UNIVERSITY OF KWAZULU-NATAL

COLLEGE OF AGRICULTURE, ENGINEERING AND SCIENCE (AES)

ADAPTIVE SEGMENTATION AND PATCH OPTIMISATION FOR MULTI-VIEW STEREO RECONSTRUCTION

The dissertation is submitted as a fulfilment of the academic requirements for the degree of

Master of Science in Engineering (MSc- Eng.)

School of Electrical, Electronic and Computer Engineering

University of KwaZulu-Natal, Durban, South Africa.

Author: Ray Leroy Khuboni

Supervisor: Mr Bashan Naidoo

DECEMBER 2015

DECLARATION OF PLAGIARISM

I, Ray Khuboni, hereby declare that:

(i) The research work reported in this dissertation, except where otherwise indicated, is my original research.

(ii) This dissertation has not been submitted for any degree or examination at any other university.

(iii) This dissertation does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

(iv) This dissertation does not contain other persons' 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, their writing has been placed inside quotation marks, and referenced.

(v) Where I have reproduced a publication of which I am an author, co-author or editor, I have indicated in detail which part of the publication was actually written by myself alone and have fully referenced such publications.

(vi) 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 sections.

Signed: ....

As the candidate's supervisor, I agree / do not agree upon the submission of this dissertation.

Signed: ....

Date of Submission: ....

DECLARATION OF PUBLICATIONS

I, Ray Khuboni, hereby declare that:

1. The research work in this dissertation submitted for publications, except where otherwise indicated, is my original research.
2. This dissertation presents the details of contributions towards publications, including publications in preparation, submitted, published, as well as of each author along with their own contributions towards the research work.

SUBMISSION 1

R. Khuboni and B. Naidoo, "Adaptive Segmentation for Multi-View Stereo", 2015

Submitted to IET - Computer Vision (Under Review)

SUBMISSION 2

R. Khuboni and B. Naidoo, "Stochastic Quasi-Newton Patch Optimization for Multi-View Stereo Model", 2015

Submitted to IET - Computer Vision (Under Review)

Publisher: Inst. Engineering Technology-IET (England, Impact Factor: 0.963)

ISSN: 1751-9632

This journal is part of Thomson Reuters/ISI Web of Science list of accredited journals.

Signed by Author: ....

As the candidate's supervisor, I agreed / did not agree upon the submission of these publications.

Signed: ....

ACKNOWLEDGEMENTS

This message serves as an expression of my gratitude to all those who accompanied me in this long journey with their support of prayers, encouragement and motivation. Most of all, I thank God for blessing me with this achievement, and for making it possible for me to reach the end mark of this degree. The challenges I experienced on the way have been the learning curves for me, and have made me stronger than I was when I started.

To my supervisor, Mr Bashan Naidoo, thank you for affording me an opportunity to prove my abilities by nurturing my potential through organising the financial support for the duration of my study towards Master's degree, as well as the mentorship you provided for my dream to come true. You have shared a lot of wisdom and proven to be a role model. I also wish to express my appreciation to all my educators, the staff members in the discipline of Electrical, Electronic and Computer Engineering for their willingness to assist me during this crucial time, the funders for providing funding, as well as my peers and friends who have inspired me to stay committed and focused, thank you.

My special thanks to my mother and family for their tireless efforts and guidance they provided throughout my childhood to this level of education. It is an honour for me to be crowned with this degree!

Portion of the building of the discipline of Electrical, Electronic and Computer Engineering was reconstructed - right outside the main entrance.

ABSTRACT

This dissertation presents two main contributions towards the Patch-based Multi-View Stereo (PMVS) algorithm. Firstly, we present an adaptive segmentation method for pre-processing input data to the PMVS algorithm. This method applies a specially developed grayscale transformation to the input to redefine the intensity histogram. The Nelder-Mead (NM) simplex method is used to adaptively locate an optimized segmentation threshold point in the modified histogram. The transformed input image is then segmented using the acquired threshold value into foreground and background data. This segmentation information is thus applied to the patch-based method to exclude the background artefacts. The results acquired indicated a reduction in cumulative error whilst achieving relatively similar results with a beneficial factor of reduced time and space complexity.

Secondly, two improvements are made to the patch optimisation stage. Both the optimisation method and the photometric discrepancy function are changed. A classical quasi-newton BFGS method with stochastic objectives is used to incorporate curvature information into stochastic optimisation method. The BFGS method is modified to introduce stochastic gradient differences, whilst regularising the Hessian approximation matrix to ensure a well-conditioned matrix. The proposed method is employed to solve the optimisation of newly generated patches, to refine the 3D geometric orientation and depth information with respect to its visible set of images. We redefine the photometric discrepancy function to incorporate a specially developed feature space in order to address the problem of specular highlights in image datasets. Due to this modification, we are able to incorporate curvature information of those patches which were deemed to be depleted in the refinement process due to their low correlation scores. With those patches contributing towards the refinement algorithm, we are able to accurately represent the surface of the reconstructed object or scene. This new feature space is also used in the image feature detection to realise more features. From the results, we noticed reduction in the cumulative error and obtained results that are denser and more complete than the baseline reconstruction.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ALGORITHMS

LIST OF ACRONYMS

2D	Two Dimensional
3D	Three Dimensional
BFGS	Broyden-Fletcher-Goldfarb-Shanno [Minimisation Method]
CIELAB	CIE Luminance channel, a-channel, b-channel [Colour Space]
CPU	Central Processing Unit [Hardware]
DAISY	Efficient Dense Descriptor
DoG	Difference of Gaussian [Feature Operator]
EMT	Edge Maximisation Threshold [Segmentation Method]
GPU	Graphics Processing Unit [Hardware]
HDT	Histogram Dependent Threshold [Segmentation Method]
HSV/HSI	Hue, Saturation, Value or Intensity [Colour Space]
IS	Iterative Snapping [Visual Hull Meshing Method]
M3C2	Multi-scale Model to Model Cloud Comparison [Model Evaluation Method]
MVS	Multiple View Stereopsis [Image-based 3D reconstruction method]
NCC	Normalized Cross Correlation [Template Matching Method]
NM	Nelder-Mead [Simplex Minimisation Method]
PCM	Phase Correlation Matching [Template Matching Method]
PMVS	Patch-based Multi-View Stereopsis [Image-based 3D reconstruction method]
PSR	Poisson Surface Reconstruction [Meshing Method]
RGB	Red, Green, Blue [Colour Space]
ROI	Region Of Interest

SBFGS	Stochastic – BFGS [Minimisation Method]
SGD	Stochastic Gradient Descent [Minimisation Method] Scale Invariant Feature Transform
SIFT	Scale Invariant Feature Transform
SQN	Stochastic Quasi-Newton [Minimisation Method]
SSD	Sum of Squared Differences [Template Matching Method]
SURF	Speeded-Up Robust Features
ZNCC	Zero Normalized Cross Correlation [Template Matching Method]

CHAPTER 1 - LITERATURE REVIEW

1.1 Introduction

More and more, modern digital cameras are becoming widely accessible and generally, they are inexpensive to procure. We now have easy access to advanced digital cameras that enable us to capture high resolution images for a range of applications. The process of acquiring three-dimensional structure from photographs is known as Image-based modelling or Multiple-View Stereopsis (MVS) reconstruction. With Multiple-View Stereopsis, we are able to easily generate three-dimensional (3D) structures from two-dimensional (2D) images; and these may be used for the purpose of viewing, preservation or value-added image-based applications. The geometric acquisition that is possible from these images allows us to obtain good enough accuracy to approximate the performance of the latest 3D laser scan technologies. The potential industrial applications of image-based reconstruction may be found, *inter alia*, in media, entertainment, scientific and engineering works. This research domain is a deeply researched area such that there exist several variations of the MVS algorithms. The Multiple-View Stereopsis (MVS) is known as a *state-of-the-art* multi-view image-based reconstruction method that uses a set of calibrated images to perform geometric acquisition. The collection of calibrated views is used for the purpose of recovering 3D information from 2D views. Each method is best suited to a particular class of problems, and each is bound to have limitations. However, most of these methods achieve results that could in fact rival 3D laser scanners. The laser scanners are active and intrusive 3D geometric acquisition systems, whereas MVS methods are passive 3D acquisition systems which are based on 3D image modelling. While laser scanners achieve relatively better geometric results when compared to MVS methods, they are rather expensive to procure. The pure laser technologies are also unable to capture object texture, and can only be used on objects that are rigid throughout the protracted scanning process. With MVS, deformable objects may be modelled by a synchronised set of cameras, and all models produced will be naturally textured and rendered in full colour by default.

1.2 Multiple-View Stereopsis (MVS) Requirements

During the initialisation process of the MVS reconstruction algorithms, an input is required, such as a set of calibrated images and some form of geometric information of the object or scene being reconstructed. The geometric information is known as the camera pose relative to the real world which can be calibrated within a controlled environment [1], [2], or auto-calibrated to estimate the actual camera parameters [3]. Some MVS algorithms may require a rough volumetric box that will serve as an initial estimate of the reconstruction [4], [5], whereas the others require a visual hull reconstructed from silhouette images which will serve as the initial outer estimate of the object's geometry [6], [7]. In most cases, some of the geometric constraints are established for the scene geometry.

1.3 Multiple-View Stereopsis (MVS) Reconstruction Algorithms

The major goal of these multiple-view stereopsis methods is to reconstruct a complete three dimensional model from a sequence of images taken from known camera viewpoints. With an increased number of highly developed algorithms, the state-of-the-art is improved rapidly. However, it becomes difficult to quantitatively compare the performance of these MVS methods when there is a lack of benchmark datasets. *Seitz et al.* [1], [2], offers a way to compare and evaluate the quantity of the different MVS reconstruction methods for a given benchmark, i.e. object datasets [2]. There are six different image datasets provided by *Seitz et al.* [1], [2] that goes along with their corresponding camera parameters, i.e. internal and external parameters. Middlebury Evaluation Survey [2], provides a quantitative comparison and the evaluation results of all the MVS methods that were submitted, this is also available online [2]. In order for MVS method to be evaluated, one of the mesh model of the benchmark datasets is submitted for evaluation against the ground truth mesh model, and the evaluation is based on the accuracy and completeness. *Strech et al.* [3], briefly provided an MVS evaluation survey for dense benchmark datasets, i.e. scene datasets. Therefore, they provided the laser scanned ground truth mesh models along with their benchmark datasets. According to *Seitz et al.* [1], [2], the MVS algorithms can be classified into four categories, i.e. the Voxel-based methods, Polygonal Meshing methods, Multiple Depth Map Fusion methods and Sparse Dense Patch-based methods. A brief description of each category of the reconstruction method is given hereunder.

1.3.1 Voxel-based Methods

These voxel methods consists of a Volumetric Graph-Cuts [8], Level-Sets [9] and Space-Carving [10] based techniques. These methods use a volumetric voxel as a surface representation to perform a single pass through a 3D volume, whilst computing the minimum cost function. Voxel cubes with a cost function value below a certain threshold are selected for reconstruction [8], [11]. The major downfall of these algorithms is that, they lack regularisation (smoothing), therefore are more prone to noisy reconstruction [8], [9], [10], [11]. Level-Sets method represents the surface as a time-varying volumetric density function and it is iteratively refined by maximisation of the photometric consistencies [10], [12], [13], [14]. Thereafter, a surface extraction method such as Matching Cubes [15] is needed to obtain the surface. However, the Level-Set method requires a large amount of computational memory to handle the high resolution images. The Graph-Cut method expresses image discrepancies as weighted links of adjacent voxels in a volumetric model [7], [9], [16], [17], [18], [19], and the surface is reconstructed based on a minimum cut of the graph (volumetric min-cut). These methods are able to handle both the object and the scene datasets, but they still require a volumetric box to initialise them.

1.3.2 Polygonal Meshing Methods

This method represents the surface of the reconstructed model as a set of polygonal surface meshes [4], [5], [20]. These algorithms extract silhouettes from the input images to construct a visual hull model, which is used to initialise the surface mesh. The surface mesh model is then refined to optimise each vertex position, using photometric and geometric consistencies. These methods produce remarkable results, however, any large error in the visual hull model cannot be recovered, hence these methods are highly reliant to an accurate visual hull [4], [5]. Therefore these methods are best suited for object datasets wherefrom the silhouettes can be accurately extracted. In a case of Poisson Surface Reconstruction (PSR) method [21], this method is used when it is difficult to extract silhouettes from the object or scene datasets. This usually occurs when there is not enough views to create an estimated bounding volume. This method takes a set of oriented points to construct a tight (closed) surface mesh model, and the mesh triangles is determined by the density of the reconstructed points. If the silhouettes are available, the Iterative Snapping (IS) [22] method can be used to compute a visual hull model to initialise the mesh model and to iteratively refine the initial mesh model to the reconstructed points by enforcing consistency. The method uses gradient descent method to optimise the position of all the vertices in the mesh model by minimizing the sum of two energy functions [22]. The Iterative Snapping method is best suited and constrained to the object datasets only.

1.3.3 Multiple Depth Map Fusion Methods

Multiple Depth Map Fusion method computes individual depth map for every single input image and then at the latter stage, all the individual depth maps are fused together to produce a single model. The common approach that is used in [23] is an independent estimation of the discrete depth range value for each pixel, whereas in [24], Markov Random Field is used to model all the individual depth maps, and the Expectation Maximisation algorithm [25] is used to fuse them into a single 3D reconstruction. The reconstruction may be prone to noise if the individual depth maps are inconsistent to each other [26] [27]. These methods are suited for both the object and the scene datasets.

1.3.4 Sparse Dense Patch-based Methods

These methods use a sparse set of unconnected surface patches to represent the surface of the reconstructed model [22], [27], [28], [29], [30]. The flexibility of the surface representation allows for the algorithm to handle both object and scene dataset with ease. These methods begin with image feature extraction and matching, and produce a set of feature points across all input views, thus resulting in a sparse set of patches, and thereafter, those patches are expanded out to produce a dense set of patches. Usually a post-processing stage is required to produce a polygonal mesh model with connected vertices [22]. These methods do not require a volumetric mesh initialisation nor a bounding volumetric box [31], [32]. The *sparse point based method* proposed by M. Lhuillier [29] is related to *patch-based methods* such as Patch-based Multi-View Stereopsis (PMVS) [22], [31]. Furthermore, a quantitative comparison and the evaluations provided by the Middlebury survey [2], show that the Patch-based MVS algorithm still produces, by far, the best overall performance for all six of its datasets in both accuracy and completeness. The work presented in this dissertation is based on this reconstruction method.

1.4 Motivation and Research Objectives

The work presented in this dissertation is aimed at building on the foundation that was proposed by Y. Furukawa and J. Ponce [22], [31]. A Patch-based MVS (PMVS) algorithm is the foundation of this dissertation. This patch-based method can handle all types of image datasets, robust to occlusions and handles the change in illumination by simply ignoring views that have bad photometric discrepancy score against the reference view. Despite the general good performance of the PMVS, the method is still subject to limitations with regard to accuracy and completeness of any reconstruction. The objective of Paper 1 was the improvement of space and time complexity while maintaining the reconstruction accuracy and completeness, and for Paper 2; we improved accuracy and completeness at the cost of the computation time.

In the first paper, we addressed the issue of spurious background features which are always unnecessarily included in the final PMVS reconstruction. Therefore, we suggested an object segmentation to pre-process the input data for the PMVS algorithm. We realised that by eliminating these spurious features at the initial seeding stage, this reduces erroneous patches that enter into iterative expansion and filtration stages. As a result, we achieved lower space and time complexity, and a more accurate and complete reconstruction became possible.

The second paper addresses the issues of the patch optimisation and specular highlights, using a specially-developed feature space to improve the patch reconstruction. The patch optimisation stage of PMVS algorithm minimises the global photometric discrepancy function, such that the 3D geometric orientation and depth information of the reconstructed patch is optimised with respect to its visible set of images. This had a significant impact on a subsequent surface reconstruction. The data is transformed to the new feature space that is designed to deal with specular highlights, and an appropriately transformed photometric discrepancy function is then minimised.

1.5 Contributions of Included Papers

All the research that has been covered in this dissertation is incorporated into the following two papers that are presented in Chapter 2. The details of the included papers are described below:

1.5.1 PAPER 1

R. Khuboni and B. Naidoo," Adaptive Segmentation for Multi-View Stereo," IET Computer Vision Journal, submitted in 2015.

Paper 1 presents an adaptive segmentation method for pre-processing of input image data to the patch-based method (PMVS). This segmentation method is applied to a specially developed grayscale transformation to the input data to redefine the intensity histogram. A heuristic simplex method is used to adaptively locate an optimized segmentation threshold point in the modified intensity histogram. The transformed input image is then segmented using the acquired threshold value into foreground and background data. This segmentation information is thus applied to the patch-based method to exclude any background artefacts that may be included in the final reconstruction. The results acquired indicated a reduction in cumulative error, whilst achieving relatively similar results with a benefit of reduced time and space complexity.

1.5.2 PAPER 2

R. Khuboni and B. Naidoo," Stochastic Quasi Newton Patch Optimization for Multi-View Stereo Model," IET Computer Vision Journal, submitted in 2015.

Paper 2 presents two main improvements towards the patch optimisation stage. Both the optimisation method and the photometric discrepancy function are changed. A classical quasi-newton BFGS method is modified to introduce stochastic objectives, whilst regularising the Hessian approximation matrix to ensure a well-conditioned matrix. The proposed method is employed to solve the optimisation of the newly generated patches, to refine the 3D geometric orientation and depth information with respect to its visible set of images. We redefine the photometric discrepancy function to incorporate a specially developed feature space in order to address the problem of specular highlights in the image datasets. Due to this modification, we are able to incorporate curvature information of those patches which were deemed to be depleted in the refinement process as a result of the low correlation scores. With the contribution of those patches, towards the refinement algorithm, we are able to accurately represent the surface of the reconstructed object or scene. This new feature space is also used in the image feature detection to promote more features. From the results, we noticed the reduction in the cumulative error and obtained the outcome that is denser and more complete than the baseline reconstruction.

1.6 Future Work

Seitz et al. [1] define four different classes of MVS algorithms, all of which are available for image-based reconstruction, and all of which exhibit limitations and drawbacks of some form. With a significant room for improvement, the new automated geometric acquisition methods are constantly in development. The reconstruction algorithms may be characterised by:

- Memory Usage,
- Accuracy,
- Completeness, and,
- Speed.

Any means of optimising one or more of these four characteristics could be considered, for example:

- **To minimise memory usage:** By using image segmentation to focus on the region of interest or using clustering method to handle the memory usage.
- **To maximise accuracy:** More advanced template matching (i.e. NCC, PCM, SSD, ZNCC) or feature descriptors, i.e. SIFT, SURF, DAISY, may be considered.
- **To maximise completeness:** Dense reconstruction techniques could be considered, i.e. to generate a 3D patch for every single pixel.
- **To maximise speed:** Parallelisation, including Cluster Processing, Many-Core or Multi-Core processors, could be introduced. Dynamic programming may be used to accomplish a higher throughout, i.e. avoid memory bottlenecks.

A major improvement to the methods such as PMVS, could be realised by parallelisation and deployment to Graphical Processing Units (GPUs); such as NVIDIA Pascal which offers a dynamic programming. With the compatibilities of the dynamic programming introduced by NVIDIA, we could implement the entire PMVS algorithm on a Pascal GPU. The dynamic programming offers a lot of flexibility when it comes to memory allocation on the GPU device, we could define, redefine or delete memory allocations made by the GPU without the interaction of the HOST, i.e. CPU. Another improvement that could be made is when you use a DAISY feature descriptor along the epipolar line to find the best or closest possible matches among the image features. This will allow for better matches and accurately match image features in the initial seeding stage of the PMVS algorithm. This should reduce erroneous patches that enter into iterative stages. In Paper 2, we proposed a method to improve the accuracy and completeness of scene reconstruction at the cost of the execution time. The proposed method showed very promising results but a more generalised 3D reconstruction method can be developed and tested to handle crowded scene dataset.

CHAPTER 2 - INCLUDED PAPERS

PAPER 1 – ADAPTIVE SEGMENTATION FOR MULTI-VIEW STEREO

Submitted to Journal of IET Computer Vision (Under review)

Ray Khuboni & Bashan Naidoo

Abstract

This paper presents an adaptive segmentation method for pre-processing input data to the Patch-based Multi-View Stereo (PMVS) algorithm. A specially developed grayscale transformation is applied to the input image data thus redefining the intensity histogram. The Nelder-Mead (NM) simplex method is used to adaptively locate an optimised segmentation threshold point in the modified histogram. The transformed input image is then segmented using the acquired threshold value, into foreground and background data. This segmentation information is thus applied to the patch-based method to exclude the background artefacts. The method is targeted at segmenting out potentially disruptive data and is able to realise a reduction in cumulative error of the reconstruction process and thus improve the final reconstruction. The performance evaluation shows simultaneous reduction in computational times and improvements to specific object reconstruction challenges. With this method, we obtain results that are relatively similar to the original patch-based method, but with reduced time and space complexity.

---

1. *Ray Khuboni*
School of Engineering, University of KwaZulu-Natal – Howard College, South Africa, Durban, 4041, 207502748@stu.ukzn.ac.za and Tel. (+2773) 913 2867

2. *Bashan Naidoo*
School of Engineering, University of KwaZulu-Natal – Howard College, South Africa, Durban, 4041, bnaido@ukzn.ac.za and Tel. (+2731) 260 2758

2.1.1 Introduction

Multi-view stereopsis uses matching across views for automated geometry acquisition in scenes and for objects. The potential applications of reconstructed models range from the use of the media and entertainment industry, to quantitative recovery of information in scientific and engineering data analyses. In accordance with the survey conducted by Seitz *et al.* [1], [2], state-of-the-art Multi-View Stereo (MVS) algorithms are able to achieve relative accuracy better than 1/200 (1mm for a 20cm wide object) from a collection of low-resolution (640x480) images. These MVS algorithms can be classified into four categories:

a) *Voxel-based methods* that require a known bounding box [4], [5],
b) *Polygonal mesh-based methods* that require visual hulls to initialize [6],
c) *Multiple depth-map methods* that are flexible but require fusing individual depth maps into a single model [20], [33], and lastly,
d) *Patch-based methods* that neither require a bounding volume nor a visual hull for initialisation [22], [31].

The later method is similar to point-based method by M. Lhuillier [29], but the patch-based method replaces the greedy expansion procedure by several iterations between expansion and filtering stages. The Middlebury benchmark [2] provides an objective quantitative comparison and evaluation. It shows that the patch-based MVS algorithm [22] still outperforms all the other methods submitted up to this day, on overall performance in both accuracy and completeness for all six datasets.

Our study of the PMVS algorithm showed that for all object reconstruction, the background is always unnecessarily included in the final reconstruction. Therefore, we suggest an object segmentation as the first step in the reconstruction algorithm. The spurious background features are sometimes consistently detected across the images, and are incorrectly incorporated into the object model. The object segmentation excludes most of these from consideration by leaving only the region of interest, i.e. the object. The elimination of spurious features at this initial stage, reduces erroneous patches entering into the cyclic expansion stages of the PMVS algorithm; and that reduces computational time in all subsequent cycles. Given that erroneous patches are no longer repeatedly expanded and filtered, the space and time complexity are reduced, and a more accurate and complete reconstruction becomes possible in a shorter time. The segmentation of the region of interest assists the algorithm by restricting the set of initial seed patches to a more plausible domain. Reduction in space complexity allows any given computing resource to solve larger scale problems that would previously have been deemed impossible, given its hardware limitations. When referring to larger scale problems, we intentionally refer to object datasets of higher image resolutions which are *lab-captured* object datasets under nominally black or white backgrounds. The proposed method of masking the region of interest in every input image constrains the PMVS method. While not explicitly reconstructed, a virtual visual hull is used to condition the matching of features across the views. This constrained PMVS approach is applicable to the object reconstruction only, where the background is controlled. Also, it modifies the classification of the original unconstrained algorithm under the analysis of *Seitz et al.* [1].

Although the application of the proposed technique to the scene reconstruction has not been fully investigated, the results obtained using datasets from *Strech et al.* [3] are very promising. A crowded scene dataset such as Colosseum, Rome, Venice and Dubrovnik [33], [34], [35], [36], can be used to develop and test generalized 3D reconstruction methods with unconstrained background and wide baseline reconstruction; other related methods included [32], [36], [37]. The PMVS may be used to reconstruct 3D scenes from such datasets. However, a great deal of time and memory is consumed, and sometimes wasted in the background reconstruction, which may ultimately be filtered out from PMVS [22]. The proposed method applies more specifically to a restricted subset of 3D reconstruction where background is controlled [1], but it also works for scene image data-set with certain distinctive colour tone, such as *Strech et al.* [3]. The intention of the object-segmentation is to exploit the potential benefits that a controlled and nominally black or white background it introduces. The PMVS does not explicitly exclude the background in its processing; this study proposes a constrained PMVS that attempts to exclude the controlled background in a pre-processing step. The proposed method looks at the most and least-dominant channels in the RGB colour space by deploying ceiling and flooring operators over the three colour channels of each pixel. The ceiling operator effectively reduces the colour image to a single channel that represents the highest intensity colour component (R, G or B) of each pixel. This causes a shift upwards in the intensity histogram and is the same as the V channel of the HSV colour space. For the least-dominant channel the floor operator similarly reduces the image to a single channel, but causes a shift downwards in the intensity histogram. These intensity transforms typically produce bi-modal histograms with one mode representing the foreground and the other representing black or white background. The ceiling operation produces good separation for images with black backgrounds while the floor operator does the same for white. The method also works for scene datasets that have distinctive background colours.

In order to improve the performance and quality of the reconstruction of the 3D patch model, we consider region of interest (ROI) segmentation using the threshold techniques. The segmentation algorithms generally exploit the basic properties of intensity value discontinuity and similarity. In the first case, the image is partitioned based on abrupt changes in intensity, such as edges in an image, whereas in the second case, it is partitioned based on predefined criteria to separate an image into foreground and background regions. In this paper, we focus on the use of the second case which is an adaptive segmentation method, and we used it as a pre-processing stage for the PMVS algorithm. The image is segmented based on its global information [38]. The PMVS algorithm [22], [31], defines foreground as being dark, and background as being white. In this approach we propose an adaptive segmentation method that will eliminate much of the background artefacts.

This paper is structured into six sections:

1. First, an introductory section is presented as an introduction of the scope of the work.
2. Second, an overview of the image segmentation techniques that are available, and the method that was used in this paper, is explained in this section.
3. Third, an overview of the patch-based method that was used as the foundation of the work.
4. Fourth, a detailed description of the proposed work.
5. Fifth, the experimental results are shown and explained in detail, from Middlebury Evaluation, Patch Evaluation, Reconstruction Efficiency and M3C2 Evaluation.
6. Sixth, the section concludes the final thoughts on the paper and the proposed work.

2.1.2 Histogram Dependent Threshold

In the initial findings of the patch-based method, it was noted that the algorithm was not efficient when it came to memory usage, and that it was impossible to process large image data-sets or higher resolution images. That is why the image segmentation techniques were considered as a memory reduction method for the patch-based method. To address any unwanted background data, since the main concern of the reconstruction algorithm, was to reconstruct the object of interest. Currently, there are many variations of image segmentation techniques available for use, such as the mean-value method, p-tile method, histogram dependent technique (HDT), edge maximization technique (EMT) and visual technique, refer to [38] for further details. This paper focuses on the histogram-based technique which is reliant on the estimation of an appropriate threshold value to separate the region of interest, i.e. the object from the background in an image. The technique may be applied to specific classes of data such as object datasets; where for example, the background is chromatically constrained to some degree. The 1D histogram of 2D gray-scale image is analysed in order to determine optimum threshold for segmentation. The histogram is acquired by sorting each pixel in the gray-scale image into ordered gray-scale bins, which range from 0 (black) to 255 (white). Once computed, the threshold may be used to binarise the image. The binary image should contain all of the essential information about the position and shape of the object of interest (foreground) [38]. The selection of an appropriate threshold can be achieved through an analysis of the histogram. This is further explained in section 4. In real applications, however, histograms may be complex, thus making threshold selection difficult. The technique is regarded as a good candidate for achieving segmentation for gray-scale images with a low computational complexity.

2.1.3 Patch-based Multi-View Stereo (PMVS) Algorithm

2.1.3.1 Overview of the PMVS algorithm

The PMVS algorithm can be decomposed into four steps: feature detection, matching of features across multiple views (stereopsis), iterative patch expansion and filtering. The last two steps are intended to improve the patch density and accuracy of the 3D model, see Algorithm 1.

a) *Feature Detection:* Features are extracted using Harris and Difference of Gaussian (DoG) feature operators.
b) *Feature Matching:* The detected features are matched across the multiple views to yield an initial set of 3D patches associated with their salient image regions.
c) *Patch Expansion:* Those initially matched set of patches are expanded to the nearby pixels to obtain a denser set of patches with fewer gaps or spaces in between.
d) *Patch Filtering:* The denser set of patches that were expanded in the previous step may contain incorrect patches. To eliminate these patches, visibility constraints are used to remove outliers that lie either in front or behind the observed surface.

This algorithm reconstructs a set of oriented and textured patches that represent the surface of an object or scene of interest. A post-processing step can be added to convert patch models into polygonal-mesh models.

2.1.3.2 Key Elements of the PMVS algorithm

2.1.3.2.1 Patch Model

The patch model is a set of 3D oriented patches which represent the reconstructed surface of an object or a scene. Each patch p is a local tangential plane on the surface. Its geometry is fully determined by its center $c(p)$, unit normal vector $n(p)$ which is oriented towards the cameras observing it in 3D space, and a *reference image* $R(p)$ in which p is visible.

2.1.3.2.2 Feature Detection

This is a process of identifying points of interest, including blobs, edges and corners [16], in an image. It is regarded as a key ingredient for automated geometric acquisition in several computer vision algorithms [6]. The PMVS algorithm uses two image feature detectors which are the Difference of Gaussian (DoG) and Harris operators to detect image features in each image [7].

2.1.3.2.3 Feature Matching

For each image I_i, there is a set of neighbouring images that is considered for stereo computation. The camera parameters of this calibrated set must be known in order to verify the geometric constraints that are satisfied. An image I_i is considered along with optical center $O(I_i)$ of the corresponding camera. For each feature f detected in I_i, potential matches f' are collected from the other images as a set F if they lie within two pixels of the corresponding epipolar line. The potential patch p associated with f is reconstructed by triangulating the pair $(f \quad f')$ in order to obtain its center $c(p)$, normal vector $n(p)$ and set the reference image $R(p)$ to be I_i. The camera orientations are used in order to verify that patch p being matched with patches across the other views satisfies the geometric and visibility constraints. The image sets are initialized as follows:

$$V(p) = \left\{ I_i \middle| n(p) * \frac{\overrightarrow{c(p)O(I_i)}}{\left|\overrightarrow{c(p)O(I_i)}\right|} > \cos(\tau) \right\} \quad (1)$$

Where τ is the visibility angle between the patch normal, $n(p)$ and the optical center, $O(I_i)$ of the camera, ($\tau = \frac{\pi}{3}$ in our experiment), and $V(p)$ is the set of images in which patch, p is visible. This is a simple assumption that a patch p is visible within a certain angle in an image I_i and those patches are collected into a set of visible patches $V(p)$ for image I_i. Since there are specular highlights or even obstacles in a scene or object data-sets, the patches with a bad photo-metric discrepancy scores are ignored or filtered out, thus we get a new set of visible patches $V^*(p)$.

$$V^*(p) = \{ I_i | I_i \in V(p), h\big(p, I_i, R(p)\big) \leq \alpha\} \quad (2)$$

Where, $V^*(p)$ is the set of images in which the patch, p is accurately visible and which satisfy the minimum photo-metric discrepancy score. α is a threshold used for the photo-metric discrepancy score (α = 0.6, before and = 0.3, after optimization).

Given an initial patch p that is visible in the set of images $V(p)$ and a reference image $R(p)$, its position $c(p)$ and normal $n(p)$ can be estimated by minimizing Photo-metric Discrepancy Function (PDF) $g^*(p)$:

$$g^*(p) = \frac{1}{|V^*(p) \backslash R(p)|} \sum_{I_i \in [V^*(p) \backslash R(p)]} h\big(p, I_i, R(p)\big) \quad (3)$$

The photo-metric discrepancy score, $h(p, I_1, I_2)$ is computed as one minus the Normalized Cross-Correlation (NCC) score between $q_1 = q(p, I_1)$ and $q_2 = q(p, I_2)$. q_1 and q_2 are sampled pixel colours of a patch p for images, I_1 and I_2. The photo-metric discrepancy score is defined as follows:

$$h\big(p, I_i, R(p)\big) = 1 - \frac{\sum_{i=1}^{\mu*\mu} \sum_{j=1}^{\mu*\mu} q_1(i) q_2(j)}{\left[\sum_{i=1}^{\mu*\mu} q_1(i)\right]^{1/2} * \left[\sum_{j=1}^{\mu*\mu} q_2(j)\right]^{1/2}} \tag{4}$$

Where $h\big(p, I_i, R(p)\big)$ is a pairwise photo-metric discrepancy score of the patch, p projection in images I_i and $R(p)$.

2.1.3.2.4 Patch Optimisation

In the reconstruction of a single patch p, there are two steps that should be followed. The initialization of the corresponding parameters of each patch p, its center $c(p)$, normal $n(p)$, set of visible images $V^*(p)$, and the reference image $R(p)$; these initial parameters are obtained from 3D triangulation from a pair of stereo views [39]. The second step, is the patch optimisation, where the initial parameters of patch p are optimised to improve their geometric components, $c(p)$ and $n(p)$, by minimising the photometric discrepancy score. To simplify computations, $c(p)$ of patch p is constrained to lie on the ray that projects from the optical center through its corresponding image point in the reference image, to the 3D space, [39], see Figure 1. This reduces the number of degrees of freedom to three by solving for depth information and orientation. The normal $n(p)$ is parameterized by Euler angles (yaw and pitch) and thus yielding an optimisation problem with three parameters which are solved using conjugate gradient methods [40]. The Patch-based Multi-View Stereo (PMVS) reconstruction algorithm uses the Nelder-Mead Simplex method [41] to solve the depth information and refine the geometric components of patch p [22], [31], [32].

After the optimisation step, the visibility image sets are updated (i.e. $V(p)$ and $V^*(p)$ to correspond to the new parameters. If $\big||V^*(p)|\big| \geq r$, where r is a minimum number of visible images with low photometric discrepancy score, then patch generation (i.e. a patch being generated from 3D triangulation) is deemed a success and saved as a possible patch

p ($r = 3$ in our experiment). After the initial feature matching, the patch expansion stage takes the existing set of initial patches, i.e. initial seeding patches), to expand to the neighbouring spaces or gaps to create a denser patch model. These newly expanded patches are also initialized and optimised to represent the 3D surface of an object or scene being reconstructed. Due to the large number of patches being expanded, there might be a large number of erroneous patches that do not belong to or are incorrectly placed in 3D space. The filtering stage removes or eliminates these patches based on their visibility constraints, which lie either in front or behind the observed surface. These erroneous patches (outliers) are illustrated in Figure 1. The patch expansion and filtering stages are iterated n times ($n = 3$ in our experiment). The parameter n was chosen to be 3 in order to replicate the patch model results obtained in [22] using the Patch-based MVS algorithm. It was noted that when n was chosen to be less than 3, the patch model became less dense and when n was greater than 3, the patch model was prone to more erroneous patches.

Algorithm 1: Patch-based Multi-View Stereo (PMVS) Algorithm

Input: Image sequence and camera parameters

Output: Sparse set of oriented rectangular patches of the reconstructed model, P

Initialize: Image sequence scale transformation

Step – 1: Feature Detection

The image features are detected using HARRIS and Difference of Gaussian (DoG) feature operators which are explained in Feature Detection section. Refer to [22] and [42] for a better explanation.

Step – 2: Feature Matching – Stereo pair matching

For each image feature f being detected in the reference image $R(f)$, a feature f' is collected along the epipolar line into a set of features F that satisfy the epipolar consistency, from the stereo view I_i.

For each feature f' along the epipolar line of the stereo view I_i, the feature f of the reference image $R(f)$ is matched to find the best correlation. This creates a potential patch p that is associated with the feature f. For further explanation refer to Feature Matching section.

2.1. $\boldsymbol{Initialise\ c(p), n(p), V(p)}$ and $\boldsymbol{V^*(p)}$

Initialisation of these parameters are obtained through 3D triangulation of stereo pairs/views. Refer to Patch Optimisation section.

2.2. $\boldsymbol{Optimise\ c(p)\ and\ n(p)}$

This stage is known as the Patch Optimisation, where the patch center and normal are geometrically optimised/ refined to ensure the projection of patch across the set of visible images is correctly aligned to estimated geometric orientation of the patch. For further understanding refer to [22] and also refer to Patch Optimisation section.

2.3. $\boldsymbol{Update\ V(p)}$ ***and*** $\boldsymbol{V^*(p)}$

The set of visible images, $V(p)$ of patch p is updated that meet the visibility angle and the set of accurately visible images, $V^*(p)$ of patch p is updated that meet the minimum photo-metric discrepancy score. Refer to Feature Matching section.

2.4. If ($\| \boldsymbol{V^*(p)} \| \geq \boldsymbol{r}$) then, Save f', Add p to the set of newly generated patches P. Else move on to the next feature. This stage is explained in Patch Optimisation section.

Step – 3: Patch Expansion

In order to create a denser patch model, this stage creates at least one patch for every image cell. The expansion procedure uses the existing patches to expand them to the neighbouring cells. This stage is explained better in [22].

Expanding newly generated patches into neighbouring patches.

Step – 4: Patch Filtering

After the patch expansion stage, the number of patches has increased significantly and some of these patches may be erroneous and therefore should be removed. This stage is subdivided into four filters.

1. First filter enforces a weak visibility consistency
2. Second filter enforces a more strictly visibility consistency
3. Third filter enforces a weak form of regularisation on a single patch
4. Fourth filter enforces a weak form of regularisation on a group of patches

This stage is also explained best in [22].

Removing the incorrect/erroneous patches as outliers

Step – 3 and – 4 are repeated n times to ensure a denser patch model, n = 3 in our experiment

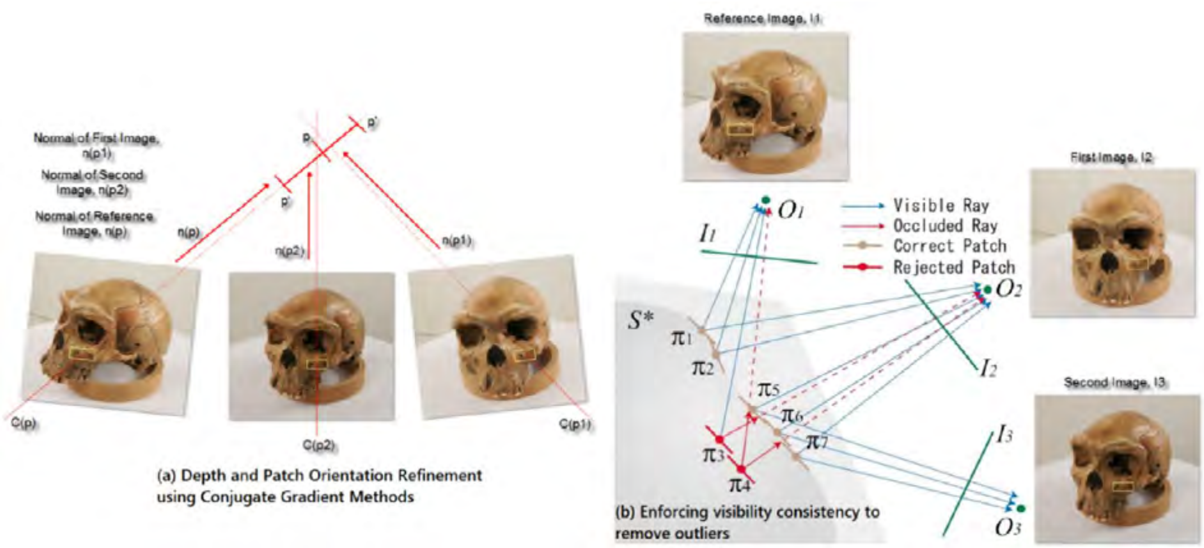

Figure 1: (a) The reference patch p along with its corresponding initial parameters, c(p) and n(p), is optimized using the parameters of the other views to refine the orientation and depth information in 3D space and (b) Patch filtering stage enforces the visibility consistency to remove outliers (red patches).

2.1.4 Proposed Method

The image segmentation is used to partition a digital image into something that is more meaningful and easier to analyse. The labels are assigned to every region or pixel, such that similar regions or pixels share the same label. The proposed method is an automated image segmentation method that is deployed as a pre-processing stage to the patch-based method. It applies to RGB encoded colour images and uses processed RGB values along with the histogram threshold. The proposed method is used to segment and remove background artefacts which often result in the reconstruction of outlier patches in 3D space. The detection of artefacts as possible image features is prevented and focus shifts itself to the object, i.e. region of interest. By removing the background features, we also reduce time and space complexity.

The patches that lie on the edge of an object in the PMVS reference image are likely to contain background artefacts. These are removed, however, they may be more accurately recovered from other views where they are not edge-aligned and contain only foreground pixels. Subsequent patch expansion to neighbouring patch spaces becomes faster and more effective because background artefacts are filtered out in the pre-processing stage. The checking for visibility of a newly expanded patch also becomes simpler because the algorithm is bounded to the region of interest, and there are fewer incorrect patches to be removed in the filtration stage. This may result in a more accurate 3D patch model.

This method assumes that all the images have been taken under a nominally black or white background conditions [2] for object datasets; and for scene-data under conditions where the background and foreground are derived from separable intensity distributions. This method is a histogram dependent threshold that analyses the intensity histogram of the original RGB image. The background and region of interest should come from reasonably separable distributions. This is why for object datasets, controlled backgrounds are nominally black or white, and whichever contrasts optimally with the foreground. The method masks out the background by using an adaptively selected intensity threshold value T. Before the image masking is performed, an initial threshold value T_0 is selected using the mean brightness or sum of all grayscale values $G(x, y)$ divided by the image resolution, i.e. $width * height$. This initial threshold value along with the dynamic range of the intensity histogram, is used to determine whether the image was taken under nominally black or white background. If the absolute difference of the initial threshold value, T_0 and the minimum dynamic range, x_{min} is greater than the absolute difference of the initial threshold value, T_0 and the maximum dynamic range x_{max}, the image is assumed to be taken under nominally white background, i.e. the majority of the pixels in the image are bright and for the opposite case, the majority of the pixels are dark. Either the RGB ceiling or floor histogram is computed and used to segment the image into foreground and background. The intensity histogram is represented in a more meaningful form to compute the local minima of the objective function. The objective function in this case, is a normalised discrete logarithmic function of the intensity histogram, i.e. either RGB ceiling or floor histogram. The optimum threshold value is finally obtained by minimising an objective function using the Nelder-Mead Simplex Method [41].

Normally, when segmenting an RGB image, the grayscale intensity image is used along with a selected threshold value to segment the pixels into foreground and background. Since the images are taken under natural conditions, some of the nominally black or white background pixels may vary significantly and might not be masked out. Similarly, some parts of the foreground will be masked out if intensities overlap with the background intensity mode. To rectify the problem, we use the RGB ceiling and floor transformations over the RGB colour-space to cause an intensity histogram shift that clearly masks out the black or white background respectively. The RGB ceiling and floor transformations are the pixel-wise selection of the most and least-dominant channels in the RGB triplet. RGB ceiling transform causes an upwards shift in the histogram by selecting the highest intensity between three colours, and RGB floor causes a downwards shift in the histogram by selecting the lowest intensity between the three colours. From these transformations, we can obtain a much clearer bi-modal histogram for nominally black and white background respectively.

After distinguishing if the image was taken under dark or bright background conditions, we proceeded with the selection of a suitable grayscale transformation (RGB ceiling or floor). Thereafter, the objective function is minimised to locate the local minima across the dynamic range of the transformed intensity histogram. Since the numerical minimisation method used is a heuristic search method, we constrain the search boundaries for the local minima, to the lower and upper bounds. The lower boundary is set to be half way between the minimum dynamic range, and the initial threshold value T_0 whereas for the upper bound, it is between the maximum dynamic range and the initial threshold value T_0. The optimum threshold value T is given by the corresponding bin value of the intensity histogram for the local minimum value of the objective function. With the selected threshold value T, we compute the binarised information of the input image data to obtain the boundary segmentation information. For the image binarisation, a suitable grayscale transformation is selected based on the background conditions for segmentation. For a nominally white background conditions, the RGB floor transform is used, whereas the opposite case, RGB ceiling transform is used. A pixel-wise binarisation is performed using the selected threshold along with transformed intensity histogram. A morphological image filling is used as a final step to remove any noisy binarised information, i.e. salt and pepper noise, within the foreground of interest. The boundary segmentation information is applied to the patch-based method. The details of the segmentation method are illustrated in Algorithm 2, also see Figure 3 to Figure 5. By setting the background pixels to null, we are able to eliminate the image features that lie on edges of the object, whilst avoiding any generation of edge-aligned patches which are associated with the eliminated image features.

Throughout the algorithm, we use the 2D boundary segmentation information to confine detection and matching of image features within the foreground, similar methods include, voxel-based method using the function costs to remove voxel-cubes which are above a certain threshold [4], [5], extracting silhouettes and fusing stereo depth maps [6] and space carving from silhouettes to create a visual hull [18]. The initial seeding patches are constrained to the 3D patch boundary so whenever there is an expansion done to the initial seeding patches, it is within the 2D boundary. For each feature **f** detected in the reference image $I(x, y)$, left image, potential matches f' are collected along the corresponding epipolar line of the right image. Those features with a high NCC score are considered to be a potential match f', as illustrated in Figure 2. We eliminate the background artefacts from the reconstruction, and thus the algorithm focuses on the foreground, see Figure 2. This improves the quality of the reconstructed object and scene model and also reduces the computational times for the reconstructed models, similar methods that exhibit similar characteristics and results include [6], [7], [15], [36], [37]. This can be seen from the results presented in Figure 6, Table 3 and Table 4.

Algorithm 2: Virtual Visual Hull Carving using Image Segmentation

Input: A set of original images and their respective camera parameters.

Output: 2D boundary segmentation information.

Step – 1: Compute the RGB flooring and ceiling grayscale transforms

a) Reading an image $I(x,y) = \{R(i,j), G(i,j), B(i,j) \mid 1 \le i \le n, 1 \le j \le m\}$ and the image size: $[m,\ n,\ 3]$

b) Compute the image into Grayscale image $G(x,y)$

c) Compute the Histogram, $H(x)$, of the Grayscale image $G(x,y)$

d) Compute the mean value $T_0 = \frac{1}{N}\sum_{i=1}^{N} H(x)$,

e) Compute the dynamic range, $x_{min} = \min(bin)$ and $x_{max} = \max(bin)$

f) If $(\|T_0 - x_{min}\| \ \ge \ \|x_{max} - T_0\|)$, Assumes a nominally white background.

Find the lowest intensities in the RGB colour space of the image $I(x,y)$,

RGB floor histogram:

$$V_{min} = floor(I(i,j)) = \{min(R(i,j), G(i,j), B(i.j)) | 1 \le i \le n, 1 \le j \le m\}$$

and the image size: $[m,\ n]$

g) If $(\|T_0 - x_{min}\| < \|x_{max} - T_0\|)$, Assumes a nominally black background.

Find the highest intensities in the RGB colour space of the image $I(x,y)$,

RGB ceiling histogram:

$$V_{max} = ceil(I(i,j)) = \{max(R(i,j), G(i,j), B(i.j)) | 1 \le i \le n, 1 \le j \le m\}$$

and the image size: $[m,\ n]$

Step – 2: Compute the appropriate threshold value using Nelder-Mead (NM) simplex method

The function minimisation is constricted to the lower and upper bound of the histogram $H(x)$. Any minimisation outside the bounding limits is not useful

$$p(x) = H([a\,..\,b])$$

Where $p(x)$ is a selected portion of the histogram, $H(x)$,

$a = \left[\frac{T_0 + x_{min}}{2}\right]$ is the lower bound and $b = \left[\frac{T_0 + x_{max}}{2}\right]$ is the upper bound of the histogram, $H(x)$.

The objective function represents the selected portion of the histogram, $H(x)$ in more meaningful way to compute the local minima of $f(x)$,

$$\text{Objective function: } f(x) = -p(x)\log(p(x))$$

The simplex method is used to search for the local minima and the local minima is the appropriate threshold.

$$\text{Threshold, } T = simplex\_min(f(x))$$

Step – 3: Compute the binarised information using the RGB grayscale transforms

Instead of using the grayscale image, $G(x, y)$ to produce the binarised information. The RGB grayscale transforms, V_{max} and V_{max} are used to clearly classify the foreground data from the background.

For nominally white: If $(V_{min}(x, y) > T)$ and $floodfill(B(x, y))$

For nominally black: If $(V_{max}(x, y) < T)$ and $floodfill(B(x, y))$

Where $B(x, y)$ is a binary image, x and y are the pixel coordinate points, $floodfill\big(B(x, y)\big)$ is a morphological filling of the binary image, $B(x, y)$ (i.e. salt and pepper remover).

Step – 4: The binarised information is used as 2D boundary information for the Patch-based MVS method.

The 2D boundary information is used throughout the stages of PMVS algorithm:

1. Feature Detection – limiting feature detection within the region of interest.
2. Feature Matching – matching is done within and matches on the edge of the foreground are discarded. Those patches that discarded from the edges can be recover within other camera views.
3. Patch Expansion – allowing the expansion to occur more within.
4. Patch Filtration – removing erroneous patches on the edges of the respective view.

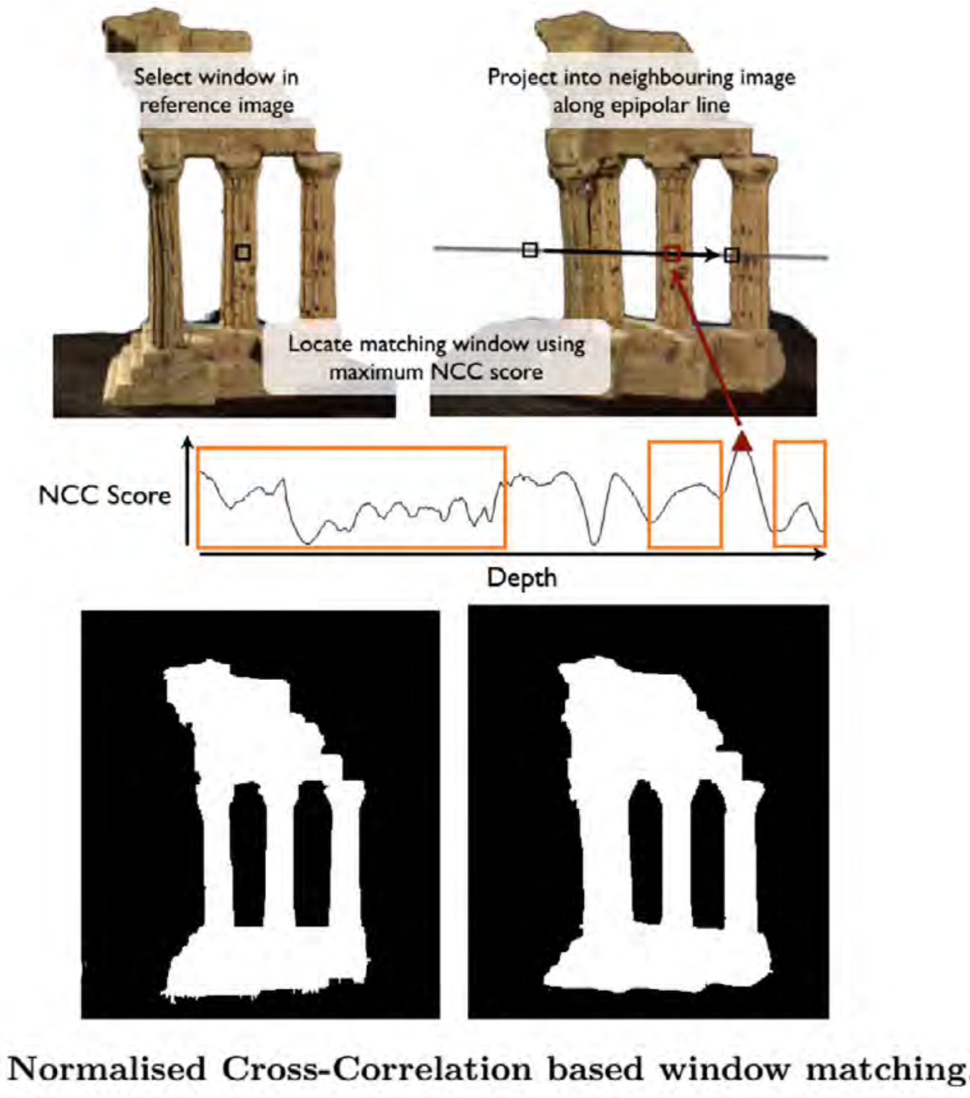

Figure 2: Template Matching is accomplished by using Normalized Cross-Correlation (NCC). (**Top-row**) By selecting a patch on the left image, we can search for the best possible match in the other views by moving our template window across the epipolar line. **(Bottom-row**) With the silhouette information obtained using the adaptive method proposed, we can reduce our search radius to the foreground only. This means that any search that is done within the orange boxes in the NCC score graph is ignored. The initial seeding and newly expanded patches that fall in the background are removed as outliers.

2.1.5 Experimental Results

The performance evaluation of our proposed methods was carried out using the Skull [22], Temple Ring, Dinosaur Ring [2], Fountain-P11 and Herz-Jesu-P8 [3] image datasets. All the experiments were conducted on an Intel® Core™ 2 Duo workstation, E7500 (2.93GHz). The post-processing is needed to turn the patch models into a surface polygonal mesh models for the image-based modelling applications. There are two mesh algorithms that are considered in this paper, namely, Poisson Surface Reconstruction (PSR) [21] and Iterative Snapping (IS) [31]. PSR is a closed form meshing algorithm that directly converts sets of oriented patches into triangulated mesh models. It produces high quality mesh models for the object and scene datasets. However, it cannot take an advantage of segmentation information associated with each input image of the object dataset. IS iteratively deforms a visual hull model towards a set of reconstructed patches. It is only applicable to the object datasets and it produces moderate mesh models [22], [31]. We aim to produce the mesh models for the object and scene datasets using a PSR. In our reconstruction pipeline, the segmentation information is acquired in the pre-processing stage to isolate the foreground from the background, thus producing a higher quality mesh model. The quality of the mesh model is directly related to the density of patch model. The denser the patch model, the finer the mesh model becomes. This can be seen from our mesh and texture mapped models, as shown in Figure 7 and Figure 8.

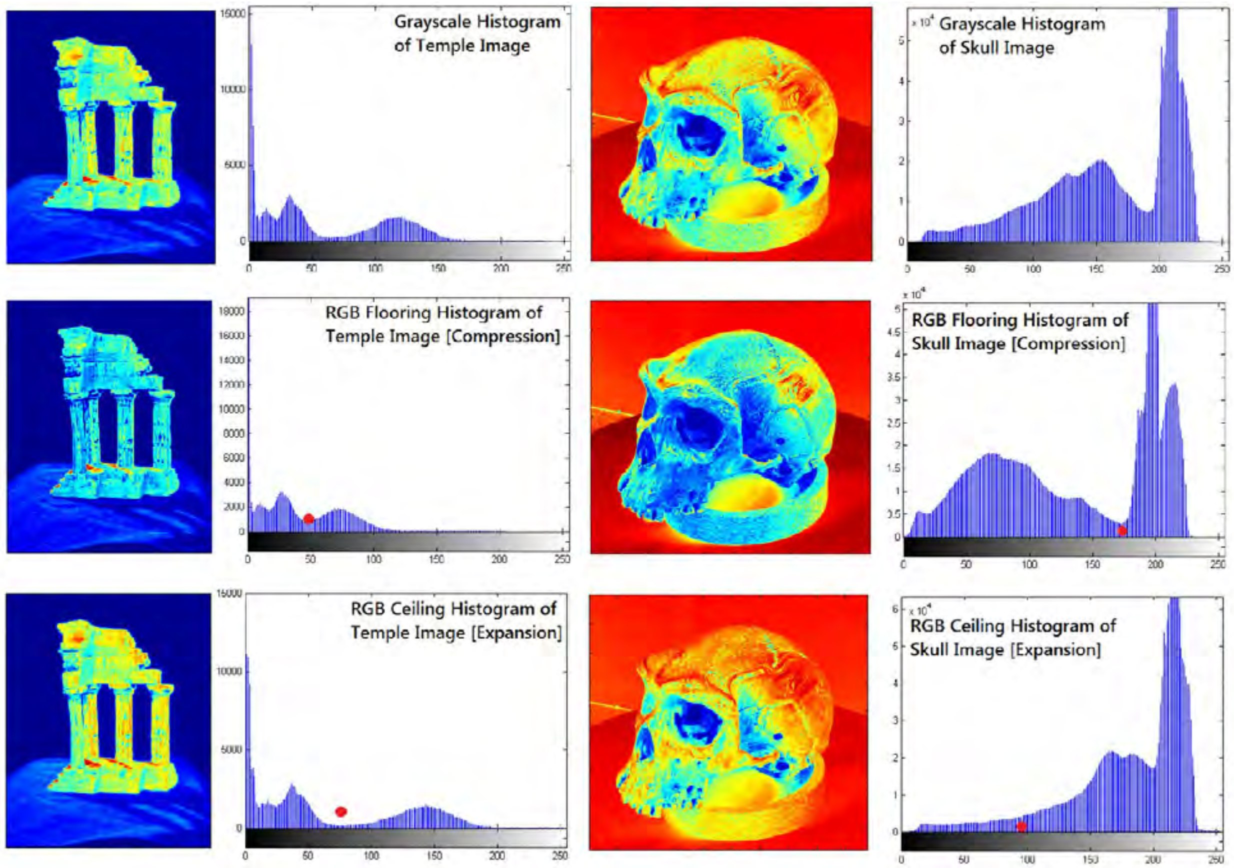

Figure 3: The images on the left were captured under nominally black background whereas on the right, were captured under nominally white background. The top-row shows a grayscale image and corresponding histogram of the Temple Ring and Skull Dataset. The middle-row is the RGB flooring and the bottom-row is the RGB ceiling grayscale transformation.

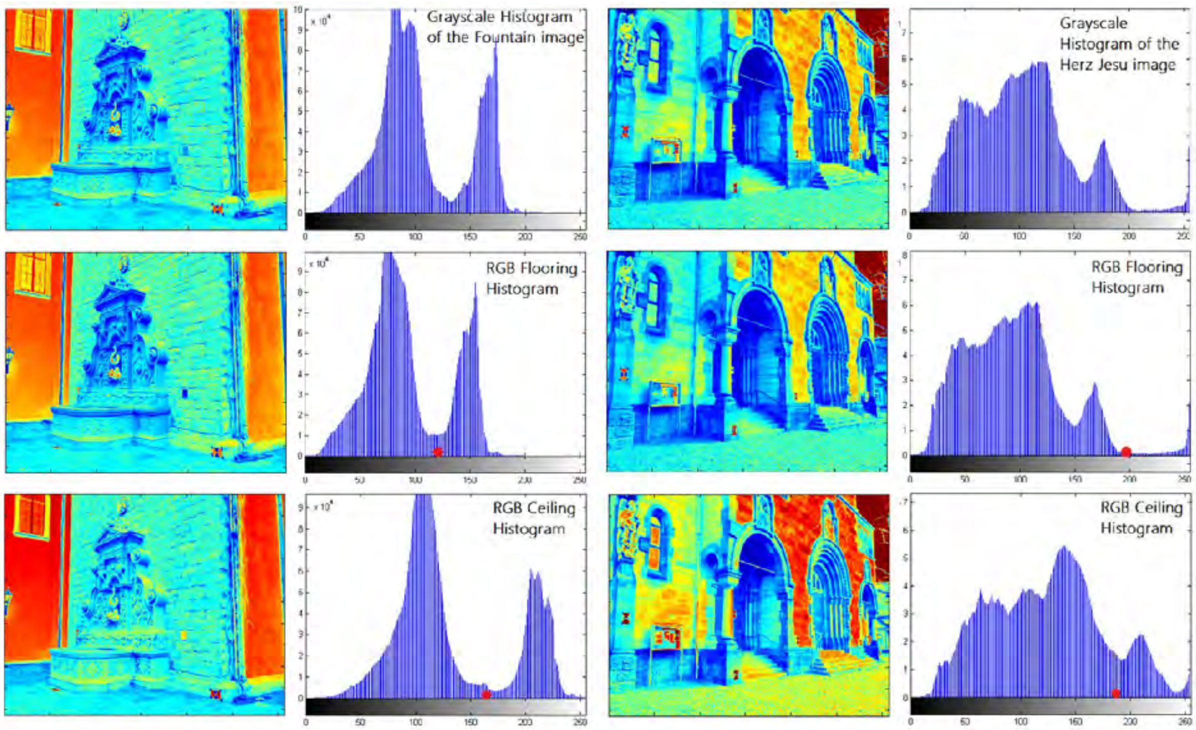

Figure 4: The images on the left and right are the Fountain and Herz-Jesu images under different grayscale transformations respectively. The top-row shows a grayscale image and corresponding histogram of the Fountain-P11 and Herz-Jesu-P8 Dataset. The middle-row is the RGB floor and the bottom-row is the RGB ceiling grayscale transformation.

2.1.5.1 Discussion of Results

Our main focus is on the initial two steps of the PMVS algorithm described in Algorithm 1. These steps are important in the initial processing stages of this reconstruction algorithm since the initial image features being detected and matched across the views are key ingredients to automated acquisition of geometry. Those image features which are correctly matched in the initial feature detection and matching stages, are used to seed the expansion and filtering stages. With a higher number of correctly matched patches, the following processing stages will expand and filter more accurately. Most of the background artefacts are removed and mainly foreground (object) data is retained to seed the expansion stage, which expands the initial set of matched patches to fill any neighbouring patch spaces (gaps). Expansion is bound to generate some incorrect patches. The subsequent filtering stage removes erroneous patches based on geometric and visibility constraints [22] of each patch. Table 4, shows the initial features that are detected and matched in six different cases. The measurement show a higher number of correctly matched patches in the pre-processed (masked) data and also sheds light on reconstruction efficiency in the three iterated stages of expansion and filtering. Most importantly, Table 4 reveals the total execution speedup for all the cases. The 3D reconstruction of the object models show that they are more complete for the pre-processed set of images, see Figure 6.

When using 2D boundary information, we noted an increase in the percentage of the correctly matched patches during the initial matching, i.e. step 2 of the algorithm, thus decreasing the chances of outliers being detected as image features in the initial stages, this is illustrated in Table 3. We therefore obtained fewer patches to check and verify, thus decreasing the processing time required for reconstruction. The initial image features are important to the entire reconstruction pipeline, so the more correct patches in the initial stages, the less incorrect patches are generated in the expansion stage, thus we obtain a smaller cumulative error.

2.1.6 Middlebury Evaluation Results

A quantitative evaluation was done on two object data-sets. We used *Temple Ring* and *Dino Ring* datasets containing 47 and 48 images respectively, taken as a ring of images around the object. Two sets of mesh models were submitted for evaluation, denoted as Khuboni-1 and Khuboni-2. These results were captured and compared against Furukawa's results in Table 2. Both of these mesh models of Khuboni-1 and Khuboni-2 were obtained using the same point-cloud model, thus leading to a conclusion that the results were directly influenced by the meshing algorithm that was used, as an example, Khuboni-1 results were obtained using Poisson Surface Reconstruction (PSR) algorithm on the Linux platform and whereas Khuboni-2 results were obtained using PSR algorithm on a windows platform, and on MeshLab, using the same parameter settings. The results obtained from Khuboni-1 and Khuboni-2 differ slightly from Furukawa-2, but our main intention was to show the speedup, i.e. lower time and space complexity that can be achieved with similar quality to original Furukawa-2. This has been achieved. Table 2 shows that relative to the Furukawa benchmark, we have achieved speedup ratios of about 11:1 and greater, and approximately 60% memory usage (40% reduction). There are noticeable improvements in the density of the patch model at the surface of the object and also at the eye sockets and nose, see Figure 6. The results are tabulated in terms of *accuracy* (distance *d* such that a percentage of the reconstructed model is within a distanced error from the ground truth model) and *completeness* (a percentage of the ground truth model that is within a specified error from the reconstructed model) [1].

A further evaluation was needed in order to understand how speedup was achieved. This was done by considering the patches being detected initially, correctly matched and finally reconstructed, see Table 3. For both datasets, *Temple* and *Dino*, we are able to restrict our image feature detection to foreground, and any features that might be detected outside or on the edge of the foreground are regarded as background artefacts. When there is no segmentation information used in the feature detection stage, we say that we attained 100 % of those image features, see Furukawa for Temple and Dino results in Table 3, but when that segmentation information is used to block out the background artefacts, we obtain 66 % of those image features that are detected by Furukawa, see Khuboni for Temple and Dino results in Table 3. We can deduce that 34% of 100% detected by Furukawa, will be eliminated or segmented out from the detection stage. The benefit of using segmentation information throughout the reconstruction pipeline is that we obtain a higher percentage of correctly matched patches among the initially detected features which lead to a denser patch reconstruction at the end of the pipeline, see Table 3. The final reconstructed patches are almost the same number of patches. This may be misleading, but the reconstructed patches by Furukawa, include the background reconstruction, whereas for Khuboni, the background reconstruction is excluded, see Figure 6. These background artefacts contribute to the number of erroneous patches and towards a larger cumulative error in the iterative expansion and filtering stages. By detecting foreground features, we were able to improve the patch matching algorithm that seeds to the iterative stages. This reduces the cumulative error, improves the accuracy and completeness of the final patch model. This can be seen in Table 3 and Table 4.

The evaluation of the scene reconstruction was done using the Multi-scale Model to Model Cloud Comparison (M3C2) algorithm [43] to measure the "goodness of fit" of the patch model with respect to the laser scanned ground truth mesh model; the results can be seen in Table 5. The M3C2 algorithm is part of a software package (Cloud Compare Ver. 2) [43] that was used for benchmarking the results which are shown in Table 5. A good fit of the reference cloud with respect to the relatively compared cloud should produce a smaller mean and a standard deviation. This means that the distances between the two clouds are smaller. The compared cloud is less prone to errors i.e. smaller mean and less noisy (smaller standard deviation). The reference cloud is the laser scanned ground truth mesh model, and we achieve results which are better than the baseline method, see Table 5. The parameter settings that were used during our experiment for the Temple and Dino Ring, Table 1:

Table 1: Parameter Settings for PMVS algorithm **[44]**

Parameter Setting	Denoted As	Initialized Value
Cell Size	csize	1
Window Template Size for NCC Matching	wsize	9
Pyramid Level (i.e. Image Scaling to Quarter, Half or Full Size Resolution)	level	0 (Full) By default 1 (Half) 2 (Quarter)
Minimum Number of Images used for Patch Generation	minImageNum	7
Template Matching Threshold	Threshold	0.85
Number of CPUs	CPU	2

Table 2: The image data-set and results captured here are from the Middlebury evaluation website **[2]** and the proposed method is compared against the benchmark patch-based method (PMVS) **[22]**.

Temple Ring	**Accuracy and Completeness**	**Total Execution Times in (H:M:S)**	**Speed-up Ratio**	**Memory Usage**
Furukawa - 2	0.55 mm / 99.1 %	6:02:40	1	100%
Khuboni – 1	0.59 mm / 98.6 %	0:25:52	14.02	~65%
Khuboni – 2	0.67 mm / 98.3 %	0:28:59	12.51	~65%
Dino Ring	**Accuracy and Completeness**	**Total Execution Times in (H:M:S)**	**Speed-up Ratio**	
Furukawa - 2	0.33 mm / 99.1 %	9:04:00	1	100%
Khuboni – 1	0.40 mm / 99.3 %	0:40:05	13.57	~60%
Khuboni – 2	0.38 mm / 99.5 %	0:47:28	11.46	~60%

Table 3: Patch Evaluation for Temple and Dinosaur Patch Models

Temple Ring	**2D boundary information**	**Detected Features # of Patches | %**	**Correctly Matched Patches**	**Final Patches**
Furukawa	No	108163 / 100 %	2466 / 2.280 %	239690 / 98.90 %
Khuboni	Yes – foreground only	71874 / 66.45 %	3079 / 4.284 %	198864 / 99.85 %
Dino Ring		**Detected Features # of Patches | %**	**Correctly Matched Patches**	**Final Patches**
Furukawa	No	137473 / 100 %	4463 / 3.246 %	280148 / 99.56 %
Khuboni	Yes – foreground only	91781 / 66.76 %	5635 / 6.140 %	255627 / 99.80 %

Table 4: Reconstruction Efficiency refers to the number of reconstructed patches after expansion and filtration stage at each iteration. There is significant improvement in the total execution times of different scaled patch models due to the fact that the background is not reconstructed, it is eliminated from the process.

	Full-Size Unmasked	**Full-Size Masked**	**Half-Size Unmasked**	**Half-Size Masked**	**Quarter-Size Unmasked**	**Quarter-Size Masked**
Detected Features	975367	565932	239126	149665	57916	40108
Correctly Matched Features	19448	16953	5278	4706	2041	1015
Iteration 1	182818	114734	17472	16719	11139	7080
Iteration 2	87478	46934	9315	8544	6042	3898
Iteration 3	93250	44382	8693	8163	5843	3725
Execution-Time	3961	2770	762	642	244	182

Table 5: The results captured here are the patch models compared against the provided 3D laser scanner ground truth models. Both the image data-sets and laser ground truth mesh models were provided by **[3]**. The Multi-scale Model to Model Cloud Comparison (M3C2) method that was used for evaluation is explained in detail in **[43]**. There is more patch variantion in the Furukawa's patch models compared to the proposed method.

	Furukawa Herz-Jesu	**Our Method Herz-Jesu**	**Furukawa Fountain**	**Our Method Fountain**
Gaussian Mean	0.018675	0.005758	0.009472	0.001574
Standard Deviation (Std. Dev.)	0.137654	0.164715	0.051030	0.040936
Average Distance from Laser Scan	-0.40 mm	-0.33 mm	-0.25 mm	-0.15 mm

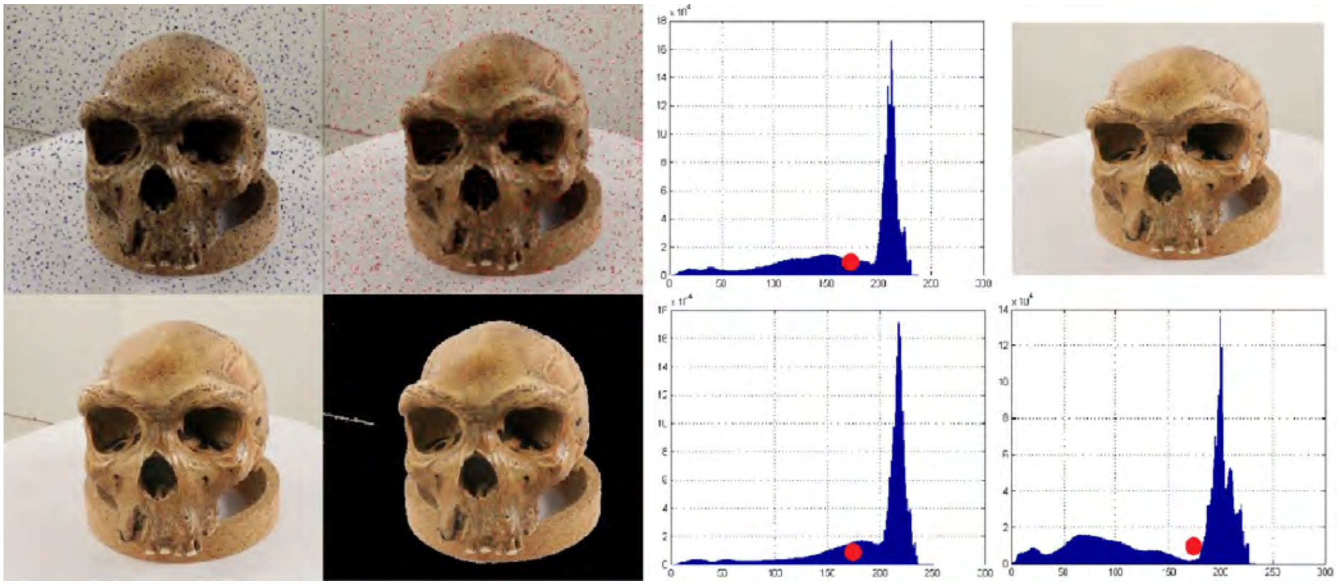

Figure 5: The images on the left show the image features that are detected by the PMVS along with the original and masked image, below. The images on the right show three different grayscale histograms with the original image and The red dot on the histogram represents the threshold value, T = 173.

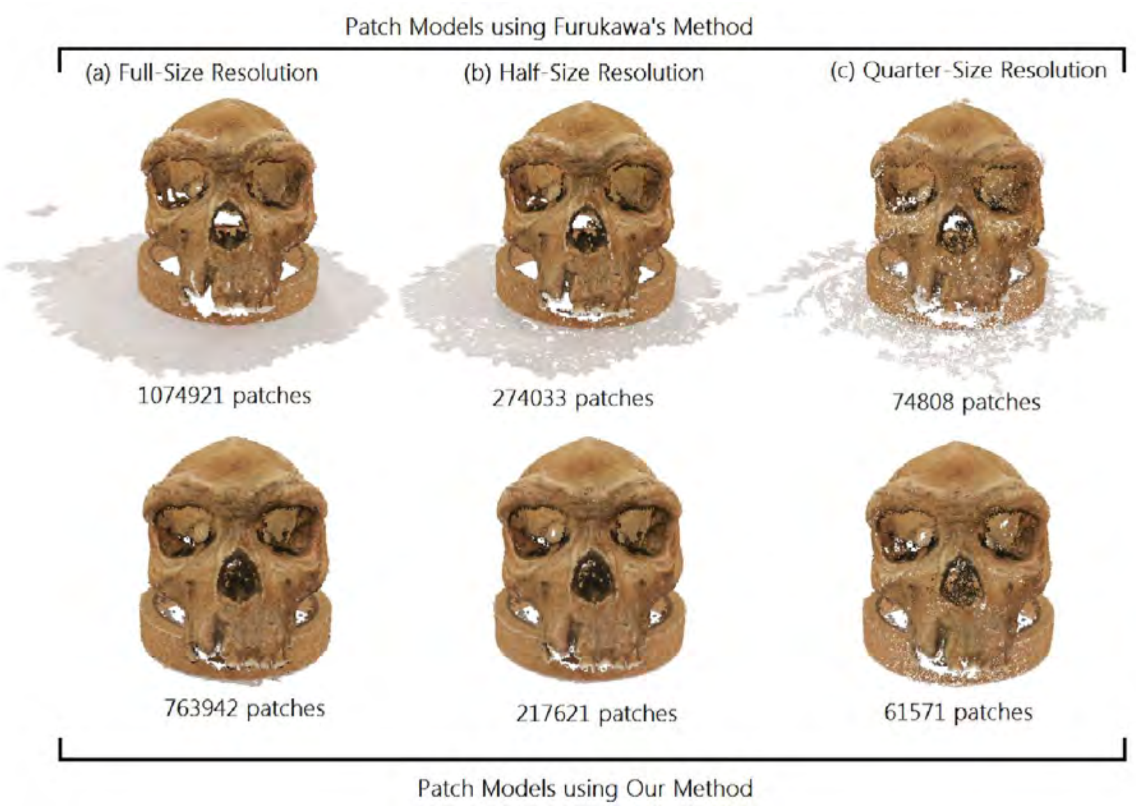

Figure 6: (a) Full-, (b) Half- and (c) Quarter-Size Image Resolution Reconstruction, the higher image resolution, the less prone to erroneous patches. The top row represented the baseline and the bottom row is the proposed method of reconstruction.

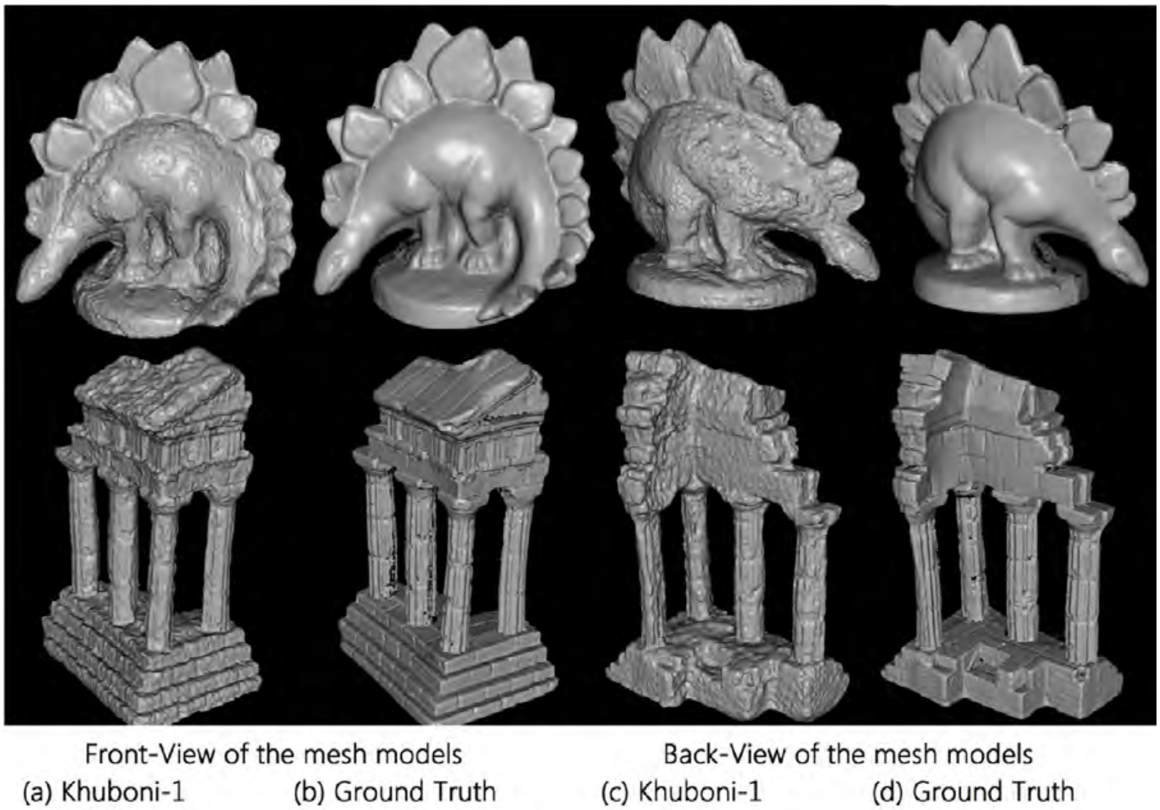

Figure 7: Our mesh models of the dinosaur and temple ring dataset, (a) and (c) were submitted for online evaluation to Middlebury website **[2]**, against the laser scanned ground truth, (b) and (d).

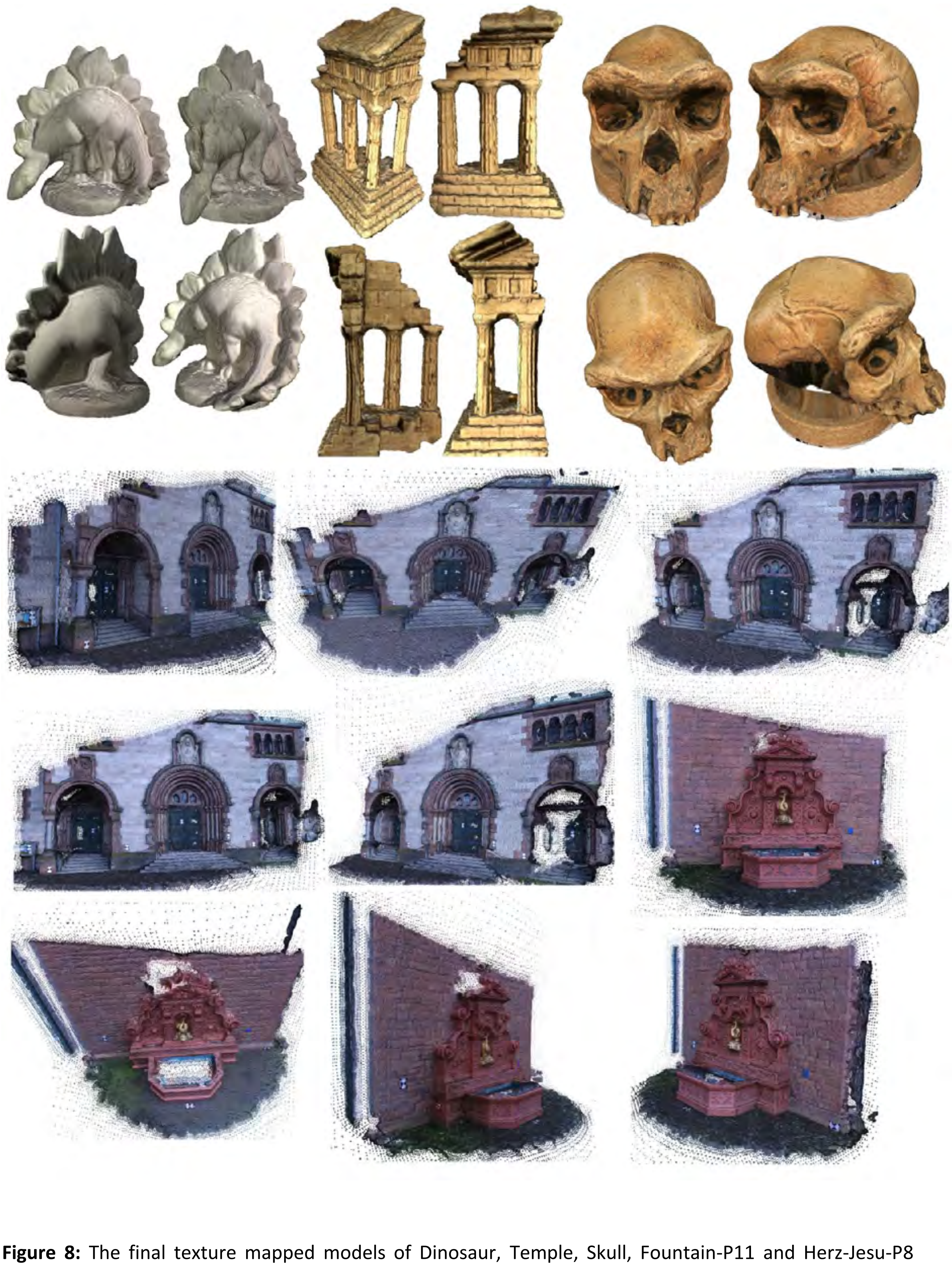

Figure 8: The final texture mapped models of Dinosaur, Temple, Skull, Fountain-P11 and Herz-Jesu-P8 datasets were acquired by using PSR meshing algorithm and texture mapping was sampled from their respective patch models.

2.1.7 Conclusion

The proposed image segmentation is a special case of segmentation that deals with the background removal, which is either nominally black or white, as in the Middlebury object data-sets [1]. This enforces the removal of most background artefacts from the final 3D reconstruction. The proposed method is time efficient and not complex to implement. It can be extended to real time segmentation for object detection or in robotics. The results show that most of the image details are retained with minimum information loss. This reduces the cumulative error of reconstructed patches, thus improving the performance of the 3D reconstruction in a shorter time interval.

The baseline reconstruction pipeline requires a visual hull which is reconstructed from the segmentation information and uses an iterative snapping algorithm to deform the visual hull model towards the set of reconstructed patches. Whereas in our proposed reconstruction pipeline, we used an adaptive threshold technique to create a 2D boundary to remove the background artefacts from the final patch model so that Poisson Surface Reconstruction (PSR) algorithm can be used to produce high quality mesh models. From our 2D boundary, we are able to constraint the patch model to a 3D patch boundary which is similar to visual hull methods. The 3D patch boundary restricts the patch model from expanding outside the bounds of the foreground. The patches that reside in 2D on edges are perfectly reconstructed from a different camera view or angle. Normally, when we use the normal reconstruction pipeline, the background artefacts of the object are unnecessarily included in the final patch model and further processing is needed to obtain the visual hull and iteratively deform the visual hull towards the set of patches. This normal process is time consuming that requires three processing stages, i.e. PMVS, Visual Hull and Iterative Snapping (IS) algorithm. Our proposed reconstruction pipeline is a two-stage process which requires the modified PMVS and PSR algorithm. This reduces time and space complexity of the reconstruction pipeline, and enabled us to replace the iterative snapping algorithm which constrained the technique for use on object datasets only. This improved pipeline that uses the PSR algorithm can be used for both object and scene image datasets to produce high-quality mesh models with similar quality to the original Furukawa-2.

PAPER 2 – STOCHASTIC QUASI-NEWTON PATCH OPTIMISATION FOR MULTI-VIEW STEREO MODEL

Submitted to Journal of IET Computer Vision (Under review)

Ray Khuboni & Bashan Naidoo

Abstract

This paper presents a classical quasi-Newton BFGS method with stochastic objectives that incorporate curvature information into the optimisation method. The BFGS method is modified to introduce stochastic gradient differences, whilst regularising the Hessian approximation matrix to ensure a well-conditioned matrix. The proposed method is employed to minimise a deterministic objective function that is founded upon a photometric discrepancy function. The discrepancy function is commonly used during patch optimisation to refine the 3D geometric orientation and depth information of newly generated patches with respect to its visibility set of views. We redefine the photometric discrepancy function to incorporate a specially developed feature space that addresses the problem of specular highlights in an object or a scene dataset. Due to this modification, we are able to incorporate curvature information of those patches which were deemed to be depleted in the refinement process due to their low correlation scores. With those patches now contributing towards the refinement algorithm, we are able to more accurately represent the surface of the reconstructed object or scene. The new feature space is also used to promote feature detections. With this method, we notice reductions in the cumulative error and obtain results that are denser and more complete than the baseline reconstruction.

---

1. *Ray Khuboni*
School of Engineering, University of KwaZulu-Natal – Howard College, South Africa, Durban, 4041, *207502748@stu.ukzn.ac.za and Tel. (+2773) 913 2867*

2. *Bashan Naidoo*
School of Engineering, University of KwaZulu-Natal – Howard College, South Africa, Durban, 4041, *bnaido@ukzn.ac.za and Tel. (+2731) 260 2758*

2.2.1 Introduction

The recent advancements in a photographic technology have enabled us to capture the high resolution images from ordinary digital cameras for a range of applications. With the Multi-View Stereopsis (MVS), we may easily generate three-dimensional structures from the images for viewing, preservation or value-added image-based applications. According to a survey by *Seitz et al.* [2], the MVS algorithms are able to attain good accuracy from the low resolution images. With MVS, the acquisition of geometry is possible from multiple-views in both scene and object datasets. The industrial application of image-based 3D reconstruction may be found, *inter alia,* in media, entertainment, scientific and engineering works. According to *Seitz et al.* [2], there are four categories of MVS algorithms:

a) *Voxel Methods* which require the use of a known bounding box [4],
b) *Polygonal Meshing Methods* that use a visual hull to initiate reconstruction [6],
c) *Multiple Depth Map Fusion Methods* that fuse individual depth maps into a single model [20], [33], and,
d) *Sparse Dense Patch Methods* which do not require volume initialisation or a bounding box [31], [32].

The *sparse point based method* proposed by M. Lhuillier [29] and the related *patch based methods* such as Patch-Based Multi-view Stereopsis (PMVS) [22], [31], fall into the latter category. The quantitative comparison and evaluations provided by the Middlebury survey [2], show that the *sparse dense patch methods* still produce the best overall performance up to this date, for all six datasets, in both accuracy and completeness.

Despite the generally good performance of PMVS, our study shows that its patch models are still subject to limitations with regard to the accuracy and completeness of the reconstructions. This paper proposes improvements to the patch optimisation algorithm and its photometric discrepancy function. The patch optimisation stage of the PMVS algorithm seeks to optimise the orientation and depth of reconstructed patches in 3D space. This has a significant impact on subsequent surface reconstruction. The patch optimisation was modified to incorporate stochastic objectives within a classical quasi-Newton approach. The photometric discrepancy function is also applied to data that is transformed to a specially developed feature space designed to address the problem of specular highlights in an object or a scene dataset.

Due to these modifications, we are able to incorporate curvature information of those patches that are deemed to be depleted in the refinement process of PMVS due to their low correlation scores. With those patches now contributing towards the refinement algorithm, we are able to represent the surface of the reconstructed object or scene more accurately. The current optimisation method excludes the use of curvature information from the estimation of patch orientation and depth. With the proposed method, we can reduce the number of erroneous patches that enter into the iterative stages of the PMVS algorithm by better estimating orientation and depth. Given that fewer erroneous patches are cycled through the iterative stages of PMVS, a more accurate and complete reconstruction becomes possible.

Although the application of the proposed method to crowded scene reconstruction has not been fully investigated, the results obtained using scene datasets from *Strech et al.* [3] are very promising. The modified PMVS approach is applicable to both object and scene reconstruction. The crowded scene datasets, such as Venice, Dubrovnik, Colosseum and Rome image sets [33], [34], [35], [36], can be used to further develop and test-generalised versions of 3D reconstruction methods with specular highlights, wide baselines, and a strongly convex nature.

We also propose a specially developed feature space to be used in the deterministic objective function of the optimisation process. This feature space facilitates the detection of image features where it is normally deemed too difficult given specular highlights for example. The feature space comprises of three channels that include the luminance L-channel from the CIELAB colour space, and the RGB ceiling and floor values of each pixel

[45]. The feature space therefore transforms the RGB information to three grayscale channels, i.e. Luminance, RGB ceiling and RGB floor. The RGB ceiling and floor operators are defined in [45].

This paper is organised into six sections:

1. An introductory section that is presented to introduce the scope of this paper.
2. An overview of the patch-based method that is used as the foundation of this paper is presented.
3. An overview of the stochastic quasi-newton method is presented, which is used in the patch optimisation of the patch-based method.
4. The proposed methodology is presented and describes its novelty towards the patch-based method.
5. The experimental results are presented and explained in detail to support the proposed work.
6. The final section concludes the findings and thoughts on the paper and the proposed work.

2.2.2 Patch-based Multi-View Stereo (PMVS) Algorithm

The *Patch-based Method* is decomposed into four stages:

a) *Image feature extraction* through the use of Harris and Difference of Gaussian (DoG) operators,
b) *Matching of features* across images (stereopsis) to yield initial seed patches,
c) *Iterative expansion* of the patch set into adjacent spaces to obtain a denser patch reconstruction, and,
d) *Iterative filtration* of erroneous patches generated in the previous step through the use of visibility constraints.

The *Patch-based Method* reconstructs a set of oriented and textured rectangular patches that represent the surface of an object or a scene in 3D space. A post-processing stage may be added to convert the reconstruction into a *polygonal-mesh model* for image-based applications [22], [31], [32].

The *Patch-based Method* employs a two-stage process to reconstruct a new patch p in 3D space. Firstly, the initialisation of parameters for each patch p, including the central point, the unit normal vector, visibility set (set of images in which this patch is visible), and a reference image from the visibility set are obtained during 3D triangulation using a multiple views of the patch. The second stage optimises parameters of patch p using a numerical conjugate gradient method, i.e. Nelder-Mead Simplex Method [40], [41], to minimise an objective function. This improves geometric orientation and depth information of the initial patch. The objective function is the photometric discrepancy function, which is computed as an average of pair-wise photometric discrepancy scores over the visibility set of patch p. The photometric discrepancy score is defined as one minus the Normalized Cross-Correlation (NCC) score between two sampled textures,

$$q_1 \;=\; q\big(p, I_1\big(R(i,j), G(i,j), B(i,j)\big) \,|\, na \le i \le nb, ma \le j \le mb\big) \qquad (5)$$

$$q_2 \;=\; q\big(p, I_2\big(R(i,j), G(i,j), B(i,j)\big) \,|\, nc \le i \le nd, mc \le j \le md\big) \qquad (6)$$

Where q_1 and q_2 are sampled RGB pixel values of patch p from image 1 and 2 with patch resolution of $n * m$. The Photometric Discrepancy Function (PDF) is defined as:

$$g(p) = \frac{1}{|V^*(p) \backslash R(p)|} \sum_{I_i \in [\, V^*(p) \backslash R(p)]} \left\{ 1 - \frac{\sum_{i=1}^{n*m} \sum_{j=1}^{n*m} q_1(i) q_2(j)}{\left[\sum_{i=1}^{n*m} q_1(i)\right]^{1/2} * \left[\sum_{j=1}^{n*m} q_2(j)\right]^{1/2}} \right\} \qquad (7)$$

Where: $V^*(p)$ is the visibility set of patches which correspond to the reference patch, $R(p)$ is a patch from the reference image and $|V^*(p) \backslash R(p)|$ is the number of accurately visible patches that correspond to the reference patch.

After the optimisation stage, the visibility set of images for the optimised patch p is updated to correspond with the new coordinate parameters. If the updated visibility set is greater than the minimum number of images required with low photometric discrepancy score, the patch is deemed a success. Otherwise it is discarded. After the initial feature matching, iterative expansion deploys new patches in the pixel spaces around seed patches to create a denser patch model. The patch refinement stage is also used to optimise the newly expanded patches. Thereafter, iterative filtration is used to remove any erroneous patches that might have been included in the iterative expansion stage. The expansion and filtration stages are iterated 3 times in all experiments [22], [31], [32]. The parameter n was chosen to be 3 in order to replicate the patch model results obtained using the Patch-based MVS algorithm. For *n* less than 3, the patch model became less dense whereas for *n* greater than 3, the patch model was prone to more erroneous patches. In the following sections, the term "objective function" may be used numerously, and in this paper it refers to the photometric discrepancy function.

2.2.3 Stochastic Quasi-Newton (SQN) Method

The benefit of quasi-Newton methods for deterministic optimisation lies in the fact that the gradient of the deterministic function or curvature information is supplied at each iteration of the optimisation process. By measuring the changes in gradients, a model of the deterministic function can be constructed at a super-linear rate of convergence [46].

The Stochastic Gradient Descent (SGD) methods are used to solve the optimisation problems of a deterministic function over a defined range of values where the expectation of deterministic function is defined by a set of random functions [46], [47]. In the classical quasi-Newton BFGS method for minimising a deterministic function, the BFGS method attempts to find the optimal point of convergence v, i.e. expected value of the deterministic function, where the first derivative of the objective function is close or equal to zero. At each iteration, the curvature information is incorporated in the gradient descend direction by using the Hessian approximation matrix which is determined by its previous approximated matrix. The Hessian approximation matrix is an approximation of a Hessian matrix which is a *second-order* partial derivative of the objective function. The Hessian matrix describes the local tangential curvature of a deterministic function with multiple variables [40]. We incorporate the curvature information of the BFGS updates within the stochastic optimisation method in order to obtain a stochastic quasi-Newton BFGS method.

The objective function $F(v)$ that is used in the optimisation algorithm is the same as the one defined in Section 2. We assume that the random function $g(v,\theta)$ of any optimisation problem is strongly convex thus making the photometric discrepancy to function strongly convex. The random samples of θ are taken from a set of convex values around the point v of the objective function [46], [47]. We then use gradient descent algorithms for the purpose of finding an optimal argument $v^* = argmin\big(F(v)\big)$, (i.e. refined 3D coordinates of patch p). We begin by computing the gradients of the random function $g(v,\theta)$ which is given by $s(v)$,

$$s(v) = \nabla_v F(v) = E_\theta[\nabla_v g(v,\theta)] \tag{8}$$

Where, $E_\theta[\nabla_v g(v,\theta)]$ is the average of the random gradients and $\nabla_v F(v)$ is the gradient of the deterministic function. Normally, when one is dealing with a large number of functions $g(v,\theta)$, the exact solution of the gradients $s(v)$ is impractical. This problem is avoided through the use of stochastic gradients over L random variables. These random variables from the distribution θ are drawn independently. We define the gradient of the stochastic method at the point v given $\tilde{\theta} = [\theta_1, \dots, \theta_L]$ as

$$\hat{s}\big(v,\tilde{\theta}\big) = \frac{1}{L}\sum_{i=1}^{L}(\nabla_v g(v,\theta_i)) \tag{9}$$

In order to compute the gradient of stochastic method $\hat{s}(v, \tilde{\theta})$, we find the gradient of the random function $g(v, \theta)$ for each θ_i component of $\tilde{\theta}$ and compute their average at manageable computational cost, i.e. small number of samples, L samples, L = 5 in our experiment. The stochastic gradient $\hat{s}(v, \tilde{\theta})$ is regarded as an unbiased estimate of the average gradient $s(v)$ [46], [47]. Now that we have defined the stochastic gradient, we can compute the iterative update for v, as

$$v_{t+1} = v_t - \epsilon_t \hat{s}(v_t, \tilde{\theta}_t) \tag{10}$$

Where t is a time index, v_1 is the initial iterate and ϵ_t is a step size sequence. Empirical results show that iteration of v_{t+1} descends along a negative gradient direction. As we gradually iterate through the time index, the stochastic gradient should converge at the optimal argument v^*. We can choose a customary step size to gradually decrease as we reach our point of convergence $\epsilon_t = \epsilon_0 \tau/(\tau + t)$, for the given parameters ϵ_0 and τ that controls the initial step size and its speed of decrease, respectively (in our experiment $\epsilon_0 = 0.01$ and $\tau = 1000$).

If the stochastic iterative update for v_{t+1} is used, the number of iteration that would be required to reach the point of convergence is large. We therefore address this problem by using quasi-Newton BFGS methods; whereby the gradient descent direction can be pre-multiplied by an inverse Hessian approximation matrix $\hat{B}_t^{-1}$. From BFGS method, the gradient descent direction d_t is defined by solving $B_t\, d_t = -s(v_t)$ and the iterative updates are given by [40].

$$v_{t+1} = v_t + \epsilon_t\, d_t \tag{11}$$

$$v_{t+1} = v_t - \epsilon_t \hat{B}_t^{-1} s(v_t) \tag{12}$$

We also add a regularisation term ΓI for gradient descent to guarantee convergence through positive definiteness of the pre-multiplier of the stochastic gradient. The optimisation trajectory descends along the direction $(\hat{B}_t^{-1} + \Gamma I)\hat{s}(v_t, \tilde{\theta}_t)$ which is moderated by step size ϵ_t (in our experiments, $\Gamma = 0.0001$ and I is an identity matrix),

$$v_{t+1} = v_t - \epsilon_t(\hat{B}_t^{-1} + \Gamma I)\hat{s}(v_t, \tilde{\theta}_t). \tag{13}$$

The main idea of selecting the matrices B_t is to closely approximate the Hessian $H(p_t) = \nabla^2 F(v_t)$ of our deterministic function. There are many different methods which are used to determine the matrices B_t and they work well, such as the BFGS method [40], [48], [49].

In BFGS method, the curvature information of the deterministic function is approximated by a finite difference. We seek to define our variable (search parameter) along with its gradient at time t. We know that by selecting the matrix B_{t+1} to be used in the next iteration, we need to satisfy the stochastic secant condition $B_{t+1}y_t = r_t$, where y_t and r_t are search parameter variation and stochastic gradient variation respectively; and that the next BFGS curvature estimate B_{t+1} can be obtained as a function of previous approximated matrix B_t. For each iteration of v_t and v_{t+1}, their corresponding gradients $s(v_t)$ and $s(v_{t+1})$ could be replaced by the gradients of the stochastic method $\hat{s}(v_t, \tilde{\theta}_t)$ in the approximation matrices and iterative descend updates. The substitution gives rise to a stochastic BFGS algorithm [46], [47], [50]. We can define the search parameter variation at time t as

$$y_t = v_{t+1} - v_t \tag{14}$$

And the stochastic gradient variation along with another regularisation term as

$$\tilde{r}_t = \hat{s}(v_{t+1}, \tilde{\theta}_t) - \hat{s}(v_t, \tilde{\theta}_t) - \delta y_t \tag{15}$$

The approximated Hessian matrix $\hat{B}_{t+1}$ for the next iteration can be defined as ($\delta = 0.0001$ and I is an identity matrix).

$$\hat{B}_{t+1} = \hat{B}_t + \frac{\tilde{r}_t \tilde{r}_t^{\,T}}{y_t^{\,T} \tilde{r}_t} - \frac{\hat{B}_t y_t y_t^{\,T} \hat{B}_t}{y_t^{\,T} \hat{B}_t y_t} - \delta I \tag{16}$$

2.2.4 Proposed Method

In many computer vision applications that require the determination of curvature information, some sort of a numerical optimisation is required to minimise a deterministic objective function. In this paper, a classical quasi-Newton BFGS method with stochastic objectives is proposed, see Algorithm 3, where the curvature information is incorporated into the stochastic optimisation. We modify the classical BFGS method to introduce stochastic gradient differences whilst regularising the Hessian approximation matrix. The proposed method is employed to solve our deterministic objective function which is described by a photometric discrepancy function. The photometric discrepancy function is simply the average of photometric discrepancy scores over the visibility set of images for that particular patch. The photometric discrepancy score is defined as one minus the Normalized Cross-Correlation score of two sampled patches, as described in Section 2 of this paper. The photometric discrepancy function is used to optimise the 3D geometric orientation and depth information of each reconstructed patch with respect to its visible set of images, in the initial seeding and iterative expansion stage. We redefine the photometric discrepancy function to incorporate a specially-developed feature space in order to address the problem of illumination changes in an object or a scene dataset. Due to this modification, we are able to incorporate curvature information of those patches which were deemed to be depleted in the refinement process due to their low correlation scores. With those patches contributing towards the refinement algorithm, we are able to accurately represent the surface of the reconstructed object or scene, see Section 5. The specially developed feature space is also used in the feature detection stage to promote the detection of more image features which can be noticed in Table 7.

Algorithm 3: Stochastic Quasi-Newton BFGS Patch Optimisation Algorithm

Input: A sparse set of 3D triangulated patches

Output: An optimised sparse set of 3D patches

Redefining the objective function to use a new feature space vector:

The deterministic objective function is defined by a Photometric Discrepancy Function (PDF), sect. 2.

Read an image $I(x,y) = \{R(i,j), G(i,j), B(i,j) \mid 1 \le i \le n, 1 \le j \le m\}$

and the image size: $[m,\ n,\ 3]$

Luminance Channel:

$$L(i,j) = CIELAB(I(i,j))$$

The following grayscale image transforms are clearly defined in [45]

RGB floor transform:

$$V_{\min}(i,j) = floor(I(i,j))$$

RGB ceiling transform:

$$V_{\max}(i,j) = ceil(I(i,j))$$

The specially developed feature space is defined as a vector of three grayscale channels that are mentioned above. Instead of using the standard RGB feature space vector, the feature space vector is now redefine as a combination of luminance channel, RGB floor transform and RGB ceiling transform:

$I^{*}(x,y) = \{L(i,j),\ V_{\min}(i,j), V_{\max}(i,j) \mid 1 \le i \le n, 1 \le j \le m\}$ and the image size: $[m,\ n,\ 3]$

Now, the sampled pixel data is: (from equation 5 and 6)

$$q_1{}^{*} = q\big(p, I_1\big(\boldsymbol{L}(\boldsymbol{i},\boldsymbol{j}), \boldsymbol{V}_{\mathbf{min}}(\boldsymbol{i},\boldsymbol{j}), \boldsymbol{V}_{\mathbf{max}}(\boldsymbol{i},\boldsymbol{j})\big) \mid na \le i \le nb, ma \le j \le mb\big)$$

$$q_2{}^{*} = q\big(p, I_2\big(\boldsymbol{L}(\boldsymbol{i},\boldsymbol{j}), \boldsymbol{V}_{\mathbf{min}}(\boldsymbol{i},\boldsymbol{j}), \boldsymbol{V}_{\mathbf{max}}(\boldsymbol{i},\boldsymbol{j})\big) \mid nc \le i \le nd, mc \le j \le md\big)$$

The Photometric Discrepancy Function (PDF) is now defined as: (from equation 7)

$$g(p) = \frac{1}{|V^{*}(p)\backslash R(p)|} \sum_{I \in [V^{*}(p)\backslash R(p)]} 1 - \frac{\sum_{i=1}^{n*m} \sum_{j=1}^{n*m} q_1{}^{*}(i) q_2{}^{*}(j)}{\left[\sum_{i=1}^{n*m} q_1{}^{*}(i)\right]^{1/2} * \left[\sum_{j=1}^{n*m} q_2{}^{*}(j)\right]^{1/2}}$$

Pre-Processing: Collecting all images which are within an angle between 5 and 60 degrees. Selecting a reference image $R(p)$ and sorting the rest of the images based on their Normalized Cross Correlation (NCC) discrepancy score.

Stochastic BFGS Patch Optimisation:

For each patch $p = [v(1), v(2), v(3)]$, Hessian approximation matrix, $B(p)$ is computed to determine the new 3D geometric orientation of the patch p w.r.t. the visible set of images. The do … while loop will continue minimising the photometric discrepancy function until convergence is reached or when the number of maximum iteration is reached. The number of iteration is set to 100.

Do…

Acquiring L independent gradient samples around the point, θ_t for the stochastic gradient computation

$$s(p, \theta_t) = [s(p, \theta_{t,1}), s(p, \theta_{t,2}), \dots, \quad s(p, \theta_{t,L})]$$

Compute the first stochastic gradient at time t

$$\hat{s}(p_t, \theta_t) = \frac{1}{L}\sum_{i=1}^{L}(s(p_t, \theta_{t,i}))$$

The gradient function, $\hat{s}(p_t, \theta_t)$ is computed as an average of the stochastic gradients, to account for even small gradient variations. The stochastic gradient thus descends along the direction of $(B_t^{-1} + \Gamma I)\hat{s}(p_t, \theta_t)$, with a pre-multiplied step size of ϵ_t. The newly minimised or optimised 3D coordinate point of the patch p^* is given by the next iteration $p_{t+1} = [v^*(1), v^*(2), v^*(3)]$. Descending from the initial point of the patch $p = p_t = [v(1), v(2), v(3)]$.

$$p_{t+1} = p_t - \epsilon_t(\hat{B}_t^{-1} + \Gamma I)\hat{s}(p_t, \tilde{\theta}_t)$$

Compute the second stochastic gradient at time $t + 1$

$$\hat{s}(p_{t+1}, \theta_t) = \frac{1}{L}\sum_{i=1}^{L}(s(p_{t+1}, \theta_{t,i}))$$

The second stochastic gradient is computed to monitor the stochastic gradient variation. The variable variation is computed to be used in the pre-multiplier approximation matrix.

$$y_t = p_{t+1} - p_t$$

The modified stochastic gradient variation is computed to determine the next pre-multiplier approximation matrix for the next gradient descend direction. The stochastic gradient variation is used to determine the point of convergence for the minimised or optimised patch coordinates.

$$\tilde{r}_t = \hat{s}(p_{t+1}, \theta_t) - \hat{s}(p_t, \theta_t) - \delta y_t$$

The approximation matrix of Hessian is updated

$$\hat{B}_{t+1} = \hat{B}_t + \frac{\tilde{r}_t \tilde{r}_t^{\,T}}{y_t^{\,T} \tilde{r}_t} - \frac{\hat{B}_t y_t y_t^{\,T} \hat{B}_t}{y_t^{\,T} \hat{B}_t y_t} - \delta I$$

The while loop will terminate once convergence is reached or the number of iterations reaches its predefined limit.

While (convergence == success || iteration < times)

Post-Processing: Collecting all new images for newly optimised patch p. The newly optimised patch p is verified to meet the three constraining requirements which are the angle, photo-metric discrepancy score and number of images a patch is visible in.

2.2.5 Experimental Results

The performance evaluation of our proposed method was conducted on the Fountain-P11 and Herz-Jesu-P8 [3] datasets. All our experiments were conducted on an Intel® Core™ 2 Duo workstation with a running processor of 2.93GHz. As a post-processing step, a meshing algorithm was used to turn our patch models into surface polygonal mesh models for image-based modelling applications. The meshing algorithm that was considered for the post-processing is the Poisson Surface Reconstruction (PSR) [21]. PSR is known to produce high quality mesh models for both the object and scene datasets. The algorithm directly converts a set of vertices, i.e. our oriented patches, into a triangulated mesh model. The quality of the mesh model is directly related to the density of patch model. The denser the patch model, the finer the mesh model becomes. This can be seen from our texture mapped mesh models as shown in Figure 11 and Figure 12.

2.2.6 Discussion of Results

Our main focus is on the patch optimisation step of the *Patch-based Method* where we can improve the geometric orientation and depth information of every newly-generated patch. The optimisation step is utilised in the initial seeding and iterative stages, i.e. patch expansion and filtration stages, of the *Patch-based Method*. These steps are important to the reconstruction algorithm since the optimisation of every patch defines the three-dimensional (3D) surface of reconstructed object or scene, and it is the key ingredient to automated geometry acquisition. The image features that are correctly matched in the initial feature detection and matching stages are used to seed the expansion and filtering stages. From the results of Table 7, we note a higher number of detected image features along with a higher number of correctly matched patches. Both are due to the modifications of the feature detection and patch optimisation stages. The use of the specially developed feature space in the feature detection stage, allowed for more image features to be detected across all views. The results shown in Table 7, show that with almost double the number of image features detected, we can obtain five times more correctly matched patches for Fountain dataset and six times more for the Herz-Jesu dataset. We also obtain more than double the patch density at a varying cost to execution time. The Fountain reconstruction slowed significantly (244%) while the Herz-Jesu reconstruction slowed only marginally (20% slower), with both yielding a significant increase in reconstruction density, see Table 7. This variation has not been fully investigated. With the use of stochastic gradients incorporated in the optimisation method, we are able to address convex problems that might arise in the patch reconstruction. During the iterative expansion stage, the initial seed patches are used to expand to any neighbouring patch spaces (gaps) in order to fill up the sparse patch model, and are optimised using the proposed approach. The expansion is known to generate some incorrect patches. The subsequent filtering stage removes erroneous patches based on geometric and visibility constraints [22] of each patch. The proposed method ensures that the seed patches are optimised to guarantee fewer incorrect patches in the iterative stages. Thus we get more correctly optimised patches to be expanded, and fewer to be filtered out. This results in smaller cumulative error and denser patch models; see Table 7.

The evaluation of the scene reconstruction was done using the Multi-scale Model to Model Cloud Comparison (M3C2) algorithm [43] to measure the "goodness of fit" of the patch model with respect to the laser scanned ground truth mesh model. The M3C2 algorithm is part of a software package (Cloud Compare Ver. 2) [43] that we used as our benchmarking software to produce the results which are shown in Figure 9 and Figure 10. The M3C2 distance algorithm refers to vertices of the reference point cloud as core points, i.e. the green distribution; and the second (estimated) point cloud is considered to be the compared cloud [43]. The compared cloud produces a Gaussian distribution, i.e. the gray distribution. A good fit for the reference cloud, measured relatively to the compared cloud will produce a small mean and standard deviation for the gray distribution. Furthermore, with a good fit, we expect the gray distribution to closely approximate the peak of the green distribution. The unit normal vectors of the laser scan ground truth vertices are used in the computations of M3C2 distance which is the computation of the distance between the two-point clouds. In Figure 9, we noticed that our results are closer to zero distance as compared to the frame of reference, whereas in Figure 10, we get an indication of how far our patch distributions from the laser scan vertices distributions. In both results, our patch distributions are closer than the relative baseline patch distributions.

In Figure 9, the estimated patch model is used as the frame of reference, and the goodness of fit of the laser scan is evaluated. In Figure 10, we evaluate the opposite case by using the laser scan as the frame of reference and the goodness of fit of the patch model is measured. Given that the laser scan and patch model have a completely different point distributions and different numbers of vertices, the two measures will not be identical and reversible, and therefore both must be computed. With clear visualisation of the gray distributions, we can see that our results show a better distribution fit with smaller standard deviations and means. In Figure 10, we perform the opposite evaluation, where the frame of reference is now a laser scanned model and we still get smaller standard deviations and means. This means that the distance from the reference model are smaller, i.e. less prone to errors, that is indicated by the smaller means and less noisy which is indicated by the smaller standard deviation. Despite the main fact that we have generated more patches than the baseline method, the additional patches did not contribute to error or noise of the reconstructed model. Thus we can say for the image datasets [3] that were used in this paper, the results show a better patch distribution than the baseline method, see Table 7, Figure 9 and Figure 10. Further tests may be needed to be done on a bigger or more representative dataset such as in [32], [33], to validate the results.

Since the quality of the mesh model is directly proportional to the density of patch model [21], we can clearly see that our patch models are more complete, see Figure 11 and Figure 12, and accurate than the baseline method, see Figure 9 and Figure 10. The parameter settings that were used during our experiment for the Fountain-P11 and Herz-Jesu-P8 reconstructions are shown in Table 6.

Table 6: Parameter Settings for PMVS algorithm **[44]**

Parameter Setting	Denoted As	Initialized Value
Cell Size	csize	2
Window Template Size for NCC Matching	wsize	7
Pyramid Level (i.e. Image Scaling to Quarter, Half or Full Size Resolution)	level	0 (Full) By default 1 (Half) 2 (Quarter)
Minimum Number of Images used for Patch Generation	minImageNum	3
Template Matching Threshold	Threshold	0.7
Number of CPUs	CPU	2

Table 7: Patch Evaluation for Fountain and Herz-Jesu Patch Models

Fountain P11	**Detected Features # of Patches**	**Correctly Matched Patches**	**Final Patches**	**Total Execution Times [s]**
Furukawa	656,932	27,821	1,409,416	4,395
Khuboni	1,517,816	162,849	3,264,535	15,097
Khuboni : Furukawa	2.31:1	5.85:1	2.32:1	3.44:1
Herz-Jesu P8				
Furukawa	504,100	18,637	1,113,559	9,671
Khuboni	1,873,245	126,448	2,426,879	11,721
Khuboni : Furukawa	3.72:1	6.78:1	2.18:1	1.21:1

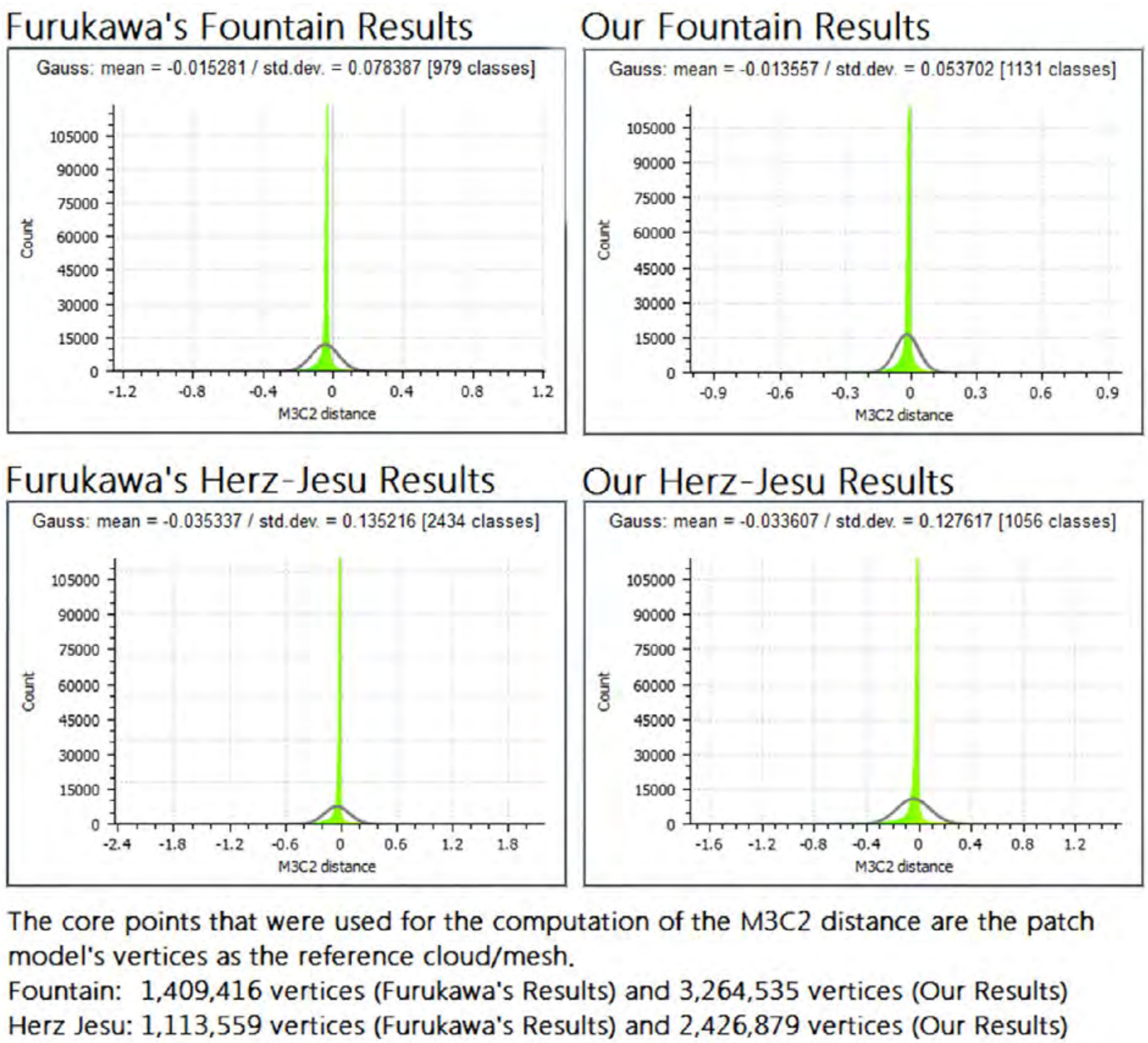

Figure 9: The core points are defined by the reference point cloud (i.e. vertices of the patch model), the green distribution and compared point cloud (i.e. vertices of the ground truth model), the gray distribution. The ground truth distribution fits better for our patch model results when compared to the baseline results. With a smaller mean and standard deviation of the ground truth distribution, we obtain better patch model results which are less prone errors and noise. These results are generated by our benchmark software, Cloud Compare Ver. 2 **[43]**.

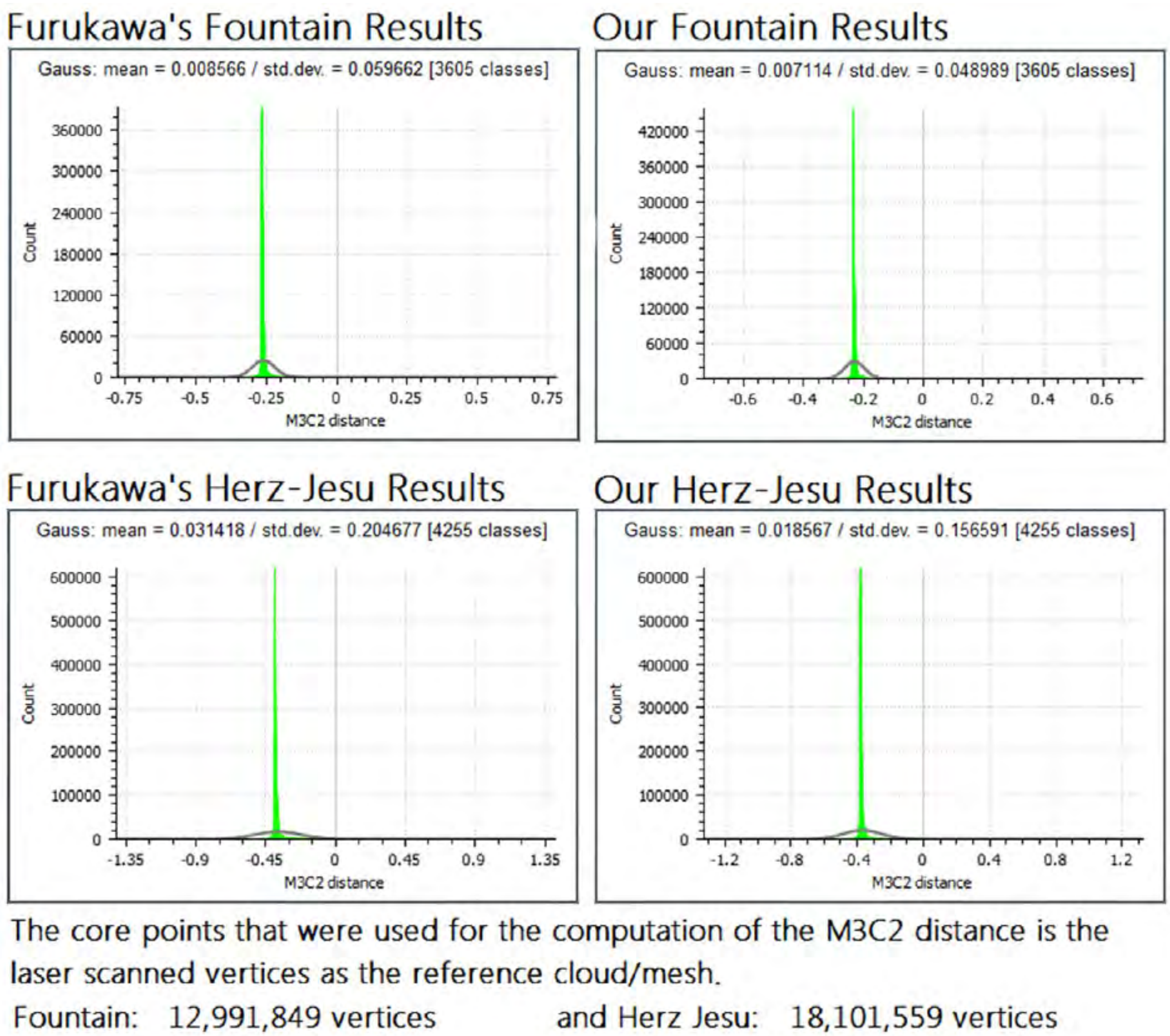

Figure 10: The core points are now defined by the vertices of the ground truth model, the green distribution and compared point cloud (i.e. vertices of the patch model), the gray distribution. The patch model distribution shows a better fit over the baseline results when compared. With a smaller mean and standard deviation of the patch distribution, we validate our patch model results to prove that they are less prone errors and noise. These results are generated by our benchmark software, Cloud Compare Ver. 2 **[43]**.

Figure 11: Our results of the Fountain reconstruction compared to Furukawa's models. The models are meshed and texture mapped from the respective patch models. The number of vertices of the patch models is indicated in the bottom of the image. It can be seen that with our method, we generate more patches than the baseline method. When we look at the Furukawa's results, we can notice some missing vertices which imply that our mesh models are more complete.

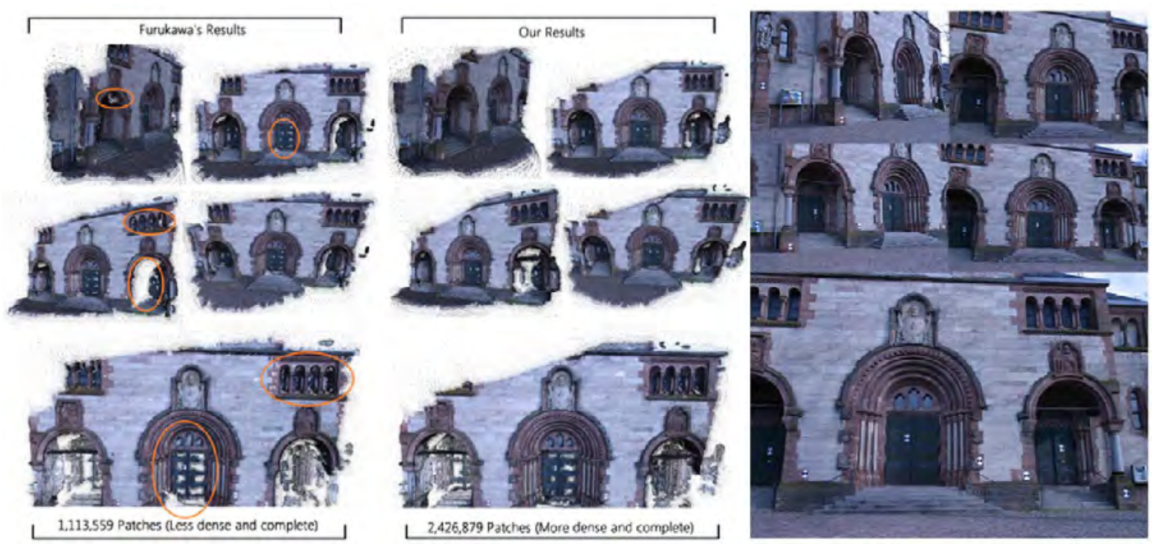

Figure 12: Our results of the Herz-Jesu reconstruction compared to Furukawa's models. Both models are meshed and texture mapped from the respective patch models. The number of vertices of the patch models is indicated in the bottom of the image. It can be seen that with our method, we generate more patches than the baseline method. We can also notice some missing vertices from Furukawa's results which imply that our mesh models are more complete.

2.2.7 Conclusion

A stochastic version of the quasi-Newton BFGS method was implemented to optimise the geometric orientation and depth information of all newly generated patches, provided that the camera coordinates are accurately obtained. The results of the proposed method show a better distribution fit with a smaller standard deviation and means for both evaluated patch models. The results indicate that for smaller means and standard deviations, the distribution is less prone to errors and noise respectively. Despite the fact that we have a denser set of patches than the baseline method, the additional patches did not contribute to the error or noise of the patch distributions. This leads to a conclusion that the quality of the generated patches is better than the baseline patch distributions. Since the quality of the mesh model is directly related to the density of patch model [22], we can say that our patch models are more complete and accurate than the baseline method.

CHAPTER 3 - CONCLUSION

The two main contributions that are described in this dissertation are presented in both Paper 1 and Paper 2. They deal with specific reconstruction challenges that relate to final model quality, time and space complexity. The performance evaluation of the modified PMVS algorithm shows the improvement to accuracy, completeness, and space and time complexity.

In Paper 1, a segmentation technique is proposed to identify and eliminate spurious background artefacts; and to reconstruct only the zone of interest. From 2D boundary information that is acquired using our adaptive segmentation technique, we are able to constrain the patch model to a 3D patch boundary which is similar to visual hull methods. The 3D patch boundary precludes the patch model expansion beyond the bounds of the foreground. The patches that reside on the edges of any view are not reconstructed using that view. However, they are perfectly reconstructed from other views in which they are not edge-aligned. In the baseline PMVS reconstruction pipeline, background artefacts are unnecessarily included in the final patch model. Furthermore, post-processing is needed to obtain the visual hull and iteratively deform the visual hull towards the set of reconstructed patches so that the surface can be approximated via Iterative Snapping. Our proposed time-efficient process requires the modified PMVS and PSR algorithm. This reduces time and space complexity of the reconstruction pipeline, and enabled us to replace the iterative snapping algorithm which constrained the technique for use on an object datasets only. The improvement to this pipeline allows for high quality mesh models to be produced using the PSR algorithm. The mesh results are of similar quality to the baseline method. Although the application of our proposed method to scene reconstruction has not been fully investigated, the results obtained using datasets from *Strech et al.* [3] are very promising. A more generalised 3D reconstruction method can be developed and tested to handle crowded scene datasets.

In Paper 2, a modification is made to the patch optimisation method that refines the geometric orientation and depth information of all newly-generated patches. The results obtained using the modified optimisation method show a good distribution fit to the laser scan. The improvements to the M3C2 means and standard deviation are noted for both of the datasets that we evaluated. The smaller means and standard deviations indicate that the optimised distribution is less prone to errors and noise respectively. Despite the fact that we generated a denser set of patches than the baseline method, the additional patches do not contribute to the error or noise of the patch distributions. This leads to a conclusion that the quality of the generated patches is better than the baseline patch distributions. The quality of the mesh model is directly related to the density of the reference patch model. We can say that our patch models are complete and more accurate than the baseline method.

In conclusion, the modification of the baseline method that is presented in both papers, show specific improvements to the final patch reconstructions. In Paper 1, we were able to reduce time and space complexity by using an adaptive segmentation technique. In Paper 2, we addressed the issue of specular highlights by using a specially developed feature-space and we improved the optimisation method used to refine geometric orientation and depth information for all reconstructed patches.

REFERENCES

[1] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein and R. Szeliski, "A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1*, Washington, DC, USA, 2006.

[2] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein and R. Szeliski, "Multi-View Stereo Evaluation," 2009. [Online]. Available: http://vision.middlebury.edu/mview/. [Accessed 24 November 2015].

[3] C. Strecha, V. H. W, V. G. L, F. P and T. U, "Dense Multi-View Stereo Evaluation," 2008. [Online]. Available: http://cvlabwww.epfl.ch/data/multiview/denseMVS.html. [Accessed 24 November 2015].

[4] R. Cipolla, G. Vogiatzis and P. H. S. Torr, "Multi-view Stereo via Volumetric Graph-cuts," in *Computer Vision and Pattern Recognition (CVPR)*, IEEE Computer Society, Washington, DC, USA, 2005.

[5] G. V. Cipolla, E. Carlos Hernández, T. Philip H. S. and T. Roberto, "Multi-view Stereo via Volumetric Graph-cuts and Occlusion Robust Photo-Consistency," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI),* vol. 29, no. 12, pp. 2241-2246, December, 2007.

[6] F. Schmitt and C. H. Esteban, "Silhouette and Stereo Fusion for 3D Object Modeling," *Journal of Computer Vision and Image Understanding,* vol. 96, no. 3, pp. 267-392, December, 2004.

[7] J. Ponce and Y. Furukawa, "Carved Visual Hulls for Image-Based Modeling," *International Journal of Computer Vision (IJCV),* vol. 8, no. 1, pp. 564--577, January, 2009.

[8] K. N. Kutulakos and S. M. Seitz, "A theory of shape by space carving," *International Journal of Computer Vision,* vol. 38, no. 3, pp. 199-218, 2000.

[9] O. Faugeras and R. Keriven, "Complete dense stereovision using level set methods," in *5th European Conference on Computer Vision (ECCV)*, Freiburg, Germany, June 2-6, page 379, 1998.

[10] P. J. Narayanan, P. W. Rander and T. Kanade, "Constructing virtual worlds using dense stereo," in *Computer Vision, 1998 Sixth International Conference*, Bombay, pages 3-10, 1998.

[11] R. Bhotika, D. Fleet and K. Kutulakos, "A probabilistic theory of occupancy and emptiness," in *Proceedings of European Conference on Computer Vision (ECCV)*, USA, Toronto, pages 112-130, 2002.

[12] H. Jin, S. Soatto and A. J. Yezzi, "Multi-view stereo beyond Lambert," in *Proceedings of IEEE Computer Society Conference on Computer Vision Pattern Recognition*, pages 171- 178, 2003.

[13] R. Keriven, "A variantional framework for shape from contours," Ecole Nationale des Ponts et Chaussees, Cermics. France, 2002.

[14] W. E. Lorensen and H. Cline, "Marching cubes: A high resolution 3D surface construction algorithm," in *Processings of the 14th annual conference on computer graphics and iteractive techniques*, pages 163-169, ACM, 1987.

[15] S. N. Sinha and M. Pollefeys, "Multi-view reconstruction using photo-consistency and exact silhouette constraints: A maximum-flow formulation," in *Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV)*, Washington, DC, USA, pages 349-356, 2005.

[16] Y. Furukawa and J. Ponce, "High-fidelity image-based modeling," University of Illinois at Urbana-Champaign, 2006.

[17] S. Tran and L. Davis, "3D surface reconstruction using graph cuts with surface constraints," in *Proceedings of European Conference on Computer Vision (ECCV)*, pages 219-231, 2006, Springer.

[18] A. Hornung and L. Kobbelt, "Hierarchical volumetric multi-view stereo reconstruction of manifold surfaces based on dual graph embedding," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 503 - 510, 2006.

[19] S. Paris, F. X. Sillion and L. Quan, "A surface reconstruction method using global graph cut optimisation," *International Journal of Computer Vision,* vol. 66, no. 2, pp. 141-161, 2006.

[20] S. M. Seitz, B. M. Goesele and B. Curless, "Multi-View Stereo Revisited," in *IEEE Computer Vision and Pattern Recognition (CVPR)*, IEEE Computer Society, Washington, DC, USA, 2006, pp2402-2409.

[21] M. Kazhdan, M. Bolitho and H. Hoppe, "Poisson surface reconstruction," in *Proceedings of the fourth Eurographics symposium on Geometry processing*, Eurographics Association, Aire-la-Ville, Switzerland, 2006, pp 61-70.

[22] J. Ponce and Y. Furukawa, "Accurate, Dense, and Robust Multiview Stereopsis," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 32, no. 8, pp. 1362-1376, 2010.

[23] C. Strecha, R. Fransens and L. Van Gool, "Combined depth and outlier estimation in multi-view stereo," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006.

[24] C. Strecha, T. Tuytelaars and L. Van Gool, "Dense matching of multiple wide-baseline views," in *Proceedings of 9th IEEE International Conference*, France, Nice, pages 1194-1201, 2003.

[25] T. K. Moon, "The expectation-maximization algorithm," *Signal Processing Magazine, IEEE,* vol. 13, no. 6, pp. 47-60, 1996.

[26] M. Habbecke and L. Kobbelt, "A surface-growing approach to multi-view stereo reconstruction," in *IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1 - 8, 2007.

[27] M. Habbecke and L. Kobbelt, "Iterative multi-view plane fitting," in *Vision, modeling, and visualization*, Aachen, Germany, page 73, November 22-24, 2006.

[28] R. Kolluri, J. R. Shewchuk and J. F. O'Brien, "Sprectral surface reconstruction from noisy point clouds," in *Proceedings of the 2004 Eurographics/ ACM SIGGRAPH symposium on geometry processing*, pages 11- 21, 2004.

[29] M. Lhuillier and L. Quan, "A quasi-dense approach to surface reconstruction from uncalibrated images," *IEEE Transactions on Pattern Analysis & Machine Intelligence,* vol. 27, no. 3, pp. 418-433, March,2005.

[30] M. Kazhdan, "Reconstruction of solid models from oriented point sets," in *Proceedings of the third Eurographics symposium on Geometry processing*, Eurographics Association, page 73, 2005.

[31] Y. Furukawa and J. Ponce, "Accurate, Dense and Robust Multi-View Stereopsis," in *IEEE Computer Vision and Pattern Recognition (CVPR)*, IEEE Computer Society, Washington,DC, USA, 2007, pp 1-8.

[32] Y. Furukawa, B. Curless, S. M. Seitz and R. Szeliski, "Towards internet-scale multiview stereo," in *Proceedings of IEEE Computer Vision and Pattern Recognition (CVPR)* , IEEE Computer Society, Washington, DC, USA, 2010.

[33] M. Goesele, N. Snavely, B. Curless, H. Hoppe and S. M. Seitz, "Multi-View Stereo for Community Photo Collections," *International Conference on Computer Vision,* pp. 265--270, 2007.

[34] S. Agarwal, N. Snavely, I. Simon, S. M. Seitz and R. Szeliski, "Building Rome in a Day," in *International Conference on Computer Vision*, Kyoto, Japan, 2009.

[35] S. Agarwal, Y. Furukawa, N. Snavely, B. Curless, S. M. Seitz and R. Szeliski, "Reconstructing Rome," *IEEE Computer,* vol. 43, no. 6, pp. 40-47, June,2010.

[36] C. Jinxiang, S. B. Kang and R. Szeliski, "Handling Occlusions in Dense Multi-view Stereo," in *IEEE Computer Vision and Pattern Recognition (CVPR)*, IEEE Computer Society, Washington, DC, USA , 2001,pp 103-110.

[37] Q. Shan, B. Curless, Y. Furukawa, C. Hernandez and S. M. Seitz, "Occluding Contours for Multi-view Stereo," in *Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, Columbus, OH, USA, 2014, pp 4002- 4009.

[38] S. Al-amri, S. Salem, N. V. Kalyankar and S. Khamitkar, "Image Segmentation by using threshold techniques," in *CORR*, 2010.

[39] R. Hartley and A. Zisserman, Multiple view geometry in computer vision, USA, Cambridge: Cambridge University Press, 2003.

[40] J. Nocedal and S. J. Wright, Numerical Optimization, 2nd ed., New York: NY: Springer-Verlag, 1999.

[41] J. Nelder and R. Mead, "A Simplex Method for Function Minimization," *Computer Journal,* vol. 7, pp. 308-313, 1965.

[42] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision (IJCV),* vol. 60, no. 2, pp. 91-110, November,2004.

[43] D. Lague, N. Brodu and J. Leroux, "Accurate 3D comparison of complex topography with terrestrial laser scanner: application to the Rangitikei canyon," *ISPRS Journal of Photogrammetry and Remote Sensing,* vol. 82, pp. 10-26, August 2013.

[44] Y. Furukawa, "PMVS (Version 2) Documentation," September 2010. [Online]. Available: http://www.di.ens.fr/pmvs/documentation.html. [Accessed 24 November 2015].

[45] R. Khuboni and B. Naidoo, "Adaptive Segmentation for Multi-View Stereo," *Journal of IET Computer Vision,* pp. 1-24, 2015.

[46] A. Mokhtari and A. Riberio, "Regularized stochastic bfgs algorithm," 2013. [Online]. Available: https://fling.seas.upenn.edu/~aryanm/wiki/index.php?n=Research.Publications.. [Accessed 24 November 2015].

[47] A. Mokhtari and A. Ribeiro, "Regularized Stochastic BFGS algorithm," in *In Proceedings IEEE Global Conf. Signal Inf. Process.*, Austin, TX, USA, December 2013.

[48] R. H. Byrd, J. Nocedal and Y. Yuan, "Global convegence of a class of quasi-newton methods on convex problems," *SIAM J. Numer. Anal.,* vol. 24, no. 5, pp. 1171-1190, October 1987.

[49] R. Fletcher, Practical Methods of Optimizations, New York: NY, USA: Wiley, 2013.

[50] N. N. Schraudolph, J. Yu and S. Gnter, "A stochastic quasi-newton method for online convex optimization," in *Proceedings 11th International Conference on Artificial Intelligence and Statistics (AIstats)*, Society of Artificial Intelligence and Statistics, 2007.

APPENDICES

Appendix A – Digital Copy of the Results

The mesh results that have been captured in this dissertation are saved in a digital format for viewing later on. The following items can be found in the digital copy (i.e. DVD):

- The digital copy of the *Turn-it-in* report with the author's response digital report.
- The digital copy of the final dissertation.
- The mesh model results of the research work presented in this dissertation.
- The source of the modified PMVS algorithm is also included.

The attached DVD

Appendix B – Texture Mapped Mesh Models

The texture mapped mesh models of that were used in this dissertation, are shown in this session as images. Dino-Ring, Temple-Ring, Skull, Fountain and Herz-Jesu reconstructions are shown below in their respective order.