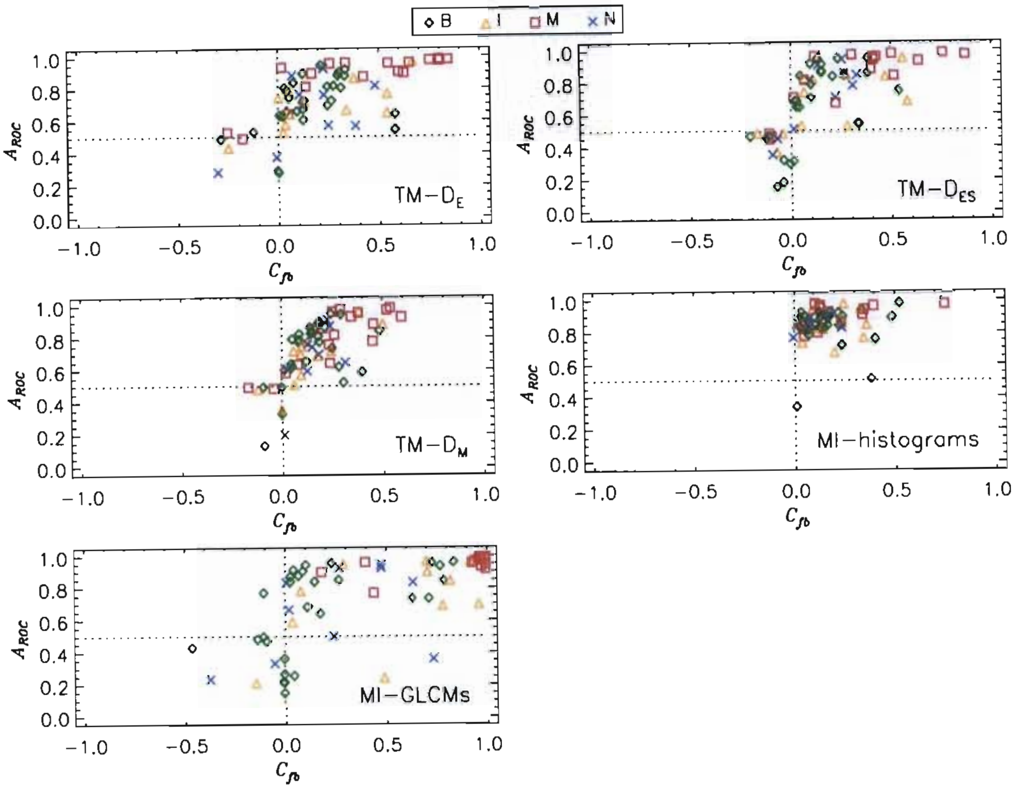


**Figure 10.30**: Scatter plot of the best results for multiple (neighbouring) reference regions applied to pairs of mammograms.

## 10.10 Summary of Effects of Matching Parameters

### 10.10.1 Effect of $w_{step}$ on Matching Accuracy

The sampling window step size was fixed at 4 pixels for the matching of mammograms, following the results of the investigation with the mosaic images. There was therefore no investigation of the effect of sampling window step size on matching accuracy.

### 10.10.2 Effect of $w$ on Matching Accuracy

Matching accuracy varied with sampling window, but results suggested that the optimal sampling window size was dependent on the scale sizes of the textures in each image.

The results of using the autocorrelation function to select the optimal window size were well correlated with the optimal window sizes obtained from the matching analysis. The optimal window sizes from TM-$D_E$ correlated best with the ACF results. The ACF results also showed that the scale widths and scale heights of most of the textures in the mammograms are similar and so justified the use of square sampling windows. However, rectangular windows would probably yield better matching results for those textures with a different scale width and height.

### 10.10.3 Effect of *nbits* on Matching Accuracy

For TM-matching, the effect of bit-depth on matching accuracy was not clear, and appeared to vary with sampling window size. For all distance similarity metrics, there were some window sizes that showed an improvement in matching accuracy as bit-depth was increased and other window sizes that showed an improvement as bit-depth was decreased. This effect might be a consequence of the result that changing the bit-depth has on the sizes of features in the image, i.e. that some features change size as the bit-depth is changed.

For MI-histograms, matching accuracy was optimal for the lower bit-depths, but this effect was only clearly seen for the larger sampling windows. Matching accuracy was independent of bit-depth for the smaller sampling windows.

For MI-GLCMs, matching accuracy was optimal for the higher bit-depths.

The results for MI-GLCMs is similar to the results obtained by Chan et al. [1995] who varied bit-depth between 4 bits and 9 bits and found the optimal bit-depth to be between 7 bits and 8 bits.

### 10.10.4 Effect of *d* on Matching Accuracy

For TM-matching, matching accuracy was generally independent of *d* for all similarity metrics.

For MI-GLCMs, matching accuracy improved as *d* was decreased.

Chan et al. [1995] found the optimal value of $d$=20 pixels at a spatial resolution of 0.1 mm per pixel. This corresponds to $d$=2 mm.

For this study, *d* was varied between 1 pixel (0.254 mm) and 10 pixels (2.54 mm), and this range could be too small to detect any dependence of matching accuracy on *d* for the TM-matching algorithm.

### 10.10.5 Effect of *nbins* on Matching Accuracy

Results showed that matching accuracy improved as fewer bins were used to calculate the histogram.

These results are consistent with those found by Tourassi et al. [2003] who used mutual information as a similarity metric for template matching in a knowledge-based mammographic CAD-system for discrimination of ROIs from normal tissue and found that results were most accurate for the fewest number of histogram bins.

### 10.10.6 Sensitivity of Matching Methods to Choice of Parameter Values

**Figure 10.31** shows all pairs of $A_{ROC}$- and $C_{fb}$-values, for all parameters and for each matching method to investigate how sensitive each method is to the choice of parameter values. If all the points are clustered in one particular region, then the method is not sensitive to the choice of values of the matching parameters, but if the points are very scattered then the choice of values of the matching parameters is critical to ensure an optimal match.

Results for TM-$D_E$ and MI-GLCM are widely scattered, indicating that these methods are sensitive to the choice of values for the matching parameters. Results for TM-$D_{ES}$ and $D_M$ and MI-histograms are more clustered, but the results are clustered about the [0.5,0] point, indicating poor matches.

The sensitivity to choice of parameter values was quantified by counting the number of pairs of $A_{ROC}$- and $C_{fb}$-values that fell into different zones, as detailed in **Figure 9.22** (page 140). These results are summarised in **Table 10.19**, with the last column showing the percentage of points that fell into zones 1, 2 and 3. Results indicate that MI-GLCMs is most accurate with the highest percentage of points falling into zone 1.

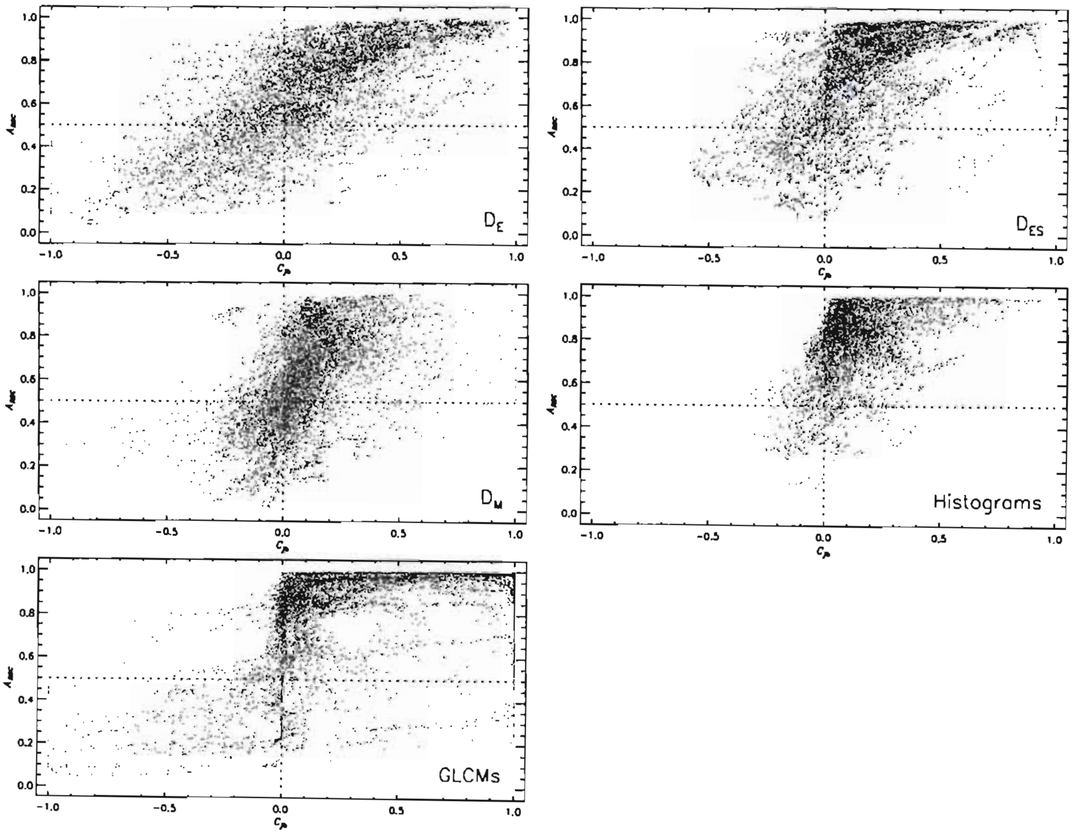


**Figure 10.31**: Scatter plot to demonstrate sensitivity of choice of matching parameter values for each method applied to pairs of mammograms.

**Table 10.19**: Results of zone analysis for pairs of mammograms to demonstrate sensitivity of choice of matching parameters for each method. An ideal method would have a high percentage of points in Zone1 and none in Zone 5.

| Method | Zone 1 (%) | Zone 2 (%) | Zone 3 (%) | Zone 4 (%) | Zone 5 (%) | Total (Zones 1,2 & 3) |
|---|---|---|---|---|---|---|
| TM-$D_E$ | 16.0 | 33.5 | 1.23 | 40.8 | 8.42 | 34.7 |
| TM-$D_{ES}$ | 8.72 | 27.7 | 1.16 | 55.4 | 7.09 | 37.0 |
| TM-$D_M$ | 0.430 | 9.04 | 1.51 | 75.3 | 13.8 | 11.0 |
| MI-histograms | 3.41 | 14.1 | 0.00 | 71.9 | 10.6 | 17.5 |
| MI-GLCMs | 44.2 | 22.5 | 5.32 | 24.9 | 3.10 | 72.0 |

## 10.11 Comparison of TM- and MI-Matching for Pairs of Mammograms

Overall, both TM-matching using distance similarity metrics and MI-matching show potential as matching schemes. TM-$D_E$ and MI-GLCMs yields the most accurate matches, followed by TM-$D_{ES}$. **Table 10.20** shows the results for each similarity metric based on ROI-type.

**Table 10.20**: Average of best matching accuracies based on diagnosis for all matching methods applied to pairs of mammograms.

| | TM-$D_E$ | TM-$D_{ES}$ | TM-$D_M$ | MI-histograms | MI-GLCMs |
|---|---|---|---|---|---|
| Benign | 0.22±0.16 | 0.14±0.13 | 0.10±0.11 | 0.12±0.13 | 0.21±0.30 |
| Indeterminate | 0.34±0.27 | 0.22±0.20 | 0.18±0.11 | 0.13±0.08 | 0.44±0.34 |
| Malignant | 0.58±0.21 | 0.45±0.20 | 0.36±0.19 | 0.15±0.18 | 0.84±0.23 |
| Normal | 0.18±0.19 | 0.14±0.13 | 0.12±0.09 | 0.08±0.06 | 0.17±0.23 |
| Overall | 0.33±0.25 | 0.24±0.21 | 0.19± 0.17 | 0.12 ± 0.13 | 0.41 ± 0.39 |

Results of significance testing for the two best methods, TM-$D_E$ and MI-GLCMs, appear in **Table 10.21** and show that the results for matching the benign and indeterminate masses and normal ROIs are similar for each method, but that the results of matching malignant masses are significantly better for MI-GLCMs.

**Table 10.21**: Results of significance ($t$-test) analysis between best matching results of TM-$D_E$ and MI-GLCMs as a function of diagnosis.

| Diagnosis | $t$-value | $p$-value |
|---|---|---|
| Benign | 0.29 | 0.78 |
| Indeterminate | -0.75 | 0.46 |
| Malignant | -3.55 | 0.00 |
| Normal | 0.04 | 0.97 |

MI-GLCMs shows clear dependence of matching accuracy on the various matching parameters so it is easier to select a set of optimal matching parameters for this algorithm. TM-$D_E$ and TM-$D_{ES}$ do not show clear dependence of matching accuracy on the various matching parameters. The optimal sampling window sizes for these three methods are, however, correlated with the ACF results, so the ACF analysis can be used to determine the size of the optimal sampling window.

TM-$D_E$ is more sensitive than MI-GLCMs to the choice of matching parameter values, but MI-GLCMs is faster to compute.

Overall, where MI-GLCMs has a weakness, TM-$D_E$ and TM-$D_{ES}$ have strengths and vice-versa. In summary, there are no significant differences between TM-$D_E$, TM-$D_{ES}$ and MI-GLCMs, and a possible hybrid matching scheme using the results of all three methods might yield good matching results.

One advantage of using mutual information or the distance similarity metrics for matching is that there is no training required which is important to the analysis of mammograms, which vary considerably from patient to patient.

## 10.12 Comparison of Matching Results to Similar Studies

Overall results for the matching of malignant masses compares favourably with those of Tourassi et al. [2003] who used mutual information as a similarity metric for template matching in a knowledge-based mammographic CAD-system for discrimination of masses from normal tissue. Tourassi et al. achieved $A_z$=0.88±0.01 for discrimination between malignant and normal tissue. The results achieved by Tourassi et al. for discrimination between benign and normal tissue of $A_z$=0.86±0.01 is considerably higher than that achieved in this study.

Filev et al. [2005] compared the effectiveness of twelve similarity metrics in matching the correspondence between masses in temporal mammograms. Filev et al. used an average template size of 17 mm×17 mm which is considerably larger than the size of the templates used in this study. Pearson's correlation coefficient, the cosine coefficient and Goodman and Kruskal's gamma coefficient performed best and were the most robust. While mutual information was robust, it did not perform so well (6th best), and the scaled mutual information consistently performed better than the unscaled mutual information. Also, mutual information did not perform well for small template sizes. This result is contrary to what was found in this study, where matching accuracy improved as sampling window size was decreased.

Chan et al. [1995] ranked the importance of GLCM-based texture measures in differentiating between masses and normal breast tissue and achieved $A_z$=0.82 with masses that had a mean diameter of 12.2 mm. Results are comparable to those achieved in this study, although the masses in this study were generally small.

## 10.13 Summary

The TM-matching and MI-matching algorithms were applied to 68 pairs of CC and MLO mammograms to confirm whether a reference region of interest identified by a radiologist in one standard mammographic view can be matched to the corresponding region in another standard mammographic view, to determine whether the matching algorithms can be utilised in a CAD-system.

TM-matching (with $D_E$ and $D_{ES}$) and MI-matching (with GLCMs) have shown great potential for use in a CAD scheme. MI-GLCMs had an overall best matching accuracy of $\kappa$=0.41 ± 0.39 corresponding to average best values of $A_{ROC}$=0.77±0.25 and $C_{fb}$=0.50±0.42. TM-$D_E$ had an overall best matching accuracy of $\kappa$=0.33±0.25 corresponding to average best values of $A_{ROC}$=0.80±0.17 and $C_{fb}$=0.46±0.26. TM-$D_{ES}$ was the third most accurate method with $\kappa$=0.24±0.21, corresponding to average values of $A_{ROC}$=0.81±0.20 and $C_{fb}$=0.28±0.25. The matching accuracy for TM-$D_M$ and MI-histograms was very low.

MI-GLCMs had the best matching accuracy for matching malignant masses ($\kappa$=0.84±0.23 corresponding to $A_{ROC}$=0.96±0.05 and $C_{fb}$=0.90±0.21), while the results for benign and indeterminate masses and normal ROIs were similar for MI-GLCMs and TM-$D_E$.

The results of the ACF analysis correlated with the optimal sampling window sizes obtained from the matching analysis. TM-$D_E$ at 8 bits and TM-$D_M$ at 5 bits had the highest correlation. The ACF-results also justified the use of square windows for the analysis of the mammogram ROIs, since most ACF-widths and ACF-heights were similar. The results for MI-GLCMs showed some correlation, but this was significantly lower than for TM-$D_E$.

While MI-GLCMs were more sensitive to matching parameters than TM-$D_E$, the overall, matching accuracies are better for MI-GLCMs than TM-$D_E$. Therefore TM-$D_E$ and MI-GLCMs are closely matched as matching schemes.

If these algorithms were to be applied in a CAD-system, it is recommended that the autocorrelation function of the reference ROI be used to determine the sampling window size. Then both, TM-$D_E$ and MI-GLCMs should be used as matching schemes. The recommended values of the matching parameters are summarised in **Table 10.22**.

**Table 10.22**: Recommended values for matching parameters for practical application of TM-$D_E$ and MI-GLCMs in a CAD-system

| Matching Parameter | TM-$D_E$ | MI-GLCMs |
|---|---|---|
| $w_{step}$ | 4 pixels | 4 pixels |
| $w$ | from ACF analysis at 8 bits | from ACF analysis at 8 bits |
| $nbits$ | 8 bits | 8 bits |
| $d$ | 10 pixels | 1 pixel |

# Chapter 11

# Matching Results: Stereotactic Biopsy Mammograms

The TM- and MI-matching algorithms have been applied to two completely different sets of images (mosaics and mammograms) and the potential to find a match to a known reference texture in a test image has been demonstrated for both image sets. In this chapter, the matching algorithms are applied to pairs of stereotactic breast biopsy mammograms.

## 11.1 Breast Biopsies

After an abnormality has been identified, through mammography, ultrasonography or magnetic resonance imaging, a sample of breast tissue is extracted to test the histology of the region and confirm whether the abnormality is malignant or benign. The procedure of extracting the tissue sample is known as a *biopsy*.

There are two types of biopsies: open and percutaneous. Open biopsies use a hooked wire, inserted into the breast under mammographic or ultrasonographic guidance, to localise the abnormality. The sample is surgically removed with the patient under a general anaesthetic. Open biopsies are more expensive because they involve a hospital stay, are generally more traumatising for the patient and leave a scar, but are reliable. Percutaneous biopsies use stereotaxis or ultrasonography to localise the abnormality in three dimensions. An incision is made through the skin and a needle is inserted to the calculated depth to extract a sample. Percutaneous biopsies are cheap, minimally invasive and can be done with the patient under a local anaesthetic, on an outpatient basis. However, percutaneous biopsies have a reputation of being less accurate than open biopsies because a very small amount of tissue (1 mm to 3 mm in diameter) is sampled [Heywang-Köbrunner et al. 1998, Verkooijen et al. 2000].

## 11.2 Stereotaxis

Stereotactic breast biopsy systems are designed to locate an abnormality in three dimensions, with the breast compressed between two parallel plates. The $x-y$ plane is located parallel to the image plane with the $x$-axis parallel to the chest wall, the $y$-axis perpendicular to the chest wall and the $z$-axis perpendicular to image plane. The location $[x_m, y_m, z_m]$ of the mass, is obtained in two stages. $x_m$ and $y_m$ are obtained from an image taken perpendicular to the direction of compression. This scout view (usually at 0°) is used to centre the lesion in the image. Two further (stereoscopic) views are taken with the x-ray source rotated to both sides of the normal by a small amount (usually 10° to 15°). $z_m$ (the depth of the lesion) is obtained by triangulation between these two views. Note that $y_m$ is the same in all views since the only movement that is assumed to have occurred is the rotation of the x-ray source resulting in a virtual shift in the $x$-direction [Hendrick & Parker 1993, Carr et al. 2001, Bushberg et al. 2002, Helbich et al. 2004].

A schematic of the geometry of the procedure is shown in **Figure 11.1** and the relationship to determine the depth, $z_m$ is given in **Equation 11-1**, with $x_{shift}$ the distance between the centroids of the mass in the stereo views.

$$z_m = \frac{x_{shift}}{2\tan\theta} \tag{11-1}$$

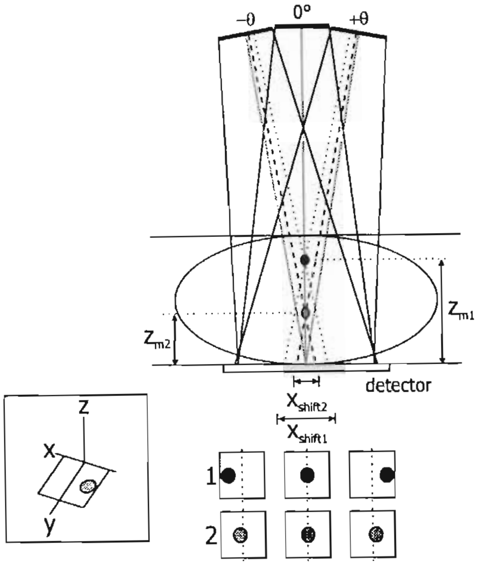


**Figure 11.1**: Schematic (not to scale) detailing the geometry of a SB system. Lesions closer to the x-ray source (1) have a greater shift from the centre of the mammogram than lesions further away from the x-ray source (2). $x_{shift}$ is the distance between the centroids of the mass in the stereoscopic views.

## 11.3 Effect of Localisation Errors on Sampling Accuracy

In stereotactic biopsy systems, any inaccuracy in localising the lesion in either stereoscopic view results in an error in the calculated depth of the lesion. Percutaneous biopsies sample a very small amount of tissue (1 mm to 3 mm in diameter) and the desired feature might be missed if the feature is not well localised.

There are many problems that result in localisation errors which lead to sampling the wrong tissue, including operator errors, patient movement, lesion movement within the breast during biopsy, the use of two different masses or calcification groups on the two stereoscopic views in the belief that they represent a single abnormality and non-visualisation of the lesion due to overlying tissue or as a result of the geometric configuration of the imaging system [Hendrick & Parker 1993, Carr et al. 2001].

All these problems cause the incorrect tissue to be sampled and can therefore lead to an incorrect diagnosis.

Carr et al. [2001] developed a computer simulation of a stereotactic breast biopsy system, based on a geometric model, to define and improve the targeting of breast lesions, and demonstrated the effect that incorrect localisation on the stereoscopic views has on the calculated depth. It was reported, that there can be an error of more than 10 mm in the calculated depth if there is a 5 mm error in the selection of the point in one stereoscopic view, for the stereotactic breast biopsy system that was used in the study.

More generally, if the true shift is $x_{shift}$ and $\Delta x$ is a deviation from the true shift value then if $\Delta\theta = 0$ and the true depth is $z_m$, the deviation from the true depth, $\Delta z$ is:

$$\begin{aligned} z_m &= \frac{x_{shift}}{2\tan\theta} \\ \Delta z &= \frac{\Delta x}{2\tan\theta} \\ \frac{\Delta z}{z_m} &= \frac{\Delta x}{2\tan\theta}\frac{2\tan\theta}{x_{shift}} \\ &= \frac{\Delta x}{x_{shift}} \end{aligned}$$

which shows that the relative deviation in $x_{shift}$ is equal to the relative deviation in $z_m$. So, if $x_{shift}$ deviates from the true shift by 10%, then $z_m$ deviates from the true depth by 10%. It is therefore vital to accurately select points that refer to the same feature in both stereoscopic views.

## 11.4 Application of Matching Algorithms to SB Mammograms

Since the identification of the point referring to the same feature in each stereoscopic view is a matching problem, the TM-matching and MI-matching algorithms were applied to SB mammograms to investigate the potential of using these algorithms to improve localisation accuracy.

The proposed solution is based on the selection of a single point in the 0° view and using the matching algorithm to locate the corresponding point in each of the other two views.

This matching of regions between SB mammograms is simpler than the problem of matching regions between two standard mammographic views as the breast compression is identical in all three SB views and the problem is reduced to one of analysing stereoscopic images.

Literature searches have yielded no results with respect to the application of CAD algorithms to SB mammograms. This study therefore represents the first steps in improving localisation accuracy by comparing the similarity of the stereoscopic views.

## 11.5 Details of Stereotactic Biopsy Mammograms

### 11.5.1 Selection of SB Mammograms and Ground Truth Data

The matching algorithms were applied to SB mammograms from 12 patients at the Inkosi Albert Luthuli Central Hospital (Durban, South Africa). The images were acquired on a Siemens

Mammomat Nova 3000 mammography machine with an add-on biopsy unit and add-on Siemens Opdima[1] unit. The latter allowed for the capture of small field-of-view digital images. The workstation consisted of a Sun Ultra 10 creator host computer. The detector cassette was a charge-coupled-device with an imaging field of 49 mm × 85 mm. The image matrix was 1024 pixels × 1792 pixels (0.048 mm per pixel) at a bit-depth of 12 bits. Images were resampled to a spatial resolution of 0.254 mm per pixel for the analysis.

There were three images per patient (0°, ±10°) resulting in a total of 36 SB mammograms. The 0° view was always used as the reference image, so the matching algorithms were applied to 24 image pairs. There were two types of features of interest in the SB mammograms, masses (m) and calcification clusters (c), each with a different format of ground truth data. The radiologist marked the boundaries of the masses in the same manner as for the standard mammograms. For the calcification clusters, the radiologist was asked to select a point in the 0° view corresponding to a calcification of interest and also to identify the same calcification in the ±10° views, as is required for a biopsy. ROIs are divided into three categories according to the diagnosis: benign (B), indeterminate (I) and malignant (M). The diagnosis was confirmed by the biopsy results, where available. If the results were not available, then the radiologist's report was used.

The reference SB mammogram label and test SB mammogram label are used in combination to label the SB mammogram pair. For example, the SB mammogram pair consisting of reference image S1 and test image S0, has the SB mammogram pair label S1S0. The labels for each SB mammogram are given in **Table 11.1**.

### 11.5.2 Characteristics of SB Mammograms

Histograms of the areas of the suspicious ROIs along with their visibilities (compared to surrounding tissue) are shown in **Figure 11.2**. Visibility was automatically determined from the original mammograms (at 0.254 mm per pixel). Visibility was defined to be the contrast of the ROI compared to the surrounding tissue and was computed from **Equation 7-10** and **Figure 7.4** (page 87). Visibility ranges between 0 for a very subtle ROI and 1 for a very obvious ROI. The histograms demonstrate that most of the ROIs are very small with very low visibilities. Full details of the sizes and visibilities of the ROIs are shown in **Table 11.1**.

### 11.5.3 Reduced Images

Reduced images of the SB mammograms used in this study are shown in **Figure 11.3** in greyscale and pseudo-colour. Magnified versions of the ground truth ROIs for the masses are shown in **Figure 11.4**.

---

[1] Search `http://www.medical.siemens.com/` for Opdima.

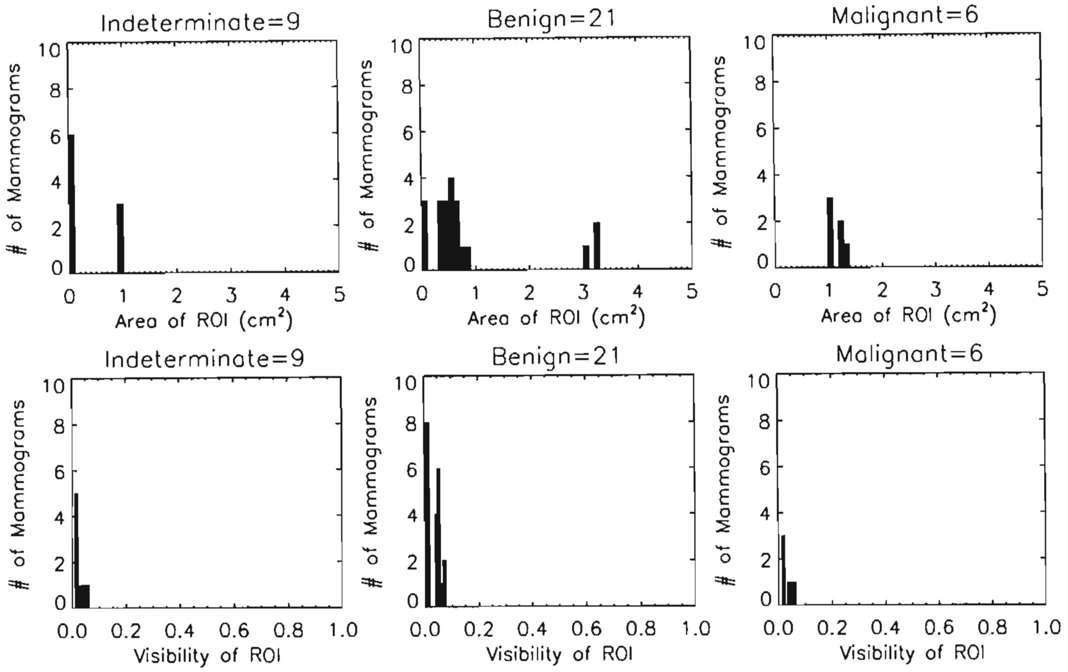


**Figure 11.2**: Histograms of areas and visibilities of the suspicious ROIs for the SB mammograms used in this study. Most of the ROIs are very small with very low visibilities.

**Table 11.1**: Characteristics of the suspicious ROIs in the SB mammograms used in this study. 'Feature' refers to whether the mammogram contains a mass (m) or a calcification (c). The possible diagnoses of the features are: benign (B), indeterminate (I) and malignant (M).

| Label | Angle (°) | Area ($cm^2$) | Visibility | Feature | Diagnosis | Label | Angle (°) | Area ($cm^2$) | Visibility | Feature | Diagnosis |
|---|---|---|---|---|---|---|---|---|---|---|---|
| S0[a] | -10 | 1.06 | 0.014 | m | M | S18 | -10 | 0 | 0.040 | c | B |
| S1[a] | 0 | 0.98 | 0.014 | m | M | S19 | 0 | 0 | 0.013 | c | B |
| S2 | 10 | 1.03 | 0.015 | m | M | S20 | 10 | 0 | 0.055 | c | B |
| S3[a] | -10 | 0.34 | 0.044 | m | B | S21[a] | -10 | 3.21 | 0.007 | m | B |
| S4[a] | 0 | 0.38 | 0.063 | m | B | S22[a] | 0 | 3.1 | 0.009 | m | B |
| S5 | 10 | 0.31 | 0.052 | m | B | S23 | 10 | 3.23 | 0.007 | m | B |
| S6[a] | -10 | 0.81 | 0.067 | m | B | S24 | -10 | 0.49 | 0.048 | m | B |
| S7[a] | 0 | 0.72 | 0.055 | m | B | S25 | 0 | 0.44 | 0.040 | m | B |
| S8 | 10 | 0.65 | 0.067 | m | B | S26 | 10 | 0.53 | 0.056 | m | B |
| S9[a] | -10 | 1.35 | 0.063 | m | M | S27[a] | -10 | 0.65 | 0.040 | m | B |
| S10[a] | 0 | 1.24 | 0.040 | m | M | S28[a] | 0 | 0.64 | 0.013 | m | B |
| S11 | 10 | 1.25 | 0.044 | m | M | S29 | 10 | 0.56 | 0.055 | m | B |
| S12 | -10 | 0 | 0.014 | c | I | S30 | -10 | 0.47 | 0.007 | m | B |
| S13 | 0 | 0 | 0.016 | c | I | S31 | 0 | 0.53 | 0.009 | m | B |
| S14 | 10 | 0 | 0.025 | c | I | S32 | 10 | 0.58 | 0.007 | m | B |
| S15 | -10 | 0 | 0.014 | c | I | S33 | -10 | 0.95 | 0.048 | m | I |
| S16 | 0 | 0 | 0.011 | c | I | S34 | 0 | 0.96 | 0.040 | m | I |
| S17 | 10 | 0 | 0.016 | c | I | S35 | 10 | 0.99 | 0.056 | m | I |

[a] Mammogram analysed for point correspondence. The exact pairs used are detailed in **Table 11.15**.

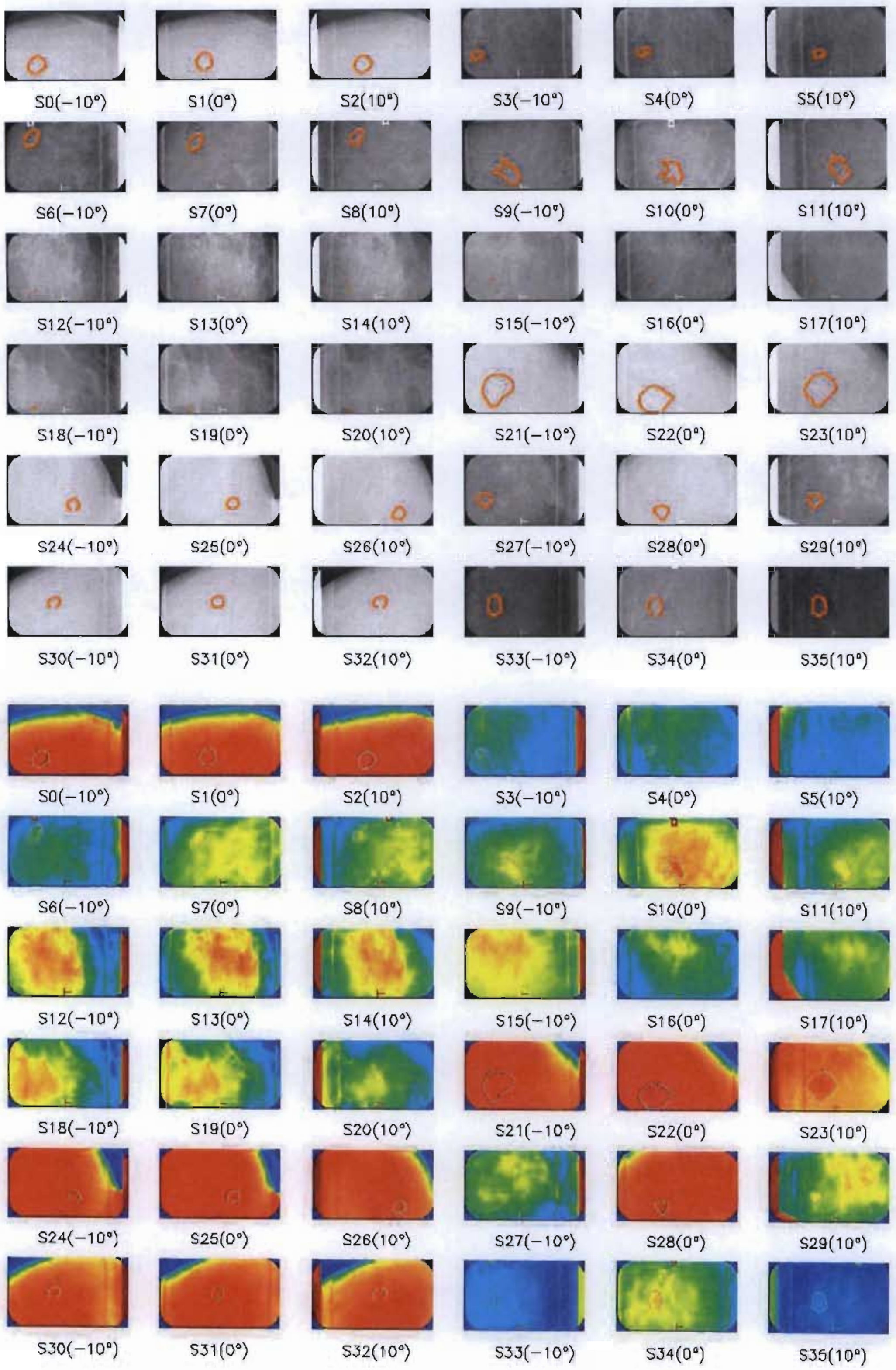


**Figure 11.3**: Reduced images of SB mammograms used in this study in the original grey-scale (top) and pseudo-colour (bottom). Images are displayed as triplets corresponding to the -10°, 0° and +10° images, from left to right, but are labelled sequentially. The mass borders are indicated in red for the grey-scale images and grey for the pseudo-colour images.

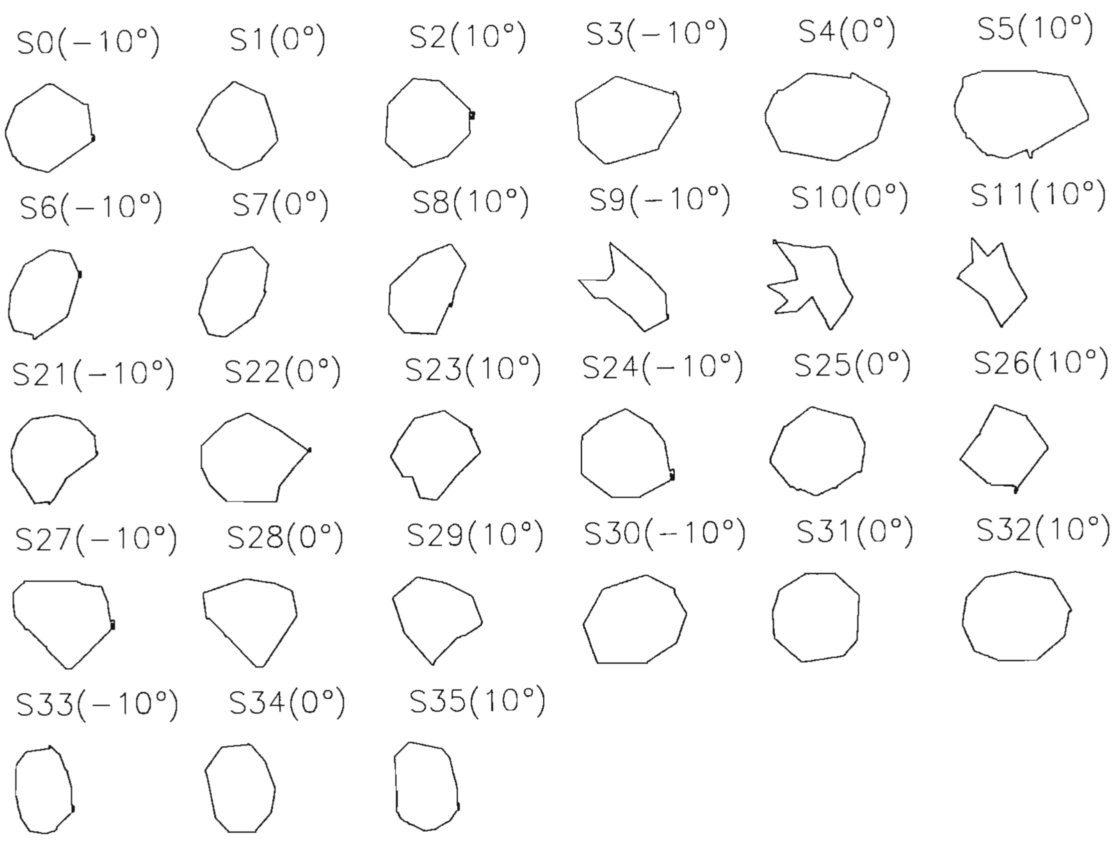


**Figure 11.4**: Magnified ROIs (not too scale) extracted from SB mammograms with masses, to demonstrate that some ROIs are spiculated, while others are almost circular. Images are displayed as triplets corresponding to the -10°, 0° and +10° images, from left to right, for each triplet.

### 11.5.4 Sampling of Reference Images

The SB mammograms with masses were sampled identically to the standard mammograms, with sampling windows centred on the centroid of the mass. Sampling windows were then increased in size (in increments of 16 pixels) around the centroid until the ROI was completely enclosed. This was chosen as the maximum sampling window size.

A single sampling window of 16 pixels was centred on the position of the calcification marked by the radiologist to sample the reference ROI containing a calcification.

## 11.6 Evaluation of Matching Accuracy for SB Mammograms

To fully test the results of the matching algorithms, experiments would have to be performed on a biopsy system to verify physically that the locations provided by the matching algorithm actually improve localisation accuracy, but this is beyond the scope of this project.

For this study, the results were evaluated in a two-stage process. Firstly, the matching accuracy was evaluated according to the method used for the mosaic images and the standard mammograms. Secondly, those results that had a good matching accuracy, were evaluated for localisation accuracy, by comparing the centroid of the region matched to the point chosen by the radiologist.

## 11.7 Selection of Matching Parameters

The aim of the investigation using the mosaic images was to evaluate the matching algorithms under ideal matching conditions and to gain an understanding of the dependence of the various matching parameters on matching accuracy. The results from that investigation indicated that the matching parameters ($w$, *nbits*, $d$ and *nbins*) depended on the individual images and so were varied for the SB mammograms containing masses.

Only *nbits* and $d$ were varied for the SB mammograms containing calcifications as the sampling window size was fixed at 16 pixels.

The sampling window step size was fixed, at 4 pixels, for all SB mammograms and all combinations of matching parameters.

## 11.8 Examples of Matching Maps

Examples of matching maps for TM- and MI-matching are shown in **Figure 11.5**. The maps were selected as those with the minimum and maximum matching accuracy, $\kappa$, over all matching parameters ($w$, *nbits*, $d$, *nbins*). Values of the actual matching parameters used for these maps are given in **Table 11.2**. Maps with a negative contrast imply that the region to be matched is darker than the background.

The example maps for the minimum $\kappa$-values appear in the left column of **Figure 11.5**. All methods except TM-$D_M$ have negative contrasts, and although $C_{fb} > 0$ for TM-$D_M$, $A_{ROC} < 0$ indicating that the ROI was not matched.

The example maps for the maximum $\kappa$-values appear in the right column of **Figure 11.5**. All methods, except TM-$D_M$, have reasonably high $A_{ROC}$-values, but the $C_{fb}$-values are very low. TM-$D_M$ has a very high value for $C_{fb}$ but an $A_{ROC}$-value close to 0.5, indicating a very poor match. All methods, except MI-GLCMs, have very low values for $\kappa$, when compared to the results for the mosaic images and the standard mammograms. The result for MI-GLCMs is comparable to the results for the mosaic images and the standard mammograms. It is interesting to note that three of the five methods have an optimal sampling window size of 96 pixels.

**Table 11.2**: Summary of matching parameters for the example matching maps of SB mammograms, listed at the maximum and minimum $\kappa$-values for image pair S1S0. Negative contrasts indicate that the background is brighter than the region to be matched.

| Method | Minimum $\kappa$ | | | | | | Maximum $\kappa$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $w$ | *nbits* | *d*/*nbins* | $A_{ROC}$ | $C_{fb}$ | $\kappa$ | $w$ | *nbits* | *d*/*nbins* | $A_{ROC}$ | $C_{fb}$ | $\kappa$ |
| TM-$D_E$ | 144 | 7 | 1 | 0.15 | -0.26 | -0.19 | 112 | 8 | 1 | 0.84 | 0.15 | 0.10 |
| TM-$D_{ES}$ | 16 | 7 | 10 | 0.36 | -0.11 | -0.03 | 96 | 5 | 10 | 0.83 | 0.09 | 0.06 |
| TM-$D_M$ | 16 | 7 | 10 | 0.48 | 1.00 | -0.03 | 96 | 5 | 1 | 0.53 | 1.00 | 0.05 |
| MI-histograms | 48 | 7 | 64 | 0.28 | -0.27 | -0.12 | 96 | 8 | 16 | 0.73 | 0.07 | 0.03 |
| MI-GLCMs | 64 | 6 | 10 | 0.90 | -0.20 | -0.16 | 32 | 7 | 1 | 0.96 | 0.86 | 0.80 |

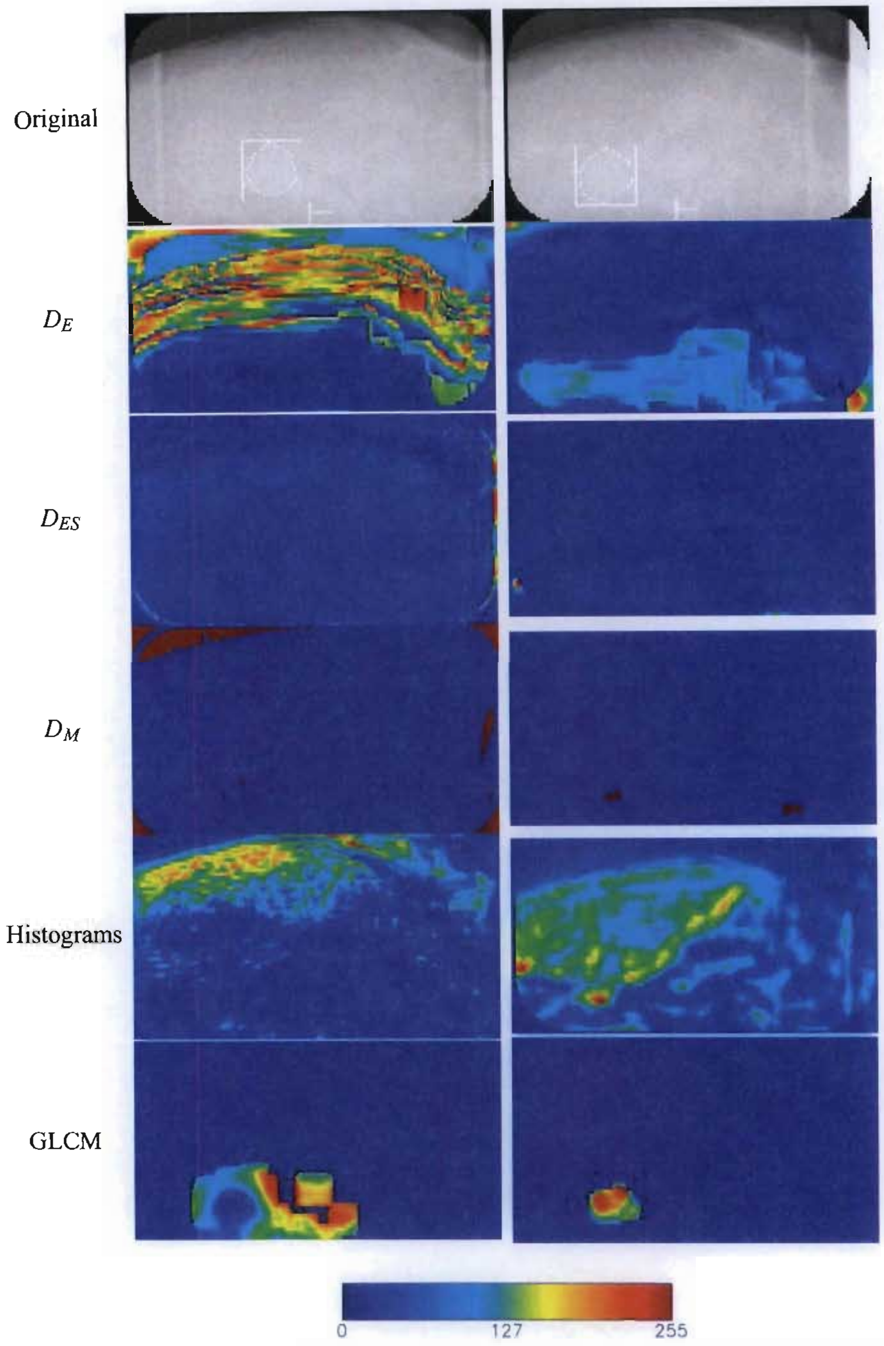


**Figure 11.5**: Examples of matching maps for a pair of SB mammograms. The maps were scaled to an intensity range between 0 and 255. The maps for the lowest and highest κ-values are shown in the left and right columns, respectively, with the colour bar at the bottom indicating the intensity scale. The reference image S1 (0°) is shown at the top left and the test image S0 (-10°) is shown at the top right.

## 11.9 TM-Matching Evaluation Results

The TM-matching algorithm (**Algorithm 8.1** on page 96) was applied to the stereotactic biopsy mammograms on page 193. The results of the dependence of matching accuracy on the matching parameters are discussed for each of the three distance similarity metrics. The results are colour-coded according to the breast tissue type of the reference ROI, benign (B), indeterminate (I) and malignant (M).

Examples of the median, average and standard deviation for each of the texture measures is shown in **Table 11.3** for image S0 to demonstrate that there is a significant difference in the ranges of the values used as inputs to the distance similarity metric. The values listed were calculated over the full range of matching parameters. Correlation, inertia, sum variance and difference variance have very large values and these texture measures probably dominate the Euclidean distance calculation. The values listed for each texture measure differ from those values for the example mosaic image in **Table 9.3** and for the standard mammogram example in **Table 10.5**. Therefore it is likely that each image has its own unique set of texture measures that best describe the image and a features selection algorithm would probably help to improve matching accuracy.

**Table 11.3**: Median, average and standard deviation of texture measures for image S0 to illustrate the different ranges of values for each texture measures. The values listed were calculated over the full range of matching parameters.

| Texture Measures for S0 | Median | Average | Standard Deviation |
|---|---|---|---|
| Entropy | 1.69 | 1.58 | 1.43 |
| Energy | 0.397 | 0.547 | 0.370 |
| Inertia | 1 680 | 5 500 | 7 880 |
| Inverse Difference Moment | 0.003 | 0.004 | 0.005 |
| Correlation | 1 670 | 34 800 | 187 000 |
| Sum Average | 58.9 | 91.8 | 95.0 |
| Sum Entropy | 1.69 | 1.58 | 1.43 |
| Difference Entropy | 1.69 | 1.54 | 1.37 |
| Sum Variance | 1 410 | 1 570 | 1 810 |
| Difference Average | 37.3 | 48.3 | 44.8 |
| Difference Variance | 292 | 1 180 | 1 690 |
| Information Measure of Correlation 1 | 1.00 | 0.779 | 0.005 |
| Maximum Probability | 0.550 | 0.632 | 0.319 |

### 11.9.1 Evaluation of TM-Matching with $D_E$

The results of applying TM-matching with the Euclidean distance similarity metric, $D_E$, to SB mammograms, are presented and discussed. The dependence of matching accuracy on *nbits*, $d$ and $w$ is examined. Results are presented in the formats described in §8.2 (page 93). Dependence on matching accuracy is indicated by separation of the different coloured plots.

### Effect of *nbits* on Matching Accuracy

Typical examples of $\kappa$ *vs.* $w$ in **Figure 11.6(a)** show how matching accuracy varies with *nbits* and $w$. None of the plots overlap completely and the plots for the two masses are very noisy. The results for the calcifications are plotted as points because only 1 sampling window was used.

The results of the autocorrelation function analysis are plotted as vertical lines for each value of *nbits*, and there is separation of the plots at the different bit-depths. The ACF-widths and ACF-heights are similar for the calcifications, but not for the masses. There appears to be poor correlation between the ACF results and the maximum values of $\kappa$. The ACF results are discussed in §11.12.

**Figure 11.6(b)** shows the average matching accuracy (over all image pairs) as a function of $d$, to highlight the general dependence of matching accuracy on *nbits*. The varying degrees of separation of the *nbits*-plots, indicates that the dependence of matching accuracy on *nbits* is influenced by the sampling window size. Matching accuracy improves as bit-depth increases for $w \leq 80$ pixels and matching accuracy improves as bit-depth decreases for $w \geq 96$ pixels.

### Effect of $d$ on Matching Accuracy

Some examples of $\kappa$ *vs.* $w$ are shown in **Figure 11.7(a)** to demonstrate how matching accuracy varies with $d$ and $w$. All plots for the different values of $d$ generally overlap almost completely, indicating that matching accuracy is independent of $d$ for these examples.

**Figure 11.7(b)** shows the average matching accuracy (over all image pairs) as a function of *nbits*, to highlight the general dependence of matching accuracy on $d$. Plots at the different values of $d$ overlap almost completely for all sampling window sizes, indicating that matching accuracy is independent of $d$.

### Effect of $w$ on Matching Accuracy

The example plots of $\kappa$ *vs.* $w$ in **Figures 11.6(a)** and **11.7(a)** show that matching accuracy varies as sampling window size varies. The variation is however unique to each image and the optimal sampling window size is most likely determined by the scale sizes of the textures in each image.

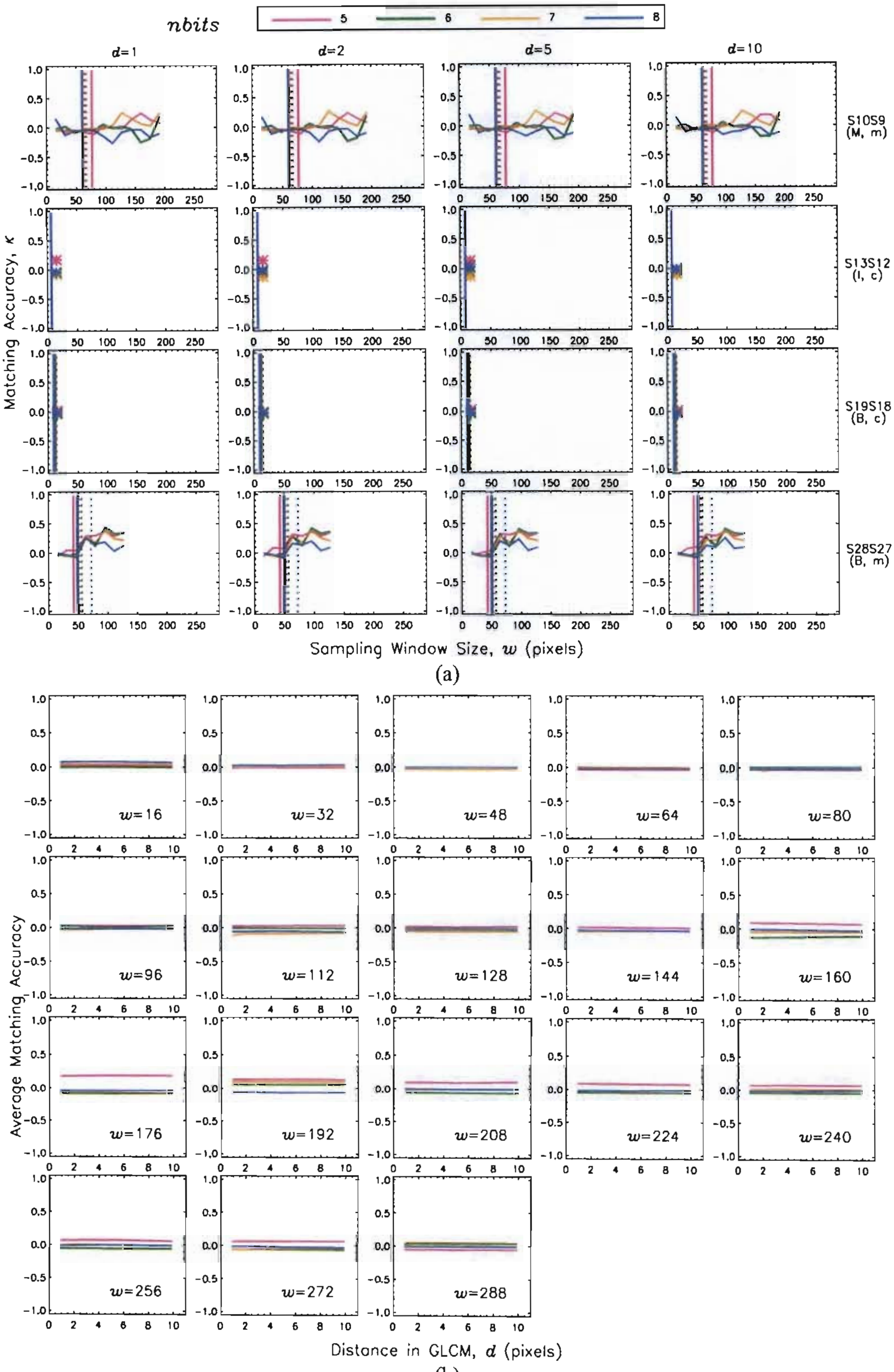
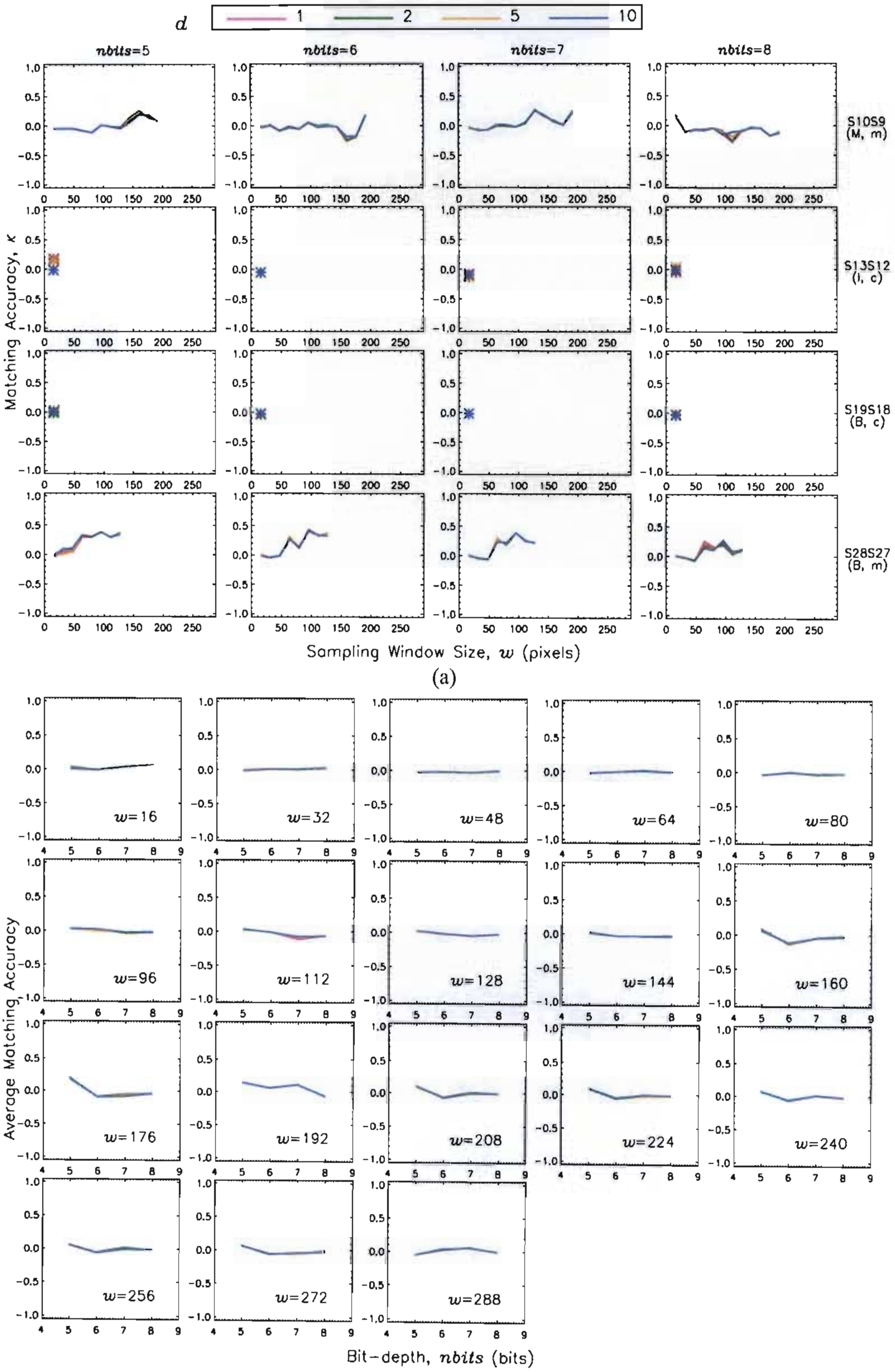


**Figure 11.6**: Effect of *nbits* on matching accuracy for TM-$D_E$ applied to pairs of SB mammograms. **(a)** Typical examples of κ *vs.* *w*. Each row contains the information for a single image pair. The vertical lines indicate the results of the ACF analysis. The solid line represents the ACF-width and the dotted line represents the ACF-height, at different bit-depths. **(b)** Average matching accuracy as a function of *d*, at different bit-depths to highlight the general dependence of matching accuracy on bit-depth for pairs of SB mammograms.



(a)

(b)

**Figure 11.7**: Effect of $d$ on matching accuracy for TM-$D_E$ applied to pairs of SB mammograms. **(a)** Typical examples of $\kappa$ *vs.* $w$. Each row contains the information for a single image pair. **(b)** Average (across *nbits*) matching accuracy.

### Summary of Best Results

**Figure 11.8** shows the best matching results for each SB mammogram pair, with results for the masses displayed on the left and those for the calcifications displayed on the right. All the masses have been matched, but one calcification has not been matched. The results are quite scattered for both masses and calcifications.

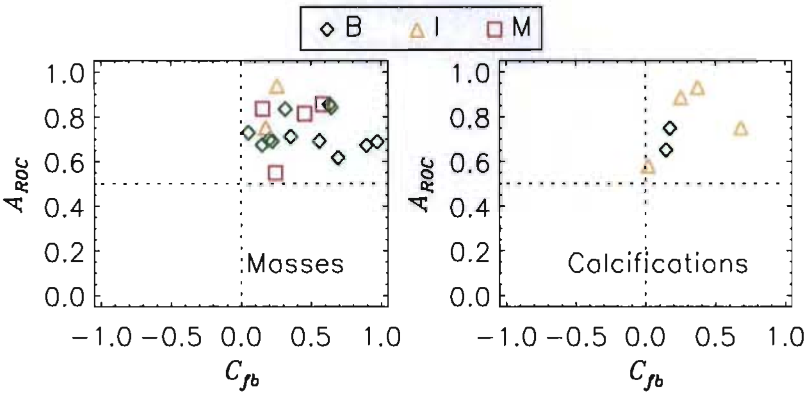


**Figure 11.8**: Scatter plot of the best results for TM-$D_E$ applied to SB mammograms, based on ROI feature.

**Figure 11.9** shows the best matching accuracies as functions of ROI area and ROI visibility, and matching accuracy is very poorly correlated with ROI area and ROI visibility. There is a range of matching accuracies for the different ROI areas and visibilities.

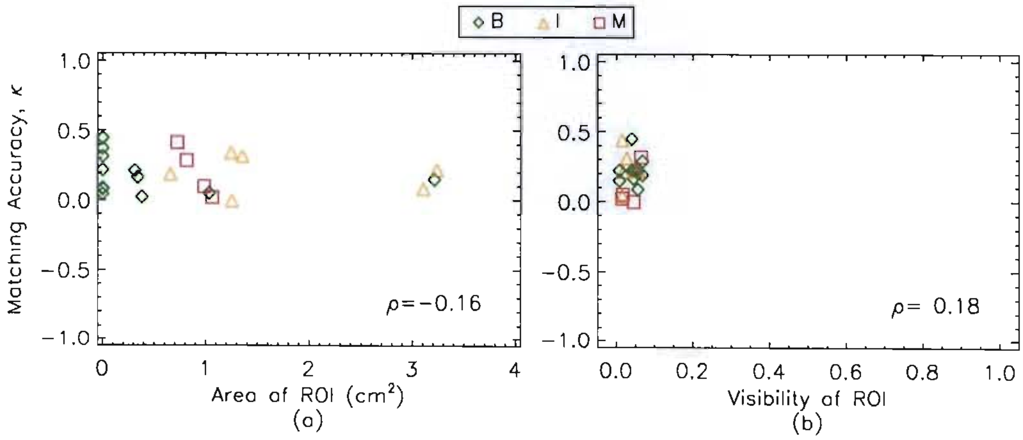


**Figure 11.9**: Best matching accuracy as functions of **(a)** of ROI area and **(b)** ROI visibility for each mammogram pair for TM-$D_E$. There is poor correlation between matching accuracy and ROI area or ROI visibility, since the correlation coefficients are very low.

**Table 11.4** shows the best matching accuracies as a function of diagnosis. There are no significant differences ($p > 0.65$) between the best average results for each diagnosis. Matching accuracies are considerably poorer than for the mosaic images and the standard mammograms. The average $A_{ROC}$-values are generally high, but all the average $C_{fb}$-values are very low.

**Table 11.4**: Average of the best matching results based on diagnosis for TM-$D_E$ for pairs of SB mammograms.

| Diagnosis | Average $A_{ROC}$ | Average $C_{fb}$ | Average $\kappa$ |
|---|---|---|---|
| | Best Results | | |
| Benign | 0.73±0.07 | 0.43±0.30 | 0.19±0.15 |
| Indeterminate | 0.81±0.14 | 0.29±0.22 | 0.20±0.13 |
| Malignant | 0.77±0.14 | 0.36±0.20 | 0.21±0.18 |
| | Best Mass Results | | |
| Benign | 0.73±0.08 | 0.47±0.30 | 0.22±0.15 |
| Indeterminate | 0.85±0.13 | 0.21±0.06 | 0.16±0.10 |
| Malignant | 0.77±0.14 | 0.36±0.20 | 0.21±0.18 |
| | Best Calcification Results | | |
| Benign | 0.70±0.07 | 0.16±0.02 | 0.07±0.03 |
| Indeterminate | 0.79±0.16 | 0.33±0.28 | 0.22±0.16 |

### 11.9.2 Evaluation of TM-Matching with $D_{ES}$

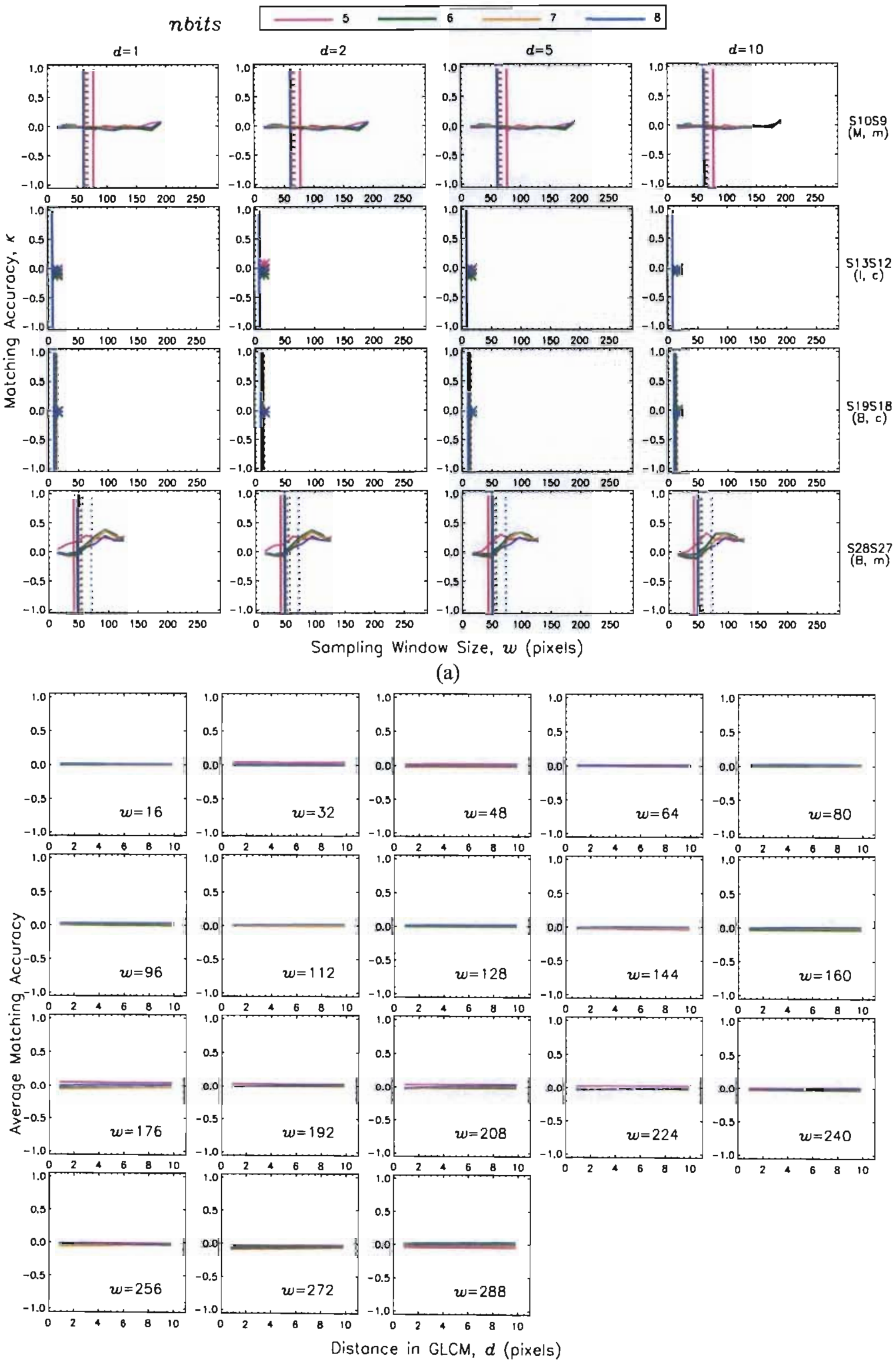
The results of applying TM-matching with the standardised Euclidean distance similarity metric, $D_{ES}$, to SB mammograms, are presented and discussed. The dependence of matching accuracy on *nbits*, $d$ and $w$ is examined. Results are presented in the formats described in §8.2 (page 93). Dependence on matching accuracy is indicated by separation of the different coloured plots.

#### Effect of *nbits* on Matching Accuracy

Typical examples of $\kappa$ *vs.* $w$ in **Figure 11.10(a)** show how matching accuracy varies with *nbits* and $w$. The plots for all the examples, except S28S27, overlap almost completely. The plots for S28S27 are separated for all values of *nbits*, but the peak for the plot at *nbits*=5 bits is shifted to the left, indicating a smaller optimal sampling window size at this bit-depth.

The results of the autocorrelation function analysis are plotted as vertical lines for each value of *nbits*, and there is separation of the plots at the different bit-depths. The ACF-widths and ACF-heights are similar for the calcifications, but not for the masses. There appears to be poor correlation between the ACF results and the maximum values of $\kappa$. The ACF results are discussed in §11.12.

**Figure 11.10(b)** shows the average matching accuracy (over all image pairs) as a function of $d$, to highlight the general dependence of matching accuracy on *nbits*. The dependence of matching accuracy on *nbits* appears to be influenced by the size of the sampling window. This is clear because the best matches occur at the highest bit-depth for some sampling window sizes (80 pixels to 160 pixels) and at the lowest bit-depth for others (32 pixels to 64 pixels, 176 pixels).



(a)

(b)

**Figure 11.10**: Effect of *nbits* on matching accuracy for TM-$D_{ES}$ applied to pairs of SB mammograms. **(a)** Typical examples of $\kappa$ *vs.* $w$. Each row contains the information for a single image pair. The vertical lines indicate the results of the ACF analysis. The solid line represents the ACF-width and the dotted line represents the ACF-height, at different bit-depths. **(b)** Average matching accuracy as a function of $d$, at different bit-depths to highlight the general dependence of matching accuracy on bit-depth for pairs of SB mammograms.

### Effect of $d$ on Matching Accuracy

Some examples of $\kappa$ *vs.* $w$ are shown in **Figure 11.11(a)** to demonstrate how matching accuracy varies with $d$ and $w$. All plots for the different values of $d$ generally overlap almost completely, indicating that there is no dependence of matching accuracy on $d$ for these examples. There is some separation of the points for S13S12 and S19S18 at the lower bit-depths.

**Figure 11.11(b)** shows the results of averaging the matching accuracy over all pairs of SB mammograms with fixed values of $w$, *nbits* and $d$, to highlight the general dependence of matching accuracy on $d$. Plots at the different values of $d$ overlap almost completely for all sampling window sizes, indicating that matching accuracy is independent of $d$.

### Effect of $w$ on Matching Accuracy

The example plots of $\kappa$ *vs.* $w$ in **Figures 11.10(a)** and **11.11(a)** show that matching accuracy varies as sampling window size varies. The variation is however unique to each image and the optimal sampling window size is most likely determined by the scale sizes of the textures in each image.

### Summary of Best Results

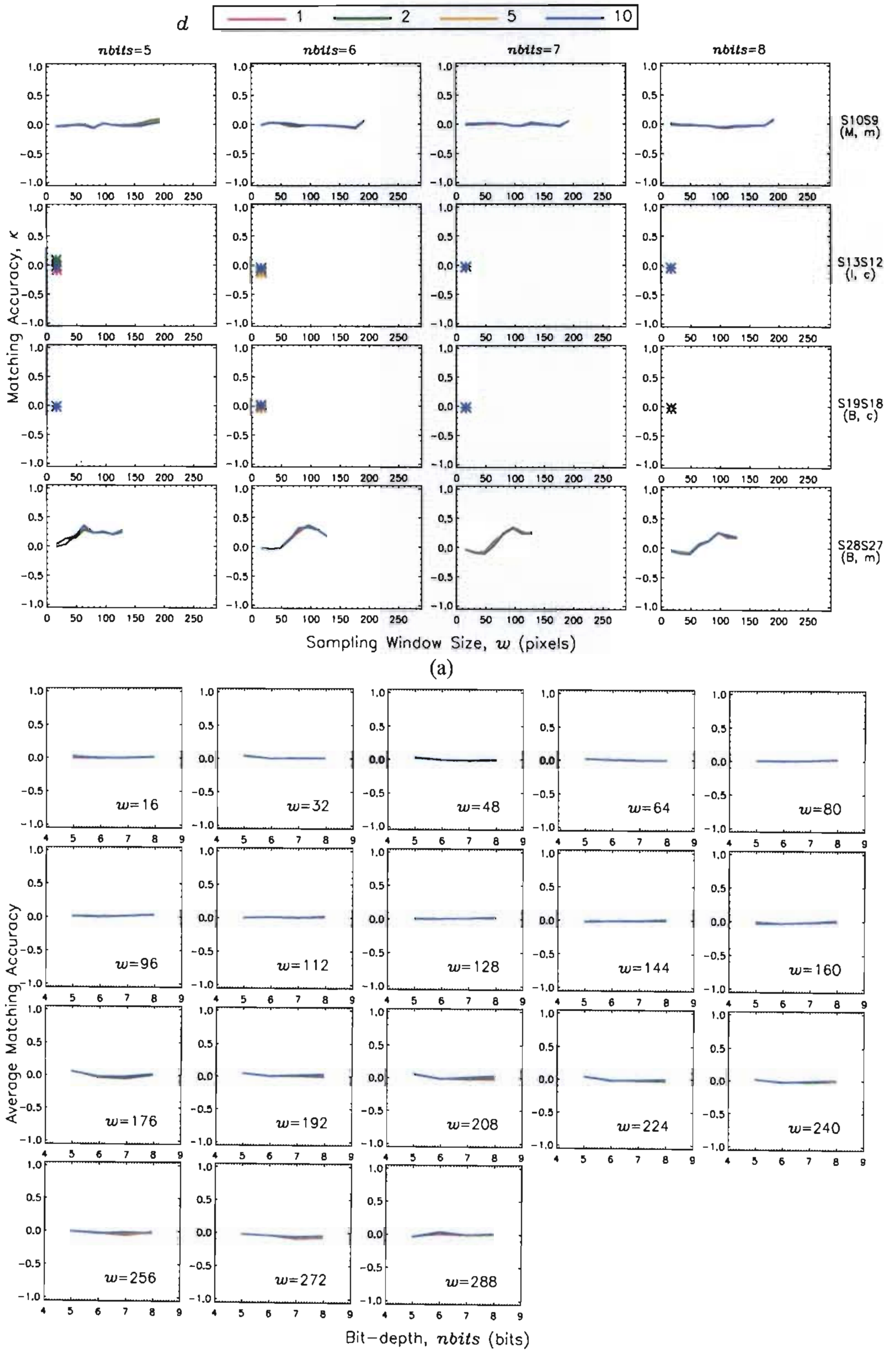
**Figure 11.12** shows the best matching results for each pair of SB mammograms according to whether the ROIs were masses or calcifications. The $A_{ROC}$- and $C_{fb}$-values are scattered over a wide range and are quite low for the calcifications. However, all masses and calcifications were matched.

**Table 11.5** shows the best matching accuracies as a function of diagnosis. The matching of the malignant masses had the lowest matching accuracy, while the indeterminate and benign ROIs were matched with slightly higher accuracies. The matching accuracies for the calcifications were very poor. All $C_{fb}$-values were very low.

**Table 11.5**: Average of the best matching results based on diagnosis for TM-$D_{ES}$ for pairs of SB mammograms.

| Diagnosis | Average $A_{ROC}$ | Average $C_{fb}$ | Average $\kappa$ |
|---|---|---|---|
| | Best Results | | |
| Benign | 0.80±0.10 | 0.29±0.20 | 0.17±0.11 |
| Indeterminate | 0.79±0.09 | 0.26±0.12 | 0.14±0.06 |
| Malignant | 0.76±0.05 | 0.22±0.13 | 0.10±0.04 |
| | Best Mass Results | | |
| Benign | 0.79±0.10 | 0.33±0.20 | 0.18±0.11 |
| Indeterminate | 0.85±0.02 | 0.31±0.07 | 0.21±0.03 |
| Malignant | 0.76±0.05 | 0.22±0.13 | 0.10±0.04 |
| | Best Calcification Results | | |
| Benign | 0.86±0.02 | 0.09±0.06 | 0.07±0.05 |
| Indeterminate | 0.77±0.10 | 0.23±0.14 | 0.11±0.03 |

**Figure 11.13** shows the best matching accuracies as functions of ROI area and ROI visibility. There is a range of matching accuracies for the small and large masses. There is also a range of matching accuracies for the ROI visibilities. This plot shows that there is poor correlation between matching accuracy and ROI area or ROI visibility, since the correlation coefficients are very low.



(a)

(b)

**Figure 11.11**: Effect of $d$ on matching accuracy for TM-$D_{ES}$ applied to pairs of SB mammograms. **(a)** Typical examples of $\kappa$ *vs.* $w$. Each row contains the information for a single image pair. **(b)** Average matching accuracy as a function of bit-depth, at different values of $d$ to highlight the general dependence of matching accuracy on $d$ for the pairs of SB mammograms.

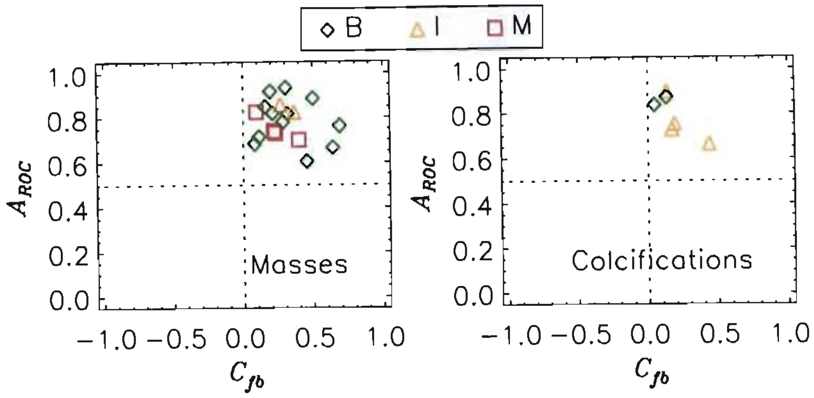


**Figure 11.12**: Scatter plot of the best results for TM-$D_{ES}$ applied to SB mammograms, based on ROI feature.

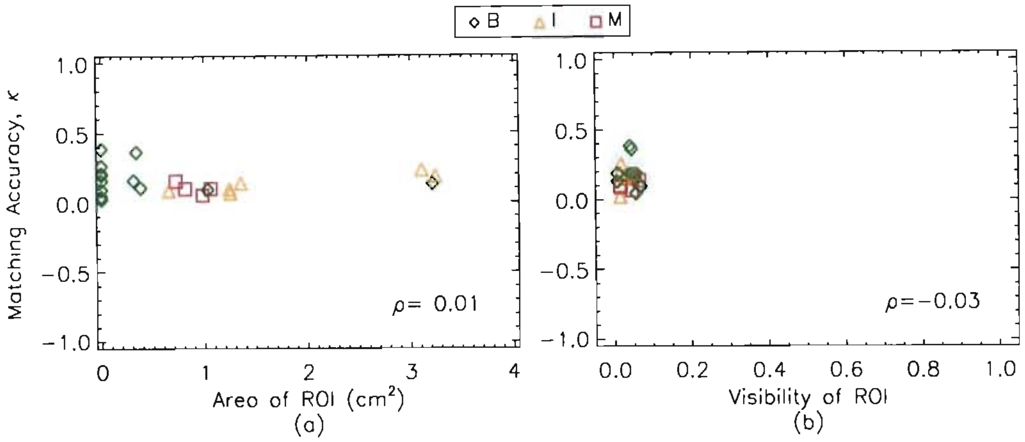


**Figure 11.13**: Best matching accuracy as functions of **(a)** ROI area and **(b)** ROI visibility for each mammogram pair for TM-$D_{ES}$. The correlation between matching accuracy and ROI area or visibility is very poor, since the correlation coefficients are very low.

### 11.9.3 Evaluation of TM-Matching with $D_M$

The results of applying TM-matching with the Mahalanobis distance similarity metric, $D_M$, to pairs of SB mammograms, are presented and discussed. The dependence of matching accuracy on *nbits*, *d* and *w* is examined. Results are presented in the formats described in §8.2 (page 93). Dependence on matching accuracy is indicated by separation of the different coloured plots.

**Effect of *nbits* on Matching Accuracy**

Typical examples of κ *vs.* *w* in **Figure 11.14(a)** show how matching accuracy varies with *nbits* and *w*. All plots for S13S12 overlap completely. For S10S9 and S19S18, all plots for *nbits*>5 bits overlap almost completely indicating no dependence of matching accuracy on bit-depth. For the 5-bit plots, S10S9 has maximal matching accuracy and S19S18 has minimal matching accuracy. All plots are separated for S28S27, with matching accuracy generally improving as *nbits* decreases.

The results of the ACF analysis are plotted as vertical lines for each value of *nbits*, and there is separation of the plots at the different bit-depths. The ACF-widths and ACF-heights are similar for the calcifications, but not for the masses. There appears to be poor correlation between the ACF results and the maximum values of $\kappa$. The ACF results are discussed in §11.12.

**Figure 11.14(b)** shows the average matching accuracy (over all image pairs) as a function of $d$, to highlight the general dependence of matching accuracy on *nbits*. The dependence of matching accuracy on bit-depth is influenced by the sampling window size. Matching accuracies are generally very low, with all plots overlapping completely for most sampling window sizes. There are some sampling window sizes (176 pixels) where there is come separation of the plots for different values of *nbits*. It appears as if optimal matching occurs at low bit-depths.

### Effect of *d* on Matching Accuracy

Some examples of $\kappa$ *vs.* $w$ are shown in **Figure 11.15(a)** to demonstrate how matching accuracy varies with $d$ and $w$. All plots for the different values of $d$ generally overlap almost completely, indicating that there is no dependence of matching accuracy on $d$ for these examples.

**Figure 11.15(b)** shows the results of averaging the matching accuracy over all pairs of SB mammograms with fixed values of $w$, *nbits* and $d$, to highlight the general dependence of matching accuracy on $d$. Plots at the different values of $d$ overlap almost completely for all sampling window sizes, indicating that matching accuracy is independent of $d$.

### Effect of *w* on Matching Accuracy

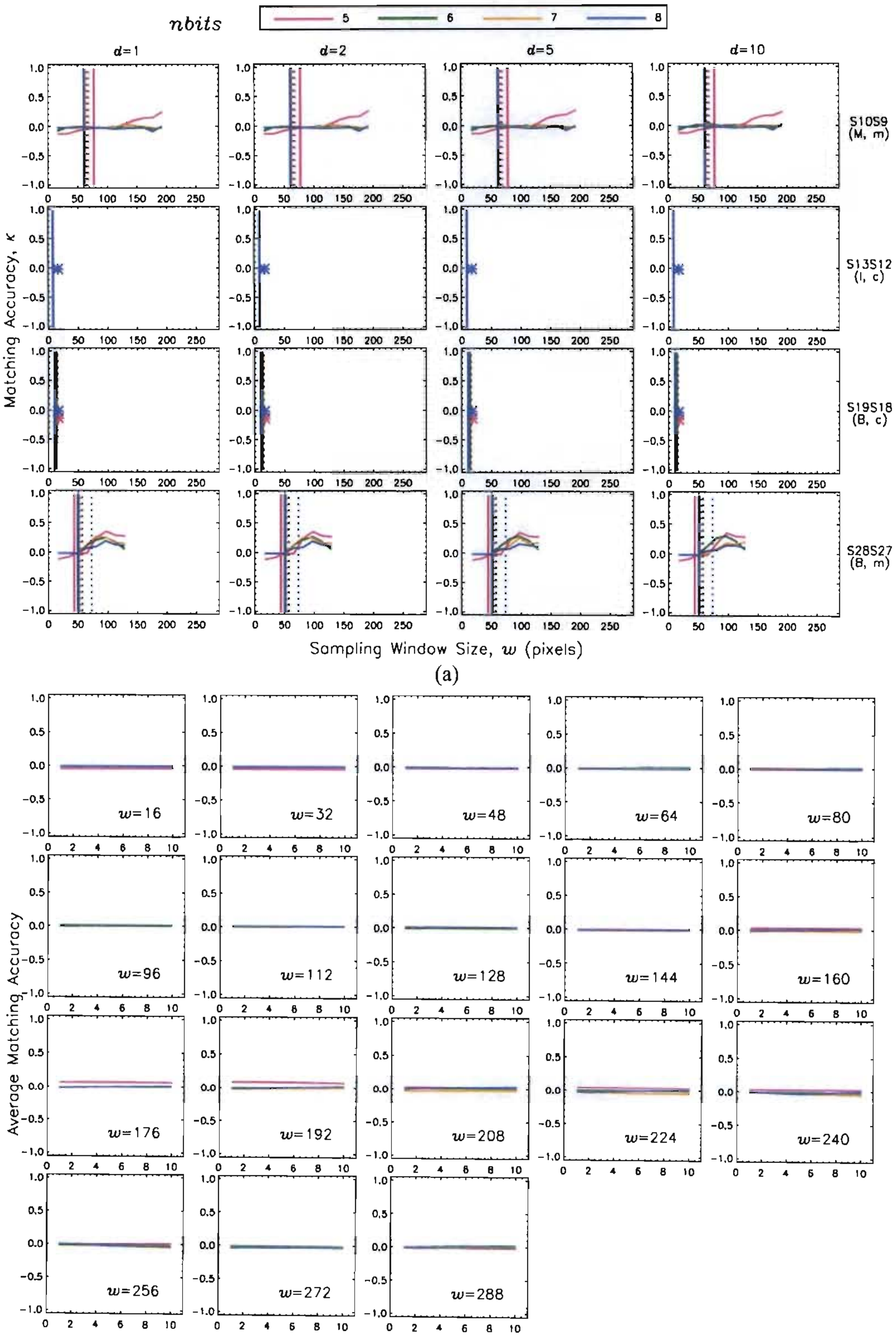
The example plots of $\kappa$ *vs.* $w$ in **Figures 11.14(a)** and **11.15(a)** show that matching accuracy varies as sampling window size varies. The variation is however unique to each image and the optimal sampling window size is most likely determined by the scale sizes of the textures in each image.

### Summary of Best Results

**Figure 11.16** shows the best results for each image pair according to whether the ROI contains a mass or a calcification. All masses were matched but no calcifications were matched. The latter is quite surprising since both TM-$D_E$ and TM-$D_{ES}$ have successfully, albeit poorly, matched the calcifications. This might be an indication that the use of $D_M$ as a similarity metric for the calcifications is inappropriate, as vital information is lost in the calculation of $D_M$.

**Table 11.6** shows the best matching accuracies as a function of diagnosis. The matching of the malignant masses had the highest matching accuracy, while the indeterminate and benign ROIs were matched with slightly lower accuracies. The poor average matching accuracies for the indeterminate and benign ROIs is most likely due to no calcifications being matched. Overall results are significantly lower than for TM-$D_E$ and TM-$D_{ES}$.

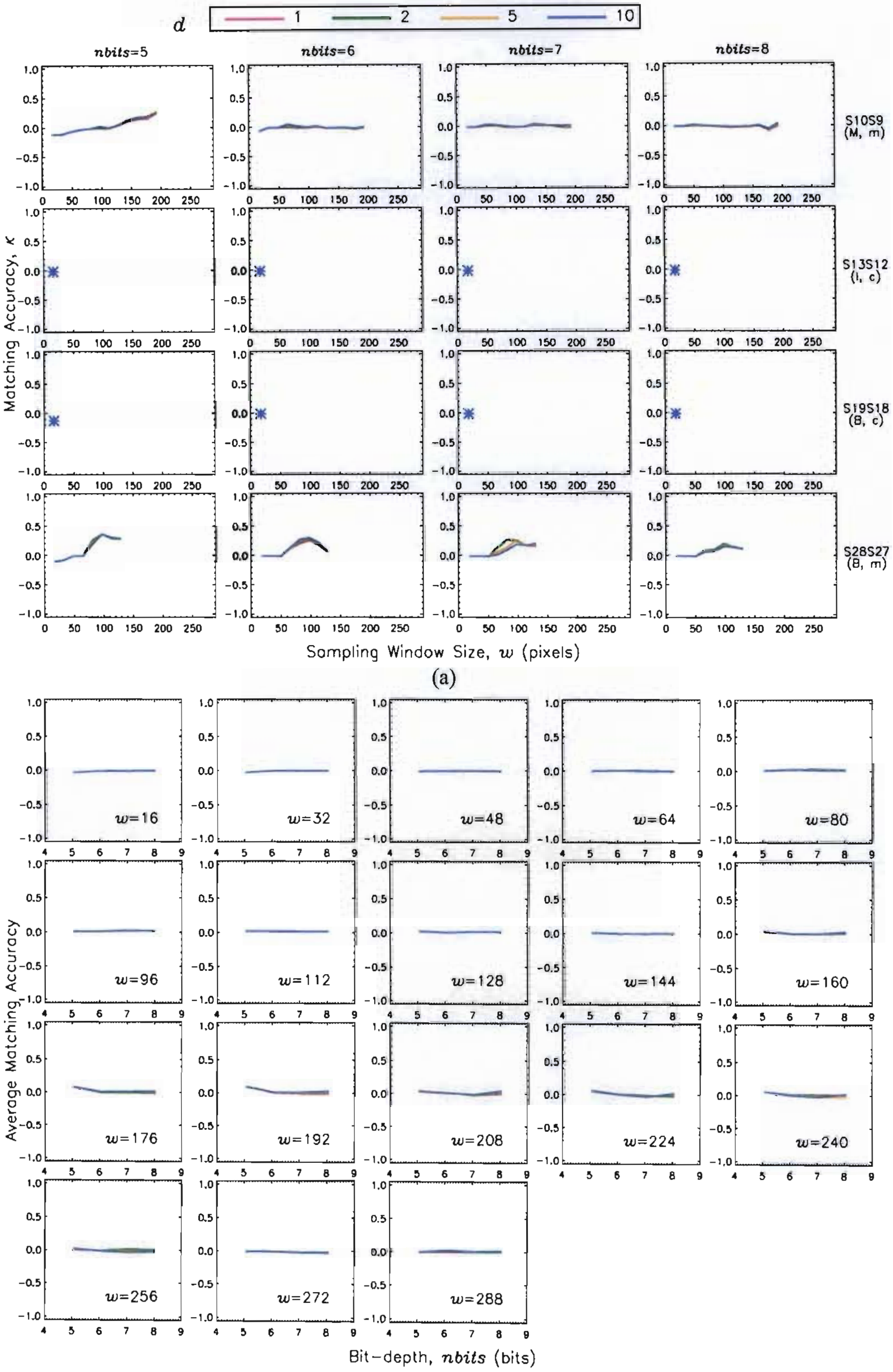
**Figure 11.17** shows the best matching accuracies as functions of ROI area and ROI visibility. Matching accuracies are generally very low irrespective of the ROI area or visibility.



(a)

(b)

**Figure 11.14**: Effect of *nbits* on matching accuracy for TM-$D_M$ applied to pairs of SB mammograms. **(a)** Typical examples of κ *vs.* *w*. Each row contains the information for a single image pair. The vertical lines indicate the results of the ACF analysis. The solid line represents the ACF-width and the dotted line represents the ACF-height, at different bit-depths. **(b)** Average matching accuracy as a function of *d*, at different bit-depths to highlight the general dependence of matching accuracy on bit-depth for pairs of SB mammograms.



(a)

(b)

**Figure 11.15**: Effect of $d$ on matching accuracy for TM-$D_M$ applied to pairs of SB mammograms. **(a)** Typical examples of κ *vs.* $w$. Each row contains the information for a single image pair. **(b)** Average matching accuracy as a function of bit-depth, at different values of $d$ to highlight the general dependence of matching accuracy on $d$ for the pairs of SB mammograms.

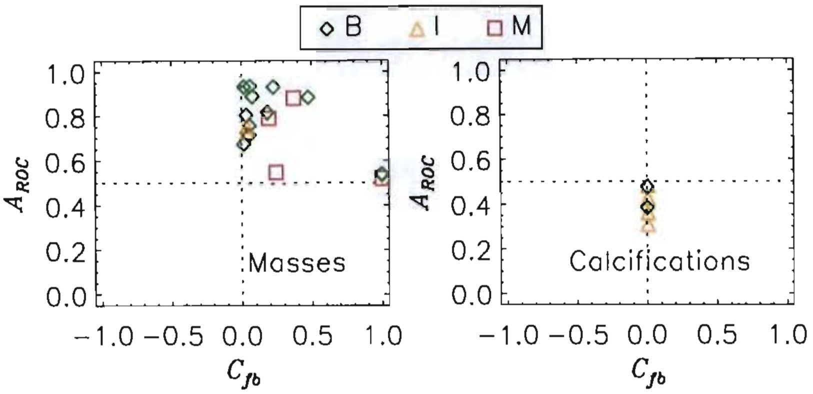


**Figure 11.16**: Scatter plot of the best results for TM-$D_M$ applied to SB mammograms, based on ROI feature.

**Table 11.6**: Average of the best matching results based on diagnosis for TM-$D_M$ for pairs of SB mammograms.

| Diagnosis | Average $A_{ROC}$ | Average $C_{fb}$ | Average κ |
|---|---|---|---|
| | Best Results | | |
| Benign | 0.74±0.19 | 0.23±0.35 | 0.08±0.10 |
| Indeterminate | 0.52±0.19 | 0.01±0.02 | 0.01±0.01 |
| Malignant | 0.68±0.17 | 0.43±0.38 | 0.10±0.08 |
| | Best Mass Results | | |
| Benign | 0.79±0.14 | 0.27±0.37 | 0.09±0.10 |
| Indeterminate | 0.75±0.02 | 0.04±0.00 | 0.02±0.00 |
| Malignant | 0.68±0.17 | 0.43±0.38 | 0.10±0.08 |
| | Best Calcification Results | | |
| Benign | 0.43±0.07 | 0.00±0.00 | 0.00±0.00 |
| Indeterminate | 0.40±0.07 | 0.00±0.00 | 0.00±0.00 |

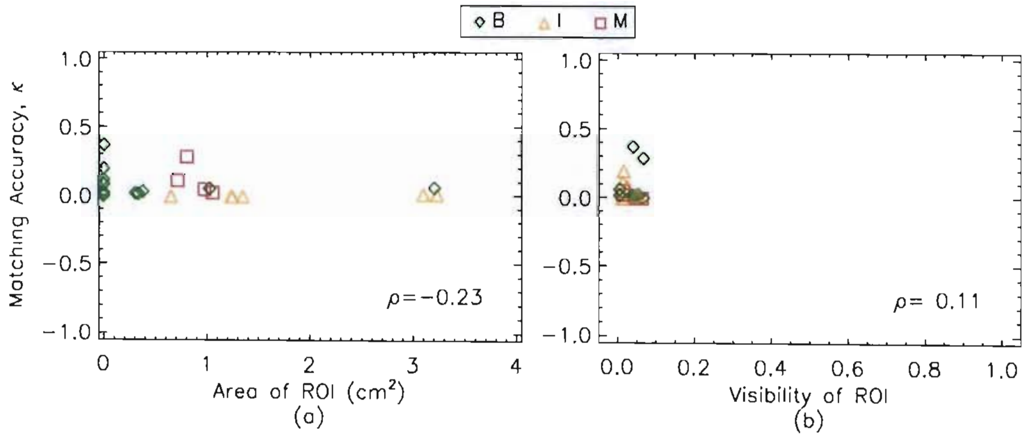


**Figure 11.17**: Best matching accuracy as functions of **(a)** ROI area and **(b)** ROI visibility for each mammogram pair for TM-$D_M$.

## 11.10 MI-Matching Evaluation Results

The MI-matching algorithm, using grey-level histograms and GLCMs to estimate the probability density functions (**Algorithm 8.2** on page 98) was applied to pairs of SB mammograms.

### 11.10.1 Evaluation of MI-Matching with Histograms

The results of applying MI-matching with histograms to pairs of SB mammograms, are presented and discussed. The dependence of matching accuracy on *nbits*, *nbins* and $w$ is examined. Results are presented in the formats described in §8.2 (page 93). Dependence on matching accuracy is indicated by separation of the different coloured plots.

#### Effect of *nbits* on Matching Accuracy

Typical examples of $\kappa$ *vs.* $w$ in **Figure 11.18(a)** show how matching accuracy varies with *nbits* and $w$. For S19S18 and S28S27, all plots overlap completely, and for S10S9 and S13S12, all plots with $nbits>5$ bits, overlap completely, indicating that matching accuracy is independent of *nbits*. The 5-bit plot has the maximal matching accuracy for S10S9 and S13S12.

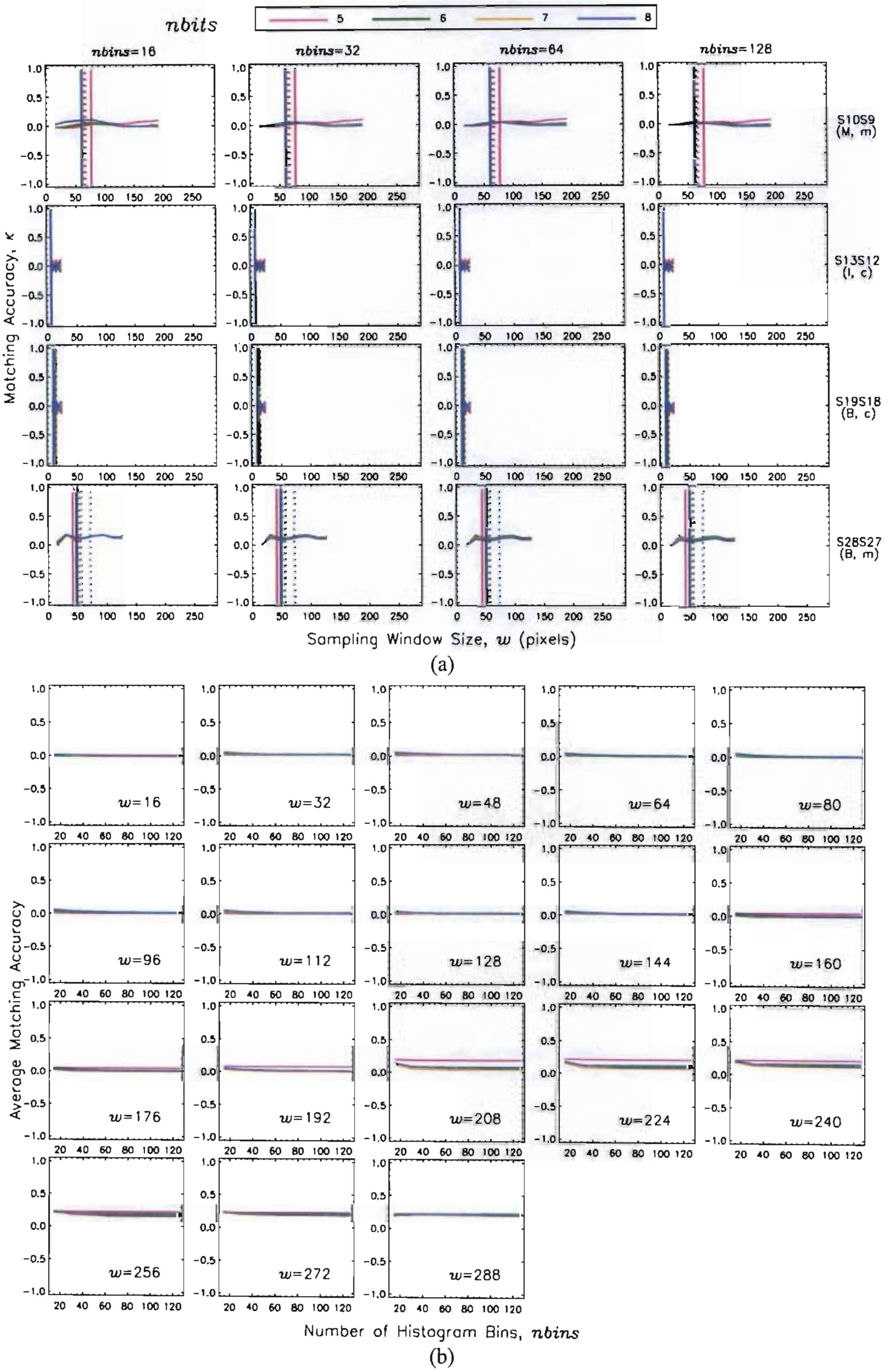
The results of the ACF analysis are plotted as vertical lines for each value of *nbits*, and there is separation of the plots at the different bit-depths. The ACF-widths and ACF-heights are similar for the calcifications, but not for the masses. There appears to be poor correlation between the ACF results and the maximum values of $\kappa$. The ACF results are discussed in §11.12.

**Figure 11.18(b)** shows the average matching accuracy (over all image pairs) as a function of $d$, to highlight the general dependence of matching accuracy on *nbits*. Most plots overlap almost completely for $w<160$ pixels. For $w\geq160$ pixels, optimal matching accuracy occurs for the lowest bit-depth.

#### Effect of *nbins* on Matching Accuracy

Some examples of $\kappa$ *vs.* $w$ are shown in **Figure 11.19(a)** to demonstrate how matching accuracy varies with *nbins* and $w$. All plots for the different values of *nbins* generally overlap almost completely, indicating that there is no dependence of matching accuracy on $d$ for these examples. However, for S10S9 there is a slight indication that the matching accuracy is best for the highest bit-depth and the fewest histogram bins.

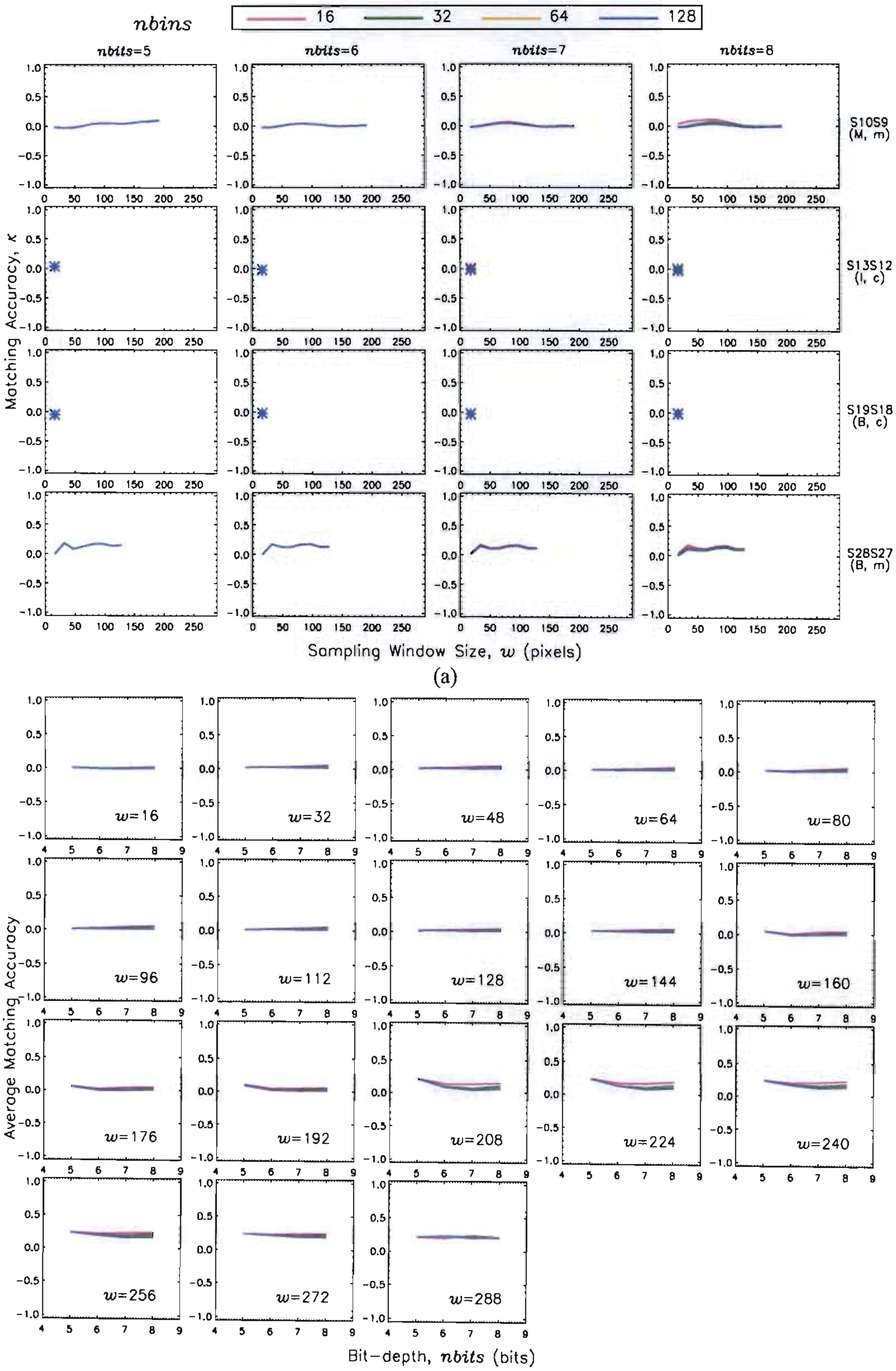
**Figure 11.19(b)** shows the average matching accuracy (over all image pairs) as a function of *nbits*, to highlight the general dependence of matching accuracy on *nbins*. The degree of separation of the *nbins*-plots increases as sampling window size increases. Matching accuracy generally improves as *nbins* decreases for all sampling window sizes.



(a)

(b)

**Figure 11.18**: Effect of *nbits* on matching accuracy for MI-histograms applied to pairs of SB mammograms. **(a)** Typical examples of κ *vs. w*. Each row contains the information for a single image pair. The vertical lines indicate the results of the ACF analysis. The solid line represents the ACF-width and the dotted line represents the ACF-height, at different bit-depths. **(b)** Average matching accuracy as a function of *nbins*, at different bit-depths to highlight the general dependence of matching accuracy on bit-depth for pairs of SB mammograms.



(a)

(b)

**Figure 11.19**: Effect of *nbins* on matching accuracy for MI-histograms applied to pairs of SB mammograms. **(a)** Typical examples of κ *vs.* *w*. Each row contains the information for a single image pair. **(b)** Average matching accuracy as a function of bit-depth, at different values of *nbins* to highlight the general dependence of matching accuracy on *nbins* for the pairs of SB mammograms.

### Effect of *w* on Matching Accuracy

The example plots of κ *vs.* *w* in **Figures 11.18(a)** and **11.19(a)** show that matching accuracy varies as sampling window size varies. The variation is however unique to each image and the optimal sampling window size is most likely determined by the scale sizes of the textures in each image.

### Summary of Best Results

**Figure 11.20** shows the best results for each pair of SB mammograms according to whether the ROI contained a mass or a calcification. Results are spread over a wide range of $A_{ROC}$- and $C_{fb}$-values, but there is a cluster around $A_{ROC}$=0.5 and $C_{fb}$=0 for the masses. Four masses were not matched and one calcification was not matched.

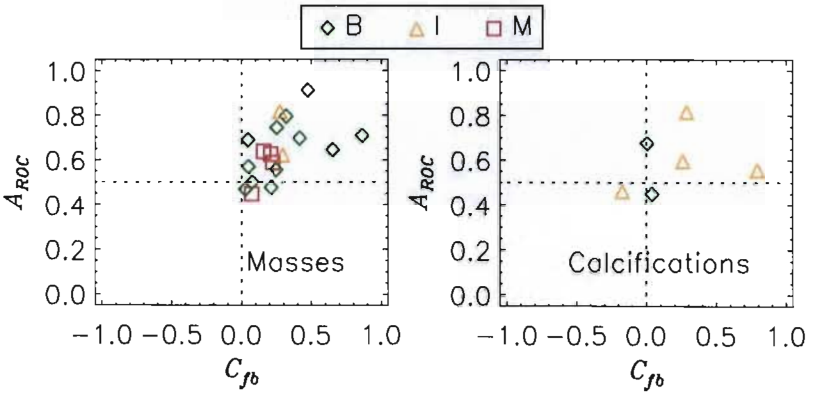


**Figure 11.20**: Scatter plot of the best results for MI-histograms applied to SB mammograms, based on diagnosis.

**Figure 11.21** shows the best matching accuracies as functions of ROI area and ROI visibility. Matching accuracies are generally very low irrespective of the ROI area or visibility. There is no correlation between matching accuracy and ROI area or ROI visibility.

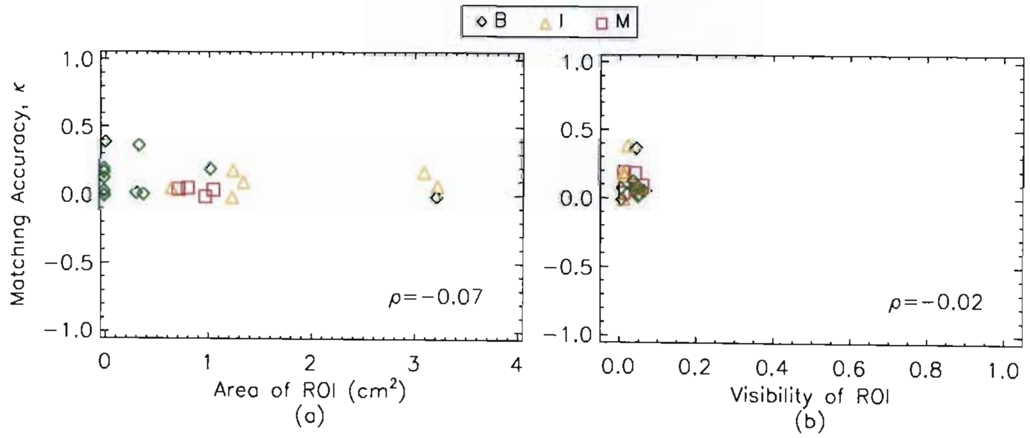


**Figure 11.21**: Best matching accuracy as functions of **(a)** ROI area and **(b)** ROI visibility for each mammogram pair for MI-histograms. There is poor correlation between visibility or ROI area and matching accuracy, since the correlation coefficients are very low.

**Table 11.7** shows the best matching accuracies as a function of diagnosis. The matching of the malignant masses had the lowest matching accuracy, while the indeterminate and benign ROIs were matched with slightly higher accuracies. Matching accuracies are very low.

**Table 11.7**: Average of the best matching results based on ROI feature for MI-histograms for pairs of SB mammograms.

| Diagnosis | Average $A_{ROC}$ | Average $C_{fb}$ | Average $\kappa$ |
|---|---|---|---|
| | Best Results | | |
| Benign | 0.64±0.14 | 0.26±0.26 | 0.10±0.14 |
| Indeterminate | 0.65±0.14 | 0.29±0.31 | 0.10±0.08 |
| Malignant | 0.58±0.09 | 0.17±0.07 | 0.03±0.03 |
| | Best Mass Results | | |
| Benign | 0.66±0.14 | 0.30±0.26 | 0.12±0.14 |
| Indeterminate | 0.73±0.14 | 0.29±0.01 | 0.13±0.07 |
| Malignant | 0.58±0.09 | 0.17±0.07 | 0.03±0.03 |
| | Best Calcification Results | | |
| Benign | 0.57±0.16 | 0.02±0.02 | 0.00±0.00 |
| Indeterminate | 0.61±0.15 | 0.29±0.40 | 0.08±0.08 |

### 11.10.2 Evaluation of MI-Matching with GLCMs

The results of applying MI-matching with GLCMs to pairs of SB mammograms, are presented and discussed. The dependence of matching accuracy on *nbits*, *d* and *w* is examined. Results are presented in the formats described in §8.2 (page 93). Dependence on matching accuracy is indicated by separation of the different coloured plots.

#### Effect of *nbits* on Matching Accuracy

Typical examples of κ *vs.* *w* are shown in **Figure 11.22(a)** to demonstrate the effect of *nbits* and *w* on matching accuracy. The plots are well separated for S10S9 and S28S27, showing that matching accuracy improves as bit-depth increases. All points overlap completely for S13S12 and there is no apparent pattern regarding the behaviour of matching accuracy as bit-depth changes. For S19S18, matching accuracy appears to improve with decreasing *nbits*.

The results of the ACF analysis are plotted as vertical lines for each value of *nbits*, and there is separation of the plots at the different bit-depths. The ACF-widths and ACF-heights are similar for the calcifications, but not for the masses. There appears to be poor correlation between the ACF results and the maximum values of κ. The ACF results are discussed in §11.12.

**Figure 11.22(b)** shows the average matching accuracy (over all image pairs) as a function of *d*, to highlight the general dependence of matching accuracy on *nbits*. The degree of separation of the *nbits*-plots increases with decreasing sampling window size. Matching accuracy improves as bit-depth increases for all window sizes, but the dependence is most apparent as the sampling window sizes get smaller.

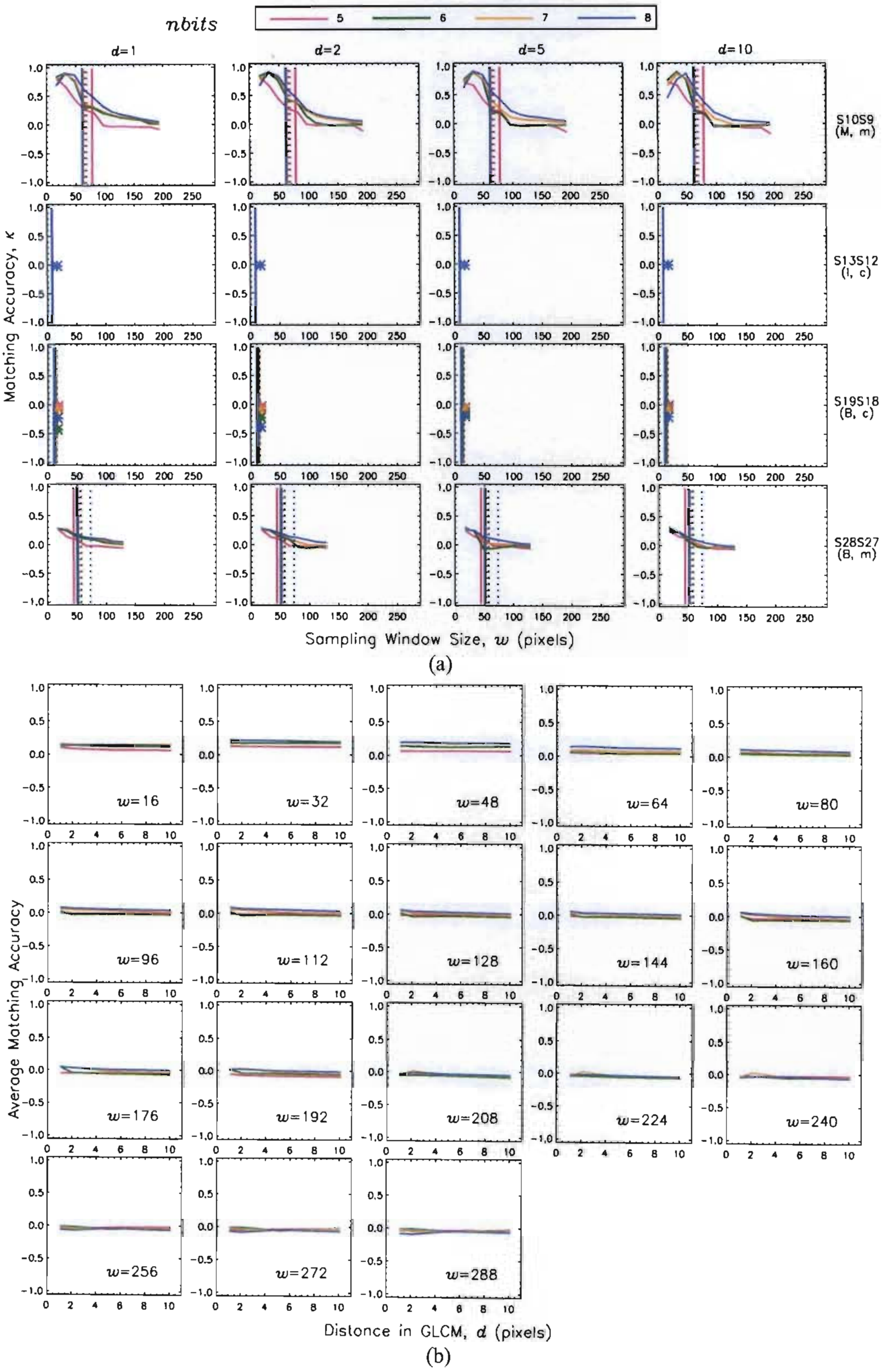


**Figure 11.22**: Effect of *nbits* on matching accuracy for MI-GLCMs applied to pairs of SB mammograms. **(a)** Typical examples of κ *vs.* *w*. Each row contains the information for a single image pair. The vertical lines indicate the results of the ACF analysis. The solid line represents the ACF-width and the dotted line represents the ACF-height, at different bit-depths. **(b)** Average matching accuracy as a function of *d*, at different bit-depths to highlight the general dependence of matching accuracy on bit-depth for pairs of SB mammograms.

### Effect of *d* on Matching Accuracy

Some examples of $\kappa$ *vs.* *w* are shown in **Figure 11.23(a)** to demonstrate how matching accuracy varies with *d* and *w*. There is some separation of the plots for S10S9 and S28S27, and it appears that optimal matching accuracy occurs at the lowest values of *d*, for the masses. There is no consistent pattern of dependence of matching accuracy on bit-depth for the calcifications, with the results for S13S12 overlapping completely and the results for S19S18 showing that matching accuracy improves with increasing *d*.

**Figure 11.23(b)** shows the average matching accuracy (over all image pairs) as a function of *nbits*, to highlight the general dependence of matching accuracy on *d*. Plots at the different values of *d* overlap almost completely for the small sampling windows ($w \leq 80$ pixels). For $w > 80$ pixels, matching accuracy appears to improve as *d* decreases.

### Effect of *w* on Matching Accuracy

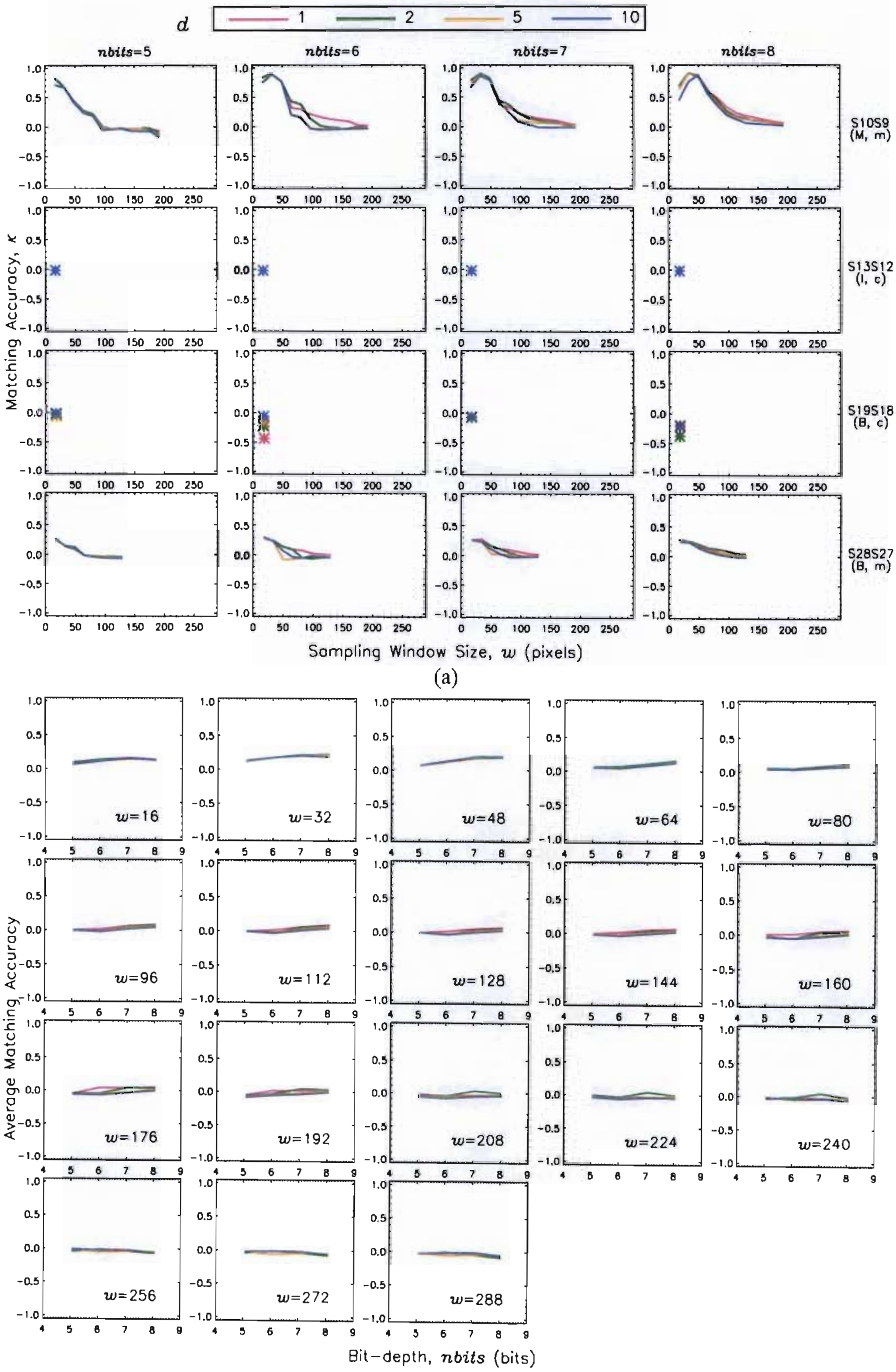
The example plots of $\kappa$ *vs.* *w* in **Figures 11.22(a)** and **11.23(a)** show that matching accuracy generally improves as sampling window size decreases.

### Summary of Best Results

**Figure 11.24** shows the best matching results for each pair of SB mammograms. Most results have $C_{fb}>0.5$ and $A_{ROC}>0.70$, but there are 6 image pairs (S22S21, S25S24, S31S30, S31S32, S34S33 and S34S35) that were not matched. If these image pairs are examined (page 193) it is obvious that the average grey-levels of the reference images vary significantly from the test images. These examples highlight one weakness of using GLCMs to estimate a probability density, since there is poor correlation in the GLCM if the average grey-levels are different in each image. This problem would probably be solved with some pre-processing to equalise grey-levels.

**Table 11.8** shows the best matching accuracies as a function of diagnosis. The matching of the malignant masses had the highest matching accuracy, considerably higher than the matching accuracies for the other matching methods. The matching results for the benign masses are comparable to the results for TM-$D_E$. However, the indeterminate ROIs and calcifications are very poorly matched.

**Figure 11.25** shows the best matching accuracies as functions of ROI area and ROI visibility. Apart from the results for the malignant masses, matching accuracies are generally very low irrespective of the ROI area or visibility.



(a)

(b)

**Figure 11.23**: Effect of *d* on matching accuracy for MI-GLCMs applied to pairs of SB mammograms. **(a)** Typical examples of κ *vs.* *w*. Each row contains the information for a single image pair. **(b)** Average matching accuracy as a function of bit-depth, at different values of *d* to highlight the general dependence of matching accuracy on *d* for the pairs of SB mammograms.

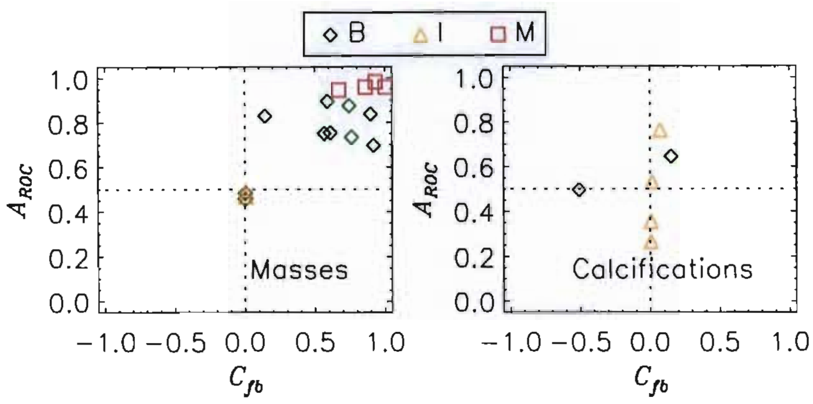


**Figure 11.24**: Scatter plot of the best results for MI-GLCMs applied to SB mammograms, based on ROI feature.

**Table 11.8**: Average of the best matching results based on diagnosis for MI-GLCMs for pairs of SB mammograms.

| Diagnosis | Average $A_{ROC}$ | Average $C_{fb}$ | Average $\kappa$ |
|---|---|---|---|
| | Best Results | | |
| Benign | 0.68±0.16 | 0.35±0.43 | 0.22±0.23 |
| Indeterminate | 0.48±0.17 | 0.01±0.03 | 0.01±0.01 |
| Malignant | 0.97±0.01 | 0.86±0.14 | 0.81±0.15 |
| | Best Mass Results | | |
| Benign | 0.69±0.17 | 0.44±0.38 | 0.26±0.23 |
| Indeterminate | 0.49±0.02 | 0.00±0.00 | 0.00±0.00 |
| Malignant | 0.97±0.01 | 0.86±0.14 | 0.81±0.15 |
| | Best Calcification Results | | |
| Benign | 0.57±0.10 | -0.18±0.46 | 0.02±0.03 |
| Indeterminate | 0.48±0.22 | 0.02±0.03 | 0.01±0.02 |

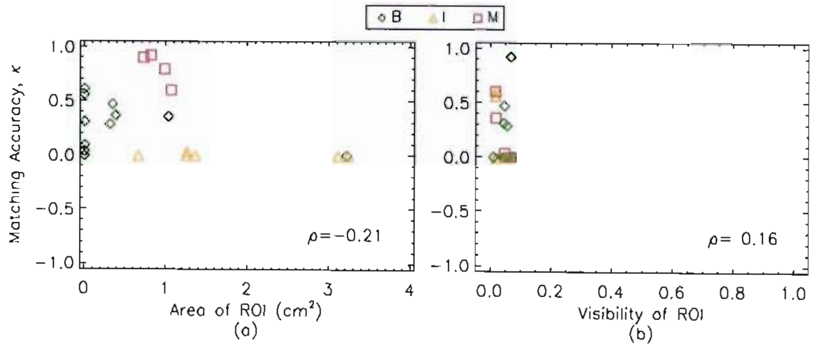


**Figure 11.25**: Best matching accuracy as functions of **(a)** ROI area and **(b)** ROI visibility for each mammogram pair for MI-GLCMs.

## 11.11 Overall Matching Results

### 11.11.1 Results

**Figure 11.26** shows the best matching results for each pair of SB mammograms. The highest $\kappa$-value was selected over all matching parameters, for each pair of SB mammograms.

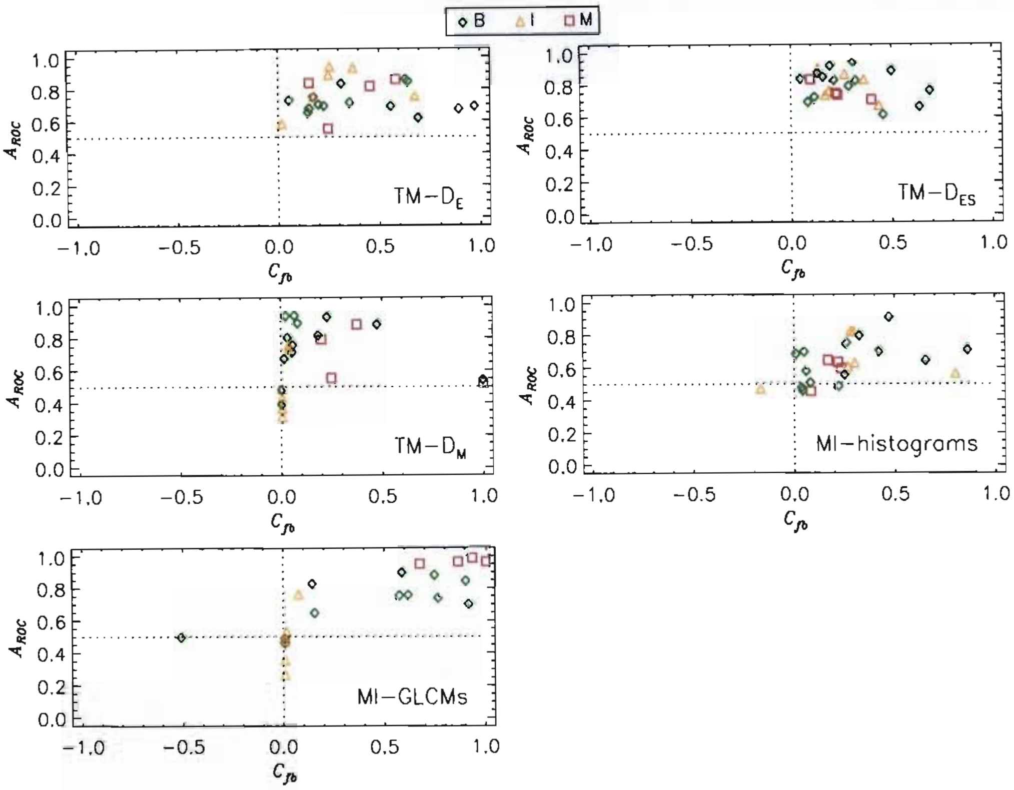


**Figure 11.26**: Scatter plot of $A_{ROC}$ *vs.* $C_{fb}$ with the best matching accuracies for all matching methods and for each SB mammogram pair. The points are colour-coded according to the diagnosis (B - benign, I - indeterminate, M - malignant).

The $A_{ROC}$- and $C_{fb}$-values are only clustered for TM-$D_{ES}$ and are generally quite scattered for the remaining four methods (TM-$D_E$, TM-$D_M$, MI-histograms, MI-GLCMs). Matching accuracies are only very high for the malignant masses matched with MI-GLCMs, otherwise, matching accuracies are very poor, with low $A_{ROC}$-values and low $C_{fb}$-values. The average of the best matching accuracies is summarised in **Table 11.9**, for each method. Results are generally lower than the equivalent values for the mosaic images and for the standard mammograms, and this is most probably because the features in the SB mammograms all have very low visibilities.

Matching accuracies for the masses are comparable to the results for the standard mammograms, but the matching accuracies for the calcifications are very low.

The combination of matching parameters that yielded the maximum $\kappa$-values is listed for each image pair in **Table A.5** for TM-matching and in **Table A.6** for MI-matching.

**Table 11.9**: Average of the best matching accuracies for all matching methods applied to SB mammograms.

| Method | $A_{ROC}$ | $C_{fb}$ | $\kappa$ |
|---|---|---|---|
| | Overall Results | | |
| TM-$D_E$ | 0.75±0.11 | 0.38±0.26 | 0.20±0.14 |
| TM-$D_{ES}$ | 0.79±0.09 | 0.27±0.17 | 0.15±0.09 |
| TM-$D_M$ | 0.67±0.20 | 0.21±0.33 | 0.06±0.09 |
| MI-histograms | 0.63±0.13 | 0.25±0.25 | 0.09±0.11 |
| MI-GLCMs | 0.68±0.21 | 0.35±0.43 | 0.27±0.32 |
| | Masses | | |
| TM-$D_E$ | 0.75±0.10 | 0.42±0.27 | 0.21±0.15 |
| TM-$D_{ES}$ | 0.79±0.09 | 0.30±0.18 | 0.17±0.10 |
| TM-$D_M$ | 0.76±0.14 | 0.28±0.35 | 0.08±0.09 |
| MI-histograms | 0.65±0.13 | 0.27±0.22 | 0.10±0.12 |
| MI-GLCMs | 0.73±0.20 | 0.48±0.40 | 0.35±0.33 |
| | Calcifications | | |
| TM-$D_E$ | 0.76±0.13 | 0.27±0.23 | 0.17±0.14 |
| TM-$D_{ES}$ | 0.80±0.09 | 0.18±0.13 | 0.09±0.04 |
| TM-$D_M$ | 0.41±0.07 | 0.00±0.00 | 0.00±0.00 |
| MI-histograms | 0.60±0.14 | 0.20±0.34 | 0.05±0.08 |
| MI-GLCMs | 0.51±0.18 | -0.05±0.23 | 0.01±0.02 |

### 11.11.2 Statistical Significance Analysis

The results of performing a paired $t$-test analysis (§8.10 on page 103) on the distribution of the best $\kappa$-values for each method are presented in **Table 11.10**.

For a significance level of 0.05, the average values of $\kappa$ for TM-$D_E$ and MI-GLCMs are not statistically different. The average values of TM-$D_{ES}$ and TM-$D_E$ as well as the average values of TM-$D_M$ and MI-histograms are also not statistically different.

**Table 11.10**: Results of significance ($t$-test) analysis for best matching accuracies from pairs of SB mammograms for the various matching methods.

| Method | TM-$D_{ES}$ | | TM-$D_M$ | | MI-histograms | | MI-GLCMs | |
|---|---|---|---|---|---|---|---|---|
| | t-value | p-value | t-value | p-value | t-value | p-value | t-value | p-value |
| TM-$D_E$ | 1.41 | 0.17 | 3.88 | 0.00 | 2.90 | 0.01 | -0.95 | 0.35 |
| TM-$D_{ES}$ | | | 3.18 | 0.00 | 1.94 | 0.06 | -1.73 | 0.09 |
| TM-$D_M$ | | | | | -0.90 | 0.37 | -2.99 | 0.01 |
| MI-histograms | | | | | | | -2.55 | 0.02 |

### 11.11.3 TM-matching

TM-$D_E$ and TM-$D_{ES}$ had the best matching accuracies with 1 mass (out of 25) not matched for TM-$D_E$ and all ROIs matched for TM-$D_{ES}$. Matching accuracies for TM-$D_M$ were very low, and no calcifications (out of 9) were matched.

The overall results are considerable lower than for the mosaic images and for the standard mammograms. This is most likely because the visibilities of all ROIs are very low. There was also no correlation between matching accuracy and ROI area or ROI visibility for the TM-matching methods.

Patients subjected to stereotactic biopsies generally have smaller, impalpable masses, most of which are poorly visualised on standard mammograms. The visibility of the masses in the SB mammograms is most likely low because these represent more difficult cases.

The dependence of matching accuracy on bit-depth was influenced by the sampling window size for all methods and there was no consistent pattern to the dependence of matching accuracy on bit-depth. Matching accuracy was independent of *d* for all methods.

### 11.11.4 MI-matching

The matching results for MI-histograms were very low with 4 masses (out of 25) and 1 calcification (out of 9) not matched. Matching accuracy improved with decreasing bit-depth and decreasing *nbins*.

The matching accuracy for the malignant masses was very high for MI-GLCMs, but there were still 6 masses (out of 25) and 4 calcifications (out of 9) not matched, the most non-matches of all five matching methods. The results for the remaining ROI-types were very low. Matching accuracy improved with increasing bit-depth and decreasing *d*.

There was no correlation between matching accuracy and ROI area or ROI visibility for both MI-histograms and MI-GLCMs.

## 11.12 Results of ACF Analysis

The autocorrelation function was used to determine the characteristic scale width and scale height of the textural features in each reference image. The ACF results were compared to the optimal sampling window sizes that were obtained from the matching analysis.

The ACF results appear in **Table 11.12** and were generated at different bit-depths because the characteristic scales of textural features are expected to change as bit-depth changes. Only a few ROIs demonstrated some variation of the ACF-widths and ACF-heights as bit-depth changed, but most ROIs had ACF-results that were independent of bit-depth. This is most likely because most of the ROIs are very small, and the features therein would not vary significantly as bit-depth changed.

The optimal sampling window sizes, for TM- and MI-matching appear in **Table 11.11**. The results of a linear Pearson correlation analysis (§8.10 on page 103) between the ACF-results and the optimal sampling window sizes are detailed in **Table 11.13**. The correlation results indicate that the ACF-results are correlated with the optimal sampling window sizes. Correlation improves as bit-depth decreases, for all methods, except MI-histograms where correlation improves as bit-depth increases.

The correlation results for ACF-widths are similar to those for the ACF-heights, which means that the scale widths and scale heights of the textures in the mammogram ROIs are similar and it is appropriate to use square sampling windows.

**Table 11.11**: Optimal sampling window sizes for SB mammograms with $D_E$, $D_{ES}$, $D_M$, histograms (GLH) and GLCMs (G).

| Image | 5-bits | | | | | 6-bits | | | | | 7-bits | | | | | 8-bits | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $D_E$ | $D_{ES}$ | $D_M$ | GLH | G | $D_E$ | $D_{ES}$ | $D_M$ | H | G | $D_E$ | $D_{ES}$ | $D_M$ | GLH | G | $D_E$ | $D_{ES}$ | $D_M$ | GLH | G |
| S1S0 | 96 | 32 | 96 | 128 | 64 | 64 | 64 | 16 | 128 | 32 | 80 | 112 | 96 | 128 | 32 | 144 | 96 | 128 | 112 | 32 |
| S1S2 | 96 | 96 | 96 | 128 | 128 | 16 | 48 | 144 | 128 | 32 | 16 | 144 | 32 | 128 | 32 | 112 | 48 | 144 | 96 | 32 |
| S4S3 | 64 | 16 | 48 | 32 | 16 | 48 | 32 | 96 | 32 | 16 | 48 | 32 | 80 | 32 | 16 | 16 | 16 | 80 | 32 | 16 |
| S4S5 | 64 | 16 | 96 | 16 | 16 | 64 | 32 | 64 | 32 | 16 | 64 | 32 | 16 | 32 | 16 | 32 | 32 | 48 | 32 | 16 |
| S7S6 | 64 | 80 | 128 | 64 | 32 | 48 | 48 | 48 | 32 | 32 | 16 | 48 | 48 | 32 | 32 | 48 | 96 | 128 | 32 | 32 |
| S7S8 | 80 | 48 | 64 | 16 | 32 | 80 | 64 | 64 | 32 | 32 | 64 | 64 | 48 | 64 | 32 | 32 | 80 | 96 | 48 | 32 |
| S10S9 | 160 | 176 | 176 | 176 | 16 | 96 | 32 | 64 | 80 | 32 | 128 | 128 | 128 | 80 | 32 | 16 | 16 | 48 | 80 | 32 |
| S10S11 | 176 | 32 | 64 | 160 | 32 | 96 | 32 | 48 | 128 | 48 | 16 | 32 | 80 | 96 | 48 | 16 | 16 | 48 | 96 | 64 |
| S13S12 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 |
| S13S14 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 |
| S16S15 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 |
| S16S17 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 |
| S19S18 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 |
| S19S20 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 |
| S22S21 | 272 | 176 | 48 | 288 | 16 | 32 | 32 | 176 | 288 | 16 | 16 | 16 | 256 | 288 | 16 | 16 | 16 | 48 | 288 | 16 |
| S22S23 | 176 | 176 | 176 | 256 | 48 | 80 | 288 | 288 | 256 | 48 | 16 | 112 | 112 | 256 | 48 | 16 | 16 | 176 | 256 | 48 |
| S25S24 | 16 | 16 | 80 | 16 | 16 | 48 | 48 | 16 | 32 | 16 | 16 | 16 | 48 | 16 | 16 | 80 | 16 | 80 | 16 | 16 |
| S25S26 | 32 | 96 | 16 | 80 | 48 | 80 | 64 | 64 | 48 | 16 | 96 | 96 | 96 | 16 | 16 | 16 | 96 | 96 | 16 | 16 |
| S28S27 | 96 | 64 | 96 | 32 | 16 | 96 | 96 | 96 | 96 | 16 | 96 | 96 | 80 | 32 | 32 | 96 | 96 | 96 | 32 | 16 |
| S28S29 | 16 | 16 | 32 | 16 | 48 | 16 | 32 | 16 | 32 | 96 | 16 | 32 | 96 | 32 | 16 | 128 | 64 | 128 | 32 | 16 |
| S31S30 | 32 | 32 | 16 | 16 | 16 | 32 | 32 | 16 | 96 | 16 | 16 | 16 | 96 | 96 | 16 | 16 | 16 | 64 | 96 | 16 |
| S31S32 | 96 | 48 | 64 | 16 | 16 | 32 | 96 | 48 | 96 | 16 | 16 | 16 | 16 | 96 | 16 | 16 | 16 | 48 | 48 | 16 |
| S34S33 | 96 | 64 | 144 | 48 | 16 | 16 | 96 | 32 | 144 | 16 | 128 | 144 | 48 | 144 | 16 | 144 | 112 | 112 | 144 | 16 |
| S34S35 | 96 | 112 | 128 | 64 | 16 | 32 | 16 | 32 | 48 | 16 | 16 | 48 | 16 | 48 | 16 | 80 | 128 | 128 | 64 | 16 |

**Table 11.12**: Results of ACF analysis for SB mammograms.

| Reference | 5-bits | | 6-bits | | 7-bits | | 8-bits | |
|---|---|---|---|---|---|---|---|---|
| | $w_{ACF}$ | $h_{ACF}$ | $w_{ACF}$ | $h_{ACF}$ | $w_{ACF}$ | $h_{ACF}$ | $w_{ACF}$ | $h_{ACF}$ |
| S1 | 88 | 114 | 76 | 67 | 54 | 66 | 62 | 65 |
| S4 | 37 | 54 | 40 | 51 | 41 | 49 | 40 | 49 |
| S7 | 101 | 82 | 100 | 77 | 98 | 77 | 98 | 78 |
| S10 | 77 | 68 | 62 | 65 | 62 | 65 | 61 | 66 |
| S13 | 8 | 7 | 8 | 8 | 8 | 8 | 8 | 8 |
| S16 | 12 | 7 | 11 | 11 | 11 | 11 | 11 | 11 |
| S19 | 13 | 12 | 12 | 15 | 10 | 12 | 10 | 12 |
| S22 | 132 | 287 | 266 | 287 | 270 | 287 | 270 | 287 |
| S25 | 48 | 80 | 34 | 46 | 44 | 50 | 44 | 49 |
| S28 | 43 | 53 | 50 | 57 | 51 | 73 | 51 | 73 |
| S31 | 44 | 39 | 47 | 38 | 47 | 40 | 46 | 40 |
| S34 | 64 | 53 | 64 | 60 | 71 | 62 | 71 | 62 |

**Table 11.13**: Correlation coefficients between ACF results and optimal sampling window sizes for SB mammograms. The linear Pearson correlation coefficient was computed between the optimal sampling window sizes and $w_{ACF}$ ($\rho_w$) and $h_{ACF}$ ($\rho_h$). Results indicate that the ACF results are correlated with the optimal window sizes for some matching methods.

| Method | 5-bits | | 6-bits | | 7-bits | | 8-bits | |
|---|---|---|---|---|---|---|---|---|
| | $\rho_w$ | $\rho_h$ | $\rho_w$ | $\rho_h$ | $\rho_w$ | $\rho_h$ | $\rho_w$ | $\rho_h$ |
| TM-$D_E$ | 0.80 | 0.78 | 0.29 | 0.29 | -0.05 | -0.04 | -0.03 | -0.01 |
| TM-$D_{ES}$ | 0.75 | 0.75 | 0.63 | 0.63 | 0.24 | 0.26 | 0.04 | -0.01 |
| TM-$D_M$ | 0.65 | 0.45 | 0.84 | 0.86 | 0.73 | 0.76 | 0.49 | 0.47 |
| MI-histograms | 0.80 | 0.87 | 0.87 | 0.87 | 0.87 | 0.87 | 0.90 | 0.90 |
| MI-GLCMs | 0.40 | 0.32 | 0.24 | 0.23 | 0.43 | 0.43 | 0.38 | 0.36 |

## 11.13 Multiple Reference Regions

The effect of using multiple reference regions was investigated by dividing each single reference region into four smaller regions, as depicted in **Figure 10.29**. The resulting matching map from each small reference region was then averaged, in an attempt to highlight different regions of the test image. The TM-matching and MI-matching algorithms were applied to the 24 pairs of SB mammograms.

**Figure 11.27** shows the best matching accuracies, κ, for each of the 24 pairs of SB mammograms.

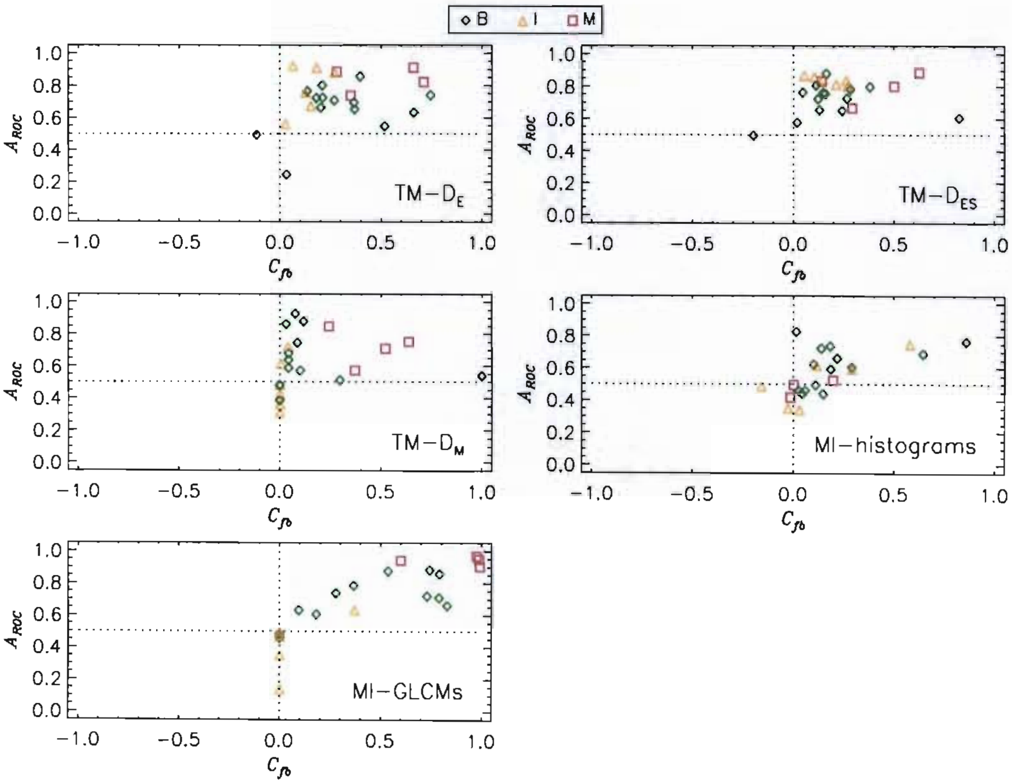


**Figure 11.27**: Summary of best matches for multiple reference regions (SB mammograms).

The $A_{ROC}$- and $C_{fb}$-values for each method are quite scattered. The matching accuracies are quite low, generally as a result of low contrast values. TM-$D_E$ performs best overall, but MI-GLCMs performs best at matching the malignant masses. The matching accuracy of TM-$D_{ES}$ is comparable to that of MI-GLCMs and the matching accuracy of TM-$D_M$ is comparable to that of MI-histograms.

Average matching accuracies are listed in **Table 11.14** for single and multiple reference regions. These results show that using multiple reference regions does not improve overall matching accuracy.

**Table 11.14**: Average of the best matching accuracies for single and multiple reference regions for all matching methods applied to pairs of SB mammograms. Results show that using multiple reference regions does not improve the overall matching accuracy.

| Method | Single Reference Region | | | Multiple Reference Regions | | |
|---|---|---|---|---|---|---|
| | Overall | | | | | |
| | $A_{ROC}$ | $C_{fb}$ | $\kappa$ | $A_{ROC}$ | $C_{fb}$ | $\kappa$ |
| TM-$D_E$ | 0.75±0.11 | 0.38±0.26 | 0.20±0.14 | 0.73±0.16 | 0.29±0.23 | 0.15±0.14 |
| TM-$D_{ES}$ | 0.79±0.09 | 0.27±0.17 | 0.15±0.09 | 0.76±0.10 | 0.22±0.21 | 0.12±0.11 |
| TM-$D_M$ | 0.67±0.20 | 0.21±0.33 | 0.06±0.09 | 0.61±0.17 | 0.19±0.30 | 0.05±0.08 |
| MI-histograms | 0.63±0.13 | 0.25±0.25 | 0.09±0.11 | 0.57±0.13 | 0.17±0.23 | 0.06±0.12 |
| MI-GLCMs | 0.68±0.21 | 0.35±0.43 | 0.27±0.32 | 0.67±0.22 | 0.39±0.38 | 0.26±0.31 |
| | Masses | | | | | |
| TM-$D_E$ | 0.75±0.10 | 0.42±0.27 | 0.21±0.15 | 0.74±0.09 | 0.36±0.21 | 0.18±0.15 |
| TM-$D_{ES}$ | 0.79±0.09 | 0.30±0.18 | 0.17±0.10 | 0.75±0.10 | 0.26±0.22 | 0.14±0.11 |
| TM-$D_M$ | 0.76±0.14 | 0.28±0.35 | 0.08±0.09 | 0.68±0.14 | 0.26±0.33 | 0.07±0.09 |
| MI-histograms | 0.65±0.13 | 0.27±0.22 | 0.10±0.12 | 0.59±0.12 | 0.22±0.24 | 0.08±0.13 |
| MI-GLCMs | 0.73±0.20 | 0.48±0.40 | 0.35±0.33 | 0.72±0.19 | 0.48±0.40 | 0.33±0.32 |
| | Calcifications | | | | | |
| TM-$D_E$ | 0.76±0.13 | 0.27±0.23 | 0.17±0.14 | 0.67±0.28 | 0.08±0.13 | 0.07±0.09 |
| TM-$D_{ES}$ | 0.80±0.09 | 0.18±0.13 | 0.09±0.04 | 0.79±0.11 | 0.10±0.07 | 0.06±0.05 |
| TM-$D_M$ | 0.41±0.07 | 0.00±0.00 | 0.00±0.00 | 0.41±0.07 | 0.00±0.00 | 0.00±0.00 |
| MI-histograms | 0.60±0.14 | 0.20±0.34 | 0.05±0.08 | 0.52±0.18 | 0.00±0.09 | 0.00±0.01 |
| MI-GLCMs | 0.51±0.18 | -0.05±0.23 | 0.01±0.02 | 0.54±0.25 | 0.11±0.14 | 0.03±0.04 |

## 11.14 Point Analysis for Improving Localisation Accuracy

The first stage of the evaluation of the matching results resulted in very few very good matches that could be analysed for selection of points. For the point analysis to be effective, it is necessary to have a good contrast as well as a good value for $A_{ROC}$. Further to that, both stereoscopic views must have good matches with the scout view. Only 3 cases satisfied these requirements for TM-$D_E$ and only 4 cases satisfied the requirements for MI-GLCMs. No other methods had sufficiently accurate matches. For stereotactic biopsies, only shifts in the $x$-direction are relevant, so the $x$-coordinate of the centroid of the ROI, $\mathrm{ROI}_x$, drawn by the radiologist and the $x$-coordinate of the pixel with the maximum intensity in the matching map (indicating the best match), $m_x$ were determined for each view. The results are shown in **Figure 11.28** and **Figure 11.29** and detailed in **Table 11.15**.

**Table 11.15**: Results from point analysis applied to best matching results of selected SB mammograms.

| Label | $w$ | $\mathrm{ROI}_x$ | $m_x$ | Label | $w$ | $\mathrm{ROI}_x$ | $m_x$ |
|---|---|---|---|---|---|---|---|
| TM-$D_E$ | | | | | | | |
| S10S9 | 16 | 133 | 88 | S10S11 | 16 | 84 | 67 |
| S22S21 | 16 | 94 | 10 | S22S23 | 16 | 59 | 61 |
| S28S27 | 16 | 84 | 37 | S28S29 | 16 | 37 | 129 |
| MI-GLCM | | | | | | | |
| S1S0 | 32 | 97 | 140 | S1S2 | 32 | 58 | 62 |
| S4S3 | 16 | 98 | 39 | S4S5 | 16 | 28 | 5 |
| S7S6 | 32 | 84 | 195 | S7S8 | 32 | 49 | 47 |
| S10S9 | 32 | 131 | 133 | S10S11 | 32 | 83 | 85 |

All results, except one, are very poor. In most cases the maximum of the matching map does not even occur within the ROI. There is only one case (S10S9 for MI-GLCMs) where the positions differ by 2 pixels for each stereoscopic view, resulting in the same shift in the matching map. However, since the matching map was generated for a sampling window size of 16 pixels, the 2 pixel-difference corresponds to 16.3 mm, which is significant on the scale of localisation accuracy.

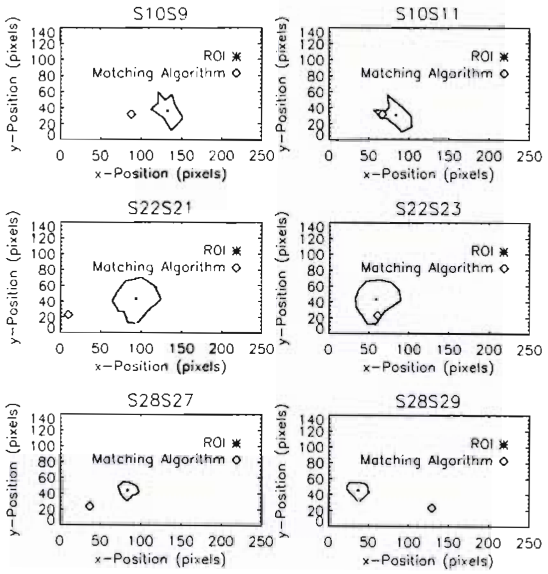


**Figure 11.28**: Results of point analysis for TM-$D_E$. The ROI drawn by the radiologist is shown together with the centroid of the ROI. The position of the pixel with the maximum intensity in the matching map is also shown. In most cases, the latter position does not even occur within the ROI.

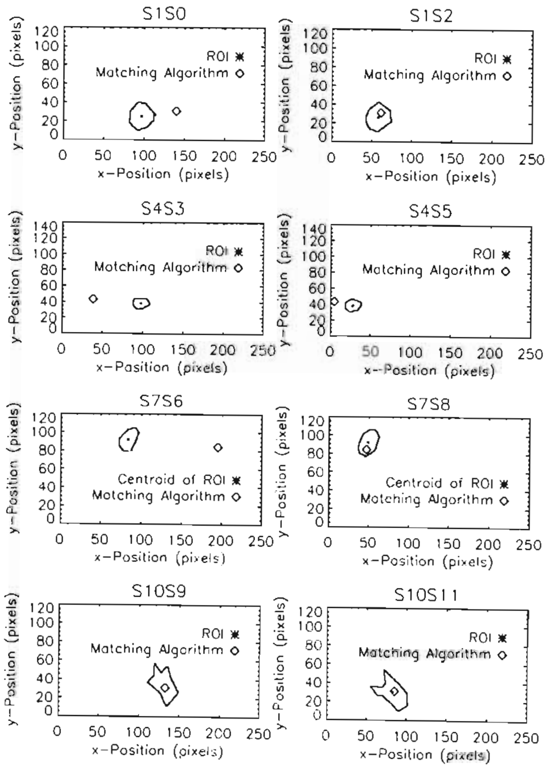


**Figure 11.29**: Results of point analysis for MI-GLCMs. The ROI drawn by the radiologist is shown together with the centroid of the ROI. The position of the pixel with the maximum intensity in the matching map is also shown. In some cases, the latter position does not even occur within the ROI.

## 11.15 Summary of Effects of Matching Parameters

### 11.15.1 Effect of $w_{step}$ on Matching Accuracy

A fixed value of 4 pixels was used for $w_{step}$.

### 11.15.2 Effect of $w$ on Matching Accuracy

Matching accuracy varies with $w$ and the optimal sampling window sizes are correlated with the ACF results, for all methods.

### 11.15.3 Effect of *nbits* on Matching Accuracy

There was no clear dependence of matching accuracy on *nbits* for TM-matching, while matching accuracy improved with decreasing *nbits* for MI-histograms and improved with increasing *nbits* for MI-GLCMs.

### 11.15.4 Effect of $d$ on Matching Accuracy

Matching accuracy was independent of $d$ for all TM-matching methods and improved with decreasing $d$ for MI-GLCMs.

### 11.15.5 Effect of *nbins* on Matching Accuracy

Matching accuracy improved with decreasing *nbins*.

### 11.15.6 Sensitivity of Matching Methods to Choice of Parameter Values

**Figure 11.30** shows all pairs of $A_{ROC}$ and $C_{fb}$, across all parameters, for each matching method. An examination of this data gives an indication of how sensitive each method is to the choice of parameter values. The sensitivity of the choice of parameter values was quantified by counting the number of pairs of $A_{ROC}$ and $C_{fb}$ that fell into different zones, as detailed in **Figure 9.22** (page 140). The results of this analysis are summarised in **Table 11.16** and shows that MI-GLCMs is most accurate for masses with the most points falling into zone 1, while TM-$D_E$ is most accurate for calcifications with the most points falling into zone 2 (there were no points in zone 1 for the calcifications). However, both TM-$D_E$ and MI-GLCMs have a wide range of data, for the masses, which indicate that these methods are very sensitive to the choice of parameter values used for the matching. Results for the calcifications are scattered for all methods.

### 11.15.7 Comparison of TM-Matching and MI-Matching for SB Mammograms

Matching results for the stereotactic biopsy mammograms are generally worse than those obtained for the mosaic images and pairs of mammograms.

TM-$D_E$ and TM-$D_{ES}$ have similar matching results for all ROIs irrespective of diagnosis or whether the ROI contained a mass or a calcification. The matching results for TM-$D_M$ and

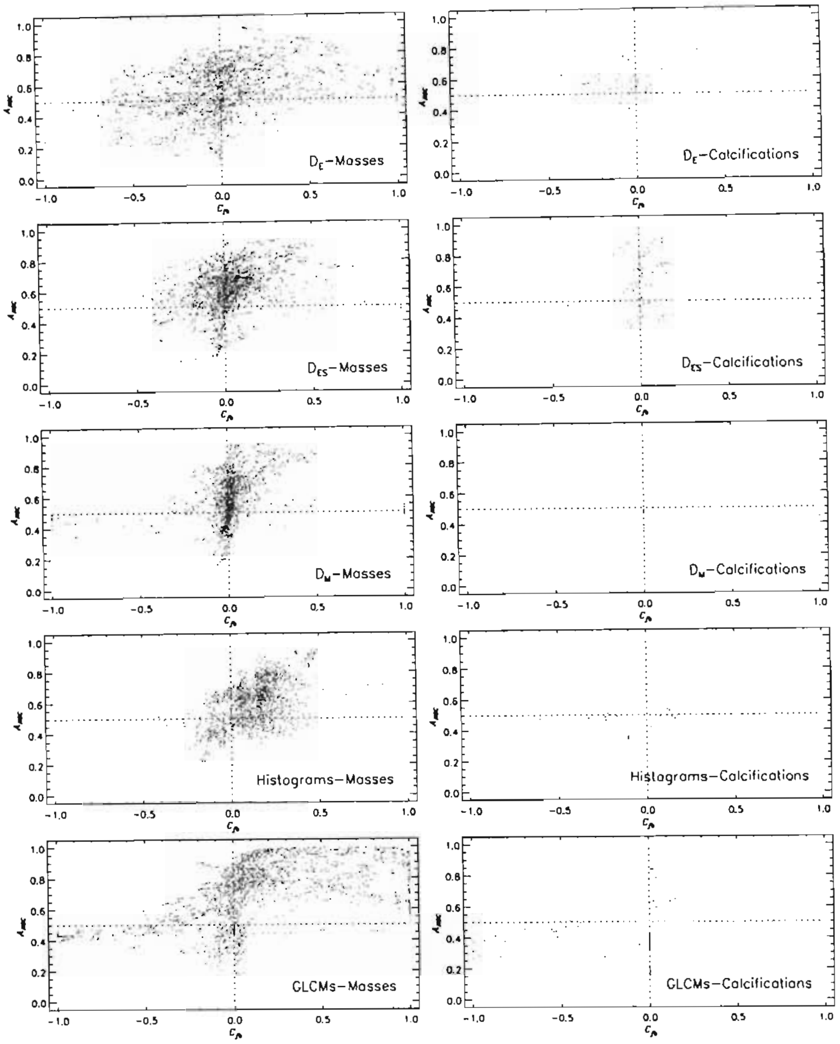


**Figure 11.30**: Scatter plot of $A_{ROC}$ *vs.* $C_{fb}$ for all combinations of matching parameters to demonstrate sensitivity of choice of matching parameter values for each method applied to SB mammograms.

MI-histograms are generally very poor. MI-GLCMs only performs very well with matching the malignant masses, but performs very poorly with matching the calcifications. Also, the MI-GLCM algorithm did not match 6 image pairs because the implementation of the GLCM appears to be sensitive to grey-level shifts.

For TM-matching, the dependence of matching accuracy on bit-depth was influenced by the size of the sampling window and there was no apparent dependence of matching accuracy on $d$. For MI-matching, matching accuracy improved with increasing bit-depth and decreasing $d$. All these patterns of behaviour are consistent with the results obtained for the mosaic images and the standard mammograms.

The results of the analysis of the sensitivity of each method to the choice of matching parameter values was similar to those for the standard mammograms with TM-$D_E$ and MI-GLCMs being most accurate, but most sensitive to the choice of matching parameter values.

Using multiple reference views did not improve the matching accuracy of any method.

**Table 11.16**: Results of zone analysis for pairs of SB mammograms to demonstrate sensitivity of choice of matching parameters for each method. An ideal method would have a high percentage of points in Zone1 and none in Zone 5.

| Method | Zone 1 (%) | Zone 2 (%) | Zone 3 (%) | Zone 4 (%) | Zone 5 (%) |
|---|---|---|---|---|---|
| Masses | | | | | |
| TM-$D_E$ | 0.000 | 8.94 | 17.3 | 52.1 | 21.7 |
| TM-$D_{ES}$ | 0.000 | 2.15 | 3.38 | 77.2 | 17.3 |
| TM-$D_M$ | 0.000 | 0.000 | 9.38 | 61.0 | 29.6 |
| MI-histograms | 0.000 | 0.284 | 2.27 | 84.0 | 13.4 |
| MI-GLCMs | 22.7 | 20.0 | 12.3 | 31.8 | 13.2 |
| Calcifications | | | | | |
| TM-$D_E$ | 0.000 | 9.04 | 0.000 | 70.1 | 20.9 |
| TM-$D_{ES}$ | 0.000 | 0.00 | 0.000 | 78.4 | 21.6 |
| TM-$D_M$ | 0.000 | 0.000 | 0.000 | 0.000 | 100. |
| MI-histograms | 0.000 | 0.000 | 34.0 | 42.6 | 23.4 |
| MI-GLCMs | 0.000 | 0.000 | 0.000 | 52.3 | 47.7 |

Since the matching results were so poor, very few cases satisfied the conditions of the second stage of the evaluation. A total of 7 pairs were analysed for point correspondence between the 3 stereoscopic views. Results were generally very poor. In most cases, the position of the pixel with the maximum intensity (indicating the best match) did not occur within the ROI drawn by the radiologist.

These results therefore indicate that the matching algorithms are not sufficiently sensitive to be used to improve the localisation accuracy for stereotactic breast biopsies.

There are two possible reasons for the poor results: quality of images and quality of ground truth data. The contrasts between the ROIs and the surrounding breast tissue in the SB mammograms were very low and there were differences in average grey-level between images in a triplet. These problems can be addressed equalising average grey-levels or by applying a contrast enhancement algorithm to the images before the matching analysis.

The second reason is the quality of the ground truth data. Since only one radiologist, retrospectively, marked the borders, it is not possible to obtain any consensus regarding the accuracy of the regions marked, as was done for the breast border analysis.

## 11.16 Summary

Stereotactic biopsies use two mammographic views of the breast, from slightly different angles to determine the depth of the abnormality to be biopsied. The sampling of the correct tissue critically depends on the point to be sampled identified correctly in each stereoscopic view. The TM- and MI-matching algorithms were applied to stereotactic biopsy mammograms to investigate whether the algorithms could be used to improve localisation accuracy, by identifying the same point in all mammographic views from using only a single view as a reference. Results were evaluated in two stages. The first stage of the analysis was identical to that for the mosaic images and the standard mammograms. The second stage only examined the matched regions to determine whether the same point had been identified in all views.

The results of the first evaluation stage were very poor, generally much poorer than the results for the mosaic images and the standard mammograms. Consequently, only 7 pairs satisfied the criteria for the second evaluation stage. However, only one case had the maximum of the matching map (indicating the best match) close to the centroid of the ROI drawn by the radiologist.

Results were most likely poor because the ROIs had a low contrast in comparison to the surrounding breast tissue, which can be addressed by applying a contrast enhancement algorithm to the images before the matching analysis. Generally contrast is very good in mammograms. However, the stereotactic biopsy mammograms represent cases that are generally more difficult than those selected for testing the algorithms on standard mammograms. Another area for concern is the quality of the ground truth data. Since only one radiologist, retrospectively, identified the ROIs, it is impossible to obtain consensus if there is any discrepancy.

These results therefore indicate that the matching algorithms are not sufficiently sensitive to be used to improve the localisation accuracy for stereotactic breast biopsies.

# Chapter 12

# Conclusions and Future Work

## 12.1 Summary of Thesis

### 12.1.1 Chapter 1: Introduction

Breast cancer is the most common form of cancer among women, worldwide. Successful treatment relies on early detection. Currently, mammography is the most widely available method of detecting breast cancer, but suffers from the problem that radiologists, in their visual interpretation of the resulting mammograms, sometimes miss the subtle signs of breast cancer. Computer-aided diagnosis was introduced as a means to assist radiologists by consistently highlighting abnormal regions for the radiologist to further analyse. The shortcomings of CAD i.e. high false-positive rates, and the current lack of systems that fully utilise all available mammographic information, form the basis of the motivation for this research.

### 12.1.2 Chapter 2: Image Processing Techniques in Mammographic CAD

The detection accuracy of CAD-systems for microcalcification clusters is higher than that for masses. Due to the varied appearance of masses on mammograms, many image-processing methods have been used to detect and classify masses as malignant or benign. Methods of segmenting suspicious regions from the surrounding breast tissue include removal of normal tissue structures, contrast enhancement, the Wavelet transform, active contours, symmetry between left and right breasts and changes in the breast over time. Once these suspicious regions have been identified, morphological, geometrical and textural information are extracted, which is used to reduce false-positives, i.e. separate any normal tissue regions from truly abnormal regions. This reduced set of regions is then passed to a classifier like linear discriminant analysis, a binary classification tree or an artificial neural network that performs the final distinction between whether the region is benign or malignant.

### 12.1.3 Chapter 3: Multiple Mammographic-View Analysis

One of the concerns regarding current CAD algorithms is the high false-positive rate, which arises as a consequence of the requirement that the algorithm have a high sensitivity. Radiologists use all available mammographic views of a single breast for diagnosis, but very little research has been done into the use of multiple single-breast mammograms for confirmation of abnormalities and false-positive reduction in CAD-systems. This study is concerned with the development of an analysis technique that uses both standard mammographic views (CC and MLO) of the same breast with the aim of confirming the presence of abnormalities and ultimately to eliminate false-positives. The algorithms developed employ standard image processing methods based on texture analysis. Regions of interest or templates are identified in a reference image and are compared to a test image to find a match. Texture is quantified with grey-level co-occurrence matrices, texture measures and grey-level histograms and is compared using distance metrics and mutual information as similarity metrics. The output of the algorithm takes the form of a matching map, which has maximal intensity at optimal match. All algorithms developed can be easily incorporated into existing CAD-systems.

### 12.1.4 Chapter 4: Pre-processing Mammograms

A novel, simple method of finding the breast edge using areas enclosed by iso-intensity contours that improves on traditional thresholding methods for segmentation, by incorporating spatial information into the segmentation, was presented. The method does not rely on models of the breast or background and borders. Results were evaluated by comparison to breast borders drawn by three radiologists in their normal working environment. The effect of various pre-processing methods on the accuracy of the automated borders was investigated. Results were generally good for those images containing clear breast edges. It was found that smoothing with a Lorentzian kernel as a pre-processing method, with the width automatically determined for each mammogram worked acceptably well for those with clear breast edges. The best results for mammograms with clear breast edges was 3.0 mm±0.3 mm.

The semi-automatic algorithm used to remove the pectoral muscle was based on the work of Karssemeijer [1998]. The arc method was used to define an annular, reduced search region, by using the position of the mass in one standard mammographic view and the positions of the nipple in both views, following the work of Paquerault et al. [2002]. The overall result of the three pre-processing steps is a significantly reduced region in the test image, which is searched for a match.

### 12.1.5 Chapter 5: Quantification of Image Texture

The analysis of texture forms an important part of image processing. Textural information must be quantified before images can be compared. The main quantification methods are statistical, structural and spectral. Statistical methods (like grey-level histograms) only incorporate the number of grey-levels in an image. Structural methods (like grey-level co-occurrence matrices) incorporate numbers of grey-levels and information about the location of the pixels in the image. Spectral methods (like the autocorrelation function) examine the periodic and scale properties of an image.

### 12.1.6 Chapter 6: Similarity Metrics

Similarity metrics quantify how similar two quantities are to each other. The similarity metrics that are used in this research are: Euclidean distance, Mahalanobis distance and mutual information. The Euclidean distance similarity metric suffers from the problem that a single input can dominate the final distance value, if this input is considerably larger than the other inputs. Standardising the inputs to a normal distribution with a variance of 1 and a mean of 0 solves this problem. However, standardisation does not consider correlations between inputs. Hence, the Mahalanobis distance similarity metric is often preferred.

Mutual information has been shown to be a robust similarity metric in image registration problems, but has also been applied to template matching, feature selection and segmentation problems. Mutual information measures a general statistical dependence between inputs compared to e.g. linear correlation coefficient.

### 12.1.7 Chapter 7: Evaluation of Results

Receiver operating characteristic (ROC) analysis is a standard method of evaluating and ranking medical diagnostic tests. To perform any evaluation, the 'truth' must be known so that it can be compared with the output of the test. The evaluation and ranking of CAD algorithms is analogous to that of a standard medical diagnostic test and is therefore perfectly suited to the use of ROC analysis.

For this study, the accuracy of the matching maps is evaluated by using ROC analysis and contrast. The area under the ROC curve, $A_{ROC}$, gives a measure of how much of the ROI has been matched while contrast, $C_{fb}$, gives a measure of how well the matched ROI stands out from the background in the matching map.

### 12.1.8 Chapter 8: Matching Methods

TM-matching uses GLCM-based texture measures to quantify texture and distance similarity metrics to compare textures between images to determine similarity. MI-matching uses grey-level histograms and GLCMs, to quantify texture and mutual information as a similarity metric to compare textures between images to determine similarity. Matching accuracy, $\kappa$, is defined as a combination of $A_{ROC}$ and $C_{fb}$. Matching parameters like sampling window size ($w$), sampling window step size ($w_{step}$), bit-depth ($nbits$), distance in the GLCM ($d$) and the number of histogram bins ($nbins$) are varied to investigate their effect on matching accuracy. The autocorrelation function was investigated as a possible independent method of extracting an optimal sampling window size.

The image processing components used in this study are summarised in **Table 12.1**.

**Table 12.1**: Summary of texture-based image processing methods used in this study

| | TM-matching | MI-matching |
|---|---|---|
| Texture Quantification | 1. GLCM-based texture measures | 1. grey-level histograms<br>2. GLCMs |
| Similarity Metric | 1. Euclidean distance ($D_E$)<br>2. Standardised Euclidean distance ($D_{ES}$)<br>3. Mahalanobis distance ($D_M$) | 1. mutual information |

### 12.1.9 Chapter 9: Matching Results: Mosaic Images

The TM-matching and MI-matching algorithms were applied to test images made up of a mosaic of single texture reference images to test matching performance under the ideal condition of knowing how the reference images are transformed in the test image.

The purpose of this chapter was two-fold: (1) to test the performance of the matching algorithms on images with clear borders between the textures, accurate ground truth data and where there were known transformations of the textures to give insight into whether the matching algorithms have any potential for identifying similar textures in mammograms; and (2) to investigate how the matching accuracy is affected by the various matching parameters.

TM-matching and MI-matching show some potential for use as matching schemes. Results from the mosaic images *do* provide evidence to support the hypothesis that using texture based image-processing methods allows a textural region to be matched with a reference texture. However, it was shown that the choice of matching parameter values can significantly affect matching accuracy, so the following parameters were still varied for the pairs of mammograms and the stereotactic biopsy mammograms: *nbits*, *d*, *w*. As a consequence of the effect of sampling window step size, $w_{step}$, on matching accuracy for the mosaic images, $w_{step}$ were fixed for the analysis of the pairs of mammograms and the stereotactic biopsy mammograms.

The three matching methods with the best matching results, based on averages of κ for the best matches, are: MI-matching with GLCMs ($\kappa$=0.63±0.20), TM-$D_E$ ($\kappa$=0.56±0.13) and TM-$D_{ES}$ ($\kappa$=0.55±0.23). However, these three results are not statistically different ($p > 0.05$).

TM-$D_M$ and MI-histograms performed poorly as matching schemes.

The recommended values of the matching parameters for implementation of the matching schemes on similar images are given in **Table 12.2**. The values for TM-$D_E$ and TM-$D_{ES}$ are obtained from the few cases where there was some dependence of matching accuracy on the matching parameters.

**Table 12.2**: Recommended values for matching parameters for application of TM-$D_E$, TM-$D_{ES}$ and MI-GLCMs to images similar to the mosaics

| Matching Parameter | TM-$D_E$ | TM-$D_{ES}$ | MI-GLCMs |
|---|---|---|---|
| $w_{step}$ | 4 pixels | 4 pixels | 4 pixels |
| $w$ | rectangular from ACF | rectangular from ACF | rectangular from ACF |
| *nbits* | 5 bits | 5 bits | 8 bits |
| $d$ | 10 pixels | 1 pixel | 1 pixel |

### 12.1.10 Chapter 10: Matching Results: Mammograms

The TM-matching and MI-matching algorithms were applied to 68 pairs of CC and MLO mammograms to confirm whether a reference region of interest identified by a radiologist in one standard mammographic view can be matched to the corresponding region in another standard mammographic view, to determine whether the matching algorithms can be utilised in a CAD-system.

TM-matching with $D_E$ and MI-matching with GLCMs showed great potential for use in a CAD scheme. MI-matching with GLCMs had an average best matching accuracy of $\kappa$=0.41 ± 0.39 corresponding to average best values of $A_{ROC}$=0.77±0.25 and $C_{fb}$=0.50±0.42. TM-matching with $D_E$ had an average best matching accuracy of $\kappa$=0.33±0.25 corresponding to average best values of $A_{ROC}$=0.80±0.17 and $C_{fb}$=0.46±0.26. The results for these two methods were not statistically different ($p$ >0.05)

TM-matching with $D_{ES}$ was the third most accurate method with $\kappa$=0.24±0.21, corresponding to average values of $A_{ROC}$=0.81±0.20 and $C_{fb}$=0.28±0.25. The matching accuracy for TM-matching with $D_M$ and MI-matching with histograms was very low. The results for these three methods were statistically different from TM-$D_E$ and MI-GLCMs.

MI-matching with GLCMs had the best matching accuracy for matching malignant masses ($\kappa$=0.84±0.23 corresponding to $A_{ROC}$=0.96±0.05 and $C_{fb}$=0.90±0.21), while the results for the other types of regions of interest (benign, indeterminate, normal) were similar for MI-GLCMs and TM-$D_E$.

There was correlation between the results of the autocorrelation function analysis and the optimal sampling window sizes obtained from the matching analysis. TM-$D_E$ at 8 bits and TM-$D_M$ at 5 bits had the highest correlations. The ACF-results also justified the use of square windows for the analysis of the mammogram ROIs, since the ACF-widths and ACF-heights were similar. The results for MI-GLCMs showed some correlation, but this was significantly lower than for TM-$D_E$. The autocorrelation function can therefore be used to determine an optimal sampling window size.

While MI-GLCMs was more sensitive to matching parameters than TM-$D_E$, the overall, matching accuracies are better for MI-GLCMs than TM-$D_E$. Therefore TM-$D_E$ and MI-GLCMs are closely matched as matching schemes.

If these algorithms were to be applied in a CAD-system, it is recommended that ACF-analysis of the reference ROI be used to determine the sampling window size. Then both, TM-$D_E$ and MI-GLCMs should be used as matching schemes. The recommended values of the matching parameters are summarised in **Table 12.3**.

### 12.1.11 Chapter 11: Matching Results: Stereotactic Biopsy Mammograms

Stereotactic biopsies use two mammographic views of the breast, from slightly different angles to determine the depth of the abnormality to be biopsied. The sampling of the correct tissue critically

**Table 12.3**: Recommended values for matching parameters for practical application of TM-$D_E$ and MI-GLCMs in a CAD-system

| Matching Parameter | TM-$D_E$ | MI-GLCMs |
|---|---|---|
| $w_{step}$ | 4 pixels | 4 pixels |
| $w$ | from ACF analysis at 8 bits | from ACF analysis at 8 bits |
| $nbits$ | 8 bits | 8 bits |
| $d$ | 10 pixels | 1 pixel |

depends on the point to be sampled identified correctly in each stereoscopic view. The TM- and MI-matching algorithms were applied to stereotactic biopsy mammograms from 12 patients to investigate whether the algorithms could be used to improve localisation accuracy, by identifying the same point in all mammographic views from using only a single view as a reference. Results were evaluated in two stages. The first stage of the analysis was identical to that for the mosaic images and the standard mammograms. The second stage only examined the matched regions to determine whether the same point had been identified in all view.

The results of the first evaluation stage were very poor, generally much poorer than the results for the mosaic images and the standard mammograms. Consequently, only 7 pairs satisfied the criteria for the second evaluation stage. However, only one case had the maximum of the matching map (indicating the best match) close to the centroid of the ROI drawn by the radiologist.

Results were most likely poor because the ROIs had a low contrast in comparison to the surrounding breast tissue in the original mammogram. This can be addressed by applying a contrast enhancement algorithm to the images before the matching analysis. Another area for concern is the quality of the ground truth data. Since only one radiologist, retrospectively, identified the ROIs, it is impossible to obtain consensus if there is any discrepancy.

These results therefore indicate that the matching algorithms are not sufficiently sensitive to be used to improve the localisation accuracy for stereotactic breast biopsies.

## 12.2 Comparison of Results between the Image Sets

The TM-matching and MI-matching algorithms were applied to three image sets (mosaics, mammograms, stereotactic biopsy mammograms) to ultimately determine whether there was potential for the matching algorithms to be used in a mammographic CAD-scheme. There were many similarities and a few differences in the results for each image set.

TM-matching with $D_E$ and $D_{ES}$ and MI-matching with GLCMs had the three best matching accuracies for each image set.

The general behaviour of the matching parameters was similar for each image set: TM-matching generally was independent of $d$ and the dependence of $nbits$ depended on the sampling window size; MI-matching with histograms had optimal matches for the lowest value of $nbits$ and the lowest value of $nbins$; and MI-GLCMs had optimal matches at the highest value of $nbits$ and at the lowest value of $d$.

The other similar characteristic concerned the results of the three distance similarity metrics. Matching accuracy always decreased from $D_E$ to $D_{ES}$ to $D_M$. The results for $D_{ES}$, while lower than those for $D_E$, were often statistically similar. Both these results were, however, very different from those for $D_M$, and strongly suggests that the use of $D_M$ for matching is not appropriate. It was surprising that the significantly different ranges of texture measure values did not affect $D_E$ as expected.

The overall matching accuracies for each image set were not similar. Matching accuracies were very high for the mosaic images, decreased slightly for the pairs of mammograms and decreased further for the stereotactic biopsy mammograms. **Table 12.4** shows that the $A_{ROC}$-values were very high for the mosaics and pairs of mammograms, but the $C_{fb}$-values, which were high for the mosaics, were very low for the two sets of mammograms.

**Table 12.4**: Average of the best matching accuracies for each matching method

| Method | Average $A_{ROC}$ | Average $C_{fb}$ | Average $\kappa$ |
|---|---|---|---|
| Mosaics | | | |
| TM-$D_E$ | 0.84±0.08 | 0.79±0.18 | 0.55±0.23 |
| TM-$D_{ES}$ | 0.91±0.04 | 0.68±0.09 | 0.56±0.13 |
| TM-$D_M$ | 0.82±0.09 | 0.45±0.18 | 0.31±0.17 |
| MI-histograms | 0.77±0.10 | 0.29±0.18 | 0.15±0.11 |
| MI-GLCMs | 0.94±0.06 | 0.70±0.16 | 0.63±0.19 |
| Mammograms | | | |
| TM-$D_E$ | 0.80±0.17 | 0.46±0.26 | 0.33±0.25 |
| TM-$D_{ES}$ | 0.81±0.20 | 0.28±0.25 | 0.24±0.21 |
| TM-$D_M$ | 0.79±0.17 | 0.25±0.20 | 0.19±0.17 |
| MI-histograms | 0.85±0.11 | 0.16±0.15 | 0.12±0.13 |
| MI-GLCMs | 0.77±0.25 | 0.50±0.42 | 0.41±0.39 |
| Stereotactic Biopsy Mammograms | | | |
| TM-$D_E$ | 0.75±0.11 | 0.38±0.26 | 0.20±0.14 |
| TM-$D_{ES}$ | 0.79±0.09 | 0.27±0.17 | 0.15±0.09 |
| TM-$D_M$ | 0.67±0.20 | 0.21±0.33 | 0.06±0.09 |
| MI-histograms | 0.63±0.13 | 0.25±0.25 | 0.09±0.11 |
| MI-GLCMs | 0.68±0.21 | 0.35±0.43 | 0.27±0.32 |

The high matching accuracy of the mosaic images is most likely because the borders between the textures in the mosaic images were clear and the ground truth data was easily and accurately extracted. Both sets of mammograms did not have clear borders around the region of interest and so the ground truth data might not be that accurate, and this would decrease the average matching accuracy. One significant problem with the ground truth data for the mammograms was that only one radiologist was used for each image set. Matching accuracy would probably be improved if at least three radiologists marked borders of regions of interest. Then the results for each region of interest could be evaluated on the basis of three sets of ground truth data, which might provide some consensus to the results.

## 12.3 Novel Features of this Study

The novel features of this study are:

1. Graphical summary of how a radiologist analyses a mammogram.
2. Flow chart summarising steps of a mammographic CAD-system.
3. A novel method of detecting the breast border using iso-intensity contours was presented.
4. Faster GLCM algorithm using the IDL functions `HIST_2D` and `SHIFT`.
5. The use of GLCM-based texture measures and distance similarity metrics to determine similarities between regions in mammographic views of the same breast.
6. The use of GLCMs and mutual information to determine similarities between regions in mammographic views of the same breast.
7. Detailed analysis of the effect of sampling window size on matching accuracy and the use of the autocorrelation function to determine optimal sampling window sizes.
8. The combined use of $A_{ROC}$ *and* contrast to evaluate matching results.
9. The application of matching methods to stereotactic biopsy mammograms.

## 12.4 Strengths of the TM- and MI-Matching Algorithms

1. The TM- and MI-matching algorithms show potential for providing more information for use in a false-positive reduction scheme in a CAD-system. The ideal solution would be to try to incorporate mutual information ideas into the texture measure method.
2. One advantage of using the distance similarity metrics and mutual information for matching is that there is no training required which is quite important for a mammographic CAD-system since breast tissue varies considerably from patient to patient.
3. The TM- and MI-matching algorithms can be applied to any image-matching problem.
4. If the object is present in both mammographic views, only one view needs to be analysed to detect the object, while the second view is analysed with information extracted from the object in the first view, for confirmation of a true object. However, if the algorithm is made more efficient (i.e. faster), then it would be best to analyse both views, from both directions. This could be used to confirm the results of the matching.

## 12.5 Weaknesses of the TM- and MI-Matching Algorithms

1. There is no guarantee that if two textures match, then they refer to the same mass.
2. The current algorithms are very time-consuming and will have to be re-designed for any implementation in a CAD-system.
3. The MI-matching algorithm using GLCMs has also been shown to perform very badly if there are any global differences in grey-levels between the reference and test images.

## 12.6 Future Work

The results of this study have shown that the following should be investigated further to improve the matching accuracy of the matching algorithms so that they can be of clinical use in a CAD-system.

### 12.6.1 Effect of Spatial Resolution

The spatial resolution of the images were kept constant in this investigation and it might be interesting to see whether processing times can be reduced by performing the matching at a spatial resolution higher than that used by radiologists.

### 12.6.2 Optimising of Breast Border Detection Algorithm

Investigations should be performed into improving the speed of the breast border detection algorithm.

### 12.6.3 Improvement of TM-Matching by use of Feature Selection

The accuracy of the TM-matching algorithm could be improved by implementing a feature selection algorithm, which selects the dominant texture measures (from the full set of 12) and only uses this reduced set in the matching.

### 12.6.4 Quality of Ground Truth Data

The performance of the matching algorithms depends critically upon the quality of the ground truth data and it might be advantageous, especially for the mammograms, to employ a method similar to double reading with arbitration employed by radiologists to obtain consensus with a diagnosis. If there is a difference in opinion between the two radiologists performing the double reading, then a third radiologist arbitrates and makes the final decision.

For the evaluation of the matching algorithms, it is therefore necessary to obtain ground truth data from at least three radiologists. The evaluation is performed separately for each radiologist and the results of the evaluation are analysed to obtain consensus in the final result.

### 12.6.5 Effect of Image Pre-processing

The only pre-processing performed, merely removed undesirable features in the image, and it was observed that the contrast of the SB mammograms was very poor as well as there being differences in average grey-level between images in a triplet.

It is therefore necessary to investigate the effect on matching accuracy of image enhancement pre-processing methods, like contrast enhancement and grey-level equalisation.

### 12.6.6 Grey-level Offset Invariance

The method used to determine the grey-level histograms and GLCMs was not invariant to constant grey-level offsets between the images, which meant that the reference and test images could not be matched if there was a difference in average grey-level.

It is therefore necessary to compute the grey-level histogram and the GLCM in such a manner that they are invariant to grey-level offsets. It is expected that this should improve the histogram results considerably.

### 12.6.7 Shape of Sampling Windows

Square sampling windows of varying sizes were used in this investigation. However, analysis of the reference images using the autocorrelation function show that the characteristic scale widths and heights of the textures are sometimes unequal. While the scale widths and heights are similar for the mammograms, the results were statistically different for the mosaic images.

It is therefore necessary to:

1. perform matching at the sampling window sizes obtained from the autocorrelation function analysis to determine matching accuracy, and
2. investigate the effect of non-square sampling windows (e.g. rectangular or elliptical) on matching accuracy.

### 12.6.8 Invariance to Directional and Scaling Factors

Normal breast tissue and benign masses are generally texturally inhomogeneous. The averaging of the GLCMs in this study strongly favoured homogeneous textures. The poor matching accuracy for non-malignant mammograms might indicate that there is a directional component to the analysis, which is not adequately considered by the averaging of the GLCMs.

It is necessary to investigate the effect on matching accuracy if, for example, a log-polar transform [Wolberg, G. and Zokai, S. 2000] or other invariant texture methods [Zhang et al. 2002, Zhang & Tan 2003] are used to remove rotational, directional and scaling effects.

### 12.6.9 Improvements to the Evaluation Algorithm

The scaling of individual matching maps (each with a different minimum and maximum) between 0 and 255 can lead to problems with the evaluation. Some maps can have a very small range of intensities, while other maps can have a very large range and each of these get scaled between 0 and 255. The scaled maps are then treated as if they are equivalent, which might explain why so many methods exhibited an independence of matching accuracy on a particular matching parameter.

Also, while the trapezoidal rule used to calculate the area under the ROC curve is simple to implement, it systematically underestimates the area [Hanley & McNeil 1983].

So, in order to confirm the accuracy of the results, it is necessary to examine the values of distances and mutual information across the range of window sizes and scale the matching maps between the minimum and maximum of all distance and mutual information values. Results might also improve if the area under the ROC curve were calculated using the binomial model.

### 12.6.10 False-Positive Reduction Algorithm

If the recommendations of improvement to the matching algorithms are successful enough to improve matching between regions in two mammographic views, it will then be necessary to design and test a false-positive reduction algorithm based on the TM- and MI-matching algorithms.

### 12.6.11 Improvement of Stereotactic Biopsy Localisation Accuracy

If the recommendations of improvement to the matching algorithms are successful enough to accurately match regions between stereotactic mammograms, it will then be necessary to design an experiment to physically test the effectiveness of TM- and MI-matching algorithms to improve localisation accuracy during stereotactic biopsies.

## 12.7 Conclusions

It is possible to match a reference region from one standard mammographic view to the corresponding region in another standard mammographic view of the same breast, using texture based image processing methods. The matching algorithms, using grey-level co-occurrence matrices, grey-level histograms, distance similarity metrics and mutual information, perform especially well at matching malignant masses. This dual-view analysis method can therefore be used to provide complementary information to a false-positive reduction scheme.

# Appendix A

# Detailed Results

Details of the matching parameters of the best matches are shown in the following tables for each set of images.

**Table A.1**: Matching parameters and evaluation results for the best matches of TM-matching applied to mosaic images, listed for each image pair.

| Images | TM-$D_E$ | | | | | | TM-$D_{ES}$ | | | | | | TM-$D_M$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $w$ | $d$ | *nbits* | $A_{ROC}$ | $C_{fb}$ | $\kappa$ | $w$ | $d$ | *nbits* | $A_{ROC}$ | $C_{fb}$ | $\kappa$ | $w$ | $d$ | *nbits* | $A_{ROC}$ | $C_{fb}$ | $\kappa$ |
| R1T1 | 176 | 10 | 6 | 0.89 | 0.94 | 0.73 | 320 | 10 | 8 | 0.92 | 0.75 | 0.63 | 400 | 2 | 6 | 0.88 | 0.60 | 0.45 |
| R2T1 | 80 | 10 | 5 | 0.85 | 0.74 | 0.51 | 80 | 1 | 5 | 0.91 | 0.75 | 0.61 | 80 | 10 | 7 | 0.88 | 0.55 | 0.41 |
| R3T1 | 160 | 1 | 5 | 0.67 | 0.58 | 0.20 | 160 | 1 | 8 | 0.87 | 0.64 | 0.47 | 176 | 1 | 6 | 0.83 | 0.59 | 0.39 |
| R4T1 | 80 | 5 | 7 | 0.85 | 0.88 | 0.62 | 320 | 1 | 5 | 0.95 | 0.80 | 0.72 | 160 | 10 | 5 | 0.90 | 0.74 | 0.60 |
| R5T2 | 128 | 5 | 5 | 0.82 | 0.86 | 0.55 | 128 | 5 | 8 | 0.90 | 0.73 | 0.59 | 128 | 5 | 7 | 0.88 | 0.33 | 0.25 |
| R6T2 | 80 | 1 | 6 | 0.84 | 0.94 | 0.65 | 160 | 10 | 7 | 0.89 | 0.57 | 0.44 | 368 | 2 | 6 | 0.78 | 0.57 | 0.32 |
| R7T2 | 224 | 1 | 8 | 0.99 | 0.99 | 0.97 | 336 | 10 | 5 | 0.98 | 0.82 | 0.80 | 368 | 1 | 8 | 0.90 | 0.46 | 0.36 |
| R8T2 | 80 | 10 | 5 | 0.76 | 0.75 | 0.39 | 128 | 2 | 5 | 0.87 | 0.64 | 0.47 | 112 | 10 | 6 | 0.85 | 0.39 | 0.28 |
| R5T3 | 128 | 10 | 5 | 0.84 | 0.78 | 0.54 | 128 | 10 | 5 | 0.88 | 0.52 | 0.39 | 240 | 1 | 7 | 0.71 | 0.19 | 0.08 |
| R6T3 | 288 | 1 | 7 | 0.82 | 0.82 | 0.53 | 288 | 10 | 7 | 0.83 | 0.62 | 0.41 | 288 | 10 | 7 | 0.68 | 0.36 | 0.13 |
| R7T3 | 160 | 2 | 8 | 0.95 | 0.96 | 0.86 | 160 | 5 | 5 | 0.93 | 0.70 | 0.60 | 176 | 1 | 5 | 0.63 | 0.12 | 0.03 |
| R8T3 | 176 | 10 | 7 | 0.86 | 0.72 | 0.52 | 304 | 10 | 5 | 0.95 | 0.68 | 0.61 | 288 | 1 | 5 | 0.91 | 0.52 | 0.42 |

**Table A.2**: Matching parameters and evaluation results for the best matches of MI-matching applied to mosaic images, listed for each image pair.

| Images | MI-Histograms | | | | | | MI-GLCMs | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $w$ | *nbins* | *nbits* | $A_{ROC}$ | $C_{fb}$ | $\kappa$ | $w$ | $d$ | *nbits* | $A_{ROC}$ | $C_{fb}$ | $\kappa$ |
| R1T1 | 400 | 16 | 8 | 0.67 | 0.43 | 0.15 | 32 | 10 | 8 | 0.91 | 0.53 | 0.44 |
| R2T1 | 368 | 32 | 8 | 0.78 | 0.69 | 0.39 | 48 | 5 | 8 | 0.97 | 0.77 | 0.72 |
| R3T1 | 400 | 16 | 6 | 0.68 | 0.35 | 0.13 | 48 | 10 | 8 | 0.96 | 0.57 | 0.53 |
| R4T1 | 16 | 32 | 8 | 0.75 | 0.11 | 0.06 | 16 | 1 | 8 | 0.98 | 0.70 | 0.66 |
| R5T2 | 16 | 16 | 8 | 0.67 | 0.08 | 0.03 | 16 | 2 | 8 | 0.97 | 0.95 | 0.90 |
| R6T2 | 96 | 32 | 5 | 0.72 | 0.34 | 0.15 | 208 | 1 | 8 | 0.99 | 0.63 | 0.62 |
| R7T2 | 80 | 16 | 6 | 0.87 | 0.15 | 0.11 | 48 | 10 | 8 | 1.00 | 0.94 | 0.93 |
| R8T2 | 64 | 128 | 8 | 0.99 | 0.17 | 0.17 | 16 | 2 | 7 | 0.88 | 0.64 | 0.49 |
| R5T3 | 384 | 16 | 5 | 0.66 | 0.24 | 0.08 | 32 | 10 | 8 | 0.98 | 0.71 | 0.69 |
| R6T3 | 96 | 32 | 5 | 0.80 | 0.35 | 0.21 | 112 | 2 | 8 | 0.80 | 0.42 | 0.25 |
| R7T3 | 176 | 128 | 8 | 0.90 | 0.43 | 0.34 | 80 | 10 | 8 | 0.98 | 0.79 | 0.75 |
| R8T3 | 64 | 64 | 7 | 0.74 | 0.14 | 0.07 | 16 | 5 | 8 | 0.91 | 0.74 | 0.60 |

**Table A.3**: Matching parameters and evaluation results for the best matches of TM-matching applied to pairs of mammograms, listed for each image pair.

| Images | $D_E$ | | | | | | $D_{ES}$ | | | | | | $D_M$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $w$ | $d$ | $nbits$ | $A_{ROC}$ | $C_{fb}$ | $\kappa$ | $w$ | $d$ | $nbits$ | $A_{ROC}$ | $C_{fb}$ | $\kappa$ | $w$ | $d$ | $nbits$ | $A_{ROC}$ | $C_{fb}$ | $\kappa$ |
| M0M1 | 80 | 10 | 8 | 0.94 | 0.48 | 0.42 | 80 | 10 | 5 | 0.98 | 0.32 | 0.30 | 80 | 1 | 7 | 0.87 | 0.32 | 0.24 |
| M1M0 | 80 | 10 | 5 | 0.99 | 0.58 | 0.57 | 80 | 5 | 5 | 0.99 | 0.47 | 0.46 | 64 | 1 | 7 | 0.96 | 0.43 | 0.39 |
| M2M3 | 16 | 1 | 8 | 0.88 | 0.49 | 0.37 | 16 | 10 | 5 | 0.87 | 0.28 | 0.21 | 16 | 1 | 8 | 0.79 | 0.19 | 0.11 |
| M3M2 | 32 | 1 | 5 | 0.89 | 0.30 | 0.23 | 48 | 1 | 8 | 0.94 | 0.37 | 0.33 | 48 | 10 | 5 | 0.74 | 0.25 | 0.12 |
| M4M5 | 16 | 5 | 5 | 0.86 | 0.65 | 0.47 | 32 | 1 | 8 | 0.92 | 0.41 | 0.35 | 48 | 5 | 7 | 0.92 | 0.45 | 0.38 |
| M5M4 | 16 | 2 | 7 | 0.71 | 0.80 | 0.34 | 16 | 10 | 5 | 0.78 | 0.55 | 0.31 | 48 | 10 | 8 | 0.85 | 0.55 | 0.39 |
| M6M7 | 32 | 1 | 5 | 0.62 | 0.55 | 0.13 | **16** | **10** | **8** | **0.64** | 0.19 | 0.05 | 48 | 10 | 8 | 0.72 | 0.09 | 0.04 |
| M7M6 | 48 | 5 | 7 | 0.74 | 0.24 | 0.12 | **48** | **2** | **5** | **0.48** | -0.04 | 0.00 | 16 | 2 | 5 | 0.50 | -0.05 | 0.00 |
| M8M9 | 32 | 10 | 8 | 0.84 | 0.60 | 0.41 | 16 | 5 | 5 | 0.92 | 0.37 | 0.31 | 16 | 10 | 8 | 0.87 | 0.07 | 0.05 |
| M9M8 | 48 | 2 | 5 | 0.86 | 0.37 | 0.27 | 48 | 1 | 6 | 0.79 | 0.20 | 0.12 | 16 | 5 | 8 | 0.71 | 0.10 | 0.04 |
| M10M11 | 16 | 10 | 6 | 0.58 | 0.12 | 0.02 | 64 | 2 | 6 | 0.87 | 0.09 | 0.07 | 48 | 10 | 6 | 0.87 | 0.11 | 0.08 |
| M11M10 | 32 | 5 | 8 | 0.92 | 0.68 | 0.57 | 32 | 1 | 8 | 0.96 | 0.37 | 0.34 | 32 | 5 | 8 | 0.93 | 0.33 | 0.29 |
| M12M13 | 16 | 5 | 5 | 0.93 | 0.46 | 0.39 | 16 | 5 | 6 | 0.55 | 0.22 | 0.02 | 32 | 2 | 8 | 0.22 | 0.00 | 0.00 |
| M13M12 | 32 | 1 | 6 | 0.72 | 0.05 | 0.02 | 32 | 1 | 6 | 0.26 | -0.04 | -0.02 | 32 | 5 | 6 | 0.84 | 0.05 | 0.04 |
| M14M15 | 48 | 10 | 7 | 0.93 | 0.56 | 0.49 | 96 | 1 | 5 | 0.98 | 0.18 | 0.18 | 48 | 10 | 8 | 0.84 | 0.30 | 0.21 |
| M15M14 | 48 | 10 | 6 | 0.88 | 0.44 | 0.34 | 64 | 10 | 5 | 0.94 | 0.26 | 0.23 | 48 | 10 | 8 | 0.87 | 0.22 | 0.16 |
| M16M17 | 16 | 10 | 7 | 0.80 | 0.43 | 0.26 | 16 | 10 | 7 | 0.87 | 0.31 | 0.23 | 64 | 1 | 5 | 0.92 | 0.14 | 0.12 |
| M17M16 | 96 | 2 | 7 | 0.85 | 0.39 | 0.28 | 32 | 2 | 8 | 0.85 | 0.37 | 0.25 | 16 | 5 | 8 | 0.53 | 0.22 | 0.01 |
| M18M19 | 16 | 2 | 8 | 0.58 | 0.15 | 0.02 | 16 | 1 | 8 | 0.89 | 0.06 | 0.05 | 32 | 5 | 5 | 0.73 | 0.03 | 0.01 |
| M19M18 | 16 | 10 | 5 | 0.70 | 0.29 | 0.12 | 16 | 10 | 6 | 0.79 | 0.09 | 0.05 | 16 | 2 | 7 | 0.81 | 0.09 | 0.06 |
| M20M21 | **80** | 10 | 6 | 0.97 | 0.29 | 0.27 | 80 | 10 | 7 | 0.99 | 0.17 | 0.17 | 64 | 2 | 6 | 0.92 | 0.38 | 0.33 |
| M21M20 | **64** | 1 | 8 | 0.85 | 0.91 | 0.64 | 64 | 2 | 8 | 0.98 | 0.42 | 0.41 | 64 | 1 | 7 | 0.91 | 0.43 | 0.36 |
| M22M23 | **96** | 2 | 7 | 0.83 | 0.64 | 0.42 | 96 | 10 | 7 | 0.87 | 0.53 | 0.39 | 96 | 10 | 6 | 0.88 | 0.40 | 0.31 |
| M23M22 | **96** | 10 | 5 | 0.95 | 0.64 | 0.58 | 112 | 1 | 5 | 0.99 | 0.48 | 0.47 | 96 | 2 | 5 | 0.81 | 0.33 | 0.21 |
| M24M25 | 16 | 5 | 7 | 0.71 | 0.43 | 0.18 | 16 | 10 | 5 | 0.85 | 0.22 | 0.15 | 32 | 10 | 8 | 0.77 | 0.09 | 0.05 |
| M25M24 | 32 | 10 | 8 | 0.26 | 0.00 | 0.00 | 32 | 1 | 5 | 0.19 | -0.07 | -0.05 | 32 | 1 | 5 | 0.19 | -0.09 | -0.06 |
| M26M27 | 16 | 1 | 5 | 0.87 | 0.53 | 0.40 | 16 | 5 | 5 | 0.90 | 0.43 | 0.34 | 32 | 1 | 8 | 0.77 | 0.21 | 0.11 |
| M27M26 | 256 | 10 | 6 | 0.93 | 0.83 | 0.71 | 256 | 10 | 5 | 0.94 | 0.64 | 0.56 | 32 | 1 | 8 | 0.93 | 0.49 | 0.42 |
| M28M29 | 128 | 10 | 8 | 0.91 | 0.48 | 0.39 | 96 | 1 | 5 | 0.77 | 0.31 | 0.17 | 80 | 2 | 5 | 0.80 | 0.29 | 0.17 |
| M29M28 | 48 | 2 | 6 | 0.71 | 0.43 | 0.18 | 48 | 10 | 6 | 0.76 | 0.60 | 0.32 | 96 | 10 | 5 | 0.76 | 0.22 | 0.11 |
| M30M31 | 112 | 10 | 7 | 0.96 | 0.75 | 0.69 | 144 | 1 | 7 | 0.98 | 0.58 | 0.56 | 144 | 1 | 5 | 0.96 | 0.65 | 0.60 |
| M31M30 | 112 | 10 | 8 | 0.98 | 0.88 | 0.85 | 144 | 1 | 7 | 0.99 | 0.66 | 0.65 | 144 | 1 | 7 | 0.96 | 0.50 | 0.45 |
| M32M33 | 32 | 10 | 6 | 0.88 | 0.70 | 0.53 | 32 | 10 | 6 | 0.95 | 0.45 | 0.41 | 48 | 10 | 8 | 0.85 | 0.37 | 0.26 |
| M33M32 | 32 | 10 | 5 | 0.82 | 0.27 | 0.17 | 80 | 1 | 7 | 0.80 | 0.12 | 0.07 | 80 | 1 | 7 | 0.77 | 0.31 | 0.17 |
| M34M35 | 64 | 10 | 6 | 0.89 | 0.30 | 0.24 | 80 | 10 | 6 | 0.96 | 0.14 | 0.13 | 80 | 1 | 5 | 0.94 | 0.20 | 0.17 |
| M35M34 | 64 | 2 | 8 | 0.86 | 0.53 | 0.38 | 64 | 2 | 8 | 0.94 | 0.31 | 0.27 | 64 | 1 | 8 | 0.85 | 0.26 | 0.18 |
| M36M37 | 32 | 5 | 7 | 0.84 | 0.30 | 0.20 | 80 | 1 | 5 | 0.96 | 0.24 | 0.22 | 80 | 10 | 5 | 0.90 | 0.19 | 0.16 |
| M37M36 | 80 | 5 | 8 | 0.89 | 0.67 | 0.53 | 96 | 10 | 7 | 0.99 | 0.37 | 0.36 | 64 | 1 | 8 | 0.84 | 0.33 | 0.22 |
| M38M39 | 64 | 1 | 8 | 0.95 | 0.89 | 0.79 | 64 | 1 | 8 | 0.98 | 0.57 | 0.54 | 144 | 10 | 5 | 0.87 | 0.43 | 0.32 |
| M39M38 | 48 | 10 | 7 | 0.96 | 0.71 | 0.65 | 48 | 5 | 7 | 0.95 | 0.35 | 0.31 | 48 | 10 | 8 | 0.75 | 0.19 | 0.10 |
| M40M41 | 64 | 10 | 7 | 0.88 | 0.36 | 0.27 | 48 | 1 | 8 | 0.97 | 0.25 | 0.23 | 48 | 2 | 7 | 0.97 | 0.24 | 0.23 |
| M41M40 | 48 | 10 | 6 | 0.80 | 0.72 | 0.44 | 48 | 5 | 5 | 0.94 | 0.19 | 0.17 | 48 | 5 | 8 | 0.94 | 0.21 | 0.19 |
| M42M43 | 64 | 10 | 7 | 0.91 | 0.66 | 0.54 | 48 | 5 | 7 | 0.88 | 0.32 | 0.24 | 48 | 10 | 7 | 0.81 | 0.27 | 0.16 |
| M43M42 | 32 | 10 | 5 | 0.87 | 0.41 | 0.30 | 16 | 10 | 5 | 0.80 | 0.24 | 0.15 | 112 | 10 | 6 | 0.95 | 0.21 | 0.19 |
| M44M45 | 176 | 10 | 7 | 0.97 | 0.88 | 0.82 | 176 | 1 | 8 | 0.97 | 0.65 | 0.61 | 176 | 1 | 7 | 0.89 | 0.61 | 0.48 |
| M45M44 | 112 | 10 | 8 | 0.95 | 0.86 | 0.78 | 128 | 10 | 7 | 0.94 | 0.54 | 0.47 | 96 | 2 | 8 | 0.81 | 0.43 | 0.27 |
| M46M47 | 32 | 2 | 6 | 0.90 | 0.38 | 0.31 | 32 | 10 | 5 | 0.95 | 0.26 | 0.23 | 32 | 10 | 5 | 0.95 | 0.19 | 0.17 |
| M47M46 | 32 | 1 | 8 | 0.80 | 0.32 | 0.19 | 32 | 1 | 6 | 0.59 | 0.02 | 0.00 | 32 | 5 | 8 | 0.84 | 0.14 | 0.10 |
| M48M49 | 32 | 1 | 5 | 0.51 | 0.10 | 0.00 | 16 | 2 | 6 | 0.63 | 0.04 | 0.01 | 32 | 10 | 5 | 0.53 | 0.08 | 0.01 |
| M49M48 | 48 | 10 | 8 | 0.78 | 0.26 | 0.15 | 48 | 10 | 8 | 0.56 | 0.07 | 0.01 | 16 | 5 | 6 | 0.66 | 0.33 | 0.11 |
| M50M51 | 16 | 1 | 6 | 0.60 | 0.14 | 0.03 | 16 | 2 | 5 | 0.50 | -0.34 | 0.00 | 32 | 10 | 6 | 0.63 | 0.03 | 0.01 |
| M51M50 | 16 | 10 | 6 | 0.60 | 0.27 | 0.05 | 32 | 5 | 5 | 0.85 | 0.09 | 0.06 | 32 | 10 | 5 | 0.81 | 0.08 | 0.05 |
| M52M53 | 16 | 1 | 6 | 0.40 | -0.10 | -0.02 | 16 | 1 | 6 | 0.64 | 0.12 | 0.03 | 16 | 5 | 7 | 0.41 | -0.01 | 0.00 |
| M53M52 | 32 | 5 | 7 | 0.73 | 0.34 | 0.16 | 32 | 10 | 7 | 0.74 | 0.06 | 0.03 | 32 | 10 | 7 | 0.80 | 0.22 | 0.13 |
| M54M55 | 176 | 10 | 7 | 0.99 | 0.85 | 0.83 | 176 | 1 | 7 | 0.99 | 0.74 | 0.73 | 176 | 1 | 7 | 0.93 | 0.98 | 0.83 |
| M55M54 | 128 | 10 | 6 | 0.98 | 0.96 | 0.92 | 32 | 5 | 5 | 0.99 | 0.95 | 0.93 | 144 | 2 | 7 | 0.95 | 0.74 | 0.66 |
| M56M57 | 32 | 1 | 5 | 0.37 | -0.17 | -0.04 | 32 | 1 | 7 | 0.64 | 0.01 | 0.00 | 32 | 5 | 5 | 0.81 | 0.23 | 0.14 |
| M57M56 | 16 | 1 | 5 | 0.75 | 0.30 | 0.15 | 32 | 2 | 5 | 0.91 | 0.22 | 0.19 | 32 | 1 | 5 | 0.75 | 0.10 | 0.05 |
| M58M59 | 16 | 10 | 5 | 0.28 | 0.03 | -0.01 | 32 | 10 | 7 | 0.29 | 0.01 | 0.00 | 16 | 1 | 6 | 0.31 | 0.01 | 0.00 |
| M59M58 | 16 | 10 | 8 | 0.70 | 0.15 | 0.06 | 16 | 2 | 7 | 0.50 | -0.28 | 0.00 | 16 | 10 | 8 | 0.74 | 0.03 | 0.01 |
| M60M61 | 64 | 10 | 5 | 0.98 | 0.73 | 0.70 | 64 | 1 | 7 | 0.99 | 0.58 | 0.56 | 64 | 1 | 8 | 0.94 | 0.32 | 0.28 |
| M61M60 | 32 | 10 | 7 | 0.80 | 0.47 | 0.28 | 32 | 2 | 5 | 0.93 | 0.21 | 0.18 | 32 | 1 | 5 | 0.91 | 0.16 | 0.13 |
| M62M63 | 16 | 5 | 5 | 0.64 | 0.62 | 0.18 | 16 | 5 | 5 | 0.70 | 0.67 | 0.27 | 32 | 10 | 8 | 0.71 | 0.51 | 0.21 |
| M63M62 | 32 | 10 | 7 | 0.66 | 0.38 | 0.12 | 48 | 10 | 8 | 0.66 | 0.15 | 0.05 | 48 | 10 | 7 | 0.66 | 0.11 | 0.04 |
| M64M65 | 16 | 10 | 5 | 0.82 | 0.43 | 0.27 | 32 | 10 | 7 | 0.94 | 0.32 | 0.28 | 16 | 5 | 7 | 0.88 | 0.26 | 0.20 |
| M65M64 | 32 | 10 | 5 | 0.63 | 0.38 | 0.10 | 32 | 10 | 5 | 0.47 | -0.01 | 0.00 | 32 | 10 | 8 | 0.49 | -0.20 | 0.00 |
| M66M67 | 32 | 2 | 6 | 0.81 | 0.09 | 0.06 | 32 | 1 | 6 | 0.50 | -0.16 | 0.00 | 112 | 10 | 6 | 0.92 | 0.41 | 0.35 |
| M67M66 | 48 | 5 | 6 | 0.75 | 0.70 | 0.36 | 48 | 2 | 7 | 0.88 | 0.45 | 0.34 | 64 | 1 | 6 | 0.82 | 0.40 | 0.26 |

**Table A.4**: Matching parameters and evaluation results for the best matches of MI-matching applied to pairs of mammograms, listed for each image pair.

| Images | MI-Histograms | | | | | | MI-GLCMs | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $w$ | $d$ | $nbits$ | $A_{ROC}$ | $C_{fb}$ | $\kappa$ | $w$ | $d$ | $nbits$ | $A_{ROC}$ | $C_{fb}$ | $\kappa$ |
| M0M1 | 16 | 16 | 8 | 0.97 | 0.13 | 0.13 | 32 | 2 | 6 | 0.96 | 0.99 | 0.91 |
| M1M0 | 16 | 16 | 8 | 0.92 | 0.06 | 0.05 | 16 | 10 | 7 | 0.99 | 0.97 | 0.95 |
| M2M3 | 16 | 16 | 6 | 0.95 | 0.33 | 0.30 | 16 | 1 | 6 | 0.87 | 0.86 | 0.64 |
| M3M2 | 32 | 32 | 5 | 0.87 | 0.23 | 0.17 | 80 | 10 | 5 | 0.49 | -0.03 | 0.00 |
| M4M5 | 16 | 16 | 5 | 0.90 | 0.18 | 0.15 | 48 | 5 | 5 | 0.97 | 0.18 | 0.17 |
| M5M4 | 32 | 32 | 5 | 0.98 | 0.57 | 0.55 | 32 | 5 | 5 | 0.81 | 0.58 | 0.36 |
| M6M7 | 48 | 16 | 5 | 0.56 | 0.29 | 0.03 | 16 | 1 | 5 | 0.29 | 0.00 | 0.00 |
| M7M6 | 32 | 16 | 5 | 0.70 | 0.25 | 0.10 | 32 | 2 | 7 | 0.50 | -0.14 | 0.00 |
| M8M9 | 16 | 16 | 6 | 0.92 | 0.15 | 0.12 | 16 | 5 | 8 | 0.94 | 0.87 | 0.77 |
| M9M8 | 32 | 16 | 5 | 0.86 | 0.14 | 0.10 | 16 | 10 | 5 | 0.84 | 0.15 | 0.10 |
| M10M11 | 48 | 16 | 5 | 0.74 | 0.12 | 0.06 | 32 | 10 | 5 | 0.88 | 0.31 | 0.24 |
| M11M10 | 32 | 16 | 5 | 0.85 | 0.18 | 0.13 | 32 | 10 | 5 | 0.85 | 0.08 | 0.06 |
| M12M13 | 16 | 16 | 5 | 0.66 | 0.31 | 0.10 | 16 | 1 | 5 | 0.36 | 0.00 | 0.00 |
| M13M12 | 16 | 16 | 5 | 0.89 | 0.21 | 0.17 | 32 | 5 | 6 | 0.57 | 0.12 | 0.02 |
| M14M15 | 16 | 16 | 7 | 0.82 | 0.05 | 0.03 | 32 | 10 | 5 | 0.99 | 0.98 | 0.96 |
| M15M14 | 16 | 16 | 7 | 0.69 | 0.03 | 0.01 | 16 | 1 | 6 | 0.95 | 1.00 | 0.90 |
| M16M17 | 32 | 128 | 8 | 0.91 | 0.07 | 0.06 | 96 | 10 | 5 | 0.85 | 0.07 | 0.05 |
| M17M16 | 32 | 32 | 5 | 0.96 | 0.55 | 0.50 | 96 | 1 | 5 | 0.23 | 0.00 | 0.00 |
| M18M19 | 16 | 16 | 7 | 0.91 | 0.04 | 0.03 | 16 | 5 | 8 | 0.87 | 0.47 | 0.35 |
| M19M18 | 16 | 32 | 6 | 0.88 | 0.09 | 0.07 | 16 | 5 | 7 | 0.65 | 0.20 | 0.06 |
| M20M21 | 16 | 16 | 8 | 0.99 | 0.10 | 0.10 | 32 | 10 | 6 | 1.00 | 1.00 | 0.99 |
| M21M20 | 16 | 16 | 7 | 0.75 | 0.03 | 0.01 | 48 | 5 | 5 | 0.93 | 0.23 | 0.19 |
| M22M23 | 16 | 16 | 8 | 0.55 | -0.15 | -0.01 | 16 | 1 | 8 | 0.79 | 0.60 | 0.35 |
| M23M22 | 16 | 16 | 8 | 0.97 | 0.08 | 0.08 | 32 | 10 | 6 | 1.00 | 1.00 | 1.00 |
| M24M25 | 16 | 16 | 5 | 0.80 | 0.22 | 0.13 | 16 | 1 | 5 | 0.48 | 0.00 | 0.00 |
| M25M24 | 32 | 32 | 5 | 0.88 | 0.17 | 0.13 | 32 | 2 | 5 | 0.10 | -0.31 | -0.25 |
| M26M27 | 16 | 16 | 8 | 0.93 | 0.35 | 0.30 | 48 | 1 | 8 | 0.97 | 0.95 | 0.89 |
| M27M26 | 16 | 16 | 8 | 0.88 | 0.24 | 0.18 | 48 | 1 | 8 | 0.99 | 1.00 | 0.98 |
| M28M29 | 16 | 16 | 8 | 0.86 | 0.02 | 0.01 | 32 | 10 | 8 | 0.97 | 0.47 | 0.44 |
| M29M28 | 16 | 16 | 8 | 0.76 | -0.01 | 0.00 | 32 | 10 | 8 | 0.94 | 0.35 | 0.31 |
| M30M31 | 16 | 16 | 8 | 0.91 | 0.17 | 0.14 | 80 | 2 | 8 | 1.00 | 0.99 | 0.99 |
| M31M30 | 16 | 16 | 8 | 0.81 | 0.11 | 0.07 | 64 | 2 | 5 | 1.00 | 0.99 | 0.98 |
| M32M33 | 16 | 16 | 7 | 0.90 | 0.03 | 0.03 | 16 | 10 | 7 | 0.96 | 0.84 | 0.77 |
| M33M32 | 16 | 16 | 8 | 0.84 | 0.02 | 0.01 | 16 | 5 | 6 | 0.86 | 0.16 | 0.12 |
| M34M35 | 16 | 16 | 7 | 0.86 | 0.02 | 0.02 | 32 | 10 | 8 | 1.00 | 0.92 | 0.92 |
| M35M34 | 16 | 16 | 8 | 0.84 | 0.04 | 0.02 | 16 | 10 | 8 | 0.90 | 0.85 | 0.68 |
| M36M37 | 16 | 16 | 5 | 0.93 | 0.12 | 0.10 | 32 | 2 | 5 | 0.94 | 0.95 | 0.84 |
| M37M36 | 16 | 16 | 7 | 0.92 | 0.04 | 0.03 | 16 | 2 | 5 | 0.99 | 0.58 | 0.57 |
| M38M39 | 16 | 16 | 8 | 0.99 | 0.26 | 0.25 | 64 | 10 | 7 | 0.99 | 0.99 | 0.97 |
| M39M38 | 16 | 16 | 8 | 0.86 | 0.10 | 0.07 | 16 | 1 | 8 | 0.98 | 0.96 | 0.92 |
| M40M41 | 16 | 16 | 8 | 0.91 | 0.06 | 0.05 | 16 | 10 | 8 | 0.82 | 0.61 | 0.39 |
| M41M40 | 48 | 32 | 5 | 0.78 | 0.20 | 0.11 | 16 | 10 | 8 | 0.67 | 0.24 | 0.08 |
| M42M43 | 16 | 16 | 7 | 0.74 | 0.06 | 0.03 | 16 | 5 | 8 | 0.85 | 0.84 | 0.60 |
| M43M42 | 16 | 32 | 5 | 0.88 | 0.22 | 0.17 | 16 | 10 | 8 | 0.90 | 0.77 | 0.61 |
| M44M45 | 16 | 16 | 8 | 0.95 | 0.35 | 0.32 | 96 | 1 | 5 | 0.93 | 0.99 | 0.86 |
| M45M44 | 16 | 16 | 8 | 0.87 | 0.06 | 0.04 | 32 | 1 | 6 | 0.92 | 1.00 | 0.84 |
| M46M47 | 16 | 16 | 7 | 0.85 | 0.04 | 0.03 | 16 | 2 | 8 | 0.84 | 0.43 | 0.30 |
| M47M46 | 16 | 32 | 6 | 0.91 | 0.13 | 0.11 | 16 | 10 | 5 | 0.62 | 0.26 | 0.06 |
| M48M49 | 32 | 16 | 5 | 0.73 | 0.24 | 0.11 | 48 | 1 | 6 | 0.13 | -0.01 | -0.01 |
| M49M48 | 16 | 32 | 6 | 0.90 | 0.29 | 0.23 | 16 | 10 | 8 | 0.29 | 0.00 | 0.00 |
| M50M51 | 32 | 16 | 5 | 0.93 | 0.12 | 0.10 | 32 | 10 | 5 | 0.90 | 0.03 | 0.02 |
| M51M50 | 16 | 32 | 6 | 0.87 | 0.08 | 0.06 | 16 | 2 | 8 | 0.61 | 0.70 | 0.16 |
| M52M53 | 32 | 32 | 5 | 0.94 | 0.24 | 0.21 | 16 | 1 | 5 | 0.75 | 0.10 | 0.05 |
| M53M52 | 16 | 16 | 5 | 0.80 | 0.08 | 0.05 | 16 | 2 | 5 | 0.53 | 0.36 | 0.02 |
| M54M55 | 16 | 16 | 7 | 0.98 | 0.40 | 0.38 | 48 | 1 | 7 | 1.00 | 0.99 | 0.99 |
| M55M54 | 16 | 16 | 7 | 0.99 | 0.75 | 0.73 | 64 | 5 | 7 | 0.99 | 1.00 | 0.98 |
| M56M57 | 32 | 32 | 6 | 0.89 | 0.15 | 0.12 | 32 | 2 | 5 | 0.25 | -0.12 | -0.06 |
| M57M56 | 16 | 32 | 6 | 0.79 | 0.18 | 0.11 | 16 | 10 | 8 | 0.41 | 0.00 | 0.00 |
| M58M59 | 16 | 128 | 8 | 0.34 | 0.01 | 0.00 | 32 | 5 | 6 | 0.23 | 0.00 | 0.00 |
| M59M58 | 16 | 64 | 7 | 0.79 | 0.08 | 0.04 | 32 | 10 | 5 | 0.86 | 0.02 | 0.02 |
| M60M61 | 16 | 16 | 8 | 0.95 | 0.09 | 0.08 | 16 | 2 | 8 | 0.98 | 0.70 | 0.67 |
| M61M60 | 16 | 32 | 8 | 0.89 | 0.01 | 0.01 | 16 | 10 | 7 | 0.97 | 0.52 | 0.48 |
| M62M63 | 16 | 16 | 7 | 0.69 | 0.27 | 0.10 | 16 | 2 | 5 | 0.68 | 0.98 | 0.36 |
| M63M62 | 16 | 64 | 6 | 0.88 | 0.17 | 0.13 | 48 | 1 | 8 | 0.60 | 0.06 | 0.01 |
| M64M65 | 16 | 16 | 7 | 0.90 | 0.08 | 0.07 | 16 | 2 | 7 | 0.86 | 0.92 | 0.65 |
| M65M64 | 48 | 16 | 5 | 0.83 | 0.16 | 0.11 | 16 | 10 | 5 | 0.62 | 0.12 | 0.03 |
| M66M67 | 48 | 32 | 5 | 0.92 | 0.27 | 0.22 | 16 | 5 | 7 | 0.88 | 0.29 | 0.22 |
| M67M66 | 48 | 16 | 5 | 0.72 | 0.32 | 0.14 | 32 | 2 | 7 | 0.72 | 0.96 | 0.42 |

**Table A.5**: Matching parameters and evaluation results for the best matches of TM-matching applied to pairs of SB mammograms, listed for each image pair.

| Images | TM-$D_E$ | | | | | | TM-$D_{ES}$ | | | | | | TM-$D_M$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $w$ | $d$ | $nbits$ | $A_{ROC}$ | $C_{fb}$ | $\kappa$ | $w$ | $d$ | $nbits$ | $A_{ROC}$ | $C_{fb}$ | $\kappa$ | $w$ | $d$ | $nbits$ | $A_{ROC}$ | $C_{fb}$ | $\kappa$ |
| S1S0 | 96 | 1 | 5 | 0.56 | 0.24 | 0.03 | 64 | 10 | 6 | 0.75 | 0.22 | 0.11 | 96 | 1 | 5 | 0.56 | 0.24 | 0.03 |
| S1S2 | 112 | 1 | 8 | 0.84 | 0.15 | 0.1 | 96 | 10 | 5 | 0.83 | 0.09 | 0.06 | 96 | 2 | 5 | 0.53 | 1 | 0.05 |
| S4S3 | 16 | 1 | 8 | 0.68 | 0.15 | 0.05 | 16 | 10 | 5 | 0.61 | 0.45 | 0.1 | 80 | 2 | 7 | 0.94 | 0.06 | 0.06 |
| S4S5 | 32 | 2 | 8 | 0.62 | 0.69 | 0.17 | 16 | 10 | 5 | 0.77 | 0.69 | 0.37 | 64 | 5 | 6 | 0.94 | 0.02 | 0.02 |
| S7S6 | 64 | 10 | 5 | 0.73 | 0.05 | 0.02 | 96 | 2 | 8 | 0.85 | 0.16 | 0.11 | 128 | 10 | 8 | 0.76 | 0.06 | 0.03 |
| S7S8 | 64 | 10 | 7 | 0.84 | 0.31 | 0.21 | 80 | 2 | 8 | 0.92 | 0.19 | 0.16 | 96 | 2 | 8 | 0.81 | 0.03 | 0.02 |
| S10S9 | 128 | 2 | 7 | 0.82 | 0.45 | 0.29 | 176 | 1 | 5 | 0.76 | 0.15 | 0.08 | 176 | 5 | 5 | 0.86 | 0.29 | 0.21 |
| S10S11 | 16 | 5 | 8 | 0.86 | 0.58 | 0.42 | 32 | 10 | 5 | 0.7 | 0.4 | 0.16 | 64 | 10 | 5 | 0.79 | 0.2 | 0.11 |
| S13S12 | 16 | 1 | 5 | 0.89 | 0.25 | 0.19 | 16 | 2 | 5 | 0.76 | 0.18 | 0.1 | 16 | 1 | 5 | 0.43 | 0 | 0 |
| S13S14 | 16 | 10 | 5 | 0.94 | 0.37 | 0.32 | 16 | 2 | 5 | 0.67 | 0.43 | 0.15 | 16 | 1 | 6 | 0.49 | 0 | 0 |
| S16S15 | 16 | 2 | 8 | 0.76 | 0.68 | 0.35 | 16 | 1 | 8 | 0.9 | 0.13 | 0.1 | 16 | 1 | 5 | 0.32 | 0 | 0 |
| S16S17 | 16 | 10 | 7 | 0.59 | 0.01 | 0 | 16 | 5 | 6 | 0.73 | 0.16 | 0.08 | 16 | 1 | 5 | 0.37 | 0 | 0 |
| S19S18 | 16 | 5 | 5 | 0.66 | 0.15 | 0.05 | 16 | 10 | 6 | 0.84 | 0.05 | 0.03 | 16 | 1 | 6 | 0.48 | 0 | 0 |
| S19S20 | 16 | 5 | 6 | 0.75 | 0.17 | 0.09 | 16 | 2 | 6 | 0.88 | 0.13 | 0.1 | 16 | 1 | 5 | 0.39 | 0 | 0 |
| S22S21 | 16 | 10 | 8 | 0.68 | 0.9 | 0.32 | 16 | 10 | 8 | 0.78 | 0.28 | 0.16 | 48 | 2 | 5 | 0.54 | 1 | 0.08 |
| S22S23 | 16 | 1 | 8 | 0.85 | 0.64 | 0.45 | 176 | 10 | 5 | 0.82 | 0.3 | 0.19 | 176 | 1 | 5 | 0.82 | 0.17 | 0.11 |
| S25S24 | 16 | 10 | 7 | 0.7 | 0.2 | 0.08 | 16 | 1 | 7 | 0.69 | 0.08 | 0.03 | 80 | 1 | 8 | 0.68 | 0.02 | 0.01 |
| S25S26 | 16 | 10 | 8 | 0.7 | 0.56 | 0.22 | 96 | 10 | 8 | 0.94 | 0.3 | 0.27 | 96 | 10 | 8 | 0.94 | 0.23 | 0.2 |
| S28S27 | 96 | 1 | 6 | 0.86 | 0.63 | 0.45 | 96 | 1 | 6 | 0.89 | 0.5 | 0.39 | 96 | 10 | 5 | 0.89 | 0.48 | 0.37 |
| S28S29 | 16 | 2 | 5 | 0.69 | 0.97 | 0.37 | 16 | 5 | 5 | 0.67 | 0.64 | 0.22 | 32 | 1 | 5 | 0.55 | 1 | 0.09 |
| S31S30 | 32 | 1 | 6 | 0.7 | 0.22 | 0.09 | 16 | 1 | 7 | 0.72 | 0.11 | 0.05 | 96 | 2 | 7 | 0.72 | 0.06 | 0.02 |
| S31S32 | 16 | 2 | 7 | 0.72 | 0.35 | 0.15 | 48 | 10 | 5 | 0.82 | 0.21 | 0.14 | 64 | 1 | 5 | 0.89 | 0.08 | 0.06 |
| S34S33 | 96 | 10 | 5 | 0.76 | 0.17 | 0.09 | 64 | 1 | 5 | 0.83 | 0.36 | 0.24 | 48 | 10 | 7 | 0.73 | 0.04 | 0.02 |
| S34S35 | 96 | 10 | 5 | 0.94 | 0.25 | 0.22 | 112 | 10 | 5 | 0.86 | 0.26 | 0.19 | 128 | 1 | 8 | 0.76 | 0.04 | 0.02 |

**Table A.6**: Matching parameters and evaluation results for the best matches of MI-matching applied to pairs of SB mammograms, listed for each image pair.

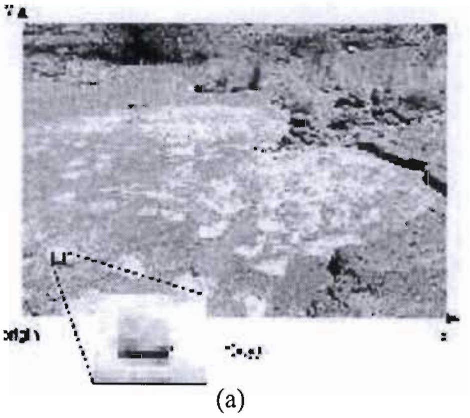
| Images | MI-Histograms | | | | | | MI-GLCMs | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $w$ | $nbins$ | $nbits$ | $A_{ROC}$ | $C_{fb}$ | $\kappa$ | $w$ | $d$ | $nbits$ | $A_{ROC}$ | $C_{fb}$ | $\kappa$ |
| S1S0 | 112 | 16 | 8 | 0.59 | 0.23 | 0.04 | 32 | 1 | 8 | 0.95 | 0.67 | 0.6 |
| S1S2 | 96 | 16 | 8 | 0.45 | 0.08 | -0.01 | 32 | 1 | 7 | 0.96 | 0.86 | 0.8 |
| S4S3 | 32 | 32 | 5 | 0.65 | 0.65 | 0.19 | 16 | 10 | 8 | 0.74 | 0.76 | 0.36 |
| S4S5 | 16 | 16 | 5 | 0.71 | 0.86 | 0.37 | 16 | 1 | 6 | 0.9 | 0.59 | 0.47 |
| S7S6 | 32 | 16 | 8 | 0.57 | 0.06 | 0.01 | 32 | 1 | 7 | 0.7 | 0.92 | 0.37 |
| S7S8 | 48 | 16 | 8 | 0.69 | 0.05 | 0.02 | 32 | 1 | 5 | 0.75 | 0.57 | 0.29 |
| S10S9 | 80 | 16 | 8 | 0.63 | 0.21 | 0.06 | 32 | 5 | 8 | 0.96 | 1 | 0.92 |
| S10S11 | 96 | 16 | 8 | 0.64 | 0.16 | 0.05 | 64 | 10 | 8 | 0.99 | 0.93 | 0.9 |
| S13S12 | 16 | 32 | 5 | 0.6 | 0.26 | 0.05 | 16 | 1 | 5 | 0.36 | 0 | 0 |
| S13S14 | 16 | 32 | 5 | 0.56 | 0.8 | 0.1 | 16 | 5 | 6 | 0.54 | 0.01 | 0 |
| S16S15 | 16 | 32 | 8 | 0.47 | -0.17 | -0.01 | 16 | 5 | 5 | 0.27 | 0 | 0 |
| S16S17 | 16 | 16 | 5 | 0.82 | 0.29 | 0.19 | 16 | 2 | 7 | 0.77 | 0.07 | 0.04 |
| S19S18 | 16 | 16 | 8 | 0.68 | 0 | 0 | 16 | 10 | 5 | 0.5 | -0.51 | 0 |
| S19S20 | 16 | 16 | 8 | 0.45 | 0.04 | 0 | 16 | 1 | 5 | 0.65 | 0.15 | 0.04 |
| S22S21 | 288 | 16 | 7 | 0.91 | 0.47 | 0.39 | 16 | 1 | 5 | 0.46 | 0 | 0 |
| S22S23 | 256 | 32 | 5 | 0.85 | 0.24 | 0.17 | 48 | 1 | 6 | 0.88 | 0.74 | 0.57 |
| S25S24 | 32 | 16 | 6 | 0.5 | 0.08 | 0 | 16 | 1 | 5 | 0.49 | 0 | 0 |
| S25S26 | 80 | 32 | 5 | 0.7 | 0.42 | 0.17 | 16 | 5 | 8 | 0.84 | 0.9 | 0.61 |
| S28S27 | 32 | 16 | 8 | 0.75 | 0.26 | 0.13 | 16 | 10 | 6 | 0.76 | 0.61 | 0.31 |
| S28S29 | 32 | 16 | 8 | 0.47 | 0.03 | 0 | 16 | 1 | 8 | 0.83 | 0.14 | 0.1 |
| S31S30 | 96 | 16 | 8 | 0.56 | 0.25 | 0.03 | 16 | 1 | 5 | 0.49 | 0 | 0 |
| S31S32 | 96 | 16 | 7 | 0.48 | 0.22 | -0.01 | 16 | 1 | 5 | 0.49 | 0 | 0 |
| S34S33 | 144 | 16 | 6 | 0.82 | 0.28 | 0.18 | 16 | 1 | 5 | 0.5 | 0 | 0 |
| S34S35 | 48 | 16 | 7 | 0.63 | 0.3 | 0.08 | 16 | 1 | 5 | 0.47 | 0 | 0 |

# Appendix B

# About Images

In this research an *image* refers to a two-dimensional light intensity function $f(x,y)$ where $x$ and $y$ denote spatial co-ordinates and $f$ is proportional to the brightness or intensity or grey-level of the image at that point. A digital image is an image that has been discretised in both spatial co-ordinates and brightness and is represented in software by a matrix whose row and column indices indicate a point in the image. The elements of the matrix are referred to as pixels [Gonzalez & Wintz 1987].

**Figure B.1** shows an example of an image together with the axis convention used. The origin is taken to be at the bottom left corner of the image, following the convention used in IDL.



$$f_{x,y} = \begin{matrix} f_{N-1,0} & \cdots & f_{N-1,M-1} \\ & \vdots & \\ f_{0,0} & \cdots & f_{0,M-1} \end{matrix}$$

(a) (b)

**Figure B.1**: **(a)** Example of an image together with the axis convention used, with the origin at the bottom left corner. **(b)** Pixels are discrete spatial elements in a digital image and are represented by the elements in a matrix.

The bit-depth of an image indicates how many shades of grey are available to describe the information in an image. The relation between bit-depth and brightness or grey-levels of the image is shown in **Table B.1**.

**Table B.1**: Bit-depth, *nbits*, of image *vs.* number of grey-levels

| number of bits *nbits* | number of grey-levels $2^{nbits}$ |
|---|---|
| 1 | 2 |
| 2 | 4 |
| 3 | 8 |
| 4 | 16 |
| 8 | 256 |
| 14 | 16 384 |
| 16 | 65 536 |

# Appendix C

# Information Theory

## Probability Theory Basics

The following is a summary of probability theory notation and terminology [Shlens 2003].

### Notation and Variables

Let $X$ and $Y$ be two discrete random variables, with a finite number of states, $n$, denoted by $[x_0, x_1, x_2, \ldots, x_{n-1}]$ and $[y_0, y_1, y_2, \ldots, y_{n-1}]$. $X$ is governed by a discrete probability distribution, that is an assignment of a probability $p(x_i)$ to each state $x_i$, denoted by $P(X)$. Similarly, $Y$ is governed by $P(Y)$. The joint probability between $X$ and $Y$ is $P(X \cdot Y)$.

### Marginal Probability, $p(x_i)$

$$p(x_i) = \sum_{j=0}^{m-1} p(x_i \cdot y_j) \tag{C-1}$$

### Conditional Probability, $p(x_i|Y)$

$$p(x_i|Y) = \sum_{j=0}^{m-1} p(y_j)p(x_i \cdot y_j) \tag{C-2}$$

## Entropy

Entropy was introduced as a measure of the uncertainty associated with a set of probabilities and is defined by Shannon [1948, p18] to be:

$$H(X) = -\sum_{i=0}^{n-1} p(x_i) \log p(x_i). \tag{C-3}$$

### Properties of Entropy

The following properties of entropy make it a reasonable measure of uncertainty [Shannon 1948]

1. $H \geq 0$ except if the outcome is certain then $H = 0$. This corresponds to the situation where only one, known outcome is possible.
2. If $p(x_i) = 1/t$ (for a given $t$) then $H = \ln t$ is a maximum. This corresponds to the most uncertain situation.
3. $H(X \cdot Y) \leq H(X) + H(Y)$ for two quantities $X$ and $Y$, with equality if and only if both quantities are independent. This shows that the uncertainty of two quantities considered jointly is less than the sum of the uncertainties of the individual quantities

4. Any change in probabilities that causes them to be more equal results in an increase in $H$

Additionally, with $H(X \cdot Y)$, the joint entropy, and $f$ a function, the following applies [Jumarie 1990]:

$$\begin{aligned}
H(X|Y)+H(Y) &= H(X \cdot Y) && \\
H(X|Y) &\leq H(X) && \text{(conditioning reduces entropy)} \\
H(X|Y) &= H(Y|X) && \text{(symmetry)} \\
H(X|Y) &\leq H(X)+H(Y) && \text{(equality iff } X,\, Y \text{ independent)} \\
H(X|Y) &\leq H(X|f(Y)) && \\
H(X|Y) &= 0 \iff X = f(Y) && \text{(special case } H(X|X)=0\text{)}
\end{aligned}$$

## Entropy Examples

Consider the tossing of a fair coin. In this situation there is an equal probability of getting heads or tails. Let $X = [x_0, x_1]$ represent the set of possible results of tossing the coin with $x_0 = \textit{heads}$, $x_1 = \textit{tails}$. The probability density function associated with $X$ is then $P(X) = [0.5, 0.5]$. The entropy associated with $X$ is:

$$H(X) = -\sum_{i=0}^{n-1} p(x_i)\log p(x_i) = -2(0.5\log_2 0.5) = 1,$$

which is maximal. If however the coin is slightly more biased to 'heads', then $P(X) = [0.6, 0.4]$ and

$$H(X) = -\sum_{i=0}^{n-1} p(x_i)\log p(x_i) = -(0.4\log_2 0.4) - (0.6\log_2 0.6) = 0.87.$$

These examples demonstrate that entropy is maximal when there is the greatest choice or when all probabilities are equal.

## Conditional Entropy

Let $X$ and $Y$ be discrete random variables, with $P(X|Y)$ the conditional probability. The conditional entropy $H(X|Y)$, which is a measure of the uncertainty in $X$ given knowledge of $Y$, is then defined as [Jumarie 1990]:

$$H(X|Y) = -\sum_{i=0}^{n-1}\sum_{j=0}^{n-1} p(x_i \cdot y_j)\log p(x_i|y_j) \tag{C-4}$$

# Appendix D

# Description of some IDL Functions

For convenience, descriptions of some IDL functions are given here.

## HIST_2D

The `HIST_2D` function is explained by means of an example. Consider two images, $I_1$ and $I_2$ and the corresponding two-dimensional histogram, $P_{I_1,I_2}$.

$$I_1 = \begin{matrix} 0 & 1 & 2 \\ 0 & 1 & 2 \\ 0 & 2 & 1 \end{matrix} \quad I_2 = \begin{matrix} 1 & 2 & 0 \\ 2 & 1 & 0 \\ 2 & 0 & 0 \end{matrix} \quad P_{I_1,I_2} = \texttt{HIST\_2D}(I_1, I_2) = \begin{bmatrix} 2 & 0 & 1 \\ 2 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix}$$

$P_{I_1,I_2}$ is built up as follows:

1. $I_1[0,0] = 0$ and $I_2[0,0] = 2$, so $P_{I_1,I_2}[0,2]$ gets incremented.
2. $I_1[1,0] = 1$ and $I_2[1,0] = 0$, so $P_{I_1,I_2}[1,0]$ gets incremented.
3. $I_1[2,0] = 0$ and $I_2[2,0] = 2$, so $P_{I_1,I_2}[0,2]$ gets incremented.
4. $I_1[0,1] = 1$ and $I_2[0,1] = 0$, so $P_{I_1,I_2}[1,0]$ gets incremented.
5. $I_1[1,1] = 2$ and $I_2[1,1] = 2$, so $P_{I_1,I_2}[2,2]$ gets incremented.
6. $I_1[2,1] = 1$ and $I_2[2,1] = 1$, so $P_{I_1,I_2}[1,1]$ gets incremented.
7. $I_1[0,2] = 2$ and $I_2[0,2] = 0$, so $P_{I_1,I_2}[2,0]$ gets incremented.
8. $I_1[1,2] = 0$ and $I_2[1,2] = 1$, so $P_{I_1,I_2}[0,1]$ gets incremented.
9. $I_1[2,2] = 2$ and $I_2[2,2] = 0$, so $P_{I_1,I_2}[2,0]$ gets incremented.

## SHIFT

From IDL help:

> The `SHIFT` function shifts elements of vectors or arrays along any dimension by any number of elements. Positive shifts are to the right while left shifts are expressed as a negative number. All shifts are circular.
>
> Elements shifted off one end wrap around and are shifted onto the other end. In the case of vectors the action of `SHIFT` can be expressed as:
>
> $$Result_{(i+s)modulation} = Array_i \text{ for } 0 \leq i < n$$
>
> where $s$ is the amount of the shift, and $n$ is the number of elements in the array.

# Appendix E

# Presentations

## $42^{nd}$ SAAPMB Congress and Winter School, 2002

30 July - 2 August 2002, Mount Amanzi, Krugersdorp, South Africa (Oral Presentation)

### Overview of artificial neural network and Texture Based Techniques for Computer Aided Diagnosis in Mammography

J. Padayachee, W. I. D. Rae*, M. Gar-Elnabi and M. J. Alport

University of Natal, Durban;*Addington Hospital, Durban

According to the Cancer Association of South Africa (http://www.cansa.co.za/), breast cancer is currently the most common cancer among women worldwide, including South Africa. Although much progress has been made in the treatment of breast cancer, the key is early detection. Mammography is currently the most effective method of detecting breast cancer in its early stages, but the analysis of mammograms is sometimes difficult due to the complex and varying structure of the human breast. Adding to the complexity are benign masses (e.g. lipomas), which may be palpable and have abnormal structure compared with the surrounding tissue. The ability to reliably diagnose these benign masses will reduce cost and patient trauma because some biopsies may be safely avoided.

Computer-aided diagnosis (CAD) systems are a vital aid in the analysis of mammograms to highlight and classify those features that may be missed by a radiologist, largely because of fatigue and distractions. Kegelmeyer et al (1994) and Zheng et al (2001) show that the performance of radiologists in analysing mammograms are enhanced when "prompted" by a CAD system.

There are currently three CAD systems that have been approved by the U. S. Food and Drug Administration (FDA) for use in screening and diagnostic mammography: ImageChecker™(R2 Technology), MammoReader™(Intelligent Systems Software, Inc) and Second Look™(CADx Medical Systems). ImageChecker™uses pattern recognition algorithms to identify spiculated masses and microcalcifications, and artificial neural networks to distinguish lesions from normal tissue. In a study of 12 860 screening mammograms, Freer and Ulissey (2001) compared radiologists performance with and without ImageChecker™and found that radiologists detected 19.5% more cancers, and there was an increase (from 73% to 78%) in the proportion of early-stage malignancies detected in a prompted environment.

In FDA (http://www.fda.gov/) clinical studies, it was found that 23% of the women diagnosed with breast cancer, who had had prior screening mammograms, could have had their cancers discovered an 14 months earlier with MammoReader™while Second Look™(using pattern recognition algorithms) reduced the number of missed cancers by 26.2%.

This paper will focus on some techniques used in the detection and classification of benign masses. Techniques that are reviewed in detail include enhancing the contrast of suspicious areas, segmentation to separate the pectoral muscle and the background (non-breast area) from the breast tissue in a mammogram, comparison of left and right breasts using symmetry and texture analysis.

1. Freer, T. W. and Ulissey, M. J. Screening mammography with computer-aided detection: prospective study of 12 860 patients in a community breast center. Radiology 220 (2001) 781-786
2. Kegelmeyer, W. P., Pruneda, J. M., Bourland, P. D., Hillis, A., Riggs, M. W. and Nipper, M. L. Computer-aided Mammographic Screening for Spiculated Lesions. Radiology 191 (1994) 331-337
3. Zheng, B., Ganott, M. A., Britton, C. A., Hakim, C. M., Hardesty, L. A., Chang, T. S., Rockette, H. E., Gur, D. Soft-copy mammographic readings with different computer-assisted detection cuing environments: Preliminary findings. Radiology 221 (2001) 633-40.

## $43^{rd}$ Annual SAAPMB Congress and Winter School, 2003

iThemba LABS, Cape Town, South Africa (Poster Presentation)

### Computer-Aided Diagnosis in Mammography - Segmenting the Breast Border

J. Padayachee[1], M. J. Alport[1], W. I. D. Rae[2]

[1] Applied Physics Group, School of Pure and Applied Physics, University of Natal, Durban, South Africa

[1] Medical Physics Department, Addington Hospital, Durban, South Africa

According to the Cancer Association of South Africa (http://www.cansa.org.za/), breast cancer is currently the most common cancer among women worldwide. Although much progress has been made in its treatment, the key is early detection. Mammography is currently the most effective method of detecting breast cancer in its early stages, but the analysis of mammograms is sometimes difficult due to the complex and varying structure of the human breast. Adding to the complexity are benign masses (e.g. lipomas), which may be palpable and have abnormal structure compared with the surrounding tissue. The ability to reliably diagnose these benign masses will reduce cost and patient trauma because some biopsies may be safely avoided.

Computer-aided diagnosis (CAD) is a vital aid in the analysis of mammograms to highlight and classify those features that may be missed by a radiologist, largely due to fatigue and distractions. Kegelmeyer et al (1994) and Zheng et al (2001) show that the performance of radiologists is enhanced when "prompted" by a CAD system.

An important step in the automatic analysis of a mammogram is the segmentation of different parts of the breast (i.e. pectoral muscle, fatty tissue, glandular tissue, etc) for further processing. One of the first steps is to segment the breast region from the rest of the mammogram. This is done to eliminate the background and to identify the breast border. Certain signs of malignancy manifest themselves in changes in the skin, causing a change to the outline of the breast, which can be identified if the breast border is known. Finding the breast border also enables the selection of reference points that are required for the registration of mammograms, especially when two mammographic views are being compared or when radiographs are being compared over time (Masek et al, 2000).

Boundary tracking (Mendez et al, 1996) and grey-level thresholding (Chandrasekhar and Attikiouzel, 2000) are two common methods of detecting the breast border. This study uses grey-level thresholding, but also an analysis of the area enclosed by contours of varying grey-levels. Various methods have been investigated to extract the breast border including an analysis of the contour bending energy and enclosed area.

Results obtained from using the contour bending energy were inconclusive, but analysing the contour area yielded positive results. The algorithms were tested on a data set that contained an extreme of mammography examples. Fitted breast borders are compared with borders determined manually by a radiologist. The radiologist's borders were absolutely registered with the fitted results and the mean error was determined by taking differences in radial distances from a central point. Results using the derivative are found to be between 0.13cm to 1.23cm of the radiologist's borders whereas the results using an empirically fitted function are found to be between 0.3cm and 1.95cm of the radiologist's borders.

The techniques are completely automatic and hence can be used to segment the breast outline without user intervention.

*References*

1. Chandrasekhar, R. and Attikiouzel, Y. Segmenting the Breast Border and Nipple on Mammograms. Australian Journal of Intelligent Information Processing Systems 6 (2000) 24-29
2. Kegelmeyer, W. P., Pruneda, J. M., Bourland, P. D., Hillis, A., Riggs, M. W. and Nipper, M. L. Computer-aided Mammographic Screening for Spiculated Lesions. Radiology 191 (1994) 331-337
3. Masek, M., Attikiouzel, Y and de Silva, C. J. S. Combining Data from Different Algorithms to Segment the Skin-Air Interface in Mammograms. CD-ROM Proceedings of the World Congress on Medical Physics and Biomedical Engineering, 23-28 July 2000.
4. Mendez AJ, Tahoces PG, Lado MJ, Souto M, Correa JL, Vidal JJ. Automatic detection of breast border and nipple in digital mammograms. Comput Methods Programs Biomed; 49 (1996) 253-262.
5. Zheng, B., Ganott, M. A., Britton, C. A., Hakim, C. M., Hardesty, L. A., Chang, T. S., Rockette, H. E., Gur, D. Soft-copy mammographic readings with different computer-assisted detection cuing environments: Preliminary findings. Radiology 221 (2001) 633-40.

## $1^{st}$ Annual Post Graduate Research Day, 2003

University of Natal, 9 August 2003, Durban, South Africa (Oral Presentation)

### Computer-Aided Diagnosis in Mammography-Segmenting the Breast Border

J. Padayachee, M. J. Alport, W. I. D. Rae[1]

Applied Physics Group, School of Pure and Applied Physics, University of Natal, Durban

[1]Medical Physics Department, Addington Hospital, Durban

According to the Cancer Association of South Africa, breast cancer is currently the most common cancer among women worldwide. Although much progress has been made in its treatment, the key is early detection. Mammography is currently the most effective method of detecting breast cancer in its early stages, but the analysis of mammograms is sometimes difficult due to the complex and varying structure of the human breast. Computer-aided diagnosis (CAD) is a vital aid in the analysis of mammograms to highlight and classify those features that may be missed by a radiologist, largely due to fatigue and distractions. An important first step in the automatic analysis of mammograms is the segmentation of the breast region from the rest of the mammogram to eliminate the background and to identify the breast border. Certain signs of malignancy manifest themselves in changes in the skin, causing a change to the outline of the breast, which can be identified if the breast border is known. Finding the breast border also enables the selection of reference points that are required for the registration of mammograms, especially when two mammographic views are being compared or when radiographs are being compared over time. This study uses grey-level thresholding by analysing various characteristics of contours at different grey-scale values. Various methods have been investigated to extract the breast border including an analysis of the contour bending energy and enclosed area. Results obtained from using the contour bending energy require further investigation, but analysing the contour area yielded positive results. The algorithms were tested on a data set that contained some extreme mammography examples with marked contrast, and some very dense structures. Fitted breast borders were compared with borders determined manually by a radiologist. The radiologist's borders were absolutely registered with the fitted results and the mean error was determined by taking differences in radial distances from a central point. Results range between 0.13cm and 1.95cm of the radiologist's borders. The techniques are completely automatic and hence can be used to segment the breast outline without user intervention.

## $49^{th}$ Annual Conference of the SAIP, 2004

29 June to 2 July 2004, Bloemfontein, South Africa (Oral Presentation)

### Matching Features in Two Mammographic Views of the Same Breast Using Image Analysis Techniques

J. Padayachee[1], M. J. Alport[1], and W. I. D. Rae[2], 1

[1] School of Pure and Applied Physics, University of KwaZulu-Natal, Durban

[2] Department of Nuclear Medicine, Inkosi Albert Luthuli Central Hospital, Durban

Breast cancer can be treated successfully if it is detected early. Mammography is used to image the breast to detect cancer. Generally two mammographic views of each breast are visually analysed by two radiologists for signs of breast cancer. For various reasons, about 10-20% of cancers are missed in current analyses. Computer aided diagnosis (CAD) is being developed in an attempt to improve diagnosis. Each mammographic view is a projection of a compressed breast and distortions and overlapping tissue means that there is no global transform to match features between both views. Towards achieving CAD, this investigation demonstrates the use of a statistical similarity metric (mutual information) to match features (e.g. tumours) between two views of the same breast. This allows meaningful comparison of areas containing comparable information and thus simultaneously assessing both views for signs of malignancy. Examples of the successful matching of similar areas are presented.

# Glossary

**acutance** Measure of the gradient between light and dark regions in an image.

**bit-depth** Maximum number of grey-levels or brightness or intensities that can be assigned to a pixel in an image. A bit-depth of $nbits$ means that $2^{nbits}$ grey-levels are available to describe the information content of an image.

**centroid** The point within an area or volume at which the centre of mass would be if the surface or body had a uniform density. For a symmetrical area or volume it coincides with the centre of mass. For a non-symmetrical area or volume it has to be found by integration.

**diagnostic mammogram** A mammogram taken of a patient experiencing physical symptoms consistent with breast cancer, or taken to evaluate a specific finding.

**fractal dimension** Rate at which the perimeter or surface area of an object increases as the measurement scale gets smaller. A small fractal dimension implies a fine texture while a large value implies a coarse texture.

**lacunarity** Measure that describes characteristics of textures that have the same fractal dimension, but different visual appearance.

**mammogram** X-ray image of human breast

**mammography** Method of using low energy x-rays to image the breast for the diagnosis of breast cancer

**matching accuracy** Combination of $A_{ROC}$- and $C_{fb}$-values used as an indication of how accurate a match is and how easily discernable the matched area is, from the background, in the matching map.

**modality** A form or method of imaging

**recall** Getting the patient to return for a repeat examination for some reason.

**screening mammogram** A mammogram of a woman not exhibiting physical symptoms of breast cancer, and who was not undergoing further evaluation for a specific finding, at the time of the exam. Screening mammograms are routinely taken.

**segmentation** The isolating of specific features

**sensitivity** Fraction of abnormal cases actually diagnosed as abnormal

**spatial resolution** Physical size that a pixel in an image represents.

**specificity** Fraction of normal cases actually diagnosed as normal

**spiculated lesion** Lesion with lines radiating from a central mass.

**ultrasonography** Diagnostic imaging in which ultrasound is used to image an internal body structure. Also known as ultrasound.

# Bibliography

Abramson, N. [1963], *Information Theory and Coding*, McGraw-Hill Book Company, New York, U.S.A.

Alberdi, E., Povyakalo, A., Strigini, L. & Ayton, P. [2004], 'Effects of incorrect computer-aided detection (CAD) output on human decision-making in mammography', *Acad. Radiol.* **11**, 909–918.

Anguh, M. & Silva, A. [1997], Multiscale segmentation and enhancement in mammograms, *in* L. de Figueiredo & L. N.M., eds, 'Proceedings of the 10th Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI97)'.

Astley, S. & Gilbert, F. [2004], 'Computer-aided detection in mammography', *Clin. Radiol.* **59**, 390–399.

Azar, F., Metaxas, D. & Schnall, M. [2001], 'A Deformable finite element model of the breast for predicting mechanical deformations under external perturbations', *Acad. Radiol.* **8**, 965–975.

Bakic, P. & Brzakovic, D. [1997], 'Application of Artificial Neural Networks in Computer aided diagnosis of breast cancer'.
`http://citeseer.nj.nec.com/404677.html`

Bassett, L. [2000], 'Digital and computer-aided mammography', *Breast J.* **6**, 291–293.

Beam, C., Conant, E. & Sickles, E. [2002], 'Factors affecting radiologist inconsistency in screening mammography', *Acad. Radiol.* **9**, 531–540.

Berg, W., D'Orsi, C., Jackson, V., Bassett, L., Beam, C., Lewis, R. & Crewson, P. [2002], 'Does training in the breast imaging reporting and data system (BI-RADS) improve biopsy recommendations or feature analysis agreement with experienced breast imagers at mammography?', *Radiol.* **224**, 870–880.

Bick, U., Giger, M., Schmidt, R., Nishikawa, R., Wolverton, D. & Doi, K. [1995], 'Automated segmentation of digitized mammograms', *Acad. Radiol.* **2**, 1–9.

Blot, L. & Zwiggelaar, R. [2001], Background texture extraction for the classification of mammographic parenchymal patterns, *in* Marlow [2001].
`http://www.cs.bham.ac.uk/research/proceedings/miua2001/`

Bocchi, L., Coppini, G., Nori, J. & Valli, G. [2004], 'Detection of single and clustered microcalcifications in mammograms using fractals models and Artificial Neural Networks', *Med. Eng. Phys.* **26**, 303–312.

Bovis, K. & Singh, S. [2000], Detection of Masses in Mammograms using Texture Features, *in* A. Sanfeliu, J. Villanueva, M. Vanrell, R. Alquezar, A. Jain & J. Kittler, eds, 'Proceedings of the 15th International Conference on Pattern Recognition, 3-7 September 2000, Barcelona, Spain', Vol. 2, IEEE, pp. 267–270.

Bovis, K. & Singh, S. [2002], 'An Evaluation of Contrast Enhancement Techniques for Mammographic Breast masses', *IEEE T. Med. Imaging (submitted, 2002)* .

Bovis, K., Singh, S., Fieldsend, J. & Pinder, C. [2000], Identification of masses in digital mammogram with MLP and RBF Nets, *in* A. Shun-Ichi, C. Giles, M. Gori & V. Piuri, eds, 'Proceedings of the IEEE-INNS-ENNS International Joint Conference on Artificial Neural Networks (IJCNN 2000), 24-27 July 2000, Como, Italy', Vol. 1, pp. 342–347.

Bracewell, R. [1965], *The Fourier Transform and Its Applications*, McGraw-Hill Book Company, New York, U.S.A.

Bradley, A. [1997], 'The use of the area under the ROC curve in the evaluation of machine learning algorithms', *Pattern Recogn.* **30**, 1145–1159.

Bradley, A., Jackway, P. & Lovell, B. [1995], Classification in scale-space: applications to texture analysis, *in* Y. Bizais, C. Barillot & R. Di Paola, eds, 'Information Processing in Medical Imaging: 14th International Conference, Ile de Berder, France, June 1995', Kluwer, Dordrecht, The Netherlands, pp. 375–376.

Bushberg, J., Siebert, J., Leidholdt, E. & Boone, J. [2002], *The Essential Physics of Medical Imaging*, Lippincott, Williams and Wilkins, Philedelphia, U.S.A.

CADx Medical Systems [2002]. 'Second Look FDA Summary of Safety and Effectiveness Data'. `http://www.fda.gov/cdrh/pdf/P010034.html`

Carr, J., Hemler, P., Halford, P., Freimanis, R., Choplin, R. & Chen, M. [2001], 'Stereotactic localization of breast lesions: how it works and methods to improve accuracy', *RadioGraphics* **21**, 463–473.

Chan, H.-P., Sahiner, B., Helvie, M., Petrick, N., Roubidoux, M., Wilson, T., Adler, D., Paramagul, C., Newman, J. & Sanjay-Gopal, S. [1999], 'Improvement of Radiologist' Characterization of Mammographic Masses by Using Computer-aided Diagnosis: An ROC Study', *Radiol.* **212**, 817–827.

Chan, H., Wei, D., Helvie, M., Sahiner, B., Adler, D., Goodsitt, M. & Petrick, N. [1995], 'Computer-aided classification of mammographic masses and normal tissue: linear discriminant analysis in texture feature space', *Phys. Med. Biol.* **40**, 857–876.

Chandrasekhar, R. & Attikiouzel, Y. [1997], Gross segmentation of mammograms using a polynomial model, *in* H. K. Boom, ed., 'Proceedings of the 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society : Amsterdam, The Netherlands, 31 October-3 November 1996', IEEE, pp. 73–74.

Chandrasekhar, R. & Attikiouzel, Y. [2000], 'Segmenting the breast border and nipple on mammograms', *Australian Journal of Intelligent Information Processing Systems* **6**, 24–29.

Chandrasekhar, R. & Attikiouzel, Y. [2001], Automatic breast border segmentation by background modelling and subtraction, *in* Yaffe [2001], pp. 1–4.

Chang, Y., Good, W., Sumkin, J., Do, B. & Gur, D. [1999], 'Computerized Localization of Breast lesions from two views: an experimental comparison of two methods', *Invest. Radiol.* **34**, 585–588.

Chang, Y., Hardesty, L., Hakim, C., Chang, T., Zheng, B., Good, W. & Gur, D. [2001], 'Knowledge based computer-aided detection of masses on digitized mammograms: a preliminary assessment', *Med. Phys.* **28**, 455–461.

Christoyianni, I., Dermatas, E. & Kokkinakis, G. [1999], Neural Classification of Abnormal tissue in digital mammography using statistical features of the texture, *in* I. Circuits & S. Society, eds, 'Proceedings the 6th IEEE International Conference on Electronics, Circuits and Systems (ICECS99), 5-8 September, 1999, Pafos, Cyprus', pp. 117–120.

De Maesschalck, R., Jouan-Rimbaud, J. & Massart, D. [2000], 'The Mahalanobis Distance', *Chemometrics and Intelligent Laboratory Systems* **50**, 1–18.

Dinnes, J., Moss, S., Melia, J., Blanks, R., Song, F. & Kleinjnen, J. [2001], 'Effectiveness and cost-effectiveness of double reading of mammograms in breast cancer screening: findings of a systematic review', *The Breast* **10**, 455–463.

Egnal, G. [2000], Mutual information as a stereo correspondence measure, Technical Report MS-CIS-00-20, Computer and Information Science, University of Pennsylvania, Philadelphia, U.S.A.

Elmore, J., Armstrong, K., Lehman, C. & Fletcher, S. [2005], 'Screening for breast cancer', *J. Am. Med. Assoc.* **293**, 1245–1256.

Elmore, J., Miglioretti, D., Reisch, L., Barton, M., Kreuter, W., Christiansen, C. & Fletcher, S. [2002], 'Screening Mammograms by Community Radiologists: Variability in False-Positive Rates', *J. Natl. Cancer I.* **94**, 1373–1380.

Ema, T., Doi, K., Nishikawa, R., Jiang, Y. & Papaioannou, J. [1995], 'Image feature analysis and computer-aided diagnosis in mammography: reduction of false-positive clustered microcalcifications using local edge-gradient analysis', *Med. Phys.* **22**, 161–169.

Faizon, F. & Sun, Y. [2000], Computerized analysis of asymmetry in digitized mammograms, *in* J. Enderle & L. Macfarlane, eds, 'Proceedings of the IEEE 26th annual Northeast Bioengineering Conference : 8-9 April 2000, University of Connecticut, Storrs, Connecticut', IEEE, Piscataway, NJ, U.S.A., pp. 73–74.

Fawcett, T. [2004]. 'ROC Graphs: Notes and Practical Considerations for Researchers'. `http://www.hpl.hp.com/personal/Tom_Fawcett/papers/ROC101.pdf`

Feig, S. [2005], 'Screening mammography controversies: resolved, partly resolved, and unresolved', *Breast J.* **11 (Supp.1)**, S3–S6.

Ferrari, R., Rangayyan, R., Desautels, J. & Frere, A. [2001], Segmentation of Mammograms: Identification of the Skin-Air Boundary, Pectoral Muscle and Fibroglandular Disc, *in* Yaffe [2001], pp. 573–579.

Filev, P., Hadjiiski, L., Sahiner, B., Chan, H. & Helvie, M. [2005], 'Comparison of similarity measures for the task of template matching of masses on serial mammograms', *Med. Phys.* **32(2)**, 515–529.

Fisher, J. & Principe, J. [1998], A methodology for information theoretic feature extraction, *in* IEEE, ed., 'Proceedings of the 1998 IEEE World Congress on Computational Intelligence (WCCI98), Anchorage, Alaska, May 4-9 1998', pp. 1712–1716.

Flach, P. [2004]. 'Tutorial on 'The Many Faces of ROC Analysis in Machine Learning''. `http://www.cs.bris.ac.uk/%7Eflach/ICML04tutorial/index.html`

Freer, T. & Ulissey, M. [2001], 'Screening mammography with computer-aided detection: prospective study of 12 860 patients in a community breast center', *Radiol.* **220**, 781–786.

Gilles, S. [1996], Description and experimentation of image matching using mutual information, Technical report, Robotics Research Group, Department of Engineering Science, Oxford University.

Gold, R., Bassett, L. & Widoff, B. [1990], 'Highlights from the History of Mammography', *RadioGraphics* **10**, 1111–1131.

Gonzalez, R. & Wintz, P. [1987], *Digital Image Processing*, Addison-Wesley Publishing Company, Reading, Massachusetts, U.S.A.

Goodsitt, M., Chan, H., Liu, B., Guru, S., Morton, A., Keshavmurthy, S. & Petrick, N. [1998], 'Classification of compressed breast shapes for the design of equalization filters in x-ray mammography', *Med. Phys.* **25**, 937–948.

Gotlieb, C. & Kreyszig, H. [1990], 'Texture descriptors based on co-occurrence matrices', *Comput. Vision Graph.* **51**, 70–86.

Greiner, M., Pfeiffer, D. & Smith, R. [2000], 'Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests', *Prev. Vet. Med.* **45**, 23–41.

Gupta, R. & Undrill, P. [1995], 'The use of texture analysis to delineate suspicious masses in mammography', *Phys. Med. Biol.* **40**, 835–855.

Gurcan, M., Chan, H., Sahiner, B., Hadjiiski, L., Petrick, N. & Helvie, M. [2002], 'Optimal artificial neural network architecture selection: improvement in computerized detection of microcalcifications', *Acad. Radiol.* **9**, 420–429.

Gurcan, M., Sahiner, B., Chan, H., Hadjiiski, L. & Petrick, N. [2001], 'Selection of an optimal artificial neural network architecture for computer aided detection of microcalcification - Comparison of automated optimization techniques', *Med. Phys.* **28**, 1937–194.

Hadjiiski, L., Sahiner, B., Chan, H., Petrick, N., Helvie, M. & Gurcan, M. [2001], 'Analysis of temporal changes of mammographic features: Computer-aided classification of malignant and benign breast masses', *Med. Phys.* **28**, 2309–2317.

Hand, D. [1981], *Discrimination and Classification*, Wiley, Chichester, UK.

Hanley, J. & McNeil, B. [1983], 'A Method of Comparing the Areas under Receiver Operating Characteristic Curves Derived from the Same Cases', *Radiol.* **148**, 839–843.

Haralick, R. [1979], 'Statistical and Structural approaches to texture', *Proceedings of the IEEE* **67**, 786–804.

Haralick, R., Shanmugam, K. & Dinstein, I. [1973], 'Texture Features for image classification', *IEEE T. Syst. Man Cyb.* **SMC-3**, 610–621.

Haralick, R. & Shapiro, L. [1985], 'Image Segmentation Techniques', *Comput. Vision Graph.* **29**, 100–132.

Helbich, T., Matzek, W. & Fuchsjäger, M. [2004], 'Stereotactic and ultrasound-guided breast biopsy', *Eur. Radiol.* **14**, 383–393.

Helvie, M., Hadjiiski, L., Makariou, E., Chan, H., Petrick, N., Sahiner, B., Lo, S., Freedman, M., Adler, D., Bailey, J., Blane, C., Hoff, D., Hunt, K., Joynt, L., Klein, K., Paramagul, C., Patterson, S. & Roubidoux, M. [2004], 'Sensitivity of non-commercial computer-aided detection system for mammographic breast cancer detection: Pilot clinical trial', *Radiol.* **231**, 208–214.

Hendrick, R. & Parker, S. [1993], Principles of stereotactic mammography and quality assurance, *in* S. Parker & W. Jobe, eds, 'Percutaneous Breast Biopsy', Raven Press, pp. 49–58.

Heywang-Köbrunner, S., Schaumlöffel, U., Viehweg, P., Hofer, H., Buchmann, J. & Lampe, D. [1998], 'Minimally invasive stereotaxic vacuum core breast biopsy', *Eur. Radiol.* **8**, 377–385.

Highnam, R., Brady, J. & Shepstone, B. [1998*a*], Estimation of Compressed Breast Thickness During Mammography, *in* 'Proceedings of 4th International Workshop on Digital Mammography, Nijmegen, Netherlands', pp. 275–278.

Highnam, R., Kita, Y., Brady, M., Shepstone, B. & English, R. [1998*b*], Determining Correspondence between Views, *in* 'Proceedings of 4th International Workshop on Digital Mammography, Nijmegen, Netherlands', pp. 111–118.

Horng, M., Sun, Y. & Lin, X. [2002], 'Texture feature coding method for classification of liver sonography', *Comput. Med. Imag. Grap.* **26**, 33–42.

Hseu, H., Bhalerao, A. & Wilson, R. [1999], Image Matching Based On The Co-occurrence Matrix, Technical Report CS-RR-358, Department of Computer Science, University of Warwick, Coventry, UK. `http://citeseer.ist.psu.edu/hseu99image.html`

Hume, A. & Thanisch, P. [1996], Microcalcification Detection for Mass screening programmes, *in* 'IEE Colloquium on Digital Mammography', Vol. IEE, U.K., Ref. No. 1996/072, pp. 7/1–7/6.

Huo, Z., Giger, M. & Metz, C. [1999], 'Effect of dominant features on artificial neural network performance in the classification of mammographic lesions', *Phys. Med. Biol.* **44**, 2579–2595.

Huo, Z., Giger, M. & Vyborny, C. [2001], 'Computerized Analysis of Multiple-Mammographic Views: Potential Usefulness of Special View Mammograms in computer-Aided Diagnosis', *IEEE Transactions on Medical Imaging* **20**, 1285–1292.

Hutton, B. & Braun, M. [2003], 'Software for image registration: Algorithms, accuracy, efficacy', *Semin. Nucl. Med.* **33**, 180–192.

Huynh, P., Jarolimek, A. & Daye, S. [1998], 'The False-negative mammogram', *RadioGraphics* **18**, 1137–1154.

Intelligent Systems Software [2002]. 'MammoReader FDA Summary of Safety and Effectiveness Data'. `http://www.fda.gov/cdrh/pdf/P010038.html`

Jiang, Y., Nishikawa, R. & Papaioannou, J. [2001], 'Dependence of computer classification of clustered microcalcifications on the correct detection of microcalcifications', *Med. Phys.* **28**, 1949–1957.

Josso, B., Burton, D. & Lalor, M. [2005], 'Texture orientation and anisotropy calculation by Fourier transform and principal component analysis', *Mech. Syst. Signal Pr.* **19**, 1152–1161.

Jumarie, G. [1990], *Relative Information*, Springer-Verlag, Berlin, Germany.

Karssemeijer, N. [1998], 'Automated classification of parenchymal patterns in mammograms', *Phys. Med. Biol.* **43**, 365–378.

Kegelmeyer, W., Pruneda, J., Bourland, P., Hillis, A., Riggs, M. & Nipper, M. [1994], 'Computer-aided mammographic screening for spiculated lesions', *Radiol.* **191**, 331–337.

Kita, Y., Highnam, R. & Brady, M. [1998], Correspondence between different view breast x-rays using a simulation of breast deformation, *in* 'Proceedings of the 1998 Conference on Computer Vision and Pattern Recognition (CVPR98), 23-25 June 1998, Santa Barbara, CA, U.S.A.', IEEE Computer Society, IEEE, pp. 700–707.

Kita, Y., Highnam, R. & Brady, M. [2001], 'Correspondence between different view breast x-rays using curved epipolar lines', *Computer Vision and Image Understanding* **83**, 38–56.

Kita, Y., Tohno, E., Highnam, R. & Brady, M. [2002], 'A CAD system for the 3D location of lesions in Mammograms', *Med. Image Anal.* **6**, 267–273.

Lado, M., Tahoces, P., Mendez, A., Souto, M. & Vidal, J. [2001], 'Evaluation of an automated wavelet-based system dedicated to the detection of clustered microcalcifications in digital mammograms', *Med. Inform.* **26**, 149–163.

Laine, A. [2000], 'Wavelets in Temporal and Spatial Processing of Biomedical Images', *Annu. Rev. Biomed. Eng.* **2**, 511–550.

Laine, A., Schuler, S., Fan, J. & Huda, W. [1994], 'Mammographic Feature Enhancement by Multiscale Analysis', *IEEE T. Med. Imaging* **13**, 725–740.

Lancaster, P. & Šalkauskas, K. [1986], *Curve and Surface Fitting: An Introduction*, Academic Press.

Lauria, A., Palmiero, R., G., F., Cerello, P., Golosio, B., Fauci, F., Magro, R., Raso, G., Tangaro, S. & Indovina, P. [2003], The CALMA system: an artificial neural network method for detecting masses and microcalcifications in digitized mammograms, *in* 'Proceedings of Frontier Detectors For Frontier

Physics, 9th Pisa Meeting on Advanced Detector, May 25-31, 2003, La Biodola, Isola d'Elba, Italy'.

Laws, K. [1980], Textured Image Segmentation, PhD thesis, University of Southern California.

Li, L., Clark, R. & Thomas, J. [2002], 'Computer-aided Diagnosis of Masses with Full-field Digital Mammography', *Acad. Radiol.* **9**, 4–12.

Liu, B., Metz, C. & Jiang, Y. [2004], 'An ROC comparison of four methods of combining information from multiple images of the same patient', *Med. Phys.* **31**, 2552–2563.

Liu, S., Babbs, C. & Delp, E. [1998], Normal Mammogram Analysis and Recognition, *in* 'Proceedings of the IEEE International Conference on Image Processing (ICIP98), 4-7 October 1998, Chicago, U.S.A.', Vol. 1, IEEE Signal Processing Society, pp. 727–31.

Liu, S., Babbs, C. & Delp, E. [2001], 'Multiresolution Detection of Spiculated Lesions in Digital Mammograms', *IEEE T. Image Process.* **10**, 874–884.

Maes, F., Collignon, A., Vandermeulen, D., Marchal, G. & Suetens, P. [1997], 'Multi-modality Image registration by maximisation of mutual information', *IEEE T. Med. Imaging* **16**, 187–198.

Maintz, J. A., Meijering, E. W. & Viergever, M. [1998], General Multimodal Elastic Registration based on Mutual Information, *in* K. Hanson, ed., 'Medical Imaging 1998: Image Processing, International Conference, San Diego, CA, U.S.A., February 23-26, 1998', Vol. 3338, SPIE, pp. 144–154.

Maintz, J. A. & Viergever, M. [1998*a*], 'A survey of medical image registration', *Med. Image Anal.* **2**, 1–36.

Maintz, J. A. & Viergever, M. [1998*b*], An Overview of Medical Image Registration Methods, Technical Report UU-CS Ext. r. no.1998-22, Utrecht University: Information and Computing Sciences, Utrecht, the Netherlands.

Majid, A., Shaw de Paredes, E., Doherty, R., Sharma, N. & Salvador, X. [2003], 'Missed breast carcinoma: pitfalls and pearls', *RadioGraphics* **23**, 881–895.

Marchette, D., Lorey, R. & Priebe, C. [1997], 'An Analysis of Local Feature Extraction in Digital Mammography', *Pattern Recognition* **30**, 1547–1554.

Marias, K., Brady, J., Highnam, R., Parbhoo, S. & Seifalian, A. [1999], Registration and matching of temporal mammograms for detecting abnormalities, *in* MIUA, ed., 'On-line proceedings, Medical Imaging Understanding and Analysis, 19-20 July 1999, Oxford, UK'.
`http://www-ipg.umds.ac.uk/miua/proceedings/IE_index.html`

Marlow, K., ed. [2001], *Proceedings of the 2001 Medical Image Understanding and Analysis Conference, 16 - 17 July 2001, University of Birmingham.*
`http://www.cs.bham.ac.uk/research/proceedings/miua2001/`

Marti, R., Zwiggelaar, R. & Rubin, C. [2001*a*], Automatic Mammographic Registration: Towards the Detection of Abnormalities, *in* Marlow [2001].
`http://www.cs.bham.ac.uk/research/proceedings/miua2001/`

Marti, R., Zwiggelaar, R. & Rubin, C. [2001*b*], Tracking Mammographic Structures over time, *in* T. Cootes & C. Taylor, eds, 'Proceedings of the 12th British Machine Vision Conference, 10-13 September 2001, Manchester, UK', pp. 143–152.

Martin, J., Moskowitz, M. & Milbrath, J. [1979], 'Breast Cancer Missed by Mammography', *Am. J. Roentgenol.* **132**, 737–739.

Marx, C., Malich, A., Facius, M., Grebenstein, U., Sauner, D., Pfleiderer, S. R. & Kaiser, W. [2004], 'Are unnecessary follow-up procedures induced by computer-aided diagnosis (CAD) in mammography?

Comparison of mammographic diagnosis with and without use of CAD', *Eur. J. Radiol.* **51**, 66–72.

Masek, M., Attikiouzel, Y. & deSilva, C. [2000], Combining data from different algorithms to segment the skin-air interface in mammograms, *in* 'CD-ROM Proceedings of the World Congress on Medical Physics and Biomedical Engineering, 23-28 July 2000'.

Masters, T. [1993], *Practical artificial neural network recipes in C++*, Harcourt Brace Jovanovich.

McInerney, T. & Terzopoulos, D. [1996], 'Deformable Models in Med. Image Anal.: A Survey', *Med. Image Anal.* **1**, 91–105.

Mendez, A., Tahoces, P., Lado, M., Souto, M., Correa, J. & Vidal, J. [1996], 'Automatic detection of breast border and nipple in digital mammograms', *Comput. Methods Programs Biomed.* **49**, 253–262.

Menut, O., Rangayyan, R. & Desautels, J. [1997], 'Parabolic Modeling and Classification of Breast Tumors', *Int. J. Shape Model.* **3**, 155–166.

Miller, P. & Astley, S. [1992], 'Classification of breast tissue by texture analysis', *Image and Vision Computing* **10**, 277–282.

Morrow, W., Paranjape, R., Rangayyan, R. & Desautels, J. [1992], 'Region-based contrast enhancement of mammograms', *IEEE T. Med. Imaging* **11**, 392–406.

Morton, A., Chan, H. & Goodsitt, M. [1996], 'Automated model-guided breast segmentation algorithm', *Med. Phys.* **23**, 1107–1108.

Mudigonda, N., Rangayyan, R. M. & Desautels, J. [2000], 'Gradient and Texture Analysis for the Classification of Mammographic Masses', *IEEE T. Med. Imaging* **19**, 1032–1043.

Mudigonda, N., Rangayyan, R. M. & Desautels, J. [2001], 'Detection of Breast Masses in Mammograms by Density Slicing and Texture Flow-field Analysis', *IEEE T. Med. Imaging* **20**, 1215–1227.

Nunes, F. S., Schiabel, H., Benatti, R., Stamato, R., Escarpinati, M. & Goes, C. [2001], A Method to Contrast Enhancement of Digital Dense Breast Images aimed to detect clustered microcalcifications, *in* 'International Conference on Image Processing, 7-10 October 2001, Greece', pp. 305–308.

O'Doherty, T. [1999], Review of the Effective Image Processing Techniques of Mammograms, Technical report, Department of Information Technology, National University of Ireland, Galway, Ireland.

Ojala, T., Nappi, J. & Nevalainen, O. [2001], 'Accurate segmentation of the breast region from digitized mammograms', *Comput. Med. Imag. Grap.* **25**, 47–59.

Paquerault, S., Petrick, N., Chan, H., Sahiner, B. & Helvie, M. [2002], 'Improvement of computerized mass detection on mammograms: Fusion of two-view information', *Med. Phys.* **29**, 238–247.

Pathmanathan, P., Gavaghan, D., Whitely, J., Brady, M., Nash, M., Nielsen, P. & Rajagopal, V. [2004], Predicting tumour location by simulating large deformations of the breast using a 3D Finite Element Model and nonlinear elasticity, *in* C. Barillot, D. Haynor & P. Hellie, eds, 'Proceedings of the 7th International Conference on Medical Image Computing and Computer Assisted Intervention, 26–30 September 2004, St Malo, France', Vol. 3217 of *Lecture Notes in Computer Science*, Springer-Verlag, pp. 217–224.

Petrick, N., Chan, H., Sahiner, B. & Helvie, M. [1999], 'Combined adaptive enhancement and region-growing segmentation of breast masses on digitized mammograms', *Med. Phys.* **26**, 1642–1654.

Petrick, N., Chan, H., Wei, D., Sahiner, B., Helvie, M. & Adler, D. [1998], 'Automated detection of breast masses on mammograms using adaptive contrast enhancement and texture classification', *Med. Phys.* **23**, 1685–1696.

PlanetMath [2006]. 'Euclidean Distance'. `http://planetmath.org/encyclopedia/EuclideanDistance.html`

Pluim, J. W., Maintz, J. A. & Viergever, M. [1998], A multiscale approach to mutual information matching, *in* K. Hanson, ed., 'SPIE Medical Imaging', Vol. 3338, SPIE press, pp. 1334–1344.

Pluim, J. W., Maintz, J. A. & Viergever, M. [2003], 'Mutual-Information-Based Registration of Medical Images: A Survey', *IEEE T. Med. Imaging* **22(8)**, 986–1004.

Qian, W., Kallergi, M., Clarke, L., Li, H. & Venugopal, P. [1995], 'Tree structured wavelet transform segmentation of microcalcifications in digital mammography', *Med. Phys.* **22**, 1247–1254.

Qian, W., Sun, X., Song, D. & Clark, R. [2001], 'Digital Mammography: Wavelet Transform and Kalman-filtering artificial neural network in Mass Segmentation and Detection', *Acad. Radiol.* **8**, 1074–1082.

R2 Technology [1998]. 'ImageChecker FDA Summary of Safety and Effectiveness Data'. `http://www.fda.gov/cdrh/pdf/p970058.pdf`

Rangayyan, R., El-Faramawy, N., Desautels, J. & Alim, O. [1997*a*], 'Measures of Acutance and Shape of Classification of Breast Tumours', *IEEE T. Med. Imaging* **16**, 799–810.

Rangayyan, R., Mudigonda, N. & Desautels, J. [2000], 'Boundary modelling and shape analysis methods for classification of mammographic masses', *Med. Biol. Eng. Comput.* **38**, 487–496.

Rangayyan, R., Paranjape, R., Shen, L. & Desautels, J. [1997*b*], 'Computer-based enhancement and analysis of mammograms', *J. Biomedical Engineering Society of India* **14**, 29–39.

Rowlands, J. [2002], 'The physics of computed radiography', *Phys. Med. Biol.* **47**, R123–R166.

Ruiter, N., Mueller, T., Stotzka, T., Gemmeke, H., Reichenbach, J. & Kaiser, W. [2002], 'Automatic Image matching for breast cancer diagnostics by 3D deformation model of the mamma', *Biomedizinische Technik* **47**, 644–647.

Russ, J. [1995], *The Image Processing Handbook*, second edn, CRC Press, Boca Raton.

Sahiner, B., Chan, H., Petrick, N., Helvie, M. & Goodsitt, M. [1998*a*], 'Computerized characterization of masses on mammograms: The rubber band straightening transform and texture analysis', *Med. Phys.* **25**, 516–526.

Sahiner, B., Chan, H., Petrick, N., Helvie, M. & Goodsitt, M. [1998*b*], 'Design of a high-sensitivity classifier based on a genetic algorithm: application to computer-aided diagnosis', *Phys. Med. Biol.* **43**, 2853–2871.

Sahiner, B., Chan, H., Petrick, N., Helvie, M. & Hadjiiski, L. [2001], 'Improvement of mammographic mass characterization using spiculation measures and morphological features', *Med. Phys.* **28**, 1455–1465.

Sahiner, B., Chan, H., Wei, D., Petrick, N., Helvie, M., Adler, D. & Goodsitt, M. [1996], 'Image Feature selection by a genetic algorithm: Application to classification of mass and normal breast tissue.', *Med. Phys.* **23**, 1671–1684.

Samani, A., Bishop, J., Ramsay, E. & Plewes, D. [1999], A 3-D Contact Problem Finite Element Model for breast shape deformation derived from MRI data, *in* 'Proceedings of the 23rd Annual Meeting of the American Society of Biomechanics University of Pittsburgh, October 21-23, 1999'.

Sanjay-Gopal, S., Rubin, C., Wilson, T., Helvie, M. & Petrick, N. [1999], 'A regional registration technique for automated interval change analysis of breast lesions on mammograms', *Med. Phys.* **26**, 2669–2679.

Shams, L., Brady, M. & Schaal, S. [2001], 'Graph matching vs mutual information maximization for object detection', *Artificial Neural Networks* **14**, 345–354.

Shannon, C. [1948], 'A Mathematical Theory of Communication', *Bell Systems Technical Journal* **27**, 379–423,623–656.
`http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf`

Shen, L., Rangayyan, R. & Desautels, J. [1993], 'Detection and Classification of Mammographic Calcifications', *Int. J. Pattern Recogn.* **7**, 1403–1416.

Shen, L., Rangayyan, R. & Desautels, J. [1994], 'Application of Shape Analysis to Mammographic Calcifications', *IEEE T. Med. Imaging* **13**, 263–274.

Shlens, J. [2003]. 'An Introduction to Information Theory: Notes'.
`http://www.snl.salk.edu/~shlens/notes.html`

Silverman, B. [1986], *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

Singh, M. & Singh, S. [2002], Spatial texture analysis: a comparative study, *in* 'Proceedings of the 16th International Conference on Pattern Recognition, Quebec, Canada, 11-15 August 2002', Vol. 1, pp. 676–679.
`http://citeseer.ifi.unizh.ch/507622.html`

Sista, S., Bouman, C. & Allebach, J. [1995], Fast Image Search using a multiscale stochastic model, *in* 'Proceedings of the International Conference on Image Processing, Washington, DC, 22-25 October 1995', pp. 2225–2228.

Softkey International, Inc. [1996], 'CD-ROM - Key Photos for Windows'.

Sonka, M., Hlavac, V. & Boyle, R. [1999], *Image Processing, Analysis and Machine Vision*, second edn, PWS Publishing.

StatSoft, Inc. [2003]. 'Basic Statistics'.
`http://www.statsoft.com/textbook/stbasic.html`

Studholme, C., Hill, D. G. & Hawkes, D. [1999], 'An overlap invariant entropy measure of 3D medical image alignment', *Pattern Recogn.* **32**, 71–86.

Suckling, J., Dance, D., Moskovic, E., Lewis, D. & Blacker, S. [1995], 'Segmentation of mammograms using multiple linked self-organizing Artificial Neural Networks', *Med. Phys.* **22**, 145–152.

Sun, X., Qian, W. & Song, D. [2004], 'Ipsilateral-mammogram computer-aided detection of breast cancer', *Comput. Med. Imag. Grap.* **28**, 151–158.

Tabar, L. & Dean, P. [1987], 'The control of breast cancer through mammography screening', *Radiol. Clin. N. Am.* **25**, 993–1005.

Tabar, L., Duffy, S., Yen, M., Warwick, J., Vitak, B., Chen, H. & Smith, R. [2002], 'All-cause mortality among breast cancer patients in a screening trial: support for breast cancer mortality as an end point', *J. Med. Screen.* **9**, 159–162.

Taft, R. & Taylor, A. [2001], 'Potential Improvement in Breast Cancer Detection with a novel computer-aided detection system', *Appl. Radiol.* **30**, 25–28.

Tahoces, P., Correa, J., Souto, M., Gomez, L. & Vidal, J. [1995], 'Computer-assisted diagnosis: the classification of mammographic breast parenchymal patterns', *Phys. Med. Biol.* **40**, 103–117.

te Brake, G. & Karssemeijer, N. [1999], 'Single and Multiscale Detection of Masses in Digital Mammograms', *IEEE T. Med. Imaging* **18**, 628–639.

te Brake, G., Karssemeijer, N. & Hendriks, J. C. L. [2000], 'An automatic method to discriminate malignant masses from normal tissue in digital mammograms', *Phys. Med. Biol.* **45**, 2843–2857.

Thurfjell, M., Lindgren, A. & Thurfjell, E. [2002], 'Nonpalpable breast cancer: mammographic appearance as predictor of histologic type', *Radiol.* **222**, 165–170.

Timp, S. & Karssemeijer, N. [2004], 'A new 2D segmentation method based on dynamic programming applied to computer aided detection in mammography', *Med. Phys.* **31**, 958–971.

Tourassi, G., Frederick, E., Markey, M. & Floyd, C. J. [2001], 'Application of the mutual information criterion for feature selection in computer-aided diagnosis', *Med. Phys.* **28**, 2394–2402.

Tourassi, G., Vargas-Voracek, R., Catarious, D. & Jr., F. C. E. [2003], 'Computer-assisted detection of mammographic masses: a template matching scheme based on mutual information', *Med. Phys.* **30**, 2123–2130.

Trochim, W. [2005]. 'The t-Test'.
`http://www.socialresearchmethods.net/kb/stat_t.htm`

Tsai, A., Wells, W., Tempany, C., Grimson, E. & Willsky, A. [2004], 'Mutual information in coupled multi-shape model for medical image segmentation', *Med. Image Anal.* **8**, 429–445.

Undrill, P., Gupta, R., Henry, S. & Downing, M. [1996], Use of texture analysis and boundary refinement to delineate suspicious masses in mammography, *in* M. Loew & K. Hanson, eds, 'Proceedings of Medical Imaging 1996: Image Processing', Vol. 2710, pp. 301–310.

US Preventive Services Task Force [2002], 'Screening for Breast Cancer: Recommendations and Rationale', *Ann. Intern. Med.* **137**, 344–346.

van Erkel, A. & Pattynama, P. T. [1998], 'Receiver operating characteristic (ROC) analysis: Basic principles and applications in Radiology', *Eur. J. Radiol.* **27**, 88–94.

Van Schalkwyk, J. [2003]. 'The magnificent ROC'.
`http://www.anaesthetist.com/mnm/stats/roc/`

Veldkamp, W. H., Karssemeijer, N. & Hendriks, H. L. [2001], 'Experiments with radiologists and a fully automated method for characterization of microcalcification clusters', *International Congress Series* **1230**, 586–592.

Verkooijen, H., Peeters, P., Pijnappel, R., Koot, V., Schipper, M. & Rinkes, I. [2000], 'Diagnostic accuracy of needle-localized open breast biopsy for impalpable breast disease', *Brit. J. Surg.* **87**, 344–347.

Viola, P. & Wells, W. I. [1995], Alignment by Maximization of Mutual Information, *in* 'Proceedings of the Fifth International Conference on Computer Vision, 20-23 June 2004', IEEE, Cambridge, Massachusetts, U.S.A., pp. 16–33.
`http://citeseer.ist.psu.edu/viola95alignment.html`

Vujovic, N. & Brzakovic, D. [1997], 'Establishing the Correspondence Between Control Points in Pairs of Mammographic Images', *IEEE T. Image Process.* **6**, 1388–1399.

Vyborny, C., Doi, T., O' Shaughnessy, K., Romsdahl, H., Schneider, A. & Stein, A. [2000], 'Breast Cancer: Importance of Spiculation in Computer-aided Detection', *Radiol.* **215**, 703–707.

Wald, N., Murphy, P., Major, P., Parkes, C., Townsend, J. & Frost, C. [1995], 'UKCCCR multicentre randomised controlled trial of one and two view mammography in breast cancer screening', *Brit. Med. J.* **311**, 1189–1193.

Wang, Z., Guerriero, A. & De Sario, M. [1996], 'Comparison of several approached for the segmentation of texture images', *Pattern Recogn. Lett.* **17**, 509–521.

Warren-Burhenne, L., Wood, S., D'Orsi, C., Feig, S., Kopans, D., O'Shaughnessy, K., Sickles, E., Tabar, L., Vybomy, C. & Castellino, R. [2000], 'Potential contribution of computer-aided detection to the sensitivity of screening mammography', *Radiol.* **215**, 554–562.

Wei, D., Chan, H., Helvie, M., Sahiner, B., Petrick, N., Adler, D. & Goodsitt, M. [1995], 'Classification of mass and normal breast tissue on digital mammograms: Multiresolution texture analysis', *Med. Phys.* **22**, 1501–1513.

Wei, D., Chan, H., Petrick, N., Sahiner, B., Helvie, M., Adler, D. & Goodsitt, M. [1997], 'False-positive reduction technique for detection of masses on digital mammograms: Global and local multiresolution texture analysis', *Med. Phys.* **24**, 903–914.

Wellman, P., Howe, R., Dalton, E. & Kern, K. [1999], Breast Tissue Stiffness in compression is correlated to histological diagnosis, Technical report, Harvard BioRobotics Laboratory.

Wells, W. I., Viola, P., Atsumi, H. & Nakajima, S. [1996], 'Multi-modal volume registration by maximization of mutual information', *Med. Image Anal.* **1**, 35–51.

Wen, H.-H., Giger, M., Horsch, K., Hendrick, R., Vybomy, C. & Lan, L. [2004], Correlation of lesions from multiple images for CAD, *in* A. Amini & A. Manduca, eds, 'Medical Imaging 2004: Physiology, Function, and Structure from Medical Images. Proceedings of the SPIE', Vol. 5370, pp. 93–96.

Wirth, M., Choi, C. & Jennings, A. [1999], A Nonrigid-body Approach to Matching Mammograms, *in* 'Image Processing and its Applications, Conference Publication No.465', IEE, pp. 484–488.

Wirth, M., Narhan, J. & Gray, D. [2002], Nonrigid mammogram registration using mutual information, *in* M. Sonka & J. Fitzpatrick, eds, 'Proceedings of Medical Imaging 2002: Image Processing', Vol. SPIE 4684, pp. 562–573.

Wirth, M. & Stapinski, A. [2003], Segmentation of the breast region in mammograms using active contours, *in* T. Ebrahimi & T. Sikora, eds, 'Visual Communications and Image Processing', Vol. 5150, SPIE, Lugano, Switzerland, pp. 1995–2006.

Wolberg, G. and Zokai, S. [2000], Robust image registration using log-polar transform, *in* 'Proceedings of the IEEE International Conference on Image Processing, Vancouver, Canada, September 2000'.

Yaffe, M., ed. [2001], *Digital Mammography: IWDM 2000, 5th International Workshop*, Med. Phys. Publishing, Madison, Wisconsin, U.S.A.

Yale [2002]. 'Yale University School of Medicine: Cardiothoracic Imaging'.
`http://info.med.yale.edu/intmed/cardio/imaging/anatomy/breast_anatomy/`

Yankaskas, B., Schell, M., Bird, R. & Desrochers, D. [2001], 'Reassessment of Breast Cancers Missed during Routine Screening Mammography: A Community Based Study', *Am. J. Roentgenol.* **177**, 535–541.

Yin, F., Giger, M., Doi, K., Metz, C., Vybomy, C. & Schmidt, R. [1991], 'Computerized detection of masses in digital mammograms: Analysis of bilateral subtraction images', *Med. Phys.* **18**, 955–963.

Zhang, B., Tomai, C. & Zhang, A. [2002], Adaptive Texture Image Retrieval In Transform Domain, *in* 'Proceedings of the IEEE International Conference on Multimedia and Expo 2002 (ICME'02),Lausanne, Switzerland, August 26-29, 2002'.

Zhang, J. & Tan, T. [2003], 'Affine invariant classification and retrieval of texture images', *Pattern Recogn.* **36**, 657–664.

Zheng, B., Ganott, M., Britton, C., Hakim, C., Hardesty, L., Chang, T., Rockette, H. & Gur, D. [2001], 'Soft-copy mammographic readings with different computer-assisted detection cuing environments:

preliminary findings', *Radiol.* **221**, 633–640.

Zwiggelaar, R. & Boggis, C. [2001], Classification of Linear Structures in Mammographic Images, *in* Marlow [2001].
`http://www.cs.bham.ac.uk/research/proceedings/miua2001/`

Zwiggelaar, R. & Rubin, C. [1999], Separating Background Texture and Image Structure in Mammograms, *in* T. Pridmore & D. Elliman, eds, 'Proceedings of the 10th British Machine Vision Conference, 13-16 September 1999, University of Nottingham, UK', pp. 362–371.

Zwiggelaar, R., Schumm, J. & Taylor, C. [1997], The Detection of Abnormal Masses in Mammograms, *in* MIUA, ed., 'On-line proceedings, Medical Imaging Understanding and Analysis, 7-8 July 1997, Oxford, UK', pp. 69–72.
`http://www.robots.ox.ac.uk/~mvl/miua97/frame_proceedings.html`