Facial Expression Recognition using Covariance Matrix Descriptors and Local Texture Patterns

Author:
Ashaylin Naidoo

Co-supervisor:
Rethabile Khutlang

Supervisor:
Prof. Jules R. Tapamo

A dissertation submitted in fulfilment of the requirements
for the degree of Master of Science in Computer Engineering

College of Agriculture, Engineering and Science, University of KwaZulu-Natal

November 2017

Preface

The research discussed in this dissertation was carried out in the College of Agriculture, Engineering and Science of the University of KwaZulu-Natal, Durban from July 2016 until November 2017 by Mr. Ashaylin Naidoo (210511253) under the supervision of Prof. Jules-Raymond Tapamo and co-supervision of Mr. Rethabile Khutlang. As the candidate's supervisor I, Prof. Jules-Raymond Tapamo agree/do not agree to the submission of this thesis.

Signed: _____

Date: _____

As the candidate's co-supervisor I, Mr. Rethabile Khutlang agree/do not agree to the submission of this thesis.

Signed: _____

Date: _____

I, Mr. Ashaylin Naidoo (210511253), hereby declare that all the materials incorporated in this dissertation are my own original work, except where acknowledgement is made by name or in the form of a reference. The work contained herein has not been submitted in any form for any degree or diploma to any institution.

Signed: _____

Date: _____

Declaration 1 - Plagiarism

I, Ashaylin Naidoo, declare that,

1. The research reported in this dissertation, except where otherwise indicated, is my original research.
2. This dissertation has not been submitted for any degree or examination at any other university.
3. This dissertation does not contain another person's data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.
4. This dissertation does not contain another person's writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
   (a) Their words have been re-written but the general information attributed to them has been referenced
   (b) Where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced.
5. This dissertation does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the References section.

Signed: _____

Date: _____

Declaration 2 - Publications

I, Ashaylin Naidoo, declare that the following publications came out of this dissertation

1. A. Naidoo, J. Tapamo, R. Khutlang, "Affective Computing: Using Covariance Descriptors for Facial Expression Recognition", *in the Third International Congress on Information and Communication Technology.*, conference proceedings by Springer AISC, London, United Kingdom, February 2018, [accepted].

Signed: _____

Date: _____

Abstract

College of Agriculture, Engineering and Science, University of KwaZulu-Natal

Master of Science in Computer Engineering

Facial Expression Recognition using Covariance Matrix Descriptors and Local Texture Patterns

by Ashaylin Naidoo

Facial expression recognition (FER) is a powerful tool that is emerging rapidly due to increased computational power in current technologies. It has many applications in the fields of human-computer interaction, psychological behaviour analysis, and image understanding. However, FER presently is not fully realised due to the lack of an effective facial feature descriptor.

The covariance matrix as a feature descriptor is popular in object detection and texture recognition. Its innate ability to fuse multiple local features within a domain is proving to be useful in applications such as biometrics. Developing methods also prevalent in pattern recognition are local texture patterns such as Local Binary Pattern (LBP) and Local Directional Pattern (LDP) because of their fast computation and robustness against illumination variations. This study will examine the performance of covariance feature descriptors that incorporate local texture patterns concerning applications in facial expression recognition. The proposed method will focus on generating feature descriptors to extract robust and discriminative features that can aid against extrinsic factors affecting facial expression recognition, such as illumination, pose, scale, rotation and occlusion. The study also explores the influence of using holistic versus component-based approaches to FER.

A novel feature descriptor referred to as Local Directional Covariance Matrices (LDCM) is proposed. The covariance descriptors will consist of fusing features such as location, intensity and filter responses, and include LBP and LDP into the covariance structure. Tests conducted will examine the accuracy of different variations of covariance features and the impact of segmenting the face into equal sized blocks or special landmark regions, i.e. eyes, nose and mouth, for classification. The results on JAFFE, CK+ and ISED facial expression databases establish that the proposed descriptor achieves a high level of performance for FER at a reduced feature size. The effectiveness of using a component-based approach with special landmarks displayed stable results across different datasets and environments.

Acknowledgements

Firstly, I would like to extend my deepest and sincerest appreciation to my supervisor Prof. Jules-Raymond Tapamo for his unwavering support and guidance. He has allowed me to grow stronger by imparting his knowledge, skills and excellent advice regarding my research and future. Secondly, I would like to extend my sincerest gratitude to Mr. Rethabile Khutlang who devoted time towards helping me complete my MSc. degree and for always being available to assist. Thirdly, I would like to thank my parents and sister for their patience, motivation and continuous support. To the Council of Scientific and Industrial Research, Modelling and Digital Sciences, Information Security Department, I would like to extend my humblest appreciations for providing me with a place to conduct my research, introducing me to new friends and for the financial assistance. Lastly, to Lulu and Zeplin thank you for motivating me to continually seek improvement and providing me comfort.

Contents

List of Figures

List of Tables

Abbreviations

AIM	**A**ffine-**I**nvariant **M**etric
ANN	**A**rtificial **N**eural **N**etwork
AU	**A**ction **U**nit
CK+	**E**xtended **C**ohn-**K**anade
FER	**F**acial **E**xpression **R**ecognition
HCI	**H**uman **C**omputer **I**nteraction
HCL	**H**ue **C**hroma **L**uminance
ICA	**I**ndependent **C**omponent **A**nalysis
ISED	**I**ndian **S**pontaneous **E**xpression **D**atabase
JAFFE	**JA**panese **F**emale **F**acial **E**xpression
LBP	**L**ocal **B**inary **P**attern
LDA	**L**inear **D**iscriminant **A**nalysis
LDCM	**L**ocal **D**irectional **C**ovariance **M**atrices
LDP	**L**ocal **D**irectional **P**attern
LEM	**L**og-**E**uclidean **M**etric
PCA	**P**rincipal **C**omponent **A**nalysis
RCM	**R**egion **C**ovariance **M**atrices
RGB	**R**ed **B**lue **G**reen
SPD	**S**ymmetric **P**ositive **D**efinite
SVM	**S**upport **V**ector **M**achine
ULBP	**U**niform **L**ocal **B**inary **P**attern
YCbCr	**l**uma **b**lue-difference **r**ed-difference

Chapter 1

Introduction

The continual advancement of computing power has made computers a fundamental ever-present part of our lives [1]. A large contribution to our day-to-day activities and work is done on computers, yet currently, these devices are indifferent to our affective states and have no perception of the user's emotional state. However, effective human-human communication depends on the ability to read emotional and affective signals. A large part of the information available in Human-Computer Interaction (HCI) is lost due to the emotional blindness of the system during the interaction [2].

Recent studies on affective computing suggest that it is beneficial to provide computers with the ability to interpret affective states of their users [3–5]. The importance of emotion in our daily lives [6] infers that to improve HCI systems we need the ability to recognise the user affect. This will allow progress in HCI by enabling affective computing to alleviate the shortcoming between the emotionally deficient computer and expressive humans [7].

Baltrusaitis [2] describes applications that can benefit from the ability to use affect in their systems, such as interfaces that do not interrupt their users when they are stressed, online learning systems that adapt the teaching if the student is confused and video games that adapt their difficulty based on player engagement. Further applications described include: assisted living environments that can monitor the user's state and report to medical professionals if the patient is feeling pain, assistive technologies for diagnosing conditions such as depression and systems that monitor drivers. Mobile technology companies such as Apple Inc. [8] have also started integrating facial analysis into the core design and usability of their devices [9]. In September 2017 they demonstrated the Face ID protocol that will be released with their flagship devices.

This protocol will replace their previous Touch ID protocol which uses fingerprint rather than face as a biometric indicator. This will have a significant impact since Apple Inc. is one of the largest developers and manufacturers of industry-leading technology with regard to human-computer interaction. This demonstrates the capacity of facial analysis, which will encourage other companies and industries to adopt more facial analysis techniques and technologies.

Successful affect-sensitive systems are dependent on reliable recognition of human emotions [10]. Humans present multi-modal affective behaviour, which makes it subtle and complex. People are skilled at using non-verbal cues, such as various hand gestures, facial expressions, vocal prosody, eye gaze, head movements and posture, for self-expression and interpreting other's behaviour [2]. All these modes of expression convey vital affective information that humans use to surmise each other's emotions [11].

Faces are the most visible social part of the human body. Therefore, the face is prioritised over the other modalities and receives great attention from both psychologists and affective computing researchers [12]. Faces reveal emotions [13], communicate intent, and help regulate social interaction [14]. Therefore, facial expressions play a crucial part of non-verbal communication. Early research in facial expression claimed that facial expressions are innate; that is, they cannot be learned and have an evolutionary meaning for survival [15]. A worldwide observational study was later conducted that determined humans of different ethnicities, ages and gender shared a certain level of universality in the appearance of emotion in the face [13, 16]. Facial expression recognition could, therefore, be applied to determine a generalised form of emotion.

1.1 Physiology of Facial Expressions

Facial expression physiology is a consequence of facial muscle activity. The muscles are known as mimetic muscles or muscles of facial expressions. They belong to the group of head muscles, that also contain muscles of the scalp and muscles of mastication, which are responsible for moving the jaw and tongue. Facial muscles are innervated by the facial nerve. This nerve branches out in the face and causes contractions when it is activated. The result is various observable movements on the face. The generally visible muscle actions are blocks of skin motion.

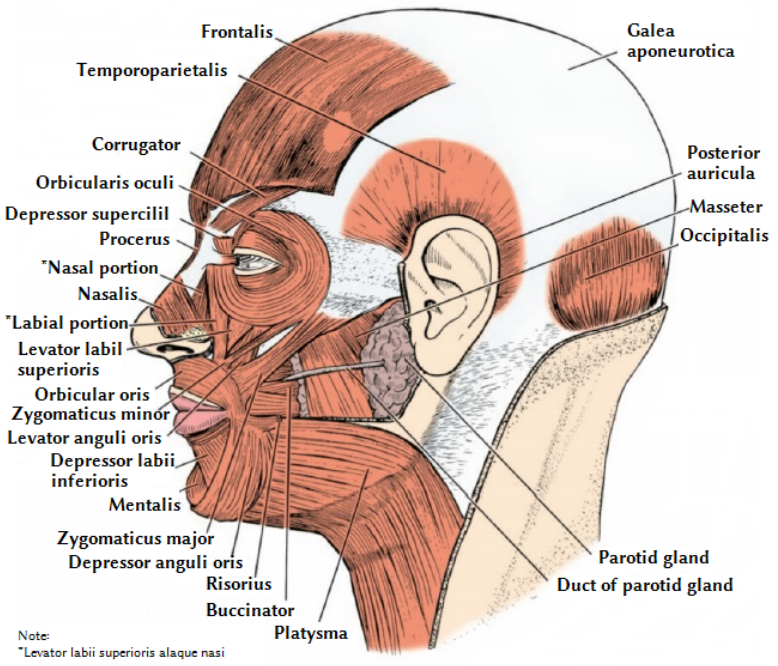


Figure 1.1: Muscles of facial expressions [17].

The eyebrows, lips, cheeks, nose and wrinkles between the eyebrows and on the forehead display these muscle actions most frequently [18].

The human face consists of 20 flat skeletal muscles [19], shown in Figure 1.1. The muscles are located under the skin and are attached to either the skin or other muscles, not the bones or joints as other muscles responsible for body movements are [19]. The muscles that are positioned near the facial orifices, that is the mouth, nose, and eyes use the facial skin to move [17]. This movement causes facial surface deformations, which result in a variable facial expression representing emotions [17, 20]. The facial muscles are designed to move in groups instead of individually and are also responsible for controlling the orifices.

According to the location, the taxonomy is partitioned into three groups: oral, nasal and orbital [19]. The oral muscles alter the shape of the oral orifice. This group is responsible for complex mouth motions and allows sophisticated shaping of the mouth. The nasal group is responsible for the compression and opening of the nostrils. The three muscles that form the orbital group are primarily responsible for the motion of the eyelid and protecting the eyes. A muscle that is critical for facial expressions, is located between the eyebrows. It pulls the eyebrows downwards, causing wrinkles over the nose and emphasising expression [17, 19].

The facial muscles do more than regulate the position and width of facial openings. They also make them more expressive. Consequently, the face is able to convey emotions and the present psychological state of a person, which play an important role in the nonverbal communication between people [19].

1.2 Expressing Different Emotion Types

Table 1.1: Emotions defined by the extent of pleasure and activation [21]

Pleasant	glad, happy, warm-hearted, delighted, pleased, cheerful
Unpleasant	sad, blue, unhappy, grouchy, miserable
High Activation	surprised, astonished, aroused, active, stimulated
Low Activation	passive, tranquil, quiet, idle, still
Pleasant + High Activation	excited, lively, enthusiastic, elated, euphoric
Unpleasant + High Activation	fearful, anxious, distressed, jittery, annoyed, nervous
Pleasant + Low Activation	calm, contented, serene, at rest, relaxed
Unpleasant + Low Activation	drowsy, droopy, dull, sluggish, tired, bored

There are roughly 200 different emotions based on extent of activation and pleasure. Table 1.1 shows some of the different emotions one can have based on these criteria [21]. These expressions change with minute facial feature changes and are identified as micro-facial expressions. The resulting emotion that humans display is often a combination of several micro-facial expressions. The combinations of these expressions were grouped into seven basic universal expression classes [16] which can be seen in Figure 1.2. The physiological signals that classify these basic expressions are described in Table 1.2.

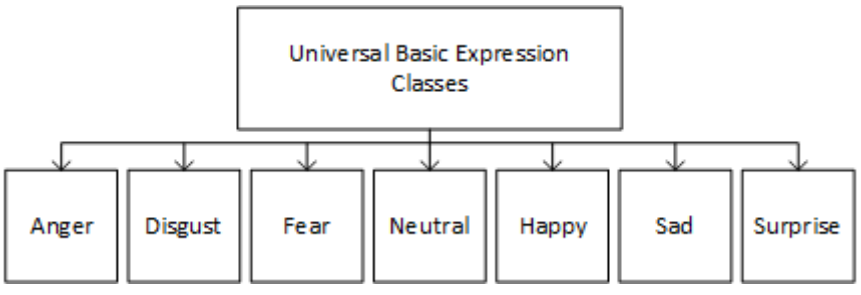


Figure 1.2: Universal basic expression classes.

Table 1.2: Physiological signal response to basic expressions[16]

Anger	**Fear**
Eyebrows lowered and squeezing together Vertical wrinkles between eyebrows Eyelids tight and straight Eyes tight and pupils narrowed Lips closed tight or gently opened	Eyebrows lifted and pulled inward Wrinkles on forehead Upper eyelids lifted Mouth open Lips are tight
Happiness	**Sadness**
Lips corners pulled back and up Mouth can be open and teeth visible Cheeks raised Wrinkles under lower eyelid Wrinkls outside eye corners	Inner parts of eyebrows pulled down Lips corners pull down Lips shake
Disgust	**Surprise**
Upper lip lifted Wrinkles on nose and under eyes Cheeks lifted Eyelids lifted but not tight Eyebrows pulled down	Eyebrows lifted and pulled inward Horizontal wrinkles appear on forehead Eyes are open wide Jaw is dropped Mouth opened and lips tight

- **Anger** is regarded as a strong emotional reaction, therefore detecting anger can be beneficial because anger is a strong predictor of violence. Frustration, physical threat, and verbal threats are common sources of anger. It can also cause an increase in blood pressure, display a red face, and cause tension in the muscles.

- **Disgust** is a negative emotion generally evoked by smell, taste, or vision. There are no universal grounds for what may cause this emotion. It is influenced by cultural or personal reasonings. The extreme physiological response is vomiting. The most significant features of the face are in the nose and mouth area.

- **Fear** is mostly induced by stressful or dangerous situations. Noticeable effects of fear on the body include an increase in heart rate and blood pressure, open eyes and wide pupils. In extreme situations, it may also induce muscle function loss such as paralysis.

- **Happiness** is a positive emotion often associated with a smile on the face.

- **Sadness** causes the facial muscles to lose tension. It is an undesirable emotion that is often caused by negative events such as death or failure.

- **Surprise** is often the most short-lived emotion as its engagement is sudden. It is usually unanticipated and often proceeds into other emotions like happiness or sadness. The typical features of surprise are lifted eyebrows, wrinkles on the forehead, widely open eyes and a dropped jaw.

1.3 Applications of Facial Expression Recognition Systems

Applications of FER systems are evident in multiple areas of HCI, including:

Affective computing: This is the study and further development of frameworks and gadgets that recognise, understand, manage and duplicate human influences. This field is interdisciplinary and consists of computer vision, machine learning, brain science and computer engineering. The system must translate the condition of its user and adapt how it behaves to them, to react appropriately to the user's feelings.

Commercial survey: Online shopping systems often require users to give feedback to determine customer satisfaction with their product or service. Facial expression recognition systems can provide a measure of the extent of satisfaction and therefore an estimation can be made as to the success of the product.

Human-Computer-Interaction: This is a relatively new and developing field that deals with communication between humans and machines. These systems use audio and visual or sensory data from the user to modify its current state for better application. A popular example used in gaming consoles is the Xbox Kinect [22], which has to do with the cognitive and affective aspects of HCI where the machine makes certain adjustments based on the user's state.

Driver state surveillance: To prevent unforeseen circumstances such as accidents, injury or death, driver state surveillance has become a leading concern for automotive industries. The monitoring of the driver's facial expressions gives sufficient insight to help avoid a collision. The most common expression on a driver's face before an accident is that of fear or surprise. If identified early preventative measures can be taken to reduce or alleviate disaster.

Treatment of Asperger's syndrome: This disorder inhibits people from recognising a speaker's words and emotions, which creates difficulty interacting with people. The use of an FER system could help them to recognise people's emotions and improve their daily communication.

1.4 Description of Features that Model Appearance

The main challenge with regard to the discussions above is in building models for the different facial expression classes to be efficiently computed. The difficulty stems from the huge variations in appearances that exist in real-world images, even for the same expression or class. These variations can occur due to different illumination conditions, viewpoints, scales, occlusions and numerous other factors [23].

To overcome these problems special attention is given to designing robust image descriptors for specific applications. These descriptors capture discriminative qualities or features within an image that can be used later on for classification. The feature selection influences different properties making it suitable for different computer vision tasks. These features are typically pixel-features, they include colour or pixel intensity, image gradients, wavelet transforms, filter responses and many others. Consequently, some features may be fast to compute, while other features might promote a robustness to noise and other issues. However, since visual entities consist of a region or continuous group of pixels, representation of such entities require *region descriptors* [23].

Theoretically, a region descriptor is a joint distribution of pixel features within a region. A histogram of features calculated at the pixels within the region is the discretised non-parametric representation of a distribution and is considered one of the earliest forms of region descriptors. Tuzel et al. [24] proposed the *covariance region descriptor*. This approach represented the region by the covariance of the features of the pixels lying within the region. This method's representation is advantageous because it allows fusion of multiple features. Therefore, such a descriptor is robust to noise in the pixel features and is partially invariant to rotation and scale of the region.

1.5 Facial Expression Recognition Systems

The prevalent approach to facial expression recognition systems is set out in Figure 1.3, and can be categorised into three steps [25]:

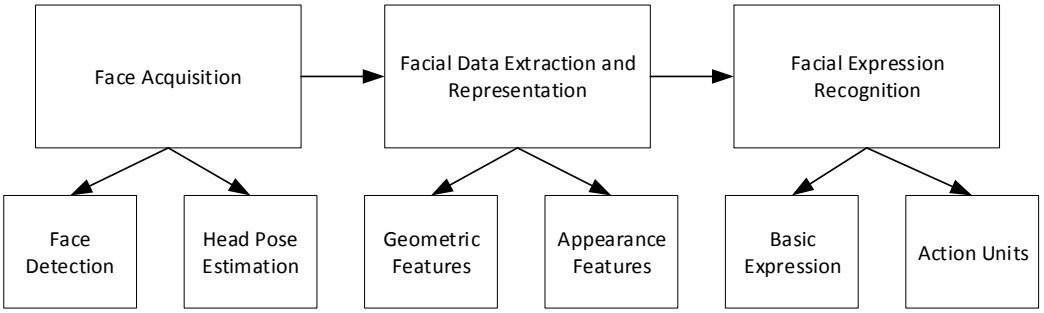


Figure 1.3: Basic structure of FER systems [25].

- Face acquisition
- Facial feature extraction and representation
- Facial Expression Recognition

The first step aims to find a facial region from the input frame images. Once the face location is determined, various facial feature extraction approaches can be used, which is done in step two. The last step involves using different classification approaches to label the expression.

This study will focus on facial feature extraction and representation in an FER system.

1.6 Motivation

The interaction between humans and computers is constantly growing. Therefore, HCI is a developing research area and an essential requirement for giving computers the intelligence to understand human behaviour and act accordingly. Interpersonal communication can be classified into both verbal and nonverbal communication. Verbal communication comprises raw voice data only, while nonverbal communication consists of intensity and voice tone as well as facial expressions and hand gestures [26]. When these modalities are combined an effective communication for understanding and interpretation is produced. The ability to recognise facial expressions in human-computer interactions is vital because it provides intuition on people's personality, psychological state and intention. When facial expression cues and gestures are combined the internal meaning of a speaker can be efficiently elicited, without any vocal data.

This forms the basis for facial expression recognition. The verbal component (spoken words) of communication makes up only 7 percent of messages, the vocal component (voice intonation) makes up 38 percent, while facial expressions of speakers make up a remarkable 55 percent of the effect of spoken messages. This shows that the major modality in human communication is facial expressions [26].

The basic steps involving facial expression recognition are described in Figure 1.3. To achieve proficient expression recognition, a fitting framework for feature extraction and representation must be chosen to be able to make the distinction between the most differentiable features that represent facial cues. This being a continuously developing topic under research, many methods and techniques to extract features were proposed and have been combined with different methodologies of classification, such as Principal Component Analysis (PCA) [27], Linear Discriminant Analysis (LDA) [28], Gabor wavelet analysis [29], and Local Binary Patterns (LBP) [30]. However, each of the above-mentioned algorithms has issues with inaccuracy or hardware complexity. Therefore, it is necessary to propose an FER system that balances both accuracy and complexity during the classification of emotions. Most of the methods listed above could not accurately classify expressions and very similar facial expression classes are still confused. Furthermore, expression in these cases was predominately identified using a holistic approach to posed expressions. In addition, the general concerns regarding facial expression recognition are that even if recognition is done in a constraint of faces specific to some culture, several factors such as the presence of facial hair or glasses, pose and facial scars increase the task complexity. Another challenge is the variation in size and orientation of the face in input images [31].

The covariance matrix as a feature descriptor is popular in object detection and texture recognition. However, its innate ability to fuse multiple local features within a domain is proving to be useful in other applications such as in biometrics. This study attempts to classify expression using holistic and component-based approaches of most distinguishable characteristics of the face using texture-based patterns and region covariance descriptors. The study additionally tests on both posed and spontaneous expression datasets and introduces a novel feature descriptor.

1.7 Research Question

This research attempts to produce new knowledge to answer the following question:

- Can facial expression be successfully classified by using region covariance descriptors?

1.8 Research Goal

The primary aim of this dissertation is to design a facial expression recognition system that can classify expressions from static images into predetermined classes accurately and efficiently. It further aims to explore the impact of holistic and component-based approaches with regards to facial expression recognition.

1.9 Research Objectives

The objectives of the research are as follows:

1. Determine the effectiveness of covariance matrix descriptors for classifying facial expression.
2. Examine the performance of a novel image descriptor for facial expression recognition called Local Directional Covariance Matrices.
3. Test holistic-based and component-based approaches to facial expression recognition.

1.10 Delineation

The bounds of this research are as follows:

1. 2D facial images from the JAFFE, Extended Cohn-Kanade, and ISED databases only are examined.
2. Existing pre-processing methods will be used where required.
3. The study will mainly focus on the feature extraction and representation framework in a facial expression recognition system.
4. Only static-based methods using the basic prototypic expressions are tested.

1.11 Contributions

This research makes five principal contributions to the field of biometrics in facial expression classification:

1. Explore the effectiveness of region covariance matrices with applications to facial expression recognition.

2. Propose a novel image descriptor for facial expression recognition referred to as Local Directional Covariance Matrices.

3. Evaluate holistic and component-based approaches for facial expression recognition.

4. Conduct research on a relatively new dataset, Indian Spontaneous Expression Database (ISED), and evaluate the proposed system's performance for various cultures and environments, in other words a cross-database evaluation.

5. Test against both posed and spontaneous facial expressions for analysis on FER.

1.12 Overview of Chapters

Chapter 2 covers the literature review that identifies the most appropriate features to classify facial expression. Chapter 3 presents the principles of RCM and the structures of different local texture patterns. Chapter 4 demonstrates the experimental results of the proposed algorithm. Chapter 5 provides the concluding remarks and outlines future works.

Chapter 2

Literature Review

In Chapter 1, we introduced the physiology and breakdown of how the face is analysed to classify the basic expressions. It also showed the applications and structure of an FER system. This chapter introduces the advances in computer vision for facial expression recognition. First, we describe the most popular and useful representation of facial expressions. Next, we expand upon the steps involved in an FER system shown in Figure 1.3. Then, after briefly presenting the different approaches and limits of facial feature extraction methods, we survey the state-of-the-art geometric and appearance-based feature methods for FER, discussing their achievements and limitations. Finally, an introduction to the psychology of visual perception of facial expression recognition by humans is presented.

2.1 Introduction

Darwin founded research in the field of human emotions in his book, "The Expression of the Emotions in Man and Animals" [15]. From this research, it was evident that one of the most significant features used in recognising human emotion is facial expression [32]. According to Huang et al. [33], facial expressions are defined as facial changes in reaction to a person's intentions, internal emotional state or social interaction. Recent advances in technology have promoted the applications of automated facial expression recognition, mentioned in Section 1.3. Other applications include sociable robotics, interactive games, data-driven animation, neuromarketing as well as numerous other HCI systems [34]. The complexity involved for a human to recognise expression is minimal, but it poses a huge challenge for computers [33]. Existing studies in the literature lack a consistent evaluation methodology, for example, tests are conducted even though there is no subject similarity in training and testing, which results in misleading high

accuracy [34]. This does not illustrate the majority of the FER issues that exist in real scenarios. Subsequently, in databases that do not have controlled environments and in cross-database evaluations, low accuracy has been reported [34]. The difficulty in FER for computers stems from the challenge in separating the expressions' features space: the facial features from one subject can exhibit similar properties to different expressions and facial features from multiple subjects with similar expressions may vary drastically. Additionally, certain expressions such as sadness and fear tend to be very similar [34]. To overcome these deficiencies a few studies attempted to enable computers to achieve the level of accuracy that humans have. This chapter aims to highlight some examples of these studies.

A simple example is presented in Figure 2.1, where six subjects displaying a happy expression show considerable variation, not only in how the subjects convey expressions, but also in other elements of the images, such as lighting, brightness, subjects' pose and the background. Another challenge that Figure 2.1 demonstrates is the training-testing scenarios that are not controlled. Training images can vary from testing images regarding environmental conditions as well as the subject ethnicity. To evaluate FER under these scenarios, cross-database techniques can be used. This implies training the method using a particular database and testing it with another one, potentially from a different ethnic group. This dissertation presents results on this approach in Section 4.6.

Figure 2.1: The happy expression displayed across six different subjects. The images are from the following databases: JAFFE [35], CK+ [36], ISED [37].

Facial Expression Recognition systems consist of two main categories: static images [38–44] and dynamic image sequences [45–48]. The temporal information is what differs between the two methods. The features in static-based methods contain information about the current input image only, while sequence-based methods use the temporal information in images captured from one or more frames to recognise expression. Automatic FER systems use static or dynamic image sequences as input and one of the basic expressions listed in Figure 1.2 is commonly output.

This dissertation will focus on static image-based methods and will consider the basic prototypic expressions for both controlled and uncontrolled scenarios. From Figure 1.3 it can be seen that the automatic facial expression analysis consists of three major components: face acquisition, facial data extraction and representation, and finally facial expression recognition. Face acquisition can then further be broken down into face detection [49–52] and estimation of head pose [53–55].

After face detection, the system extracts the arrangement of facial features caused by facial expressions. These features are represented either by geometric-based [52, 56–58] or appearance-based [39–42, 44, 56] methods.

Expression recognition can be performed after acquiring the facial features. According to Liu et al. [38], expression recognition systems comprise of a training procedure that has three stages: feature learning, feature selection and then classifier construction. In the feature learning stage the extraction of all features associated with the facial expression takes place. In the feature selection stage the best feature to represent the facial expression is selected. The aim of feature learning and selection is to minimise variations that occur intra-class and maximise those that occur inter-class [39]. The challenge in minimising the intra-class expression variation is that images of different individuals with the same expressions are far from each other in terms of pixel's space. Maximising the variation that occurs inter-class is also challenging because images of the same person with varying expressions may be very close to one another in terms of the pixel's space [59]. After the features are learned and selected into a fitting representation, the facial expression is inferred by using a classifier.

2.2 Face Detection

The most popular approaches used for face detection in images include:

Skin-colour-based segmentation schemes [60]. This method uses the area and colour of the

skin to classify between face and non-face regions. The images are represented in RGB, HCI, and YCbCr colour models. This procedure loses effectiveness with movement and varying illumination. It is also susceptible to many false detections when the background is of the similar colour to the skin. Another drawback of skin-colour-based segmentation is that it varies across ages and is not uniform for all races.

The Viola-Jones algorithm [61], also referred to as boosted cascade of simple features, is a method that uses the AdaBoost learning algorithm; it is remarkably fast and can detect frontal view faces rapidly. This method achieves exceptional performance by using new methods that determine features very quickly and then separate the background from the face rapidly [61]. This method is used in this study for pre-processing.

2.3 Expression Classification

2.3.1 The Facial Action Coding System (FACS)

Early research on facial behaviour relied on human observers to scrutinise the subjects' faces and to subsequently make their analysis. But this can cause ambiguity, leading to inaccurate and unreliable results. Ekman et al. recognised this problem and, by bringing forward the influence of context to the observer, questioned how valid these observations are [62]. The context may give prominence to voice instead of face and the observations may further be made incoherent between different cultures due to misinterpretation [62]. To overcome these limitations, the FACS system [62] was created to represent facial expressions and behaviours with reference to a fixed set of facial parameters. This framework voids the facial behaviour of the face as a whole and focuses on the individual parameters. Facial Action Coding is an approach that is muscle-based and identifies the variety of facial muscles that cause changes in facial behaviours, whether the muscle movements are individual or in groups. These facial behaviour changes are called Action Units (AUs). The FACS consist of a few such action units. For example:

AU 1 is the action of raising the Inner Brow.

AU 2 is the action of raising the Outer Brow.

AU 26 is the action of dropping the Jaw.

There are additive and non-additive AUs. If the appearance of each AU is independent, the AUs are additive. But if the AUs modify each other's appearance, they are non-additive AUs [26].

Thus, expressions can be demonstrated by a combination of any number of additive or non-additive AUs. Figure 2.2 shows instances of upper and lower AUs as well as some of the facial movements they produce when combined.

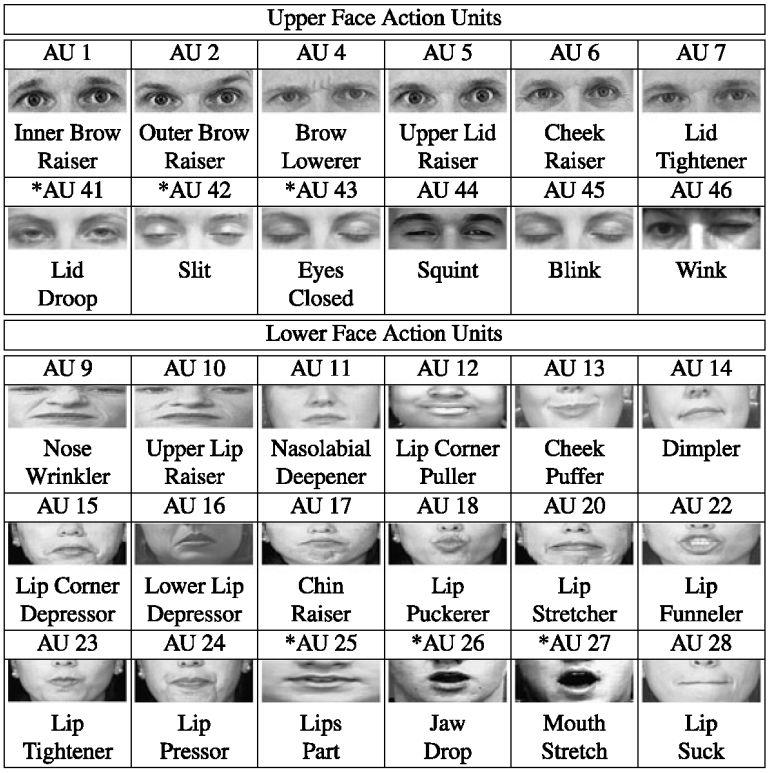


Figure 2.2: Upper and Lower Face AUs as well as their combinations [62].

2.3.2 The Six Prototypic Expressions

The six prototypic expressions are presented and labelled in Figure 2.3: anger, disgust, surprise, happiness, sadness and fear. When compared to the other facial expression possibilities, these six are the best researched. The observations made on these expressions show that the six prototypic expressions are not always mutually distinguishable; there therefore exists confusion between classifying among the different expressions. Sebe et al. [63] found that the most commonly confused pairs of emotions are anger and disgust, and fear and surprise. This confusion arises because these sets of expressions have the same types of facial movements. Modern-day FER systems still experience these issues [64–66]. However, the confusion between surprise and fear was not always apparent. Newer studies showed more confusion between happiness and fear [67, 68] and fear with anger [69, 70]. Also, sadness was confused with anger [65, 67, 68].

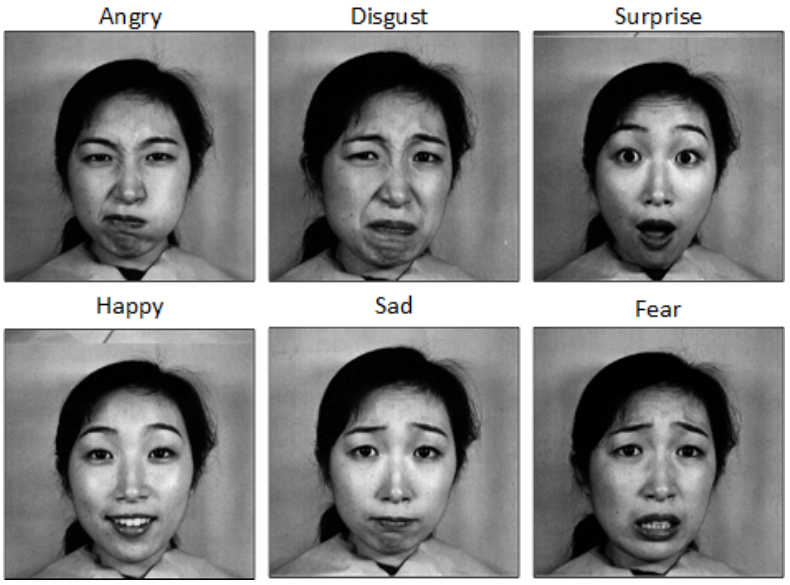


Figure 2.3: The six basic facial expressions from one subject of JAFFE dataset.

Artificial expressions that subjects make when instructed to are referred to as posed expressions. This instruction is generally given under normal test conditions or when subjects will be observed in a fixed environment. Spontaneous expressions follow opposite circumstances; these are expressed naturally without the subject being consciously aware thereof. Until recently, most research on FER systems focused on posed expressions only. This is due to the difficulty of obtaining datasets consisting of spontaneous expression classes. The most common method researchers use to elicit natural expressions from subjects is done by showing them films and clips that evoke emotions. Researchers discovered that eliciting spontaneous expressions of sadness and fear in subjects using this method of retrieval was difficult [68]. However, Cohn et al. [71] reported that eliciting fear was difficult but that was not the case for sadness. This difference can be accounted for by a variation in videos used in the studies. Anger was found to be the most difficult to elicit using videos because it requires more personal involvement to trigger [71]. The degree of ground truth involved with spontaneous expressions can also be challenging due to the context of the environment subjects use. This environment can cause suspicion, changing the natural expressions of the subjects [26]. Nevertheless, many psychologists believe that spontaneous expression recognition is superior to posed expression recognition. This study tests on both posed and spontaneous expression datasets.

2.4 Feature Classification Methods

The common feature classification methods used in FER are mentioned below:

- The unknown test subject is classified to a class using the minimum distance classifier. The classifier minimises the distance between the test subject data and the class multi-feature space, and the distance is expressed in terms of an index of similarity. This is done so that the minimum distance and the maximum similarity are identical. [72].
- Support Vector Machines (SVM) [64] are based on the idea of decision planes, which establish decision boundaries. A decision plane differentiates between a set of objects that have varying classes.
- Artificial Neural Network (ANN) [73] is a method used to classify samples that have numerous attributes. It is capable of modelling and processing nonlinear relationships between inputs and outputs in parallel.

2.5 Feature Extraction and Representation

The face representation for FER systems based on still and dynamic facial images spawned a variety of approaches over the past decades. This created many categorization classes for facial feature representation, the most general is the geometric-based and appearance-based approaches. Both of these approaches are summarised briefly in Tables 2.1 and 2.2 respectively. More can be read about the other categorisations in [12, 58, 74–76].

2.5.1 Geometric-based Feature Methods

Geometric features have shown good performance in FER and have proven to be efficient, but they are severely compromised by inadequate facial landmark detection and tracking [58, 76–78]. The geometric-based interpretation of the facial structure of various facial expressions can be defined as the following:

1) Using the position of points of facial features as visual information [56, 76].

2) Measuring facial feature points' geometric displacement [69, 70].

3) Forming a geometric graph that represents faces [79, 80].

The position of geometric points allows for an easy means of directly measuring faces' contours. This method was proposed by Zhang et al. [56] where 34 fiducial points were used to reflect the facial geometry of images that are still, conveyed into a feature vector. Rudovic et al. [76] made use of 39 facial landmark points in still images to portray facial expression in multiple views. Recently more attention was given to this method [58, 77, 78], and its use was integrated to dynamic facial images. Shin et al. [78] used 18 main feature point, as defined in MPEG-4 files, and then used the dense optical flow method for tracking the feature points in sequential frames. Another approach, introduced by Jain et al. [58], used Generalized Procrustes analysis to locate a minimum of 18 facial points, which were subsequently formed into a 136-dimensional feature vector for every facial image. This feature vector was used to describe the geometric structure of every frame, over time, in a video clip of a facial expression. Using temporal model classifiers, like Hidden Markov models (HMMs) and Dynamic Bayesian networks (DBN), these features can be modelled according to the facial expression dynamics.

A more generalised approach for the formulation of geometrics is quantifying facial movements. This is done by measuring facial points' displacement between a facial image and the reference image, which is the facial image that has a neutral expression. The analysis of the deformation of the face is considered similar to human observations of facial activities [69, 70, 80]. The process can be generalised into three steps: (1) the grids are tracked in consecutive frames over time through grid tracking and deformation, (2) the difference of node coordinates is calculated by comparing the image with the neutral expressions with the peak expression frame, and (3) these differences are used in the stage when classification takes place.

An alternative approach that used a parametric setup [13] was formed to complement the Action Unit protocol [79, 80]. To describe the facial representation it extracts facial components' shape, including brows, eyes, cheeks and lips. Distances between each facial component were calculated to determine feature vectors that defined the geometric structure of the face [81]. The feature vectors were then used in an ANN for classification, achieving good performance.

The main advantage of using geometric features is the low dimensionality and simplicity. However, all the methods used for constructing the geometric features are challenged by variations in lighting, non-rigid motion, image registrations error sensitivity as well as motion discontinuities [82]. Therefore, difficulty rests in creating a deterministic physical model of facial expressions that are better representations of facial geometrical properties and muscle movements for all facial expressions [83].

Table 2.1: Geometric-based feature methods research summary [83]

Reference	Features	Tracking method	Dynamic	Classifier
Zhang et al. (1998) [56]	34 facial feature points	Manual labelling	No	Two-layer perception
Tian et al. (2002) [79]	15 parameters of geometric features	Multi-state models	Yes	Three-layer neural network
Zhang & Ji (2005) [80]	Geometric deformation	Kalman filtering	Yes	Dynamic Bayesian networks
Kanajia & Metaxas (2006) [77]	78 facial feature points	Modified active shape model	Yes	Conditional random fields (CRF)
Kotsia & Pitas (2007) [69]	Geometric deformation feature	Kanade-Lucas-Tomasi (KLT) tracker	Yes	Support Vector Machine (SVM)
Kotsia et al (2008) [70]	Geometric deformation feature	-	No	SVM
Shin & Chun (2008) [78]	18 facial feature points	Dense optical flow	Yes	Hidden Markov models
Jain et al. (2011) [58]	68 facial points	Generalized proscrustes analysis	Yes	Latent-dynamic CRF
Rudovic et al. (2013) [76]	39 facial feature points	Active appearance model	No	SVM
Durmusoglu et al. (2016) [81]	18 facial feature points	Manual labelling	No	Artificial Neural Network

2.5.2 Appearance-based Feature Methods

Appearance-based features are recognised as more stable during image spatial transforms, especially for inaccurate misalignment and images with low-resolution in comparison to geometric features. Appearance-based features generally characterise an image of the face in terms of variation of pixel intensity or low-level features. Many studies during the past decade have introduced different approaches using appearance-based features for FER. Two of the most represented features used in FER are Gabor [56, 79] and LBP [39, 84, 85].

The Gabor feature implementation is popular due to its relation to how human's visual perception system works. A Gabor feature is made up of a sinusoid signal modulated by a Gaussian function, and each one defines the applicable frequency for the filter. Gabor energy filters are robust to contrast polarity and image alignment errors, hence they have produced some of the most successful FER systems thus far [83]. Littlewort et al. [86] used spatial Gabor energy filters as the main type of feature and achieved respectable performance classifying the seven basic emotions on the Cohn-Kanade dataset. Zhang et al. [56] preserved low dimension while retaining suitable performance by efficiently applying Gabor wavelets to 34 facial points instead of the whole image.

Local Binary Patterns have shown a high degree of accuracy in texture-based recognition tasks while still being easy to implement and having fast computation. In recent years, there has been a substantial influence of LBP in facial expression recognition [84, 87]. The LBP operator's main advantage in practical applications is its invariance against monotone gray-level changes, which are caused by illumination variations. A further considerable advantage is the simplicity of its computation, which allows for images to be analysed even in demanding real-time situations.

The LBP was found to be more discriminative and efficient than Gabor features by Shan et al. [39] when they used it to represent salient micro-patterns of face images to express facial expressions.

Succeeding Shan et al. [39], many efforts have been made to utilise variants of LBP directly, to estimate the intensity of facial expressions [88], facial AUs [89], multi-view facial expression recognition [85] and 3D facial AUs detection [90]. Other variants of LBP have emerged to attempt to enhance the representability to further improve performance or efficiency [91, 92]. Jabid et al. [91] proposed the Local Directional Pattern (LDP) for recognising facial expressions. They calculated the edge response values of each pixel point for all directions, and then a code was generated according to the relative magnitude strength. Numerous studies [93–95] present that LDP is superior to LBP for face and facial expression recognition. Srikrishna et al. [94] evaluate a method using LDN, which encodes the face's directional pattern and produces a code that is more selective. The directional patterns are computed using the compass mask and the information is encoded by means of the foremost direction indices and signs that differentiate between similar structural features.

Faces can be divided into several blocks; each block has the LDP operator applied to it and the individual feature vector of each block is later concatenated to form a single discriminative feature vector. Jun et al. [92] encoded compact LBP through the maximisation of the shared information of features and class labels. Ryu et al. [96] put forth the local directional ternary pattern for FER; they encode the shapes of the emotion-related features of the face, such as eyes, eyebrows, upper nose and mouth, by making use of the directional information. The face image is described spatially by using active patterns and sub-regions, which improve discriminability of emotion-related features. Further uses of LBP in the recognition of facial expression are mentioned in several studies [33, 89, 92, 97–100].

In dynamic facial expression recognition or video-based sequences, a recognised approach is using dense optical flow [101]. The method involves computing movements in the rectangular areas to estimate each face region's level of activity by catching the smooth flow and global information. This method is also capable of getting accurate time derivatives by using more than two frames. Later, Lien et al. [102] suggested a spatial-temporal descriptor that integrated dense optical flow, feature point tracking and high-gradient component analysis, and then HMMS were applied to classify fifteen different AUs. Optical flow was then proposed [103, 104]; it used horizontal and vertical components to represent motion patterns of facial expression. However,

optical flow efficiency is reduced due to it's sensitivity to image misalignment errors.

Image filter and techniques for texture descriptors for still images have recently gained popularity for recognising dynamic facial expressions. Examples of common image filters for face analysis are Haar features [57], Gabor wavelet representation [105] and independent component analysis (ICA) [106]. The temporal descriptor was designed using Gabor representations. The Gabor motion-energy filters were made as a representation that was biologically inspired for dynamic facial expressions [105]. Facial expressions can be decomposed into non-Gaussian signals using ICA. Recently, Long et al. [106] used ICA in natural videos to learn spatiotemporal filters, and subsequently constructed feature representations for input videos based on learned filters.

Zhao et al. [107] extended the LBP operator to dynamic images. They used LBPs to describe the temporal motion and the texture of appearance to obtain an effective description of dynamic facial expression. Almaev & Valstar [108] achieved good results for emotion recognition in unrestricted conditions by using LBP to encode the Gabor filters' multi-scale and multi-orientation templates, and they called this method the Local Gabor Binary Pattern from Three Orthogonal Planes (LGBP-TOP). Furthermore, Jiang et al. [109] proposed a new method that made use of local phase quantisation to describe facial actions' temporal information. They also noted that recent LBP studies showed that dynamic LBP outperforms describing facial expressions' temporal variation when compared to the temporal model making use of deep belief networks or HMMs [83].

2.5.3 Covariance-matrix-based Feature Methods

In recent years there has been an expansion of covariance descriptors in computer vision applications. The covariance descriptor was first proposed by Tuzel et al. [24] to represent an image region in applications of object detection and texture classification. The study used pixel locations, colour values and derivatives of the intensities as feature vectors for the covariance descriptor. Several studies [110–116] used Gabor filter features for tracking, person identification, and face recognition. Facial recognition using Gabor filters as features were implemented first [110] and then improved upon by using a kernel Gabor-based weighted region covariance matrix [117]. They constructed a weighted matrix by computing the similarity of each pixel within a face sample to emphasize features. Gabor features could carry more discriminative information and display strong characteristics of scale, spatial locality and orientation selectivity [82], but

Table 2.2: Appearance-based feature methods research summary [83]

Reference	Features	Dynamic	Classifier
Yacoob & Davis (1996) [101]	Optical flow	Yes	A rule based system
Zhang et al. (1998) [56]	Gabor	No	Two-layer perception
Tian et al. (2002) [79]	Gabor	Yes	Neural network
Buciu et al (2003)	Independent component analysis (ICA) and Gabor	No	Maximum correlation classifier
Feng et al. (2005) [84]	Local Binary Pattern(LBP)	No	Linear programming
Shan et al. (2005) [30]	LBP	No	SVM
Littlewort et al. (2006) [86]	Gabor	Yes	SVM
Yesin et al. (2006) [103]	Optical flow	Yes	Hidden Markov models
Zhao & Pietikainen (2007) [107]	Local Binary Pattern from three orthongonal planes	Yes	SVM
Shan et al. (2009) [39]	LBP and Boosted-LBP	No	SVM
Jabid et al. (2010) [91]	Local directional pattern	No	SVM
Wu et al. (2010) [105]	Gabor motion energy filters	Yes	Linear SVM
Moore & Bowden (2011) [85]	Variants of LBP	No	SVM
Jun et al. (2011) [92]	Compact LBP	No	Nearest neighbor classifier
Sanchez et al. (2011) [104]	Differential optical flow	Yes	SVM
Long et al. (2012) [106]	Spatiotemporal features based on ICA	Yes	SVM
Almaev & Valstar (2013) [108]	Local Gabor binary pattern from three orthongonal planes	Yes	SVM
Feng et al. (2013) [87]	LBP on key points	No	SVM
Jiang et al. (2014) [109]	Local Phase Quantization from three orthongonal planes	Yes	SVM
Srikrishna et al. (2015) [94]	Local Directional Number Pattern	No	Nearest neighbour classifier
Vishnudharan et al. (2016) [95]	LDNP	No	SVM
Ryu et al. (2017) [96]	Local Directional Ternary Pattern	No	SVM with RBF

convolving face images with multi-banks of Gabor filters to extract multi-scale and orientational Gabor coefficients is computationally expensive [118].

Another application covariance descriptors became popular in is action recognition [119–123], where motion-related feature vectors such as optical flow and locations of 3D joints are used. Wang et al. [119] verified the covariance descriptor representative effectiveness on the classification of image sets. An image set is defined as a collection of images that belong to the same class, but with variation, such as images of different views of the same object. The image set, instead of the individual object in the image, is classified [82]. In this case, each image of its respective class is vectorised into a feature vector and the covariance matrix of these feature vectors is computed to represent this set of images [82]. Guo et al. [118] introduced a novel feature descriptor using the covariance matrix for facial expression recognition. It used LBP features instead of Gabor features to improve the discriminating ability and to decrease the computational cost. An important factor to consider when using covariance descriptor is the feature selection

as well as the metric used for classification [124]. In an attempt to uncover the key attributes of the covariance descriptor, Faulkner et al. [124] characterised the interdependence between the choice of features and distance measures. They concluded that the region covariance descriptor would prove useful for methods that perform image super-resolution, deblurring, and denoising based on matching and retrieval of image patches from an image dictionary.

2.6 Visual Perception of Facial Expression Recognition

Facial expression recognition has improved significantly since its inception in computer vision. However, it still remains a considerable challenge for computer vision systems [125]. The human visual system remains superior by a great margin. Two fundamental questions that can help to improve FER systems are:

1) What makes humans remarkably adept at recognising facial expressions?

2) How can we take advantage of this skill?

The first question's answer lies in the study of visual perception, which shows that different visual cues are used by human observers to recognise different facial expressions [125]. Boucher & Ekman [126] stated that the whole face is used by human observers to recognise anger and surprise, while only the lower half of the face is looked at to recognise happiness and disgust. Gouta & Miyamoto's [126] research shows that mostly the top half of the face is used to recognise anger, fear, surprise and sadness, where as the lower half is better suited for disgust and happiness. Bassili's [127, 128] work differed in that the entire face was found to be useful in recognising basic facial expression (74.4%) versus only the lower (64.9%) or top (55.1%) part of the face. The importance of facial features, such as eyebrows, eyes, mouth and wrinkles, was studied by Smith et al. [129] and Roy et al. [130] for the static and dynamic recognition of the six basic expressions. The studies highlighted the exact facial features that human observers use to recognise each basic prototypic facial expression for static and dynamic facial expressions. They reported humans use the mouth instead of eyes for happiness and conversely the eyes instead of the mouth for fear. In other cases, humans use transient features like nasolabial furrow and wrinkles on the forehead for disgust and sadness respectively.

Future implementations of FER systems should take advantage of these human traits and their relative importance to improve performance, given that humans easily outperform machines at recognising facial expressions in everyday situations. Hammal et al. [131] attempted to utilise

this information by comparing their model to human performances in the same experimental condition as Smith et al. [129]. Their study tested performance using their model on partially occluded facial parts. The results showed that the model compared favourably to human-based performance but also showed differences between the visual cues used by the model and the human observers. Consequently, relative weights associated with each facial feature and its respective facial expression were derived to further refine the classification. These additional processing steps can be considered as fundamental improvements to FER systems that aim to get closer to human performances. In Section 4.4.1 we compare the mouth versus the eye regions of the face and their impact on the different facial expressions.

2.7 Conclusion

It is evident, from the literature, that the recognition of facial expression consists of multiple facets. Each facet has its own challenges and complexities. The covariance-matrix-based descriptor has minimal research concerning applications to facial expression recognition. This dissertation will contribute towards discovering the effects of using a covariance-matrix-based descriptor for feature representation. The study will also test the theories mentioned in Section 2.6 concerning the different segments of the face for expression classification. The use of local texture patterns such as LBP is prevalent throughout facial analysis. The study will use these local patterns and its variants with the covariance-matrix to propose a novel feature descriptor to classify the six prototypic expressions. This work will focus more on controlled environments but use datasets that consist of both posed and spontaneous expressions.

Chapter 3

Local Texture Patterns and Region Covariance Matrices

In this chapter, we explore the methods and algorithms implemented in this study to create an image descriptor for Facial Expression Recognition. The local texture patterns, such as LBP and LDP, are described in detail, followed by a brief overview of the construction of RCM, emphasising its properties and limitations. We also introduce the different metrics used to analyse RCM. Finally, we conclude with the implementation of the proposed LDCM operator.

3.1 Introduction

The use of local photometric descriptors has seen wide applications across texture recognition [132], object recognition [133, 134], wide baseline matching [135], image retrieval [136, 137], mining of video data [138], building panoramas [139], and recognising object categories [140–143]. These local descriptors are computed in terms of the regions of interest of an image because they are distinctive and robust to occlusion. Recently, studies have been aimed at making these descriptors invariant to image transformation [144]. The traditional approach is to discern image regions covariant to a class of transformations; these are then used as support regions to create invariant descriptors [144]. Given an invariant region detector, the choice of the most appropriate descriptor to characterise the region must then be considered. There are numerous possible descriptors and associated distance measures, which emphasis various image properties, like pixel intensity, colour, texture and edges. In this study, we focus on descriptors computed on gray-value images.

The accuracy of classification depends greatly on the information contained in the feature representation, thus an effective and discriminative feature set is the most important constituent of a successful FER system [145]. The best classifiers are still prone to fail if supplied with inconsistent or inadequate features. Nevertheless, in real-world applications, facial images are vulnerable to distortion by different factors, such as variations in lighting condition, pose, aging, alignment, and occlusion [146]. Hence, designing a robust feature extraction method that can perform consistently in changing environment is still a challenging task.

The Local Binary Patterns and Local Directional Patterns are local descriptors that have been used for FER because of their discriminability of facial features. The RCM popularity stems from its invariance to rotation and scale, making it robust to image transformations. This chapter further analyses RCM and local texture patterns. In Section 3.4, the methods, properties, and construction of RCM are examined. The algorithms and structures of the different local texture patterns are covered in Section 3.2.

3.2 Local Texture Patterns

3.2.1 The Principle of Edge Detection

The edges contained in digital images are the group of pixels whose gray values have step changes or the regions where the brightness of the image changes considerably [147]. The gray profile observed in these regions is typically considered as one step. In other words, in a minute buffer area of the image a gray value changes promptly to vastly different value. The edges within an image commonly exist between objects and objects, objects and backgrounds, and primitives and primitives (image element, such as an arc, from which more complicated images can be constructed). The edge of an object is highlighted in the interruption of the gray value [147]. The general approach to achieve effective edge detection is therefore to study the changes of one image pixel in a gray area. The edge detection is principally the measurement, detection and location of changes in image gray values [147]. When viewing images the most basic observable features are the edges and lines of images. According to the edge and line composition, the object structure can intuitively be determined. Hence, edge extraction is a vital method in processing images and extracting features.

3.2.2 Local Binary Pattern

The Local Binary Pattern operator is an image operator; it converts an image into an array or image of integer labels that describe the small-scale nature of the image. Further image analysis is done using these labels or their statistics (usually represented by histograms). The most widespread types of the operator are intended for still, monochromatic images, but have also been extended to colour images, videos ad volumetric data [148]. This section deals with the fundamentals [148] and the different versions of the LBP operator in the spatial domain [30, 84].

3.2.2.1 Original Basic LBP

Ojala et al. [149] introduced the LBP operator. It is founded on the assumption that texture has two locally complementary aspects: pattern and its accompanying strength. The original LBP operator works in a pixel block of an image that measures 3×3. The outer pixels of this block are thresholded by the value of its centre pixel, assigning '1' for a value greater than the threshold, and '0' for a value lower than the threshold. Starting from a reference pixel-point, the new outer thresholded values form a binary pattern consisting of $t_{p0}, ..., t_{p7}$, where t_p is the thresholded pixel. The LBP code is obtained by converting the binary pattern to a decimal number, as $LBP_{centre} = \sum_{p=0}^{7} t_p 2^p$. The new decimal code becomes the centre pixel value of it's neighbourhood. The 8-neighbourhood is considered for the evaluation of LBP code, therefore a total of $2^8 = 256$ various labels can be attained, depending on the relative gray values of the centre gray pixel as well as that of its 8-neighbourhood. An example of the LBP coded image is shown in Figure 3.1 and its code generation of one-pixel in a 8-neighbourhood is presented in Figure 3.2

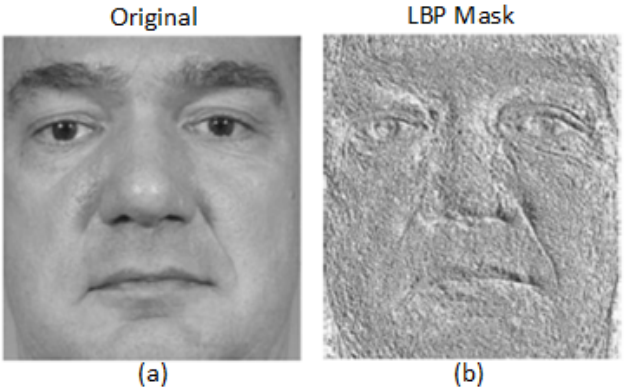
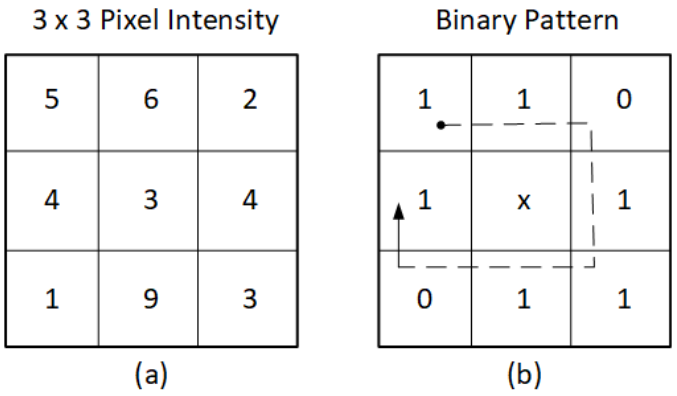


Figure 3.1: Example of (a) Input Image and (b) LBP mask image [148].



LBP = 11011101

Decimal =

$$(1 \times 2^0) + (0 \times 2^1) + (1 \times 2^2) + (1 \times 2^3) + (1 \times 2^4) + (0 \times 2^5) + (1 \times 2^6) + (1 \times 2^7) = 221$$

Figure 3.2: Example of LBP code generation, (a) Intensity Pixel Mask and (b) Threshold values.

3.2.2.2 Derivation of Generic LBP Operator

The original basic LBP operator was defined to operate using only pixels in the 8-neighbourhood. Thus, it was later transformed into a more generalised solution [150] to remove the limitations of the size of the neighbourhood or the number of sampling points. The generic LBP is derived below as formulated in [150–152].

Take a monochrome image I and let g_c signify the gray level of an arbitrary pixel (x, y), thus, $g_c = I(x, y)$. Furthermore, let g_p signify the gray value of a particular sampling point in an evenly spaced circular neighbourhood of P sampling points and radius R around point (x, y):

$$g_p = I(x_p, y_p), \quad p = 0, ..., P - 1. \tag{3.1}$$

$$x_p = x + Rcos(2\pi p/P), \tag{3.2}$$

$$y_p = y - Rsin(2\pi p/P), \tag{3.3}$$

Figure 3.3 demonstrates examples of local circular neighbourhoods.

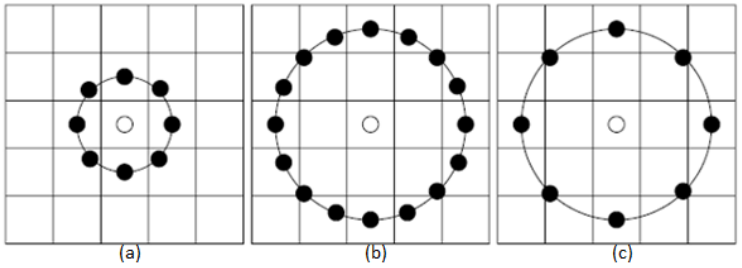


Figure 3.3: The circular (a) - (8,1), (b) - (16,2) and (c) - (8,2) neighbourhoods [150].

If we assume that the local texture of the image $I(x, y)$ is characterised by the joint distribution of gray values of $P + 1$ $(P > 0)$ pixels:

$$T = t(g_c, g_0, g_1, ..., g_{p-1}). \tag{3.4}$$

The value of the centre pixel can be subtracted from the neighbourhood without any loss of information:

$$T = t(g_c, g_0 - g_c, g_1 - g_c, ..., g_{p-1} - g_c). \tag{3.5}$$

The joint distribution is then approximated by assuming the centre pixel to be statistically independent of the differences. This allows for factorisation of the distribution:

$$T \approx t(g_c)t(g_0 - g_c, g_1 - g_c, ..., g_{p-1} - g_c). \tag{3.6}$$

Therefore, the first-factor $t(g_c)$ is the intensity distribution over the image I. When evaluating local textural patterns, factor $t(g_c)$ does not contain useful information. Thus, the joint distribution of differences

$$t(g_0 - g_c, g_1 - g_c, ..., g_{p-1} - g_c). \tag{3.7}$$

can be used to model the local texture. However, difficulty lies in achieving reliable estimation in this multidimensional distribution of image data. To alleviate this shortcoming, vector quantisation of the distribution was proposed [153]. To reduce the dimensionality of the high dimensional feature space, they implemented learning vector quantisation using a codebook of 384 codewords. Hence, this operator that is based on signed gray-level differences can be seen as a texton (fundamental micro-structures in natural images) operator [154]. Consequently, the

learning vector quantisation-based approach exhibits certain downfalls that reduce its effectiveness [148]: a) the differences $g_p - g_c$ are invariant to the image's mean gray value but not to other differences in gray levels, b) to use it for texture classification the codebook must have similar training to other texton-based methods. Therefore, only the signs of the differences are considered to solve these challenges:

$$t(s(g_0 - g_c), s(g_1 - g_c), ..., s(g_{p-1} - g_c)). \tag{3.8}$$

where $s(z)$ is the thresholding (step) function

$$s(z) = \begin{cases} 1, & z \geq 0 \\ 0, & z < 0 \end{cases} \tag{3.9}$$

The generic LBP operator is derived from this joint distribution. It is obtained by summing the threshold differences weighted by powers of two, similar to the basic LBP. The $\text{LBP}_{P,R}$ operator is defined as

$$\text{LBP}_{P,R}(x_c, y_c) = \sum_{0}^{P-1} s(g_p - g_c)2^p \tag{3.10}$$

The main differences between the basic LBP and $\text{LBP}_{8,1}$ operators, is that the neighbourhood in the generalised LBP is indexed circularly, which promotes the creation of texture descriptors that are rotation invariant. Also, the diagonal pixels in the 3×3 neighbourhood are interpolated in $\text{LBP}_{8,1}$ [148].

3.2.2.3 Uniform Local Binary Pattern

The Uniform Local Binary Pattern (ULBP) [150] operator is an extension of the LBP operator. It is advantageous over LBP because it inherently creates features that are invariant or robust to rotations of the input image. The $\text{LBP}_{P,R}$ patterns are generated by sampling circularly around the pixel in the center of the neighbourhood, which causes two effects should the input image be rotated. Firstly, each local neighbourhood is rotated so that they are in other pixel locations. Moreover, the sampling points on the circle that surround the centre point in each neighbourhood are rotated so that they are in different orientations [148].

The ULBP is defined by uniform patterns when the bit pattern is considered to be circular. A pattern is determined to be regular or uniform when the number of bitwise transitions between 0 and 1 in either direction is two at most. For example, consider a 3 x 3 pixel block, where the centre pixel has 8 neighbours. This produces an 8-bit pattern; the patterns 00000000 (no transitions), 11111111 (no transitions), 00000100 (2 transitions), 11111101 (2 transitions) are uniform, whereas 01010011 (6 transitions) and 11001001 (4 transitions) are not uniform. Each uniform patten has separate output labels and a single assigned label for all non-uniform patterns. Thus, there are $P(P-1)+3$ various output labels to map patterns of P bits. Hence, a neighbourhood of 8 sampling points produces 59 distinctive uniform output labels and a neighbourhood of 16 sampling points give 243 unique uniform labels.

The ULBP operator is beneficial to FER because majority of the LBPs in natural images are uniform [150]. In texture images 90% of all patterns are uniform, when using the $(8,1)$ neighbourhood and around 70% in the $(16,2)$ neighbourhood. It was found, using facial images, that 90.6% of the patterns in the $(8,1)$ neighbourhood and 85.2% of the patterns in the $(8,2)$ neighbourhood are uniform [148]. The ULBP easily detects different texture primitives, such as edges, corners, flat regions, line ends and spots (examples of these patterns can be seen in Figure 3.4), thus ULBP efficiently represents local facial features, making it coherent when representing the facial texture that adheres to expression recognition.

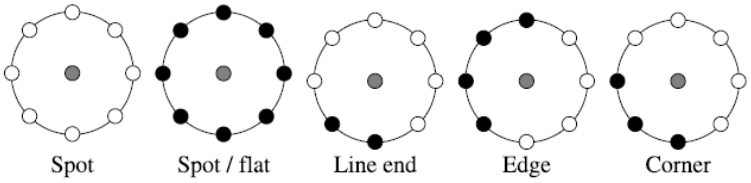


Figure 3.4: Different texture primitives detected by ULBP where black spots represents 1 and white spots represents 0 [148].

3.2.3 Local Directional Pattern

The Local Directional Pattern describes local image features by computing the values of the edge response to all its neighbours; that is in all 8 directions for all the pixel positions. A code is then generated from the relative strength magnitude, Jabid et al. [91] established that edge responses are more stable than intensity values when noise and non-monotone illumination changes are

present. Local Directional Pattern, therefore, performs better in these environments as compared to its predecessor LBP.

The LDP is made up of an eight-bit binary code that is assigned to each pixel in an input image. This pattern is encoded using an edge response value of pixels in various directions. There are different edge detectors such as Kirsch, Prewitt and Sobel that can be utilised for this. The Kirsch edge detector is more proficient at detecting directional edge responses because it considers all eight neighbours as compared to the others [91]. Kirsch masks are shown in Figure 3.5 and an example of the output images for the Kirsch edge responses are shown in Figure 3.6.

$$\begin{bmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{bmatrix} \quad \begin{bmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{bmatrix} \quad \begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix} \quad \begin{bmatrix} 5 & 5 & -3 \\ 5 & 0 & 5 \\ -3 & -3 & -3 \end{bmatrix}$$

East M_0 North East M_1 North M_2 North West M_3

$$\begin{bmatrix} 5 & -3 & -3 \\ 5 & 0 & 3 \\ 5 & -3 & -3 \end{bmatrix} \quad \begin{bmatrix} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{bmatrix} \quad \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ 5 & 5 & 5 \end{bmatrix} \quad \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & 5 \\ -3 & 5 & 5 \end{bmatrix}$$

West M_4 South West M_5 South M_6 South East M_7

Figure 3.5: Kirsch edge response masks in eight directions [91].

Figure 3.6: Example of output images for the Kirsch edge response directions [155].

Each mask $(M_i)_{i=0,1,...,7}$ represents a different orientation. For each mask M_i we compute the response m_i. In total we obtain response values $m_0, m_1, ..., m_7$, each representing the edge significance in their corresponding directions. The higher the response value, the more significant the edge is in that direction. Local Directional Pattern code is generated by using the k most prominent directions. The bits corresponding to the k most significant directional responses $|b_i|$ are set to 1 and the $(8-k)$ bits that remain are set to 0. The code LDP_k is then computed as

$$LDP_k = \sum_{i=0}^{7} b_i(m_i - m_k) \times 2^i \tag{3.11}$$

$$b_i(n) = \begin{cases} 1 & n \geq 0 \\ 0 & n < 0 \end{cases} \tag{3.12}$$

where m_k is the kth most significant response.

Figure 3.7 demonstrates positions of the eight directional edge responses and the positions of the LDP binary bits. Figure 3.8 is an example of the LDP code generation using 3 prominent directions $(k = 3)$.

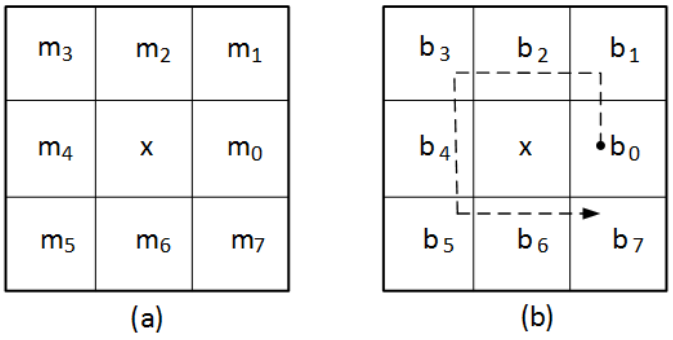


Figure 3.7: (a) The 8 directional edge response positions, (b) The LDP binary bit positions.

85	32	26
53	50	10
60	38	45

$\Rightarrow$

Mask Index	m_7	m_6	m_5	m_4	m_3	m_2	m_1	m_0
Mask Value	161	97	161	537	313	97	-503	-393
Rank	6	7	5	1	4	8	2	3
Code Bit	0	0	0	1	0	0	1	1
LDP Code	19							

Figure 3.8: Example of LDP code generation with the left matrix being the intensity mask of a region in an image and k=3.

3.2.3.1 Robustness of LDP

The LDP pattern is less susceptible to pattern changes when noise and non-monotone illumination changes are present, as compared to LBP. This is due to the fact that edge responses are more stable than the intensity values [91]. An example is given in Figure 3.9, where LBP and LDP codes are generated after Gaussian white noise is added to the original image. From the corresponding image after the addition of noise (Figure 3.9 (b)), the 5th bit of LBP changed from 1 to 0, thus LBP pattern converted from uniform code to non-uniform code. However, the LDP code remained unchanged, showing that edge response values exhibit greater stability than gray values when noise and non-monotone illumination changes are present.

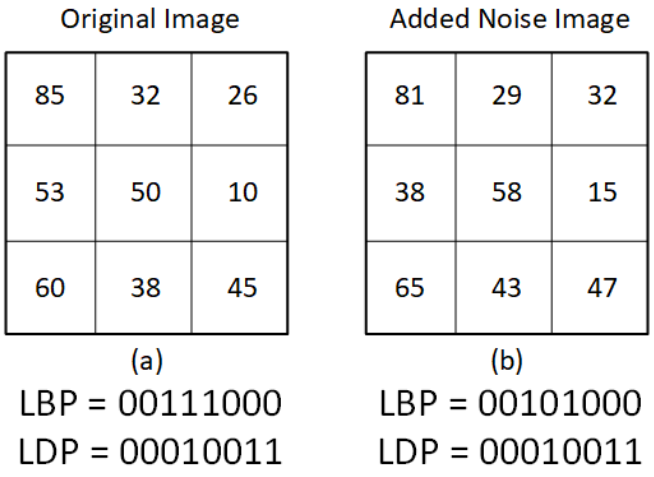


Figure 3.9: Superior stability of LDP shown where (a) Original Image and (b) Added Noise Image [91].

3.3 Edge Detection Model: Sobel Operator

The advantage of using the Sobel operator for edge detection is that it has somewhat of a smoothing effect on the image's random noise, and because it is the differential of two rows and columns, both sides of the the elements of the edge are enhanced, making them appear thick and bright [147]. The Sobel operator can be considered an orthogonal gradient operator. The gradient relates to the first derivative and gradient operator corresponds to a derivative operator. The Sobel operator can be derived following Gao et al. [147].

For a continuous function $f(x, y)$, in the position (x, y), the gradient can be expressed as a vector, where the two components are two first derivatives which are along the X and Y directions respectively:

$$\nabla f(x, y) = [G_x \quad G_y]^T = \begin{bmatrix} \frac{\delta f}{\delta x} \\ \frac{\delta f}{\delta y} \end{bmatrix} \tag{3.13}$$

The magnitude and direction angle of the vector are respectively:

$$mag(\nabla f) = |\nabla f_{(2)}| = [G_x^2 + G_y^2]^{\frac{1}{2}} \tag{3.14}$$

$$\phi(x, y) = arctan(\frac{G_x}{G_y}) \tag{3.15}$$

For each pixel location, the partial derivatives of the formulas above are calculated. A small area template convolution is used for approximation. In Figure 3.10, the two 3 x 3 templates used by Sobel are shown. These kernels convolute with every point in the image. Each kernel has a maximum response to either the vertical edge or the level edge. The point's output bit is determined by the maximum value of the two convolutions, resulting in an image of edge amplitude.

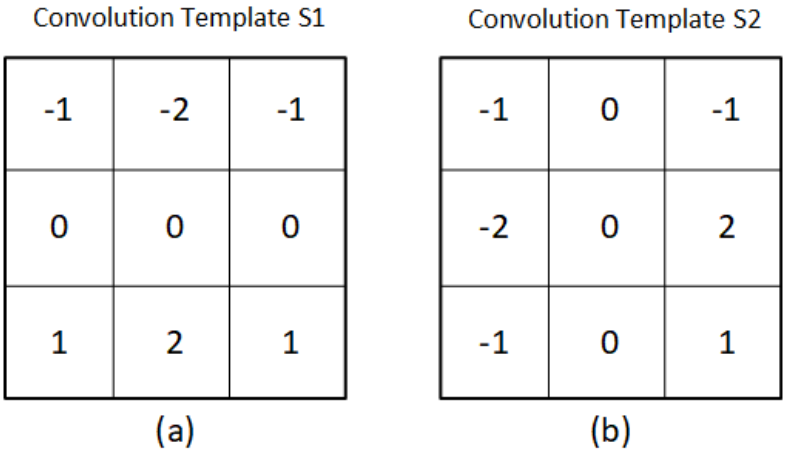


Figure 3.10: The Sobel Edge Operator [147].

Their convolution is as follows:

$$g_1(x,y) = \sum_{k=-1}^{1}\sum_{l=-1}^{1} S_1(k,l)f(x+k,y+l) \tag{3.16}$$

$$g_2(x,y) = \sum_{k=-1}^{1}\sum_{l=-1}^{1} S_2(k,l)f(x+k,y+l) \tag{3.17}$$

$$g(x,y) = g_1^2(x,y) + g_2^2(x,y) \tag{3.18}$$

If $g_1(x,y) > g_2(x,y)$, that means there is an edge that has a vertical direction that passes through the point (x,y), alternatively, if $g_1(x,y) < g_2(x,y)$, then there is an edge that has a level direction that passes through the point (x,y). If the pixel value of the point (x,y) is $f(x,y)$, this point is determined as an edge point, if $f(x,y)$ satisfies one of the following two conditions[147]:

a) Condition 1:

1. $g(x,y) > 4 \times \sum_{i=1}^{row}\sum_{j=1}^{col}\frac{g^2(i,j)}{row \times col}$
2. $g_1(x,y) > g_2(x,y)$
3. $g(x,y-1) \leq g(x,y)$
4. $g(x,y) \geq g(x,y+1)$

b) Condition 2:

1. $g(x,y) > 4 \times \sum_{i=1}^{row}\sum_{j=1}^{col}\frac{g^2(i,j)}{row \times col}$
2. $g_1(x,y) > g_2(x,y)$
3. $g(x-1,y) \leq g(x,y)$
4. $g(x,y) \leq g(x+1,y)$

In the formulas above, row and col indicate the number of rows and columns in the image, respectively.

3.4 Region Covariance Matrices

3.4.1 Properties of Covariance Representations

A covariance matrix belongs to the set of symmetric positive-definite (SPD) matrices.

The set of SPD matrices with the size $d \times d$ can be defined as:

$Sym_d^+ = \{A|A = A^T, \forall x \in \mathbb{R}^d, x \neq 0, x^T Ax > 0\}$. Using any two SPD matrices A_1 and A_2 and two positive scalars α and β, then $\alpha A_1 + \beta A_2$ is also SPD. Therefore, SPD matrices form a convex cone, which is a Riemannian manifold in the Euclidean space [113]. A Riemannian manifold is a real smooth manifold that is differentiable and equipped with an inner product that is smoothly varying for each tangent space. An illustration is provided in Figure 3.11. A difficulty posed by this manifold structure is that it becomes challenging to process and analyse SPD matrices. To evaluate between SPD matrices, a similarity or dissimilarity metric is needed. A serious issue for SPD matrices is how to effectively and efficiently measure the similarity between them. In Figure 3.11 it is demonstrated that for accurate measurement, methods that promote the geodesic distance properties are preferred to Euclidean [82]. Devising such distance measures remains an open issue.

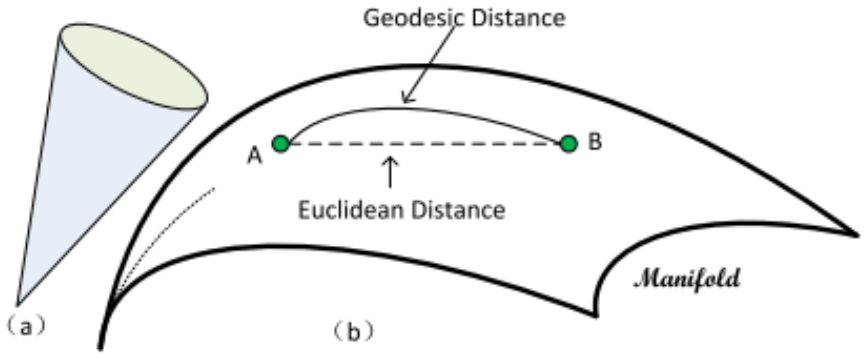


Figure 3.11: Visualisation of Riemannian manifold of SPD matrices. (a) Sym_d^+ forms a closed, self-dual convex cone, which is a Riemannian manifold in the Euclidean space $\mathbb{R}^{d \times d}$ [82].

3.4.2 Methods for Covariance Representations

The covariance descriptor was originally proposed as a region descriptor for feature representation. The feature vector of a given image region is extracted from each pixel to describe the pixel properties, such as location, gradient, filter response, etc. The covariance matrix is computed using these feature vectors to characterise this region. As introduced in Section 2.5.3, the covariance descriptor has gained much interest in various computer vision and image processing tasks. Successively, the uses of RCM has branched out to wider applications, promoting research on how to process and improve covariance representations. The research relative to these improvements has been categorised into three areas: covariance representation, similarity measures, and classification methods.

Covariance representation aims to improve the quality of the covariance matrix for better expression of features. Gabor features extract more discriminable information as compared to the first- and second-order gradient features. Hence, it was preferred to compute the covariance matrix for face recognition [110]. When used in object tracking [112, 114], pixels are weighted in calculating the covariance matrix. The pixels that are further from the centre of a region are lower in weight. In action recognition [121], to limit background pixels, the covariance matrix is computed using only the pixels whose temporal gradients are above a certain threshold.

Similarity measures for covariance representation are a fundamental issue in the analysis of SPD matrices. Since SPD matrices reside on a Riemannian manifold [113], commonly used Euclidean-based measures lack efficiency because they do not consider the manifold structure in computation. A proposed solution to this problem is using Affine-Invariant Metric (AIM) [156] for comparing covariance matrices. Although AIM improves similarity measurement, it involves using matrix inverse and square rooting, resulting in high computational cost when the dimensions of SPD matrices are large. The past decade has offered more contributions in an attempt to produce effective similarity metrics for SPD matrices. One contribution mapped the manifold to a Euclidean space [112], that is the tangent space at the mean point. However, these approaches suffer from two main limitations [82]: a) mapping the points between the manifold and the tangent space or vice-versa is computationally expensive, and b) the tangent space is only a local approximation of the manifold at the mean point, thus, it may lead to a suboptimal solution. To address these issues, kernel-based methods have been generalised to handle SPD data residing on a manifold [82]. A point X on a manifold M is mapped to a feature vector $\phi(X)$

in some feature space *F*. The mapping is implicitly induced by a kernel function $k : (M, M) \rightarrow \mathbb{R}$, which defines the inner product in *F*, that is $k(X_i, X_j) = \langle \phi(X_i), \phi(X_j) \rangle$. The advantages of this approach are:

1. By selecting an efficient kernel, the computational cost can be reduced.
2. The manifold structure can be well incorporated in the embedding.
3. Euclidean algorithms like SVM can be used.

Classification methods. The ability to gain a proficient similarity measure helps greatly in classifying. For example, when a respectable similarity measure is available, the k-nearest neighbour (KNN) classifier will be able to achieve excellent classification performance. Furthermore, if a valid kernel function is available, then SVM classification can be applied.

3.4.3 Construction of RCM

Tuzel et al. originally proposed the RCM feature descriptor [24]. Let I be a one-dimensional intensity (grayscale) or three-dimensional colour image (RGB, HSV infrared, depth images) and F be the $W \times H \times d$ dimensional feature image extracted from I, we have

$$F(x, y) = \phi(I, x, y), \tag{3.19}$$

where the function ϕ can be any mapping such as colour, intensity, filter responses, gradients, etc. For a given rectangular region $R \subset F$, let $\{Z_i\}_{i=1..S}$ be the d - dimensional feature points inside R. The region R is represented with the $d \times d$ covariance matrix of feature points

$$C_R = \frac{1}{S-1} \sum_{i=1}^{S} (z_i - \mu)(z_i - \mu)^T, \tag{3.20}$$

where μ is the mean of the feature vector z_i,

$$\mu = \sum_{k=1}^{s} z_i \tag{3.21}$$

The covariance matrix structure represents the diagonal entries as the variance of each feature, and the non-diagonal entries are their respective correlations. This inherent representation provides multiple advantages to the region covariance descriptor. It allows the fusing of different types of features that share some correlation with each other. Its robustness allows matching in

different views and poses from a single covariance matrix extracted from a region. Noise from the sample is reduced considerably during the computation of the covariance due to the average filter. RCM are low-dimensional and as a result of symmetry; C_R has only $\frac{(d^2+d)}{2}$ different values [112].

The use of common machine learning methods on a standard covariance matrix is prohibited as it does not lie on Euclidean space [24]. To be able to classify between these symmetric positive definite matrices a dissimilarity metric was developed [157], which calculates the distance between feature points of two covariance matrices, C_1 and C_2. It can be described as

$$\rho(C_1, C_2) = \sqrt{\sum_{i=1}^{d} ln^2 \lambda_i(C_1, C_2)} \tag{3.22}$$

where $\{\lambda(C_1, C_2) \mid i = 1, 2, ..., d\}$ are the generalised eigenvalues of C_1 and C_2, computed from

$$\lambda_i C_1 u_i = C_2 u_i, \quad i = 1, 2, ..., d \tag{3.23}$$

and $u_i \neq 0$ are the generalised eigenvectors.

3.4.4 Log-Euclidean Metric on SPD Manifold

The SPD manifold is a topological space that is locally similar to Euclidean space. It has a globally defined differential structure, which allows the derivatives of the curves on the manifold to be defined. From Huang et al. [158], using the logarithm map $\log_{S_1} : \mathbb{S}_+^d \rightarrow T_{S_1}\mathbb{S}_+^d (S_1 \in \mathbb{S}_+^d)$, the derivatives at point S_1 on the manifold lie in a tangent space $T_{S_1}\mathbb{S}_+^d$, which has an inner product $\langle , \rangle_{S_1}$. The Riemannian metric of the manifold is the collection of inner products that are on all tangent spaces. Hence, the geodesic distance between two points $S1$ and $S2$ on the SPD manifold can be calculated by $\langle log_{S_1}(S_2), log_{S_1}(S_2) \rangle_{S_1}$.

The AIM [156] and Log-Euclidean Metric (LEM) [159] are the two most popularly used Riemannian metrics on the SPD manifold, because of their smoothly varying inner product and their qualification to derive true geodesic on the SPD manifold. As mentioned previously, the AIM is computationally too expensive to work in practice, due to the curvature of the SPD manifold. Contrastingly, LEM only requires Euclidean computations in the domain of matrix logarithms.

Therefore, it results in a strong reduction in computation time. In this study, we compare the LEM and the Geodesic Equation 3.22 on the manifold of SPD matrices for FER.

3.4.4.1 Derivation of LEM for SPD Manifold

In the study [159], LEM for the SPD manifold $\mathbb{S}_+^d$ is derived by capitalising on the Lie group structure under the group operation $S_1 \odot S_2 := exp(log(s_1) + log(S_2))$ for $S_1, S_2 \in \mathbb{S}_+^d$ where $exp(\cdot)$ and $log(\cdot)$ denote the matrix exponential and logarithm operators.
From [158], LEM on the Lie group of SPD matrices relates to a Euclidean metric in the SPD matrix logarithmic domain. Using LEM on $\mathbb{S}_+^d$, the scalar product between two elements T_1, T_2 in the tangent space at a point, S is given by:

$$\langle T_1, T_2 \rangle_S = \langle D_S log.T_1, D_S log.T_2 \rangle \tag{3.24}$$

where $D_S log.T$ represents the directional derivative of the matrix logarithm at S along T. The logarithmic and exponential maps associated with the metric can be shown in terms of matrix logarithms and exponential:

$$log_{S_1}(S_2) = D_{log(S_1)} exp.(log(S_2) - log(S_1)), \tag{3.25}$$

$$exp_{S_1}(T_2) = exp(log(S_1) + D_{S_1} log.T_2) \tag{3.26}$$

Further details for Equations 3.24, 3.25 and 3.26 are found in study [159].

Using Equations 3.24, 3.25 and 3.26, the geodesic distance between two SPD matrices is achieved by LEM:

$$\begin{aligned} D_{le}(S_1, S_2) &= \langle log_{S_1}(S_2), log_{S_1}(S_2) \rangle_{S_1} \\ &= ||log(S_1) - log(S_2)||_F^2 \end{aligned} \tag{3.27}$$

which corresponds to a Euclidean distance in the logarithmic domain; that is the tangent space at identity matrix. The distance between any two points on the SPD manifold using the LEM framework is obtained by propagating by translation the scalar product in the tangent space at identity matrix. Therefore, the space of SPD matrices is reduced to a flat Riemannian space using LEM [159].

3.5 Local Directional Covariance Matrix

The success of a region covariance matrix as a descriptor relies on the pixelwise features chosen for its specified operation. The LDP and RCM operators are designed to detect textures. Facial expression of a person can be regarded as a texture of the face. Pixel location and intensity are used in the RCM as it improves its discrimination ability [113]. The pixelwise mask of the LDP-generated image will also be incorporated into the RCM. Thus, we form a novel mapping function that is founded on local directional feature defined as

$$\phi(I, x, y) = [x \quad y \quad I(x, y) \quad LDP(x, y)]^T \tag{3.28}$$

The feature vector in region R can now be defined as $z_k = \phi(I, x_R, y_R)$, $z_k \in R^d, k = 1, 2, ..., n$, and the covariance matrix C_R can be derived by substituting (3.28) into (3.20).
The LDCM mapping has a total dimension of $d = 4$ and the consequent covariance matrices are 4×4 in size. This feature descriptor is considerably smaller than other methods, such as LDP or LBP. The advantage of LDCM is that it is more compact than traditional LBP or LDP. The incorporation of the LDP features versus LBP makes it more stable in the presence of noise, and the inherent structure of the region covariance matrix makes it rotation and scale invariant. The summary of the computation for the LDCM descriptor is depicted in Algorithm 1.

Algorithm 1 LDCM generation

Input: Image: $I(x, y)$
Output: Covariance Matrix Descriptor: C_{4x4}
Detect and Crop the face F from I
Compute LDP mask of image $I(x, y)$ = $LDP_{mask}(x, y)$
Generate feature matrix: $\phi(I, x, y) = [x \quad y \quad F(x, y) \quad LDP_{mask}(x, y)]^T$
Calculate covariance matrix C of feature vector $\phi(I, x, y)$

3.6 Conclusion

This chapter has outlined the methods used in this study to create a robust image descriptor for FER. It presented the detailed mathematical and algorithmic descriptions of the techniques implemented, that is, local texture patterns (LBP, ULBP, and LDP) and RCM. The study also shows that specific distance metrics, such as Geodesic and Log-Euclidean, are the preferred choice to optimise classification when using RCM. The intuition to take advantage of the individual properties of the studied methods and systems led to the creation of the proposed LDCM descriptor. The proposed LDCM method incorporates the benefits of local descriptors and RCM by fusing them into a single descriptor. The advantages include compactness, stability in the presence of noise, and rotation and scale invariance. In the next chapter, we evaluate the methods and systems proposed, showing their effectiveness in real-world scenarios for FER.

Chapter 4

Experimental Results

This chapter covers the implementation of the methods previously described in Chapter 3 with application to FER. Two main concepts are researched: effectiveness of proposed LDCM descriptor and holistic versus component-based approaches. Further tests examine efficiency of LEM for classification, and use cross-database evaluations with posed and spontaneous expressions to simulate real-world scenarios.

4.1 Introduction

This dissertation investigates whether facial expressions can be successfully classified by using covariance descriptors with various pixel-level features. An FER system using RCM and local texture patterns was presented to specifically address issues in challenging facial expression recognition. Using the flow diagram in Figure 4.1, we describe the presented system as follows. An input face image is pre-processed by converting the image to grayscale and cropping the face area using the Viola-Jones algorithm. Two approaches are discussed in this study: holistic-based and component-based. The holistic approach uses the entire face while the component-based segments the face into smaller regions. The next step for both methods involves generating the pixel feature matrices from different image masks. Figure 4.2 shows examples of the different image masks of the face used to obtain the feature matrix. Once the feature matrix is compiled using Equation 3.28 then the covariance matrix can be computed by Equation 3.20. Using the covariance matrix as a facial expression descriptor, facial expression can then be classified.

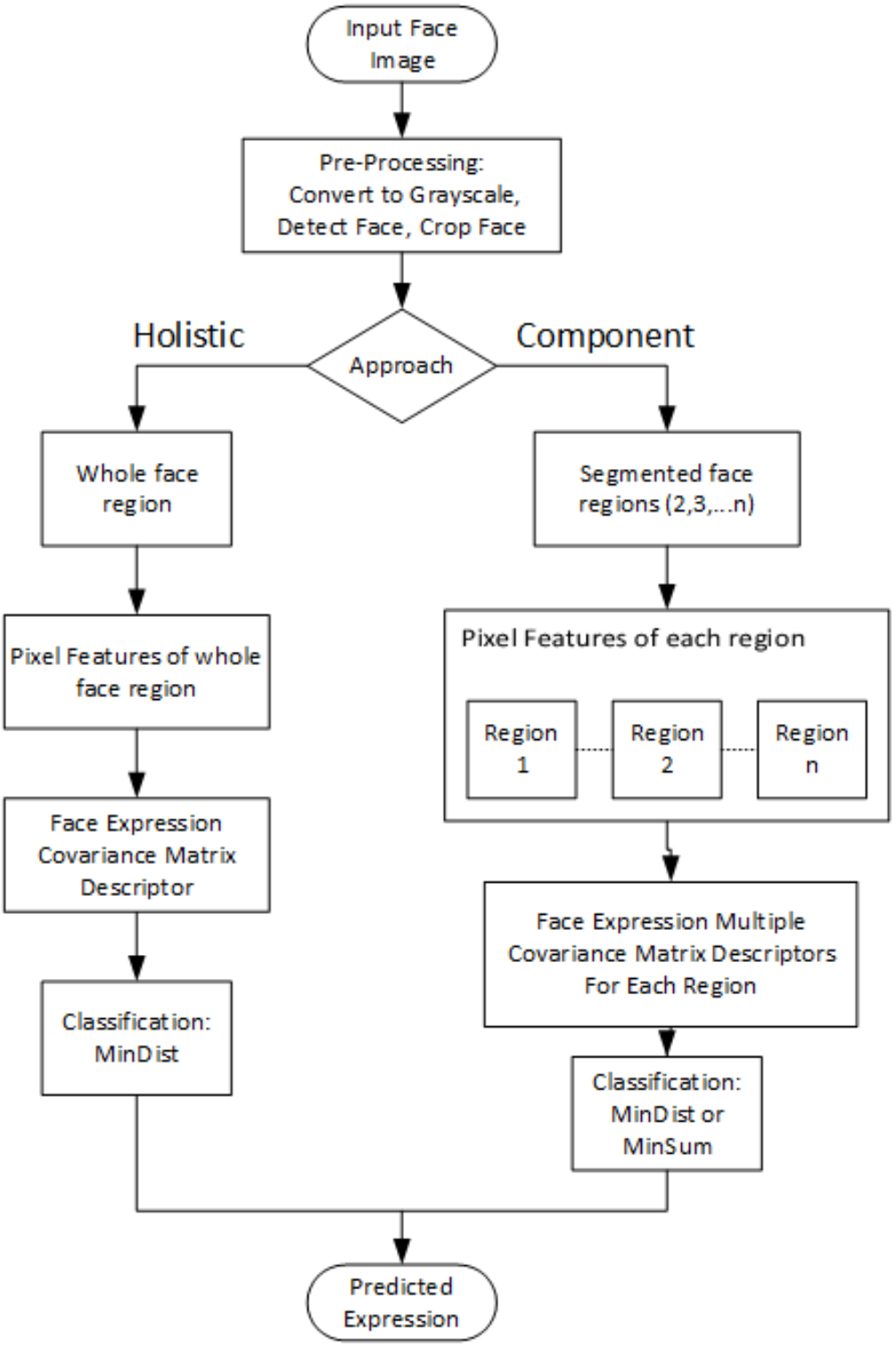


Figure 4.1: Flow Diagram of proposed FER system.

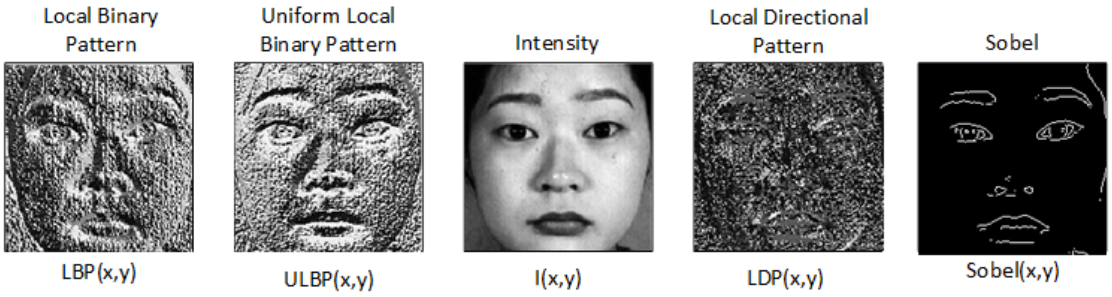


Figure 4.2: The pixelwise feature masks used in the RCM structure.

The classifier has two methods of determining classification: MinDist and MinSum.

Method 1: MinDist uses the minimum distance between covariance matrix descriptors based on a medium K-nearest neighbour with distance metric shown in Equation 3.22.

Method 2: MinSum uses the minimum sum of total regions of covariance matrices. The minimum sum of distances finds the minimum distance for each special region (eyes, nose, and mouth) then adds them together. The smallest sum is chosen for classification. This method utilises the discriminable features of all special regions instead of just the most dominant one. It also uses a medium KNN with distance metric shown in Equation 3.22.

Figure 4.3 helps to visualises the two classification methods used.

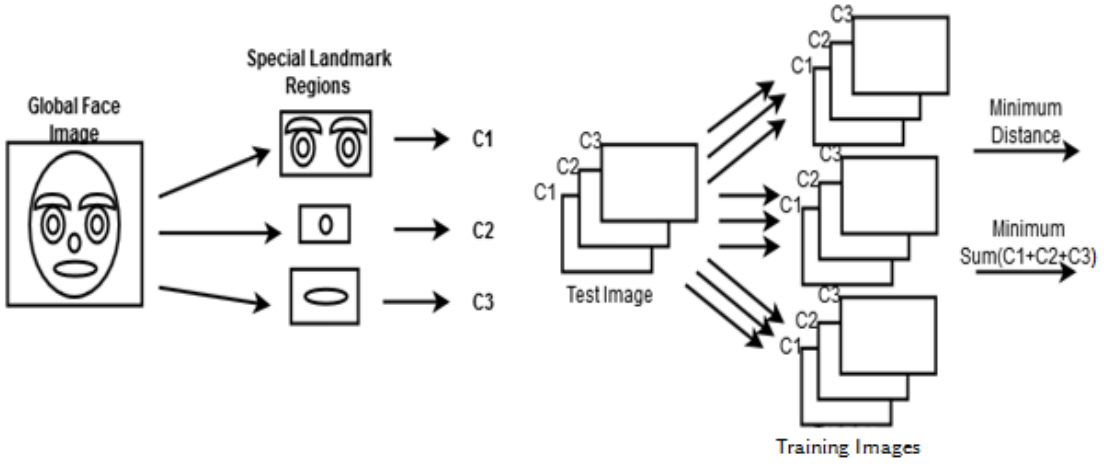


Figure 4.3: Classification using MinDist or MinSum methods.

Expression classification evaluation is performed with the leave-one-out-cross-validation technique to determine the accuracy per expression class for each dataset used. This is described in Algorithm 2. The expression classes consist of different facial expressions in each dataset,

for example, the happy class. The expression class happy then consists of $T_{i=1,\ldots,n}$ happy images. It is evaluated against the training class consisting of $T_{j=1\ldots m}$ mixed expression class images. Each image in the test expression class is evaluated against all images in the training class. The total number of correct classifications (T_c) is divided over the total number of images in the expression class used (T_n) giving the overall accuracy of the expression being evaluated, $T_c/T_n \times 100 = A_h$ (accuracy of expression class happy). The mean class average of the dataset can then be calculated by summing all of the individual expression accuracies (Ai) and dividing by the total number of expression classes (Te), $(\sum_{i=1}^{T_e} A_i)/T_e \times 100$.

Algorithm 2 Leave-One-Out-Cross-Validation

Input: Expression Class
Output: Expression Class Accuracy %
  for Each Test Image in Expression Class $T_{i=1,\ldots,n}$ **do**
    Get Covariance Descriptor for T_i
    for Each Image in Training Class $T_{j=1,\ldots,m}$ **do**
      Get Covariance Descriptor for T_j
      Compute Covariance Distance Similarity Metric $d(T_i, T_j)$
    end for
    Classify using MinDist or MinSum
    if Classification = true **then**
      increment T_c
    end if
  end for
  Accuracy = $(T_c/T_n) \times 100$

In this chapter, the review of the performance of the proposed algorithm for facial expression recognition, on JAFFE [35] , Extended Cohn-Kanade [36] and ISED [37] facial expression databases is presented. The analysis is divided into five experiments. Firstly, the performance of various covariance-based features tested against the whole face region is examined. Next, we explore the impact of segmenting the face into equal-sized regions for FER. We then focus on using special landmarks of the face with emphasis on eye versus mouth region. Additional tests review the efficiency of LEM for classification and simulate real-world environment scenarios by using cross-database evaluation of posed and spontaneous expressions.

4.2 Experiment 1: Global Face Covariance Features

In this experiment, LDCM is used to analyse the face holistically and to determine its effectiveness against different facial expressions in the databases mentioned below. We also incorporate other feature patterns into the covariance matrix like LBP and Sobel mask and compare them to conventional LDP and LBP methods, which use histograms as feature vectors. From experimentation, it was shown that using a value of three for most prominent directions for LDP proved to give the best results. Accordingly, all LDP methods use three most prominent directions. When an LBP operator is chosen that covers a large number of neighbours with various labels, it creates a sizeable feature vector and therefore the calculation of the distance between covariance matrices reduces performance [118]. Therefore, the LBP and ULBP methods used an (8,1) radius filter. The Sobel mask was generated using the method discussed in Section 3.3.

4.2.1 JAFFE Database

The Japanese Female Facial Expression (JAFFE) database has 213 images of 7 facial expressions consisting of six basic and one neutral that 10 Japanese female models posed. All experiments carried out on the JAFFE database use an average of 30 images per class tested against an average of 60 random images consisting of 7 classes. Figure 4.4 illustrates examples of the JAFFE database's normalised images. These images are cropped automatically to make two eyes align at the same position and are then resized to 160 x 160.

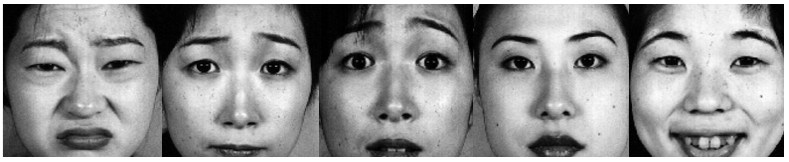


Figure 4.4: Cropped images from JAFFE database [35].

4.2.2 Extended Cohn-Kanade Database

The extended Cohn-Kanade Database (CK+) consists of 593 sequences from 123 subjects. These sequences begin from a neutral position and ends with expression peak. The database comes with 327 validated emotion labels consisting of six basic (happiness, anger, fear, disgust, surprise and sadness) plus contempt expressions. In our analysis, contempt is left out. Twenty-five images are chosen per class and test against an average of 75 random images consisting of 6 classes.

The images are adjusted and cropped to ensure two eyes align and are then resized to 256 x 256. An example of the images used can be seen in Figure 4.5.

Figure 4.5: Cropped images from CK+ database [36].

4.2.3 ISED Database

The ISED database is relatively new, thus limited research has been conducted based on this dataset. The database consists of 428 segmented video clips of spontaneous facial expressions of 50 participants. The database consists of labelled peak expressions of 4 classes: happiness, sadness, disgust and surprise. The database features mixed images of people with glasses, non-cohesive poses, and other varying uncontrolled environmental factors. The images are cropped by using a facial detector and then resized to 256 x 256. Examples of the cropped images are shown in Figure 4.6. An average of 48 images per class was tested against an average of 93 random images consisting of 4 classes.

Figure 4.6: Cropped images from ISED database [37].

Tables 4.1, 4.2 and 4.3 illustrate the results when using the global face experiment. This experiment established the effectiveness of the LDCM method compared to the original based LBP and LDP methods. The LDCM method gives good performance accuracy of 90% and 71% using JAFFE and CK+ datasets, respectively. The LDCM and the other covariance descriptor variation methods were outperformed by LBP and LDP using CK+ database. With JAFFE database LDCM was also outperformed, but marginally. The LDCM performed the best with an impressive 97% using ISED database; the covariance feature-based methods performed better than the

LBP and LDP methods. This can be due to the fact that the ISED database contains more random images that have partial occlusions and more pronounced pose variations of the face. The covariance descriptor proves to be more robust for these conditions.

The Sobel-based feature descriptors had the lowest performance, yet it is still reasonable. Its smoothening effect could cause the minute details of facial features to get lost, giving it a lack of discriminability. The ULBP method performed worse than standard LBP; it could not capture all the face's texture with its reduced histogram feature vector. It is also noteworthy that the LBP and LDP feature vectors consist of [1 x 16 348] feature points versus the much smaller [4 x 4] feature descriptor of the covariance matrix. The covariance descriptor is able to produce similar or more effective results at a far lower computational cost in terms of feature size. The confusion matrices in Table 4.4 indicate that the *Happy* expression is the most easily recognised in JAFFE and ISED databases using the proposed descriptor. In the CK+ dataset *Happy* gets misinterpreted as 'Fear' or 'Surprise'. From Table 4.4 we also note a strong correlation between *Anger* and *Surprise* in CK+ dataset.

Table 4.1: JAFFE database Global-Face accuracy

Features	**Acc.%**	Neutral	Happy	Sad	Surprise	Anger	Disgust	Fear
LDCM	**90**	90	100	90	90	87	79	91
LBCM	**89**	90	100	90	90	87	76	91
LDP+Sobel+COV	**88**	90	100	87	90	83	72	91
LBP+Sobel+COV	**86**	87	100	87	83	83	72	91
Sobel+COV	**86**	90	100	84	83	83	72	91
LDP	**92**	100	97	81	87	93	100	84
LBP	**93**	97	100	87	87	93	93	91
ULBP	**88**	94	99	88	89	94	93	92

Table 4.2: CK+ database Global-Face accuracy

Features	**Acc.%**	Anger	Disgust	Fear	Happy	Sad	Surprise
LDCM	**71**	76	76	84	56	72	64
LBCM	**73**	76	76	88	60	76	60
LDP+Sobel+COV	**68**	60	76	88	64	68	52
LBP+Sobel+COV	**71**	76	76	88	52	76	56
Sobel+COV	**70**	76	72	84	68	76	44
LDP	**85**	100	68	88	96	76	80
LBP	**87**	100	80	88	96	84	72
ULBP	**83**	90	78	84	92	84	70

Table 4.3: ISED database Global-Face accuracy

Features	**Acc.%**	Happy	Surprise	Sad	Disgust
LDCM	**97**	100	100	98	92
LBCM	**96**	100	96	96	92
LDP+Sobel+COV	**96**	100	96	96	94
LBP+Sobel+COV	**96**	100	94	96	94
Sobel+COV	**96**	100	98	94	94
LDP	**94**	92	94	96	94
LBP	**91**	96	83	96	88
ULBP	**90**	94	86	96	92

Table 4.4: Confusion matrices for holistic face on JAFFE, CK+, and ISED datasets

JAFFE

	Hap.	Fear	Dis.	Ang.	Sad	Neut.	Sur.
Hap.	**100**	0	0	0	0	0	0
Fear	9	**91**	0	0	0	0	0
Dis.	3	7	**79**	0	7	0	3
Ang.	10	3	0	**87**	0	0	0
Sad	10	0	0	0	**90**	0	0
Neut.	6	0	0	0	3	**90**	0
Sur.	7	3	0	0	0	0	**90**

CK+

	Ang.	Dis.	Fear	Hap.	Sad	Sur.
Ang.	**76**	8	8	8	0	0
Dis.	16	**76**	0	8	0	0
Fear	4	4	**84**	0	4	4
Hap.	8	4	16	**56**	4	12
Sad	8	16	4	0	**72**	0
Sur.	12	8	8	4	4	**64**

ISED

	Hap.	Sur.	Sad	Dis.
Hap.	**100**	0	0	0
Sur.	0	**100**	0	0
Sad	2	0	**98**	0
Dis.	2	0	6	**92**

4.3 Experiment 2: Segmented Face

We begin this experiment by following the principles of Guo et al. [118]. They used the LBCM operator with KNN to evaluate the impact of dividing the image into k equally sized rectangular regions. Figure 4.7 demonstrates the segmentation of the image, where the image is divided horizontally and vertically into equally sized partitions. The JAFFE dataset is used with LBP consisting of an (8, 1) neighbourhood and radius. We also evaluate the proposed MinSum classifier method. Tables 4.5 and 4.6 illustrate the 7-class and 6-class mean recognition accuracy for regions ranging from 1 to 8. The 7-class includes the *neutral* expression while the 6-class excludes it.

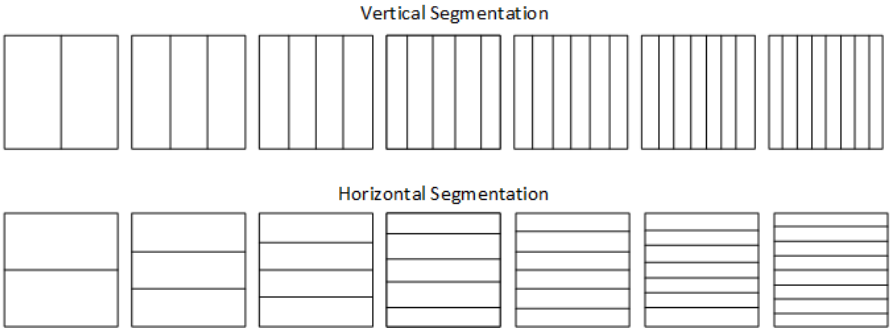


Figure 4.7: The face is divided using both vertical and horizontal segmentation ranging from 2 to 8 regions.

Table 4.5: Horizontal segmentation using JAFFE database with MinDist and MinSum classifiers using LBCM

Regions	7-Class Mean Recognition % Accuracy		6 -Class Mean Recognition % Accuracy	
	MinDist	MinSum	MinDist	MinSum
1	83	83	84	84
2	75	75	82	79
3	79	71	79	71
4	80	69	84	72
5	79	75	84	78
6	79	69	79	73
7	79	71	79	73
8	74	65	76	68

Table 4.6: Vertical segmentation using JAFFE database with MinDist and MinSum classifiers using LBCM

Regions	7-Class Mean Recognition % Accuracy		6 -Class Mean Recognition % Accuracy	
	MinDist	MinSum	MinDist	MinSum
1	83	83	84	84
2	85	77	85	77
3	85	77	85	77
4	83	75	86	82
5	81	72	81	74
6	79	74	81	78
7	82	68	87	76
8	78	75	82	74

The 7-class mean recognitions from Tables 4.5 and 4.6 are graphed in Figure 4.8.

The results show that when the region number becomes greater than 3 for the MinDist and 4 for the MinSum methods, the accuracy begins to drop. This could be due to the statistic

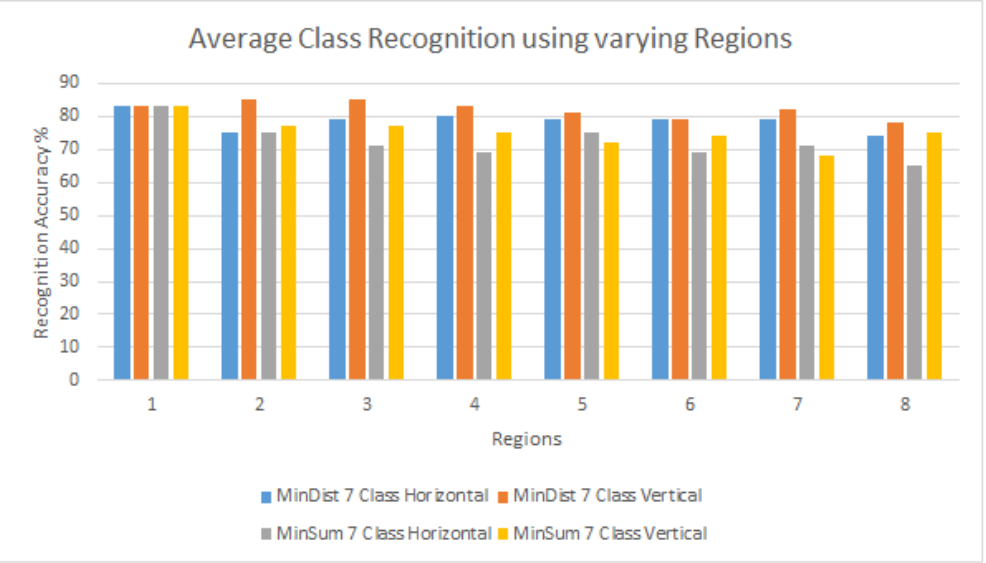


Figure 4.8: The average recognition accuracies for different region numbers on the JAFFE database.

characteristics of covariance matrices. With the increase of region numbers, the number of pixels in each region decreases, resulting in statistic bias in computing the covariance matrices. The results also show that the holistic face (1 region) performed better in the horizontal segmentation but the vertical segmented approach had superior results for 2 and 3 region segmentations. The MinSum was outperformed by the MinDist, suggesting that individual components of the face are better recognised.

The next step for this experiment consisted of determining the individual region accuracies for the horizontal and vertical segmentation of the face. This will allow some insight into how to better choose regions when classifying the facial expression. The results of this experiment can be seen in Tables 4.7 and 4.8. Interestingly, the results conform to Pantic et al. [26]; the vertical segmentation across the centre of the face shows identical accuracy for both regions using the 6-class mean. This entails that the symmetry of the face allows computation to be reduced by only using half of the data required. We also see that the centre regions of the face perform weaker than the outer regions, and for the horizontal segmentation, the lower half of the face performs better than the top half. A further analysis is done in Section 4.4.1, where the eye region is tested against the mouth.

Table 4.7: LBCM 7-Class and 6-Class individual region mean accuracy % for JAFFE database using **Vertical Segmented Regions**

7-Class Mean Acc. %							
Region Number							
1	2	3	4	5	6	7	8
81	76						
79	80	74					
80	78	78	73				
75	74	81	73	72			
77	77	78	73	76	71		
77	76	74	79	74	73	76	
74	73	75	78	78	70	70	71

6-Class Mean Acc. %							
Region Number							
1	2	3	4	5	6	7	8
79	79						
80	84	78					
83	79	84	75				
78	75	83	77	74			
82	78	82	78	79	72		
81	79	79	79	77	77	77	
77	75	74	84	81	71	71	74

Table 4.8: LBCM 7-Class and 6-Class individual region mean accuracy % for JAFFE database using **Horizontal Segmented Regions**

7-Class Mean Acc. %							
Region Number							
1	2	3	4	5	6	7	8
76	80						
74	75	77					
76	73	74	80				
74	78	74	70	80			
69	80	65	70	70	78		
71	71	76	70	73	72	78	
68	73	72	66	72	74	74	82

6-Class Mean Acc. %							
Region Number							
1	2	3	4	5	6	7	8
79	85						
76	76	80					
79	75	75	82				
77	82	76	75	83			
70	83	70	70	73	80		
72	76	79	72	79	73	81	
71	75	75	70	72	78	76	82

The last phase of this experiment was testing the component-based approach using LDCM. The global face image is segmented into equal-sized regions of [1 x 2], [2 x 1], [2 x 2], [3 x 3]. Figure 4.9 demonstrates a representation of how the face is divided. To classify between segments each region in the test image is compared to its like region in the training images. The region that has the minimum distance is chosen for classification (MinDist).

The results from Tables 4.9, 4.10 and 4.11 show that the holistic approach performs better than the component-based approach using LDCM in CK+ and ISED databases. This could be due to the fact that when the face is divided into smaller random segments it loses important discriminable information. However, the JAFFE database performed the best using the proposed method compared to the holistic approach, receiving a recognition accuracy of 96%. It is also evident that certain regions exhibit greater performance than other regions. In CK+ the best performing segment was the [1 x 2] split whereas in the JAFFE dataset it was the [2 x 2] split and the ISED

dataset the [2 x 1] split. The information from the random segments can be improved upon by targeting specific regions of the face.

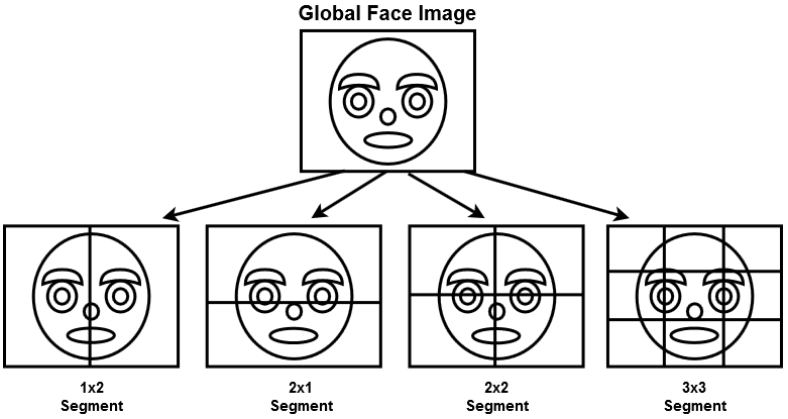


Figure 4.9: Segmentation of face into different regions.

Table 4.9: JAFFE segmented regions accuracy %

Segments	**Acc.%**	Neutral	Happy	Sad	Surprise	Anger	Disgust	Fear
1x2	**90**	90	100	84	87	93	76	97
2x1	**89**	90	94	84	87	83	93	94
2x2	**96**	97	100	100	90	90	97	97
3x3	**93**	100	100	94	87	93	86	91

Table 4.10: CK+ segmented regions accuracy %

Segments	**Acc.%**	Anger	Disgust	Fear	Happy	Sad	Surprise
1x2	**57**	56	72	80	40	56	40
2x1	**53**	40	64	64	52	48	52
2x2	**54**	36	60	80	44	64	40
3x3	**54**	36	64	76	40	60	48

Table 4.11: ISED segmented regions accuracy %

Segments	**Acc.%**	Happy	Surprise	Sadness	Disgust
1x2	**88**	90	85	94	83
2x1	**89**	90	92	94	79
2x2	**88**	88	94	94	77
3x3	**88**	90	85	90	85

4.4 Experiment 3: Special Landmark Regions

From the previous experiment of segmenting the face into random blocks, there is a hint that certain regions possess more discriminable properties for classification than others. In this experiment, we use eyes, nose and mouth detector to first extract the regions of interest from the face. The pair of eyes including the eyebrows is used, the extracted eye region is expanded to include the eyebrows. The special regions from the databases are segmented as follows:
CK+ dimensions: Eye - [80 x 160], Nose - [56 x 60], Mouth - [50 x 90]
JAFFE dimensions: Eye - [60 x 130], Nose - [40 x 50], Mouth-[40 x 60]
ISED dimensions: Eye - [70 x 200], Nose - [80 x 80], Mouth - [70 x 120]
The LDCM descriptor is then applied to each region and MinDist and MinSum classifiers are used for classification. Figure 4.10 shows the regions of interest.

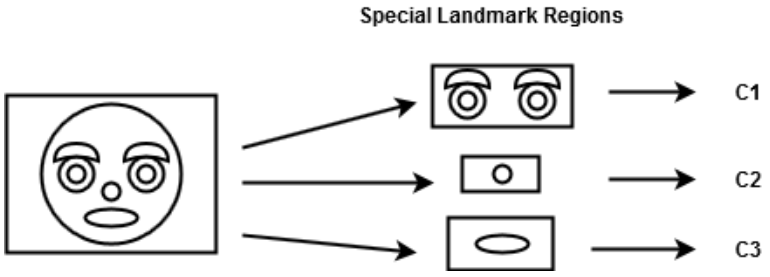


Figure 4.10: Landmark region extraction into region covariance descriptors.

Tables 4.12, 4.13 and 4.14 validate that the proposed method achieved the best results on CK+ database. Tracking the performance on CK+ dataset, from the global face we get an accuracy of 71% versus the split segments achieving 57% and finally attaining 82% using special landmark regions. For all datasets, a high facial expression recognition accuracy is achieved using special landmark regions and LDCM. The datasets JAFFE and ISED scored an average of 94% across both classification methods. The minimum sum classification method achieved a mean of 89% across all datasets, it outperformed the minimum distance classification method for CK+ and ISED databases. This could propose that for spontaneous expressions a holistic-approach is superior to a component-based approach.

Table 4.12: JAFFE special landmark regions accuracy %

Method	**Acc.%**	Neutral	Happy	Sad	Surprise	Angry	Disgust	Fear
1 - MinDist	**95**	97	100	90	97	97	90	94
2- MinSum	**91**	93	97	97	97	83	69	100

Table 4.13: CK+ special landmark regions accuracy %

Method	**Acc.%**	Anger	Disgust	Fear	Happy	Sad	Surprise
1 - MinDist	**75**	84	76	76	68	80	64
2- MinSum	**82**	92	88	80	80	76	76

Table 4.14: ISED special landmark regions accuracy %

Method	**Acc.%**	Happy	Surprise	Sadness	Disgust
1 - MinDist	**94**	96	96	92	92
2- MinSum	**95**	96	98	92	94

4.4.1 Eye versus Mouth Region

The aim of this test is to determine how closely the proposed classifier operates compared to the human observer approach concerning FER, mentioned in Section 2.6. For this test, the face is divided into the top half to encapsulate the eye area, and the lower half to capture the mouth area, and then each region is classified exclusively using LDCM on JAFFE, CK+, and ISED. The recognition results for each region are presented in Table 4.15 and the confusion matrices in Table 4.16.

Table 4.15 shows that in the CK+ dataset, *anger, surprise,* and *fear* are enhanced by the top of the face, while the bottom of the face is more suited to classify *sadness*. In the JAFFE database, *surprise, anger, disgust* and *fear* had greater performance using the lower half of the face. For both CK+ and JAFFE *happy* expression results showed that using the top half of the face helps achieve recognition. All four expressions: *happiness, surprise, sadness*, and *disgust* of the ISED database achieved equal or superior accuracy using the lower half of the face. For CK+ dataset superior performance is achieved using eye (89%) and mouth (82%) regions instead of the global (71%) face region. The confusion matrices in Table 4.16 demonstrate that *sad* and *surprise*, as well as *happiness* and *disgust* are misclassified on the top half of the face. On the lower half, *disgust, sadness* and *surprise*, as well as *fear* and *disgust*, suffer confusion.

Table 4.15: LDCM accuracy % of **Eye and Mouth** components for FER on CK+, JAFFE, and ISED datasets

Datasets	**Acc.%**	Anger	Disgust	Fear	Happy	Sad	Surprise	
Cohn Eye	**89**	100	88	96	80	72	100	
Cohn Mouth	**82**	92	88	80	72	92	68	
Datasets	**Acc.%**	Neutral	Happy	Sad	Surprise	Angry	Disgust	Fear
JAFFE Eye	**83**	80	97	87	83	73	72	88
JAFFE Mouth	**88**	90	94	87	90	83	83	91
Datasets	**Acc.%**	Happy	Surprise	Sadness	Disgust			
ISED Eye	**93**	98	96	92	88			
ISED Mouth	**97**	100	96	94	98			

Table 4.16: Confusion matrices for **Eye and Mouth** components using LDCM

Cohn: Eye Confusion Matrix

	Ang.	Dis.	Fear	Hap.	Sad	Sur.
Ang.	**100**	0	0	0	0	0
Dis.	8	**88**	0	0	0	4
Fear	4	0	**96**	0	0	0
Hap.	8	4	4	**80**	0	4
Sad	8	0	8	0	**72**	12
Sur.	0	0	0	0	0	**100**

Cohn: Mouth Confusion Matrix

	Ang.	Dis.	Fear	Hap.	Sad	Sur.
Ang.	**92**	0	0	8	0	0
Dis.	4	**88**	0	8	0	0
Fear	4	12	**80**	4	0	0
Hap.	4	4	16	**72**	4	0
Sad	0	4	4	0	**92**	0
Sur.	0	12	0	8	12	**68**

JAFFE: Eye Confusion Matrix

	Hap.	Fear	Dis.	Ang.	Sad	Neut.	Sur.
Hap.	**97**	0	3	0	0	0	0
Fear	9	**88**	3	0	0	0	0
Dis.	24	0	**72**	0	3	0	0
Ang.	13	3	10	**73**	0	0	0
Sad	10	3	7	0	**80**	0	0
Neut.	6	3	3	0	0	**87**	0
Sur.	13	0	3	0	0	0	**83**

JAFFE: Mouth Confusion Matrix

	Hap	Fear	Dis.	Ang	Sad	Neut.	Sur.
Hap.	**94**	0	0	0	0	3	3
Fear	6	**91**	3	0	0	0	0
Dis.	3	14	**83**	0	0	0	0
Ang.	10	0	3	**83**	0	0	3
Sad	7	0	0	0	**90**	3	0
Neut.	10	0	0	0	3	**87**	0
Sur.	7	0	3	0	0	0	**90**

ISED: Eye Confusion Matrix

	Hap.	Sur.	Sad	Dis.
Hap.	**98**	0	0	2
Sur.	4	**96**	0	0
Sad.	6	0	**92**	2
Dis.	0	0	13	**88**

ISED: Mouth Confusion Matrix

	Hap.	Sur.	Sad	Dis.
Hap.	**100**	0	0	0
Sur.	4	**96**	0	0
Sad	4	0	**94**	2
Dis.	2	0	0	**98**

4.5 Experiment 4: LEM Distance Classifier

The LEM distance metric has gained appreciation in recent years for its use with SPD matrices. It is computationally efficient and allows the covariance matrix to be represented in a form where traditional classification methods can be computed. The LEM from Section 3.4.4 is used with the MinDist classifier for FER on the JAFFE, Extended Cohn-Kanade and ISED databases. The holistic approach is used with the different covariance features. It is compared to the distance metric from Equation 3.22.

Table 4.17 gives the mean class accuracy for each dataset. The results show that LEM is equivalent to, if not better than, AIM. It achieved the same mean class accuracy for CK+ and ISED using LDCM. It performed better across all datasets using the Sobel pixel feature. A future development that studies LEM with SVM for FER using covariance descriptors will be highly beneficial to further the advancement of FER.

Table 4.17: LEM and Riemann distance metric mean class % accuracy on CK+, JAFFE and ISED databases

Features	6-Class Mean Recognition CK+		7-Class Mean Recognition JAFFE		4-Class Mean Recognition ISED	
Distance Metric	LEM	Riemann	LEM	Riemann	LEM	Riemann
LDCM	71%	71%	89%	90%	97%	97%
LBCM	73%	73%	89%	89%	95%	96%
LDP+Sobel+COV	65%	68%	90%	88%	97%	96%
LBP+Sobel+COV	70%	71%	87%	86%	95%	96%
Sobel+COV	71%	70%	88%	86%	97%	96%

4.6 Experiment 5: Cross-Database Environment

As stated previously, the challenge of FER is increased in uncontrolled environments. To simulate these conditions, a cross-database approach is used. By using different databases for your training and testing tests, you invite more randomness and provide more difficulty to the FER system. You can also formulate new patterns between different ethnic groups if your datasets support it.

The results using posed and spontaneous expressions, alternating between testing and training images from JAFFE, ISED, and Extended Cohn-Kanade datasets are displayed in Tables 4.18, and 4.19. The tests used the global face region and classified against the six prototypic basic expressions for JAFFE and CK+ cross-database evaluation. When evaluating the ISED database with JAFFE and CK+ databases, only four expressions were classified, those are, happy, sad, surprise, and disgust. The proposed descriptor (LDCM) and K-NN using distance metric 3.22 were applied with the leave-one-out-cross-validation technique. The training and testing samples of each respective dataset were the same used in Sections 4.2.1, 4.2.2, and 4.2.3. From Table 4.18, it can be observed that LDCM achieved an accuracy of 92% using JAFFE expression test

Table 4.18: Posed cross-database FER using JAFFE and CK+ datasets

Test Images	Training Images	**Acc.%**	Ang.	Dis.	Fear	Hap.	Sad	Sur.
JAFFE	Extended Cohn	**92**	73	93	100	84	100	100
Extended Cohn	JAFFE	**86**	80	84	92	92	84	84

images and CK+ training images. The results also improved when using CK+ test images and JAFFE training images, where a six class mean expression recognition rate of 86% is achieved. Evidently, from Tables 4.1, 4.2 and 4.18, it is proven that the cross-database results are better than the standard JAFFE and CK+ dataset results using the proposed descriptor. When using the standard JAFFE facial expression test class images against JAFFE facial expression training class images and CK+ facial expression training class images the results were 90% (Table 4.1) and 92% (Table 4.18) respectively. Subsequently, when using the standard CK+ expression test images against CK+ expression training images and JAFFE expression training images, we achieved 71% (Table 4.2) and 86% (Table 4.18) respectively. The increase in performance can account for the difference of facial features between subjects used or, in this case, a higher variance between different persons or expressions.

Contrarily, when using ISED, a spontaneous dataset, as the training sample in cross-database evaluation with JAFFE and CK+ test samples, there is a significant decrease in accuracy. The cultural differences between the training and testing subjects is one motivation for the reported small accuracy as in Table 4.19. Although, when using the ISED dataset as the test sample and JAFFE and CK+ as training samples, it achieves a recognition of 95% and 82% respectively, shown in Table 4.19. The difference between the high and low recognition accuracy underlines that it is more suitable to train with a posed expression set as opposed to a spontaneous dataset, in these testing scenarios. The ISED data consists of facial images that contain non-cohesive pose, partial occlusions like glasses, and other varying uncontrolled factors, while the CK+ and JAFFE predominately consist of cohesive, posed and non-occluded images. Another factor may include the difference in the number of subjects trained and tested. The ISED consists of 50 subjects whereas the JAFFE only consists of 10.

Table 4.19: Spontaneous cross-database FER using ISED, JAFFE, and CK+ datasets

Test Images	Training Images	**Acc.%**	Hap.	Sad	Sur.	Dis.
JAFFE	ISED	**34**	100	0	7	28
Extended Cohn	ISED	**36**	100	0	28	16
ISED	JAFFE	**95**	100	92	88	100
ISED	Extended Cohn	**82**	58	100	94	77

4.7 Conclusion

The use of covariance features for facial expression recognition is not commonly practised. Its applications originate from object detection and tracking. Newer studies are proving that RCM performs adequately if you select appropriate features for specific tasks. This study has proposed a new local feature of facial expression based on LDP codes and region covariance matrices. The LDP code holds local information by encoding the texture of the face and the covariance descriptor contains the global information. From the results obtained, it is established that the proposed descriptor achieves a high level of performance for FER at a reduced feature size. We have also investigated the effect of holistic vs component-based approaches to FER using LDCM. It was found that by focusing on special regions of the face such as eyes, nose and mouth, stable results across different datasets and environments were achieved. In the future, more tests with images that incorporate more noise and partial occlusions can be beneficial to determine the contribution of LDP and LBP when they are used in the covariance structure. Covariance descriptors are also limited with regards to using standard machine learning methods. Transforming covariance structure to accommodate standard machine learning methods is an interesting research direction.

Chapter 5

Conclusion

The present chapter summarises the work presented in this dissertation, which was centered around FER using covariance descriptors and local texture patterns. A brief discussion is presented in Section 5.1. In Section 5.2 we provide a summary of contributions made in this dissertation and then proceed to discuss possible directions and challenges for future research in Section 5.3.

5.1 Discussion

Recognising facial expression is an effective procedure to gauge the emotional state of human beings. In this dissertation, we have studied covariance descriptors and local texture patterns for facial expression recognition, and attempted to propose a state-of-the-art FER system using a novel image descriptor.

In establishing the foundation of using a covariance approach, we acknowledge some of the advantages of using covariance descriptors as follows:

- **Compactness.** A dimensionality reduction without any significant loss in recall has been shown in Section 3.4.3, which implies lower computational and storage requirements.
- **Flexibility.** The addition of new features in the covariance descriptor has been shown in Section 3.4.3 to be straight-forward. It does not cause any significant change in memory or computation time. Hence, it provides an advantage compared to histogram-based approaches, in which the addition of a feature amounts to the addition of a dimension in the histogram cube.

- **Parameter-freeness.** There is no requirement to tune parameters such as bin size or bin number.
- **Distinctiveness.** The covariance descriptor possesses an inherent ability to remove common features available in the data and consider only the discriminative information.

An important question regarding covariance descriptors that has not yet been addressed is: *Which data type is effective and when can it be used?* Though this question is subjective and depends exclusively on the current application, there are a few data properties that can assist in decision making. Covariance valued data can be used in situations where the discriminating or recognising properties of the data are founded on feature correlations instead of on the frequency with which certain patterns occur, which is the case with feature histograms.

This dissertation aim of determining the effectiveness of covariance descriptors for facial expression recognition was positively achieved. In Chapter 2, the literature review examined the different features, such as geometric and appearance-based methods used in FER. It further discussed the benefits of using local texture patterns and covariance matrices because of their robustness and fast computation; hence, there was viability to use the covariance descriptor for FER. From Chapter 3, the advantages and disadvantages of the different operators, such as LBP, LDP and Sobel, were discussed. The properties and methods of the covariance matrix were described in detail and the LDCM image descriptor was proposed. Analysing the results in Chapter 4 proved that the covariance descriptor successfully classified facial expressions. The JAFFE posed expression dataset, achieved 90% accuracy and the ISED spontaneous expression dataset achieved 97% accuracy using the holistic approach, validating the LDCM operator.

The face was segmented into equal-sized horizontal and vertical regions ranging from 2 to 8 regions using LBCM. It was observed that a loss of accuracy is obtained when the region sizes begin to get too small, due to the statistic bias in computing the covariance matrices. The individual region accuracies were also examined showing that the symmetry of the face can provide for faster computation without losing significant precision by choosing to use half of the vertical face. The special regions of the face were then examined to determine which regions perform best for classification. The findings illustrated that the bottom half of the face is superior for recognising *happiness, surprise, sadness,* and *disgust* for spontaneous expressions. The confusion matrices show that *sadness* and *surprise*, as well as *happiness* and *disgust* are misclassified

on the top half of the face. On the lower half, *disgust, sadness* and *surprise*, as well as *fear* and *disgust* suffer confusion.

The special landmarks did better than the holistic approach on JAFFE and CK+ datasets, as accuracies of 95% and 82% are achieved respectively. The ISED database was marginally lower at 95%. The results show that the component-based method is superior for posed expressions while the holistic approach is preferred to spontaneous expressions. The cross-database evaluation made it more challenging by replicating real-world environments. It further approved the effectiveness of the proposed descriptor. Lastly, the LEM similarity metric was evaluated for FER using the covariance feature descriptors. It showed promising results and provides alternate means for classification.

5.2 Contributions

This dissertation helped contribute to affective computing research by:

- Establishing the state-of-the-art in facial expression recognition. In-depth analysis of the geometric-based and feature-based methods was covered in Section 2.
- Introducing covariance descriptors from image processing into facial expression recognition. Covariance descriptors have proven to be very successful in other application areas within image processing. This dissertation tries to continue the success in the domain of FER.
- Developing different features using the covariance structure for FER. The study proposed an effective and efficient novel image descriptor, Local Directional Covariance Matrix, as well as other variants seen in Section 4.2. The proposed descriptor performed better than traditional methods for spontaneous expressions.
- Comparing the benefits of a component-based versus holistic-based approach for classification. The component-based approach might lead to the loss of some of the benefits that the covariance matrix introduced, by increasing computation and memory.
- Testing for special regions of interest that could lead to increase FER accuracy, if applied in a component-based or rule-based approach. The eyes, nose and mouth were specifically targeted.
- Using both posed and spontaneous expressions to provide a complete study of all types of expressions.

- Demonstrating the possibility of real-world scenarios by using cross-database evaluation on posed and spontaneous datasets.
- Evaluating the performance of an alternative similarity metric (LEM) for SPD matrices.

5.3 Limitations and Future Directions

5.3.1 Limitations

There are existing open-ended questions to facial expression analysis that highlight limitations when trying to develop a robust FER system.

How do humans correctly recognise facial expressions?

In Section 2.6, we discuss the human visual perception of how facial expressions are recognised, but how humans recognise facial expressions is still not clear. Further understanding of the parameter types and how they are processed is needed. By comparing how humans and machines recognise facial expressions, new ways of improving recognition can be researched.

Is there any better way to code facial expressions for computer systems?

The majority of research concentrated on using either emotion-specific expressions or FACS coded action units. The emotion-specific expressions label expressions at a rather basic level and are not always adequate for all applications [160]. The intended design of FACS is to observe subtle changes in facial features, but it is a system based on human observers and has a limited ability to differentiate between variation in intensity. Challenges still exist in the design of a computer-based system for coding facial expressions, which has more quantitative definitions [160].

How do we obtain reliable ground truth?

Emotion-specific expressions are inadequately defined: one label may be applicable to various expressions. Consequently, one expression may have various labels that refer to it, which confuses system comparisons. A further issue is that label reliability is not known. For example, investigators fail to disclose inter-observer reliability and facial expression validity of the expressions they analysed. This creates uncertainty in knowing if subjects truly had the target expression or if only judges determined that subjects had that expression.

How do we recognise facial expressions in real life?

There is much more difficulty in the analysis of real-life expressions compared to posed actions. The elements that make facial expression analysis complex in real life are the motion of the head, input images that are of low resolutions, the lack of a neutral expression for comparison, and expressions of low intensity [160].

Expressions consist of a blend of multiple micro expressions. Therefore, one expression class is insufficient to ascertain an expression. Making the differentiation between these micro-facial expressions and being able to increase the number of standard emotion classes requires a higher research scope.

The databases available to conduct research, lack a variety of different conditions, such as illumination, occlusion, and pose. It is also difficult to obtain authentic spontaneous expression databases because the environment is not in a natural setting when the subject is being observed.

5.3.2 Future Directions

Multi-modal. Human emotion comprises of facial expressions, gestures, and vocal data. For a complete analysis of human emotion, data from all these domains should be incorporated. Hence, multi-modal methods are an interesting research direction.

Real-Time Dynamic. It is possible for the suggested descriptor to be extended to a real-time dynamic facial expression implementation using spatiotemporal methods on larger databases that include a variety of races and ages.

Metric Learning on Covariance Matrices. The feature vectors collected from an image region are a heuristic combination used in covariance descriptors, under the assumption that they will provide an optimal solution for its intended application. Mechanisms that learn each feature's effective weights or contributions to the overall performance of the application, provide intuition on why a certain feature combination performs well as well as what should not be used.

Kernal-Based Method. As discussed in Section 3.4.2, by embedding the manifold to tangent spaces further analysis is simplified significantly, but some of the manifold structure has to be disregarded to achieve this.

Convolutional Neural Networks. An interesting area worth investigation is the combination between kernel representations and Convolutional Neural Networks (CNNs). CNNs incorporates feature extraction, feature representation, and label prediction into a unified framework. They

are considered as the state-of-the-art feature learning approach. It can be expected that applying kernel representation to CNN features could generate more promising performance.

References

[1] Z. Wang, L. Xie, and T. Lu, "Research progress of artificial psychology and artificial emotion in China," *CAAI Transactions on Intelligence Technology*, vol. 1, no. 4, pp. 355–365, 2016.

[2] T. Baltrusaitis, "Automatic facial expression analysis," *Pattern recognition*, p. 220, 2014.

[3] M. Pantic, A. Pentland, A. Nijholt, and T. S. Huang, "Human computing and machine understanding of human behavior: A survey," *Lecture Notes in Computer Science*, vol. 4451, pp. 47–71, 2007.

[4] P. Robinson and R. el Kaliouby, "Computation of emotions in man and machines." *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, vol. 364, no. 1535, pp. 3441–3447, 2009.

[5] R. W. Picard, "Affective Computing," *MIT press*, no. 321, pp. 1–16, 1995.

[6] J. F. Cohn, "Foundations of human computing: Facial expression and emotion," *Lecture Notes in Computer Science*, vol. 4451, pp. 1–16, 2007.

[7] S. D'Mello and R. A. Calvo, "Beyond the basic emotions," *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, p. 2287, 2013.

[8] L. Indvik. (2011) Apple Now World's Most Valuable Brand. [Online]. Available: http://mashable.com/2011/05/09/apple-google-brandz-study

[9] Seun Hollister. (2017) iPhone X: How Face ID works - CNET. [Online]. Available: https://www.cnet.com/news/apple-face-id-truedepth-how-it-works/

[10] R. Picard and J. Klein, "Computers that recognize and respond to user emotion: theoretical and practical implications," *Interacting with Computers*, vol. 14, no. 1945, pp. 141–169, 2002.

[11] N. Ambady and R. Rosenthal, "Thin Slices of Expressive Behaviour as Predictors of Interpersonal Consequences: A Meta-Analysis," *Psychological Bulletin*, vol. 111, no. 2, pp. 256–274, 1992.

[12] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 1, pp. 39–58, 2009.

[13] P. Ekman and E. L. Rosenberg, "What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)," *Oxford University Press*, pp. 1–672, 2012.

[14] K. L. Schmidt and J. F. Cohn, "Human facial expressions as adaptations: Evolutionary questions in facial expression research," *American Journal of Physical Anthropology*, vol. 116, no. 33, pp. 3–24, 2001.

[15] C. Darwin, P. Ekman, and P. Prodger, "The Expression of the Emotions in Man and Animals," *Oxford University Press*, 1998.

[16] P. Ekman, "Universal Facial Expressions of Emotions," *California mental health research digest*, vol. 8, no. 4, pp. 151–158, 1970.

[17] M. J. Fehrenbach and S. W. Herring, "Illustrated Anatomy of the Head and Neck," *Elsevier Health Sciences*, 2015.

[18] P. Husak, "Emotive Facial Expression Detection," *Master thesis, Center for Machine Perception, Czech Technical University in Prague*, 2017.

[19] A. Karunaharamoorthy. (Accessed on: 2017-10-11) Facial muscles - Anatomy, Function and Pathology. [Online]. Available: https://www.kenhub.com/en/library/anatomy/the-facial-muscles

[20] C.-H. Hjortsjo, "Man's face and mimic language," *Studentlitteratur*, 1969.

[21] J. A. Russell, "Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies." *Psychological Bulletin*, vol. 115, no. 1, pp. 102–141, 1994.

[22] Z. Zhang, "Microsoft kinect sensor and its effect," *IEEE Multimedia*, vol. 19, no. 2, pp. 4–10, 2012.

[23] A. Mitra, "Efficient Covariance-based Algorithms for Appearance Modeling in Computer Vision," *Master of Engineering thesis, Faculty of Engineering, Department of Computer Science and Automation, Indian Institute of Science*, 2010.

[24] O. Tuzel, F. Porikli, and P. Meer, "Region covariance: A fast descriptor for detection and classification," *Lecture Notes in Computer Science*, vol. 3952, pp. 589–600, 2006.

[25] H. Soyel and H. Demirel, "Facial Expression Recognition Using 3D Facial Feature Distances," *Image Analysis and Recognition*, vol. 4, pp. 831–838, 2008.

[26] M. Pantic, S. Member, and L. J. M. Rothkrantz, "Automatic Analysis of Facial Expressions : The State of the Art," *Analysis*, vol. 22, no. 12, pp. 1424–1445, 2000.

[27] V. Kshirsagar, M. Baviskar, and M. Gaikwad, "Face recognition using Eigenfaces," *3rd International Conference on Computer Research and Development*, vol. 2, pp. 302–306, 2011.

[28] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[29] M. J. Lyons, "Automatic classification of single facial images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 12, pp. 1357–1362, 1999.

[30] P. McOwan, "Robust facial expression recognition using local binary patterns," *IEEE International Conference on Image Processing*, p. 370, 2005.

[31] P. Rathod, L. Gagnani, and K. Patel, "Facial Expression Recognition : Issues and Challenges," *International Journal of Enhanced Research in Science Technology & Engineering*, vol. 3, no. 2, pp. 108–111, 2014.

[32] Y. Wu, H. Liu, and H. Zha, "Modeling facial expression space for recognition," *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1968–1973, 2005.

[33] T. Huang, Z. Xiong, and Z. Zhang, "Face recognition applications," *Handbook of Face Recognition*, pp. 617–638, 2011.

[34] A. T. Lopes, E. de Aguiar, A. F. De Souza, and T. Oliveira-Santos, "Facial expression recognition with Convolutional Neural Networks: Coping with few data and the training sample order," *Pattern Recognition*, vol. 61, pp. 610–628, 2017.

[35] M. J. Lyons, J. Budynek, and S. Akamatsu, "Automatic classification of single facial images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 12, pp. 1357–1362, 1999.

[36] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 94–101, 2010.

[37] S. L. Happy, P. Patnaik, A. Routray, and R. Guha, "The indian spontaneous expression database for emotion recognition," *IEEE Transactions on Affective Computing*, vol. 8, no. 1, pp. 131–142, 2017.

[38] P. Liu, S. Han, Z. Meng, and Y. Tong, "Facial Expression Recognition via a Boosted Deep Belief Network," *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1805–1812, 2014.

[39] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on Local Binary Patterns: A comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, 2009.

[40] K. Y. Chang and C. S. Chen, "Facial Expression Recognition via Discriminative Dictionary Learning," *IEEE International Conference on Internet of Things (iThings), and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom)*, pp. 464–466, 2014.

[41] I. Song, H. J. Kim, and P. B. Jeon, "Deep learning for real-time robust facial expression recognition on a smartphone," *IEEE International Conference on Consumer Electronics (ICCE)*, pp. 564–567, 2014.

[42] P. Burkert, F. Trier, M. Z. Afzal, A. Dengel, and M. Liwicki, "DeXpression: Deep Convolutional Neural Network for Expression Recognition," *Pattern Recognition Letters*, pp. 1–8, 2015.

[43] M. Liu, S. Li, S. Shan, and X. Chen, "Au-inspired deep networks for facial expression feature learning," *Neurocomput.*, vol. 159, pp. 126–136, 2015.

[44] G. Ali, M. A. Iqbal, and T.-S. Choi, "Boosted nne collections for multicultural facial expression recognition," *Pattern Recogn.*, vol. 55, no. C, pp. 14–27, 2016.

[45] Y.-H. Byeon and K.-C. Kwak, "Facial expression recognition using 3d convolutional neural network," *International Journal of Advanced Computer Science and Applications(IJACSA)*, vol. 5, no. 12, pp. 107–112, 2014.

[46] J. J.-J. Lien, T. Kanade, J. F. Cohn, and C.-C. Li, "Detection, tracking, and classification of action units in facial expression," *Robotics and Autonomous Systems*, vol. 31, no. 3, pp. 131–146, 2000.

[47] X. Fan and T. Tjahjadi, "A spatial-temporal framework based on histogram of gradients and optical flow for facial expression recognition in video sequences," *Pattern Recognition*, vol. 48, no. 11, pp. 3407–3416, 2015.

[48] W. Zhang, Y. Zhang, L. Ma, J. Guan, and S. Gong, "Multimodal learning for facial expression recognition," *Pattern Recognition*, vol. 48, no. 10, pp. 3191–3202, 2015.

[49] C. R. Chen, W. S. Wong, and C. T. Chiu, "A 0.64 mm2 real-time cascade face detection design based on reduced two-field extraction," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 19, no. 11, pp. 1937–1948, 2011.

[50] C. Garcia and M. Delakis, "A neural architecture for fast and robust face detection," *16th International Conference on Pattern Recognition, Proceedings*, vol. 2, no. 11, pp. 44–47, 2002.

[51] Z. Zhang, D. Yi, Z. Lei, and S. Z. Li, "Regularized transfer boosting for face detection across spectrum," *IEEE Signal Processing Letters*, vol. 19, no. 3, pp. 131–134, 2012.

[52] M. S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Recognizing facial expression: machine learning and application to spontaneous behavior,"

IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 568–573, 2005.

[53] P. Liu, M. Reale, and L. Yin, "3d head pose estimation based on scene flow and generic head model," *IEEE International Conference on Multimedia and Expo*, pp. 794–799, 2012.

[54] W. W. Kim, S. Park, J. Hwang, and S. Lee, "Automatic head pose estimation from a single camera using projective geometry," *8th International Conference on Information, Communications Signal Processing*, pp. 1–5, 2011.

[55] M. Demirkus, D. Precup, J. J. Clark, and T. Arbel, "Multi-layer temporal graphical model for head pose estimation in real-world videos," *IEEE International Conference on Image Processing (ICIP)*, pp. 3392–3396, 2014.

[56] Z. Z. Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu, "Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron," *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 454–459, 1998.

[57] P. Yang, Q. Liu, and D. N. Metaxas, "Boosting coded dynamic features for facial action units and facial expression recognition," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–6, 2007.

[58] A. Jain and S. Z. Li, "Facial expression analysis," *Handbook of face recognition*, pp. 487–519, 2005.

[59] S. Rifai, Y. Bengio, A. Courville, P. Vincent, and M. Mirza, "Disentangling Factors of Variation for Facial Expression Recognition," *12th European Conference on Computer Vision, Proceedings, Part VI*, pp. 808–822, 2012.

[60] S. K. Singh, D. S. Chauhan, M. Vatsa, and R. Singh, "A robust skin color based face detection algorithm," *Tamkang Journal of Science and Engineering*, vol. 6, no. 4, pp. 227–234, 2003.

[61] P. Viola and M. M. J. Jones, "Robust Real-Time Face Detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[62] P. Ekman and W. Friesen, "Facial action coding system: Investigator's guide part 1," *Facial Action Coding System: Investigator's Guide*, no. 1, 1978.

[63] N. Sebe, I. Cohen, A. Garg, and T. Huang, "Application: Facial Expression Recognition," *Machine Learning in Computer Vision*, pp. 187–209, 2005.

[64] P. Michel and R. El Kaliouby, "Real time facial expression recognition in video using support vector machines," *ICMI'03: Fifth International Conference on Multimodal Interfaces*, pp. 258–264, 2003.

[65] P. S. Aleksic and A. K. Katsaggelos, "Automatic facial expression recognition using facial animation parameters and multistream HMMs," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 1, pp. 3–11, 2006.

[66] I. Cohen, N. Sebe, A. Garg, L. S. Chen, and T. S. Huang, "Facial expression recognition from video sequences: Temporal and static modeling," *Computer Vision and Image Understanding*, vol. 91, no. 1-2, pp. 160–187, 2003.

[67] J. Wang and L. Yin, "Static topographic modeling for facial expression recognition and analysis," *Computer Vision and Image Understanding*, vol. 108, no. 1-2, pp. 19–34, 2007.

[68] N. Sebe, M. S. Lew, Y. Sun, I. Cohen, T. Gevers, and T. S. Huang, "Authentic facial expression analysis," *Image and Vision Computing*, vol. 25, no. 12, pp. 1856–1863, 2007.

[69] I. Kotsia and I. Pitas, "Facial expression recognition in image sequences using geometric deformation features and support vector machines," *IEEE Transactions on Image Processing*, vol. 16, no. 1, pp. 172–187, 2007.

[70] I. Kotsia, I. Buciu, and I. Pitas, "An analysis of facial expression recognition under partial facial image occlusion," *Image and Vision Computing*, vol. 26, no. 7, pp. 1052–1067, 2008.

[71] J. Cohn, Z. Ambadar, and P. Ekman, "Observer-based measurement of facial expression with the Facial Action Coding System," *The handbook of emotion elicitation and assessment, Oxford University Press Series in Affective Science*, 2006.

[72] (Accessed on: 2017-10-20) 11.6 Minimum Distance Classifier. [Online]. Available: http://wtlab.iis.u-tokyo.ac.jp/{~}wataru/lecture/rsgis/rsnote/cp11/cp11-6.htm

[73] S.-C. Wang, "Artificial Neural Network," *Interdisciplinary Computing in Java Programming*, pp. 81–100, 2003.

[74] B. Fasel and J. Luettin, "Automatic facial expression analysis: A survey," *Pattern Recognition*, vol. 36, no. 1, pp. 259–275, 2003.

[75] A. Jain and S. Z. Li, "Facial expression analysis," *Handbook of face recognition*, pp. 247–275, 2005.

[76] O. Rudovic, M. Pantic, and I. Patras, "Coupled Gaussian processes for pose-invariant facial expression recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1357–1369, 2013.

[77] A. Kanaujia and D. Metaxas, "Recognizing facial expressions by tracking feature shapes," *Proceedings - International Conference on Pattern Recognition*, vol. 2, pp. 33–38, 2006.

[78] G. Shin and J. Chun, "Spatio-temporal Facial Expression Recognition Using Optical Flow and HMM," *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*, pp. 27–38, 2008.

[79] Y. L. Tian, T. Kanade, and J. F. Cohn, "Evaluation of Gabor-wavelet-based facial action unit recognition in image sequences of increasing complexity," *Proceedings - 5th IEEE International Conference on Automatic Face Gesture Recognition, FGR*, pp. 229–234, 2002.

[80] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local Gabor Binary Pattern Histogram Sequence (LGBPHS): A novel non-statistical model for face representation and recognition," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 786–791, 2005.

[81] A. Durmuşoğlu and Y. Kahraman, "Facial expression recognition using geometric features," *International Conference on Systems, Signals and Image Processing (IWSSIP)*, pp. 1–5, 2016.

[82] J. Zhang, "Developing advanced methods for covariance representations in computer vision," *Doctor of Philosophy thesis, School of Computing and information Technology, University of Wollongong*, 2016.

[83] X. Huang, "Methods for facial expression recognition with applications in challenging situations," *Doctor of Philosophy thesis, Faculty of Information Technology and Electrical Engineering, Department of Computer Science and Engineering, University of Oulu*, 2014.

[84] X. Feng, M. Pietikainen, and A. Hadid, "Facial expression recognition with Local Binary Patterns and Linear Programming," *Pattern Recognition And Image Analysis*, vol. 15, no. 2, pp. 546–548, 2005.

[85] S. Moore and R. Bowden, "Local binary patterns for multi-view facial expression recognition," *Computer Vision and Image Understanding*, vol. 115, no. 4, pp. 541–558, 2011.

[86] G. Littlewort, M. S. Bartlett, I. Fasel, J. Susskind, and J. Movellan, "Dynamics of facial expression extracted automatically from video," *Image and Vision Computing*, vol. 24, no. 6, pp. 615–625, 2006.

[87] X. Feng, Y. Lai, X. Mao, J. Peng, X. Jiang, and A. Hadid, "Extracting Local Binary Patterns from Image Key Points: Application to Automatic Facial Expression Recognition," *Image Analysis: 18th Scandinavian Conference, Proceedings*, pp. 339–348, 2013.

[88] K. Y. Chang, C. S. Chen, and Y. P. Hung, "Intensity rank estimation of facial expressions based on a single image," *IEEE International Conference on Systems, Man, and Cybernetics, SMC, Proceedings*, pp. 3157–3162, 2013.

[89] A. Yuce, M. Sorci, and J.-P. Thiran, "Improved local binary pattern based action unit detection using morphological and bilateral filters," *10th IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, pp. 1–7, 2013.

[90] N. Bayramoglu, G. Zhao, and M. Pietikainen, "CS-3DLBP and geometry based person independent 3D facial action unit detection," *International Conference on Biometrics, ICB, Proceedings*, pp. 1–5, 2013.

[91] T. Jabid, M. H. Kabir, and O. Chae, "Facial expression recognition using Local Directional Pattern (LDP)," *17th IEEE International Conference on Image Processing*, pp. 1605–1608, 2010.

[92] B. Jun, T. Kim, and D. Kim, "A compact local binary pattern using maximization of mutual information for face analysis," *Pattern Recognition*, vol. 44, no. 3, pp. 532–543, 2011.

[93] A. R. Rivera, J. R. Castillo, and O. Chae, "Local directional number pattern for face analysis: face and expression recognition." *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 22, no. 5, pp. 1740–1752, 2013.

[94] A. Srikrishna, C. S. H. Priya, and P. A. Kiran, "Face recognition with varying facial expression using local directional number pattern," *IEEE Power, Communication and Information Technology Conference (PCITC)*, pp. 615–619, 2015.

[95] B. Vishnudharan and M. T. Student, "A Discriminative Model For Facial Expression Recognition Using Local Directional Number Pattern," *International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)*, no. 1, pp. 349–352, 2016.

[96] B. Ryu, A. R. Rivera, J. Kim, and O. Chae, "Local Directional Ternary Pattern for Facial Expression Recognition," *IEEE Transactions on Image Processing*, vol. 7149, no. c, pp. 1–13, 2017.

[97] R. S. Smith and T. Windeatt, "Facial Expression Detection using Filtered Local Binary Pattern Features with ECOC Classifiers and Platt Scaling," *Machine Learning Research*, vol. 11, pp. 1–7, 2010.

[98] M. F. Valstar, B. Jiang, M. Mehu, M. Pantic, and K. Scherer, "The first facial expression recognition and analysis challenge," *Face and Gesture*, pp. 921–926, 2011.

[99] K. Yu, Z. Wang, L. Zhuo, J. Wang, Z. Chi, and D. Feng, "Learning realistic facial expressions from web images," *Pattern Recognition*, vol. 46, no. 8, pp. 2144–2155, 2013.

[100] A. Majumder, L. Behera, and V. K. Subramanian, "Facial Expression Recognition with Regional Features Using Local Binary Patterns," *Computer Analysis of Images and Patterns: 15th International Conference, Proceedings, Part I*, pp. 556–563, 2013.

[101] Y. Yacoob and L. S. Davis, "Recognizing human facial expressions from long image sequences using optical flow," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 6, pp. 636–642, 1996.

[102] J. J. Lien, J. F. Cohn, T. Kanade, and C. C. Li, "Automated facial expression recognition based on FACS action units," *3rd IEEE International Conference on Automatic Face and Gesture Recognition, Proceedings*, pp. 390–395, 1998.

[103] M. Yeasin, B. Bullot, and R. Sharma, "Recognition of facial expressions and measurement of levels of interest from video," *IEEE Transactions on Multimedia*, vol. 8, no. 3, pp. 500–507, 2006.

[104] A. Sanchez, J. V. Ruiz, A. B. Moreno, A. S. Montemayor, J. Hernandez, and J. J. Pantrigo, "Differential optical flow applied to automatic facial expression recognition," *Neurocomputing*, vol. 74, no. 8, pp. 1272–1282, 2011.

[105] T. Wu, M. S. Bartlett, and J. R. Movellan, "Facial expression recognition using Gabor motion energy filters," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Workshops*, pp. 42–47, 2010.

[106] F. Long, T. Wu, J. R. Movellan, M. S. Bartlett, and G. Littlewort, "Learning spatiotemporal features by using independent component analysis with application to facial expression recognition," *Neurocomputing*, vol. 93, pp. 126–132, 2012.

[107] G. Zhao and M. Pietikainen, "Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 915–928, 2007.

[108] T. R. Almaev and M. F. Valstar, "Local gabor binary patterns from three orthogonal planes for automatic facial expression recognition," *Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII, Proceedings*, pp. 356–361, 2013.

[109] B. Jiang, M. Valstar, B. Martinez, and M. Pantic, "A dynamic appearance descriptor approach to facial actions temporal modeling," *IEEE Transactions on Cybernetics*, vol. 44, no. 2, pp. 161–174, 2014.

[110] Y. Pang, Y. Yuan, and X. Li, "Gabor-based region covariance matrices for face recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 7, pp. 989–993, 2008.

[111] R. Sivalingam, D. Boley, V. Morellas, and N. Papanikolopoulos, "Tensor Sparse Coding for Region Covariances," *Proceedings of the 11th European Conference on Computer Vision: Part IV*, pp. 722–735, 2010.

[112] O. Tuzel, F. Porikli, and P. Meer, "Pedestrian detection via classification on Riemannian manifolds," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 10, pp. 1713–1727, 2008.

[113] O. Tuzel and Porikli, "Human Detection via Classification on Riemannian Manifolds," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2007.

[114] S. Paisitkriangkrai, C. Shen, and J. Zhang, "Fast pedestrian detection using a cascade of boosted covariance features," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 8, pp. 1140–1151, 2008.

[115] M. T. Harandi, C. Sanderson, R. Hartley, and B. C. Lovell, "Sparse coding and dictionary learning for symmetric positive definite matrices: A kernel approach," *Lecture Notes in Computer Science*, vol. 7573, pp. 216–229, 2012.

[116] G. Cheng and B. Vemuri, "A novel dynamic system in the space of SPD matrices with applications to appearance tracking," *SIAM journal on imaging sciences*, vol. 6, no. 16, pp. 1–24, 2013.

[117] H. Qin, L. Qin, L. Xue, and Y. Li, "A kernel Gabor-based weighted region covariance matrix for face recognition," *Sensors (Switzerland)*, vol. 12, no. 6, pp. 7410–7422, 2012.

[118] S. Guo and Q. Ruan, "Facial expression recognition using local binary covariance matrices," *4th IET International Conference on Wireless, Mobile & Multimedia Networks*, pp. 237–242, 2011.

[119] R. Wang, H. Guo, L. S. Davis, and Q. Dai, "Covariance Discriminative Learning : A Natural and Efficient Approach to Image Set Classification Institute for Advanced Computer Studies," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2496–2503, 2012.

[120] K. Guo, P. Ishwar, and J. Konrad, "Action recognition from video using feature covariance matrices." *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 22, no. 6, pp. 2479–94, 2013.

[121] K. G. K. Guo, P. Ishwar, and J. Konrad, "Action Recognition Using Sparse Representation on Covariance Manifolds of Optical Flow," *7th IEEE International Conference on Advanced Video and Signal Based Surveillance*, pp. 188–195, 2010.

[122] C. Yuan, W. Hu, X. Li, S. Maybank, and G. Luo, "Human Action Recognition under Log-Euclidean Riemannian Metric," *9th Asian Conference on Computer Vision, Revised Selected Papers, Part I*, pp. 343–353, 2010.

[123] M. E. Hussein, M. Torki, M. A. Gowayyed, and M. El-Saban, "Human Action Recognition Using a Temporal Hierarchy of Covariance Descriptors on 3D Joint Locations," *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, pp. 2466–2472, 2013.

[124] H. Faulkner, E. Shehu, Z. L. Szpak, W. Chojnacki, J. R. Tapamo, A. Dick, and A. Van Den Hengel, "A Study of the Region Covariance Descriptor: Impact of Feature Selection and Image Transformations," *International Conference on Digital Image Computing: Techniques and Applications, DICTA*, pp. 1–8, 2016.

[125] Z. Hammal, "From face to facial expression," *Advances in Face Image Analysis: Techniques and Technologies*, vol. 3, pp. 217–238, 2010.

[126] J. D. Boucher and P. Ekman, "Facial areas and emotional information," *Journal of Communication*, vol. 25, no. 2, pp. 21–29, 1975.

[127] J. N. Bassili, "Facial motion in the perception of faces and emotional expression," *Journal of experimental psychology. Human perception and performance*, vol. 4, pp. 373–379, 1978.

[128] J. N. Bassili, "Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face," *Journal of personality and social psychology*, vol. 37, pp. 2049–2058, 1979.

[129] M. Smith, G. Cottrell, F. Gosselin, and P. Schyns, "Transmitting and Decoding Facial Expressions," *Psychological science*, vol. 16, pp. 184–189, 2005.

[130] S. Roy, C. Roy, Z. Hammal, D. Fiset, C. Blais, B. Jemel, and F. Gosselin, "The use of spatio-temporal Information in decoding facial expression of emotions," *Journal of Vision*, vol. 8, no. 6, pp. 707–707, 2010.

[131] Z. Hammal, M. Arguin, and F. Gosselin, "Comparing a novel model based on the transferable belief model with humans during the recognition of partially occluded facial expressions," *Journal of vision*, vol. 9, pp. 19–22, 2009.

[132] S. Lazebnik, C. Schmid, and J. Ponce, "A sparse texture representation using affine-invariant regions," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 319–324, 2003.

[133] V. Ferrari, T. Tuytelaars, L. Van Gool, and L. V. Gool, "Simultaneous Object Recognition and Segmentation by Image Exploration," *Toward Category-Level Object Recognition*, vol. 4170, pp. 145–169, 2006.

[134] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[135] T. Tuytelaars and L. Van Gool, "Matching widely separated views based on affine invariant regions," *International Journal of Computer Vision*, vol. 59, no. 1, pp. 61–85, 2004.

[136] K. Mikolajczyk and C. Schmid, "Indexing based on scale invariant interest points," *Proceedings Eighth IEEE International Conference on Computer Vision*, vol. 1, pp. 525–531, 2001.

[137] C. Schmid and R. Mohr, "Local grayvalue invariants for image retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 5, pp. 530–535, 1997.

[138] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," *Toward Category-Level Object Recognition*, no. Iccv, pp. 1470–1477, 2003.

[139] M. Brown and D. G. Lowe, "Recognising panoramas," *Proceedings Ninth IEEE International Conference on Computer Vision*, pp. 1218–1225, 2003.

[140] G. Dorko and C. Schmid, "Selection of scale-invariant parts for object class recognition," *Proceedings Ninth IEEE International Conference on Computer Vision*, pp. 634–639, 2003.

[141] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 264–271, 2003.

[142] B. Leibe and B. Schiele, "Interleaving object categorization and segmentation," *Cognitive Vision Systems: Sampling the Spectrum of Approaches*, pp. 145–161, 2006.

[143] A. Opelt, M. Fussenegger, A. Pinz, and P. Auer, "Weak hypotheses and boosting for generic object detection and recognition," *8th European Conference on Computer Vision, Prague, Czech Republic, Proceedings, Part II*, pp. 71–84, 2004.

[144] K. Mikolajczyk, K. Mikolajczyk, C. Schmid, and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.

[145] F. Ahmed and E. Hossain, "Automated Facial Expression Recognition Using Gradient-Based Ternary Texture Patterns," *Chinese Journal of Engineering*, vol. 2013, 2013.

[146] X. Tan and B. Triggs, "Recognition Under Difficult Lighting Conditions," *IEEE transactions on image processing*, vol. 19, no. 6, pp. 1635–1650, 2010.

[147] W. Gao and X. Zhang, "An Improved Sobel Edge Detection," *3rd International Conference on Computer Science and Information Technology*, pp. 67–71, 2010.

[148] M. Pietikainen, A. Hadid, G. Zhao, and T. Ahonen, "Computer Vision Using Local Binary Patterns," *Springer Science and Business Media*, vol. 40, 2011.

[149] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognition*, vol. 29, pp. 51–59, 1996.

[150] T. Ojala, M. Pietikäinen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine Intelligence*, vol. 24, pp. 971–987, 2002.

[151] J. Huang, J. Zhao, W. Gao, C. Long, L. Xiong, Z. Yuan, and S. Han, "Local binary pattern based texture analysis for visual fire recognition," *3rd International Congress on Image and Signal Processing*, vol. 4, pp. 1887–1891, 2010.

[152] T. Maenpaa, "The local binary pattern approach to texture analysis – extensions and applications," *Infotech Oulu, Department of Electrical and Information Engineering, University of Oulu*, 2003.

[153] T. Ojala, K. Valkealahti, E. Oja, and M. Pietikäinen, "Texture discrimination with multi-dimensional distributions of signed gray-level differences," *Pattern Recognition*, vol. 34, no. 3, pp. 727–739, 2001.

[154] M. Varma and A. Zisserman, "A statistical approach to material classification using image patch exemplars," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 11, pp. 2032–2047, 2009.

[155] (Accessed on: 2017-10-10) Digital Image Processing - Kirsch Compass Mask. [Online]. Available: https://www.tutorialspoint.com/dip/Krisch_Compass_Mask.htm

[156] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, "Geometric means in a novel vector space structure on sysmetric positive-definite matrices," *SIAM J. Matrix Analysis Applications*, vol. 29, pp. 328–347, 2006.

[157] W. Forstner and B. Moonen, "A Metric for Covariance Matrices," *Quo vadis geodesia*, vol. 66, pp. 113–128, 1999.

[158] Z. Huang, R. Wang, S. Shan, X. Li, and X. Chen, "Log-Euclidean Metric Learning on Symmetric Positive Definite Manifold with Application to Image Set Classification," *Proc.ICML*, vol. 37, pp. 720–729, 2015.

[159] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, "Log-euclidean metrics for fast and simple calculus on di usion tensors," *Magnetic resonance in medicine*, vol. 56, pp. 411–21, 2006.

[160] Y. L. Tian, T. Kanade, and J. F. Cohn, "Facial expression analysis," *Handbook of Face Recognition*, pp. 247–275, 2005.