

Biometric Fusion Methods for Adaptive Face Recognition in Computer Vision



Mahammad Majed Fakhir

Newcastle University

United Kingdom

A thesis submitted for the degree of

Doctor of Philosophy

2017

To my ...

... Parents, for their prayers;

... Supervisors, for their patience:

... Friends, for their support;

.... And future wife for her love.

Declaration

I, Mahammad Majed Fakhir, hereby declare that this thesis is my own work and effort that it has not been previously submitted anywhere for any award.

Signature:

Student: Mahammad Majed Fakhir

Date:

Acknowledgement

I would like to express my sincere appreciation to my main supervisor, Prof. Satnam Dlay, for his continued support in my PhD study and related research, and for his patience, motivation, and immense knowledge. His guidance helped me throughout the period of research and writing this thesis. I could not have imagined having a better supervisor and mentor for my PhD study.

My sincere thanks also to Dr Wai Lok Woo and Prof. Jonathon Chambers, who supported me with advice and guidance.

I would like to thank my parents for allowing me to realise my own potential. All the support they have provided me with over the years is the greatest gift I have ever received.

To all my colleagues and my friends: thank you very much for what you have done for me. I thank you all for the companionship that made this journey much easier. In fact, I do not need to list your names because I am sure that you know who you are. Special thanks for Dr. Mahammad and Dr. Ahmad for been part of my earliest experiments in this thesis.

I would like to thank my future wife Sarah for her love and constant support, and for keeping me happy over the past few months. Thank you for being my muse, editor, proofreader, and my best listener after my mother ☺. But most of all, thank you for being my best friend.

Finally, I dedicate this research to my home country of Iraq.

ABSTRACT

Face recognition is a biometric method that uses different techniques to identify the individuals based on the facial information received from digital image data. The system of face recognition is widely used for security purposes, which has challenging problems. The solutions to some of the most important challenges are proposed in this study. The aim of this thesis is to investigate face recognition across pose problem based on the image parameters of camera calibration. In this thesis, three novel methods have been derived to address the challenges of face recognition and offer solutions to infer the camera parameters from images using a geometric approach based on perspective projection. The following techniques were used: camera calibration CMT and Face Quadtree Decomposition (FQD), in order to develop the face camera measurement technique (FCMT) for human facial recognition.

Facial information from a feature extraction and identity-matching algorithm has been created. The success and efficacy of the proposed algorithm are analysed in terms of robustness to noise, the accuracy of distance measurement, and face recognition. To overcome the intrinsic and extrinsic parameters of camera calibration parameters, a novel technique has been developed based on perspective projection, which uses different geometrical shapes to calibrate the camera. The parameters used in novel measurement technique CMT that enables the system to infer the real distance for regular and irregular objects from the 2-D images. The proposed system of CMT feeds into FQD to measure the distance between the facial points. Quadtree decomposition enhances the representation of edges and other singularities along curves of the face, and thus improves directional features from face detection across face pose. The proposed FCMT system is the new combination of CMT and FQD to recognise the faces in the various pose.

The theoretical foundation of the proposed solutions has been thoroughly developed and discussed in detail. The results show that the proposed algorithms outperform existing algorithms in face recognition, with a 2.5% improvement in main error recognition rate compared with recent studies.

CONTENTS

Abstract.....	v
List of Figures.....	x
List of Tables	xiii
List of Symbols.....	xiv
Abbreviations/Acronyms.....	xv
Chapter 1 : Introduction.....	1
1.1 Chapter Outline	1
1.2 Overview of biometrics systems	3
1.1.1 <i>Biometrics Authentication</i>	4
1.1.2 <i>Measurement requirements</i>	5
1.1.3 <i>Modules of biometrics systems</i>	6
1.2 Biometric Facial Recognition	8
1.2.1 <i>Geometrical Facial Measurement</i>	9
1.2.2 <i>Approaches of Face Recognition</i>	9
1.3 The challenges in face recognition and this work.....	11
1.4 Work Motivation.....	12
1.4.1 <i>Geometric parameters</i>	12
1.4.2 <i>Face recognition across poses</i>	13
1.4.3 <i>Quadtree decomposition</i>	14
1.5 Contributions.....	15
1.6 Thesis structure	17
Chapter 2 : Camera Calibration and Face Recognition	19
2.1 Chapter outline.....	19
2.2 Introduction.....	20
2.2.1 <i>Perspective projection in camera</i>	26
2.2.2 <i>Pinhole camera model</i>	29
2.3 Camera and projective parameters	30
2.4 Classification of face recognition techniques	32
2.4.1 <i>Face recognition across poses</i>	34
2.4.2 <i>The one sample per person problem</i>	36
2.5 Image quadtree decomposition in face recognition	37
2.6 Summary	39
Chapter 3 :Camera calibration based on perspective projection.....	40

3.1	Introduction.....	40
3.1.1	<i>Camera Calibration Matrix</i>	41
3.2	Camera calibration characteristics	42
3.2.1	<i>Extrinsic parameters</i>	43
3.2.2	<i>Intrinsic parameter</i>	44
3.2.2.1	<i>The central vanishing points</i>	46
3.3	Perspective Projection.....	49
3.3.1	<i>Two Vanishing Points Used to Infer the Intrinsic Parameters of the Camera</i>	50
3.3.2	<i>Focal Length from Vanishing Points</i>	51
3.4	Results and experiment	55
3.4.1	<i>Camera Parameters</i>	55
3.4.2	<i>Focal Length and Application</i>	57
3.4.3	<i>Discussion</i>	62
3.4	Summary	64
Chapter 4	: 3-D Information measurement from 2-D image	65
4.1	Introduction.....	66
4.1.1	<i>The Challenges</i>	67
4.2	Numerical methods	68
4.3	Test of the system	71
4.3.1	<i>Tilt Angle Variation</i>	72
4.3.2	<i>Cross Pose Angle Variation</i>	73
4.3.3	<i>Swing Angle Variation</i>	75
4.4	Inferring dimensions from an image using automatic calibration.....	77
4.4.1	<i>Regular Shape</i>	78
4.4.2	<i>Irregular Shape</i>	81
4.5	Results and Experiments.....	82
4.5.1	<i>Measurement of Dimensions for Regular Shapes</i>	82
4.5.2	<i>Measurement of Dimension of Irregular Shapes</i>	86
4.6	Discussion.....	87
4.7	Summary	88
Chapter 5	: Novel Method of Face Recognition from Various Poses.....	90
5.1	Introduction.....	91
5.1.1	<i>Face Recognition Approaches</i>	92
5.2	Face tracking And Decomposition.....	94
5.2.1	<i>Quadtree Decomposition</i>	95

5.2.2	<i>Face Tracking</i>	96
5.2.3	<i>FQD</i>	97
5.3	FCMT.....	98
5.3.1	<i>Physical and Mathematical Model of FCMT</i>	101
5.3.2	<i>Coordinate System of Camera Perspective</i>	102
5.4	Facial database	103
5.5	Results and Experiments.....	104
5.5.1	<i>Real-time Face Recognition</i>	104
5.5.2	<i>Face Tracking</i>	106
5.5.3	<i>Face Recognition</i>	109
5.6	Summary	115
Chapter 6	: Conclusion and Future work.....	116
1.1	Conclusion	116
6.2	Future Work	118
Reference	122

LIST OF PUBLICATIONS

- 1- M. M. Fakhir, W. L. Woo, and S. S. Dlay, "Inferring dimensions from an image using automatic calibration," in *Communication Systems, Networks & Digital Signal Processing (CSNDSP), 2014 9th International Symposium on*, 2014, pp. 776-780. Uk-Manchester
- 2- M. M. Fakhir, W. L. Woo, and S. S. Dlay, "Face Recognition Based on Features Measurement Technique," in *Modelling Symposium (EMS), 2014 European*, 2014, pp. 158-162. Italy-Pisa.
- 3- M. M. Fakhir, W. L. Woo, J. A. Chambers, and S. S. Dlay, "Perspective projection for variance pose face recognition from camera calibration," in *SPIE Photonics Europe*, 2016, pp. 98961L-98961L-7. Belgium Brussel
- 4- M. M. Fakhir, W. L. Woo, J. A. Chambers, and S. S. Dlay, "Novel Method of Face Recognition from Various Pose," *the International Conference on Pattern Recognition Systems (ICPRS)*, 2016.Chile – Talca.
- 5- M. M. Fakhir, W. L. Woo and S. S. Dlay, " Face Recognition from Various Poses Based on Geometrical Technique of Computer Vision," *Submitted to Pattern Recognition Elsevier*, 2017.

LIST OF FIGURES

Figure 1.1: Physiological biometrics	2
Figure 1.2: Behavioural biometrics.....	3
Figure 1.3: the classification of a biometric system.....	6
Figure 1.4: PCA eigenfeatures for eyes, nose, and the edge of the mouth [25].....	11
Figure 1.5: Face recognition system	13
Figure 2.1: Axis PTZ IP [41]	21
Figure 2.2: Classification of geometric projection.....	24
Figure 2.3: The Holy Trinity by Masaccio[67].....	25
Figure 2.4: Landscape and the horizon	26
Figure 2.5: Standard perspective projection in camera scene	27
Figure 2.6: Pinhole camera structure	29
Figure 2.7: Volume holographic image recognition system [135]	38
Figure 3.1: Camera calibration characteristic.	42
Figure 3.2: Intrinsic parameter transformation	43
Figure 3.3: Elements of rotation	44
Figure 3.4: Vanishing point from one direction.....	46
Figure 3.5: Two-vanishing point perspective from image plane.	48
Figure 3.6: Camera perspective showing projection of a point in space.....	50
Figure 3.7: Coordinate system of the distance between the centres of two vanishing points and the centre of the lens	51
Figure 3.8: The nearest point on the vanish line to the original point	52
Figure 3.9: Recovered projection points from the corner shape	54
Figure 3.10: The pinhole camera	57
Figure 3.11: Four pictuers captured from different distances in cms	58
Figure 3.12: Coordinate system of focal length.....	60
Figure 3.13: Effect of noise (dB) on focal length (mm)	61
Figure 3.14: The error rate of focal length.....	62
Figure 4.1: Camera perspective showing projection of a point in space.....	68
Figure 4.2: Distance between camera and references varied from 500 mm to 2m.....	71

Figure 4.3: The average error of the depth and length in tail variation for three different geometrical shapes of calibration.....	73
Figure 4.4: Changing the pan angle from 0 to 180° degrees (left to right),.....	74
Figure 4.5: Average error for the depth and length with pan angle from 0 to 180°, for three geometrical shapes.....	75
Figure 4.6: Swinging for the camera around 180°, when $\theta_1 = 90^\circ$ and $\theta_2 = 0^\circ$	76
Figure 4.7: Average error in the depth and length with swing angle varying from 0 to 180°, for three geometrical shapes.....	77
Figure 4.8: Three different geometrical shapes.....	79
Figure 4.9: shows size reference paper 110mm × 148mm.....	79
Figure 4.10: Size of the reference paper 135mm × 135mm.....	80
Figure 4.11: Isosceles triangle and equilateral triangle references.	81
Figure 4.12: Two different irregular shapes to be measured.	82
Figure 4.13: The results of CMT for the standard size of shapes.	83
Figure 4.14: Pictures captured from different distances by 2 MP Camera.	84
Figure 4.15: Alteration in error with distance -2 MP.....	84
Figure 4.16: Alteration in error with distance -3 MP.....	85
Figure 4.17: Pictures captured from different distances by 8 MP Camera.	85
Figure 4.18: Alteration in error with distance -8 MP.....	86
Figure 4.19: Measurement of an irregular shape.	86
Figure 4.20: Measurement of two different irregular shapes.....	87
Figure 5.1: The theory of the proposed method of 3D artificial human face.....	91
Figure 5.2: Harvesting dimensions ear to ear from three difference points.	93
Figure 5.3: Concept of the proposed method with front, left and right side references of various poses from FERET database.	94
Figure 5.4: The process of features measurement applied with Quadrature decomposition and retrieval information image to the original.	95
Figure 5.5: Example of face tracking.	96
Figure 5.6: Three side of calibration for cube (A, B, C) shows the frontal and two sides of view.	97
Figure 5.7: The output of FQD on one object from FRET database.....	98
Figure 5.8: The FCMT system.	100
Figure 5.9: Camera perspective showing projection of a point in space.....	101

Figure 5.10: FCMT coordinator system.....	102
Figure 5.11: The experiment of real-time face recognition for different subjects	104
Figure 5.12: Names of ear points.....	105
Figure 5.13: Accuracy of face recognition based on CMT	106
Figure 5.14: Sample of ORL database facial image.	107
Figure 5.15: Performance in tracking faces using images from the ORL database for 40 people where for each person there are 10 different pictures.	107
Figure 5.16: Sample of FERET database facial images.....	108
Figure 5.17: Performance in tracking face images from the FERET database, of 100 people where person has 9 pictures in different poses.....	109
Figure 5.18: Performance of FCMT compared to typical cross-pose face recognition methods on the FERET database.....	110
Figure 6.1: 3D reconstruction from 2 views image (right and left)	119
Figure 6.2: the 3D reconstruction of face recognition from one view using ORL database....	119
Figure 6.3: MCT future work.....	120
Figure 6.4: FQD applied for three points.....	121
Figure 6.5: FQD free poses points detection.....	121

LIST OF TABLES

Table 2.1: 2-D techniques for face recognition across pose	35
Table 2.2: Face recognition across pose with assistance of 3-D models	35
Table 3.1: Intrinsic Camera Parameters from Manual Point Selection.....	56
Table 3.2: Intrinsic Camera Parameters from Automatic Point Selection	57
Table 3.3: The internal focal length of the camera from four distances	59
Table 4.1: The result of measurement for the geometrical shapes.....	83
Table 4.2: Summary of measurement error	88
Table 5.1: IPD values (mm) [219]	99
Table 5.2: Results of the actual distances and CMT	105
Table 5.3: The error percentage in recognition between $\pm 60^\circ$ angles.....	111
Table 5.4: Confusion matrix of recognition rate and relative mean error for each vector, Horizontal (H) and Vertical (V).....	113
Table 5.5: Face recognition rate per person across poses	114

LIST OF SYMBOLS

n	Surface normal vector
Π	Image plan
f	Focal length
o	Principal points
\sim	Approximately
Z''	Identification Card
λ	Lambda
\sum_i^N	Summation from index I to i=N
θ	Theta (the pan angle)
Z'	Cordnate of prespective projection
Z''	The real word coordinate
z	The original coordinate
Σ	Sigma (one of the intrinsic camera parameters)
Δ	Delta (changing of the image parameters)
Φ	Phi (the tilt angle)
x^*	The array of 3-D points
$. $	Absolut value
ξ	The swing angle

ABBREVIATIONS/ACRONYMS

2-D	Two Dimensional
3-D	Three Dimensional
FCMT	Face Camera Measurement Technique
CMT	Camera Measurement Technique
PIN	Personal Identification Number
ID	Identification Card
PCA	Principal Component Analysis
LDA	Linear Discriminant Analysis
FDA	Fisher discriminant analysis
FQD	face quadtree decomposition
2.5D	Two and a half dimensional
PTZ	Pan–tilt–zoom camera
AAM	Active appearance model
1-D	One Dimensional Vector
AAMs	Active appearance models
ASMs	Active shape models
MNPP	Multi-linear neighbourhood are preserving projection
NPP	Neighbourhood Preserving Projection
LBP	Local Binary Pattern
VP	Vanish point
PPo	Perspective projection Original point
RGB	Red, Green and blue light in the image
MP	Megapixel
RMSE	Root Mean Square Error

Chapter 1 : INTRODUCTION

1.1 CHAPTER OUTLINE

Camera surveillance and face recognition are useful tools in identifying criminal activities and avoiding crime. This problem is part of the grand challenge of face recognition variation in pose and camera calibration. Therefore, the surveillance system and the processing of the facial matching is a critical research topic for artificial intelligence research, at the same time there are justified reasons for automatic attribute analysis to take advantage of the benefits of present technologies. Recently there have been many advances and developments in communications technology and camera mobility, which help in estimating real dimensions from virtual 2-D images. This link between communications systems and camera devices makes life much easier, and more accuracy is possible due to recent research in computer vision and camera calibration; for example in the field of computer vision in cars, scene 360 cameras, GPS, navigation devices in ships and aircraft, predictive analytic, building construction and for medical purposes.

Therefore, this thesis proposes a system of face recognition from various poses. The concept is based on a geometric method from various face poses that is efficient in handling the range of pose variations within $\pm 60^\circ$ of rotation, and the method is called the Face Camera Measurement Technique (FCMT). The approach aims to calibrate the camera for a 2-D image using fixed and robust facial landmarks to ensure the reliable estimation of real dimensions. Therefore, the image is decomposed into four equal-sized square blocks to use the final retrieval information image as the output in the next step of feature extraction and selection. Extensive and systematic experimentation using the FERET and ORL database showed that that proposed method consistently outperforms single task-based baselines, as well as state-of-the-art methods for the problem of variations in pose.

In this chapter the biometric face recognition technology, which has recently received significant attention, is discussed. Biometric systems and their applications are introduced with particular reference to facial recognition problems. Then the key technical challenges in solving such problems are outlined. Next, the motivation for this study is described. Popular subspace learning techniques are reviewed, with a focus on camera

calibration from one 2-D image. Finally, the contributions of this study are highlighted, and the structure of the thesis is explained.

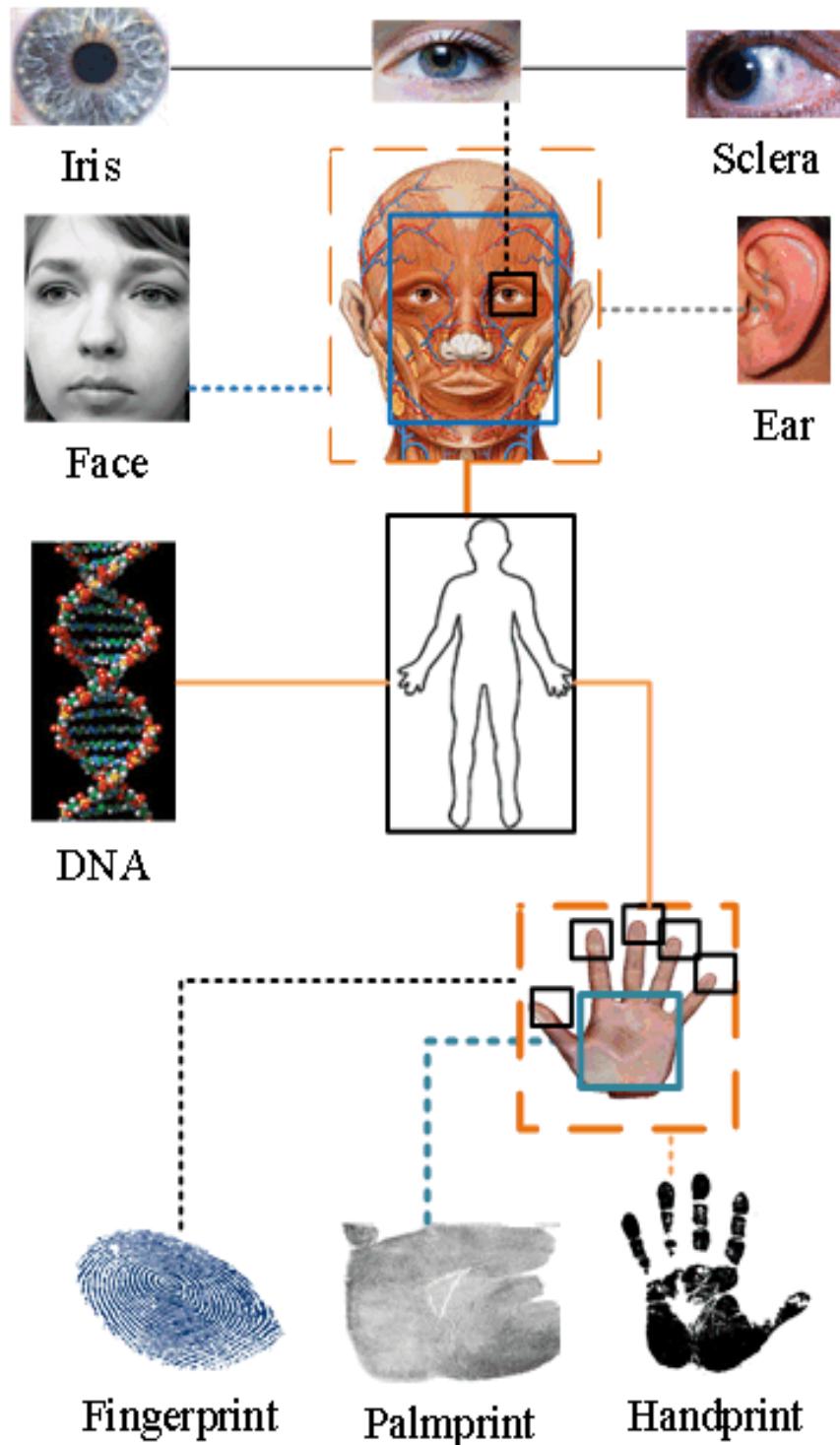


Figure 1.1: Physiological biometrics

1.2 OVERVIEW OF BIOMETRICS SYSTEMS

A biometrics system can be defined as a system for the measurement of the physical characteristics of a person, where these parameters are used to identify the person. Hence, such systems are used in security and access control applications. These measurements are used to analyse distinctive human characteristics so as to determine people's identity for authentication purposes [1]. The use of biometrics as a method of identification helps to avoid the problems related to traditional methods, such as requiring personal identification cards, keys or passwords. Thus, biometrics systems use a computer technology to identify individuals depending on physiological features, as shown in Figure 1.1 and Figure 1.2

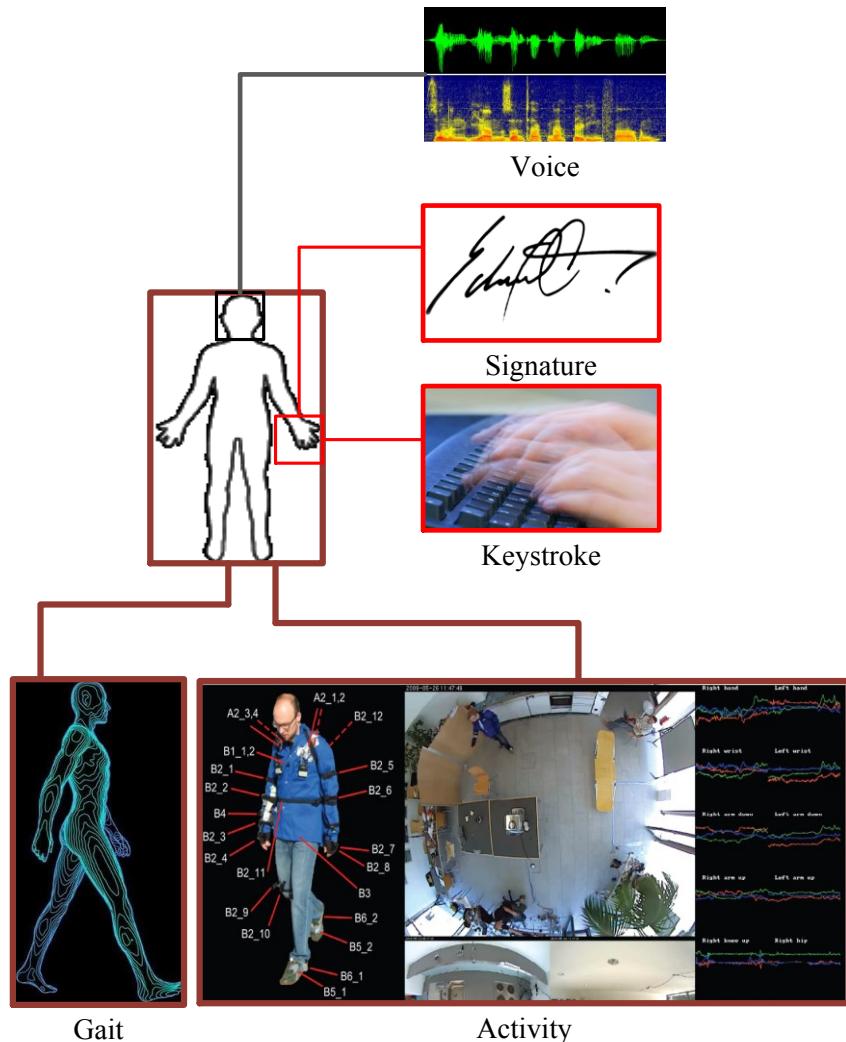


Figure 1.2: Behavioural biometrics

Better authentication techniques became necessary in the wake of heightened concerns about security in the world, especially after the terrorist attack of 9/11. Furthermore, progress in networking, communication, and mobility has encouraged research into biometrics. Hence, biometric systems have now been expanded for use in various commercial, civilian, and forensic applications as a means of establishing identity.

1.1.1 *Biometrics Authentication*

The identification of people via biometric systems uses on the characteristics of the body or behavioural traits and is increasingly being used instead of, or with other forms of identification based on data the user has or knows such as IDs, passwords and PINs. Furthermore, such systems can work in conjunction with traditional identification to confirm security. The biometric system starts by possessing biometric data collected from the subject using sensor models. A typical biometric system operates by comparing the recorded person's features with the biometric data samples in a database to determine the identity of the individual (identification) or to authenticate a requested identity (verification) [1]. In a biometric identification system, the class label indicates the identity of the individual; while in a verification system the class label is a match (genuine) or a non-match (impostor). In both modes of operation, a reject label results when the system does not match the subject with a valid class. In general, the aim of a biometric identification system can be defined as the recognition of an individual by using information about certain physical characteristics or personal behaviours that could be gathered by the measurement of features in any of the three following groups:

1 Intrinsic biometrics

Intrinsic biometrics identify the individual's genetic make-up, for example from fingerprints, iris and sclera patterns.

2 Extrinsic biometrics

Extrinsic biometrics involves the individual's learned behaviour, such as signatures and keystrokes.

3 Hybrid biometrics

Hybrid biometrics systems are based on a combination of the individual's physical characteristics, and personal traits such as face recognition.

In biometric face recognition modalities, the facial recognition system measures and analyses the overall structures in the terms of shapes and sizes of features; for example the distance between the eyes, nose, mouth, and jaw edges, the upper outline of the eye openings, the sides of the mouth, the location of the nose, eyes and areas near the cheekbones.

1.1.2 *Measurement requirements*

Nowadays, many applications use biometric characteristics and these have become quite popular. These biometrics systems have advantages and disadvantages. Hence, the choice of a biometric feature for a precise application depends on a range of factors as well as its comparable performance requirements. However, the biological measurements needs to qualify for biometric use in biomedical systems, thus, there are seven criteria which have been recognised to determine the suitability of physical or a behavioural measurements to be used in biometric applications [2].

1. Universality: Everyone should have the characteristic.
2. Acceptability: The degree to which people are willing to accept the use of a particular form of biometric identification in daily life.
3. Measurability: the characteristic should be measurable and its measurement achievable. Furthermore, the acquired raw data should be open to processing to extract representative feature sets.
4. Performance: The accuracy of features matching should be acceptable.
5. Stability: The characteristic should be suitably invariant over a period.
6. Circumvention: How easily the system can be tricked using artificial methods such as fake fingers for physical features or in simulating traditional behavioural functions.
7. Distinctiveness: The features used should be sufficiently different across individuals.

Moreover, there are other important characteristics of the use of biometrics, such as privacy, where the process should not infringe on the privacy of the person. What is more,

with all appropriate biometric characteristics it should be possible to decrease the quality of the data to a state that makes it numerically similar to others samples of faces.

Whereas no biometric characteristics may be ideal because they do not completely meet all of the requirements such as accuracy and cost, some are acceptable in practice. Among the numerous biometric technologies being considered, the characteristics that best satisfy the above conditions are shown in Figure 1.1 and Figure 1.2.

1.1.3 *Modules of biometrics systems*

A) *Models classifications*

A biometric system can be classified into two models [3-5]:

- Enrolment model: this includes training and database preparation.
- Verification model: this includes a matcher that is matching the authentication and share it with the database preparation. Both of the models have to end with a decision.

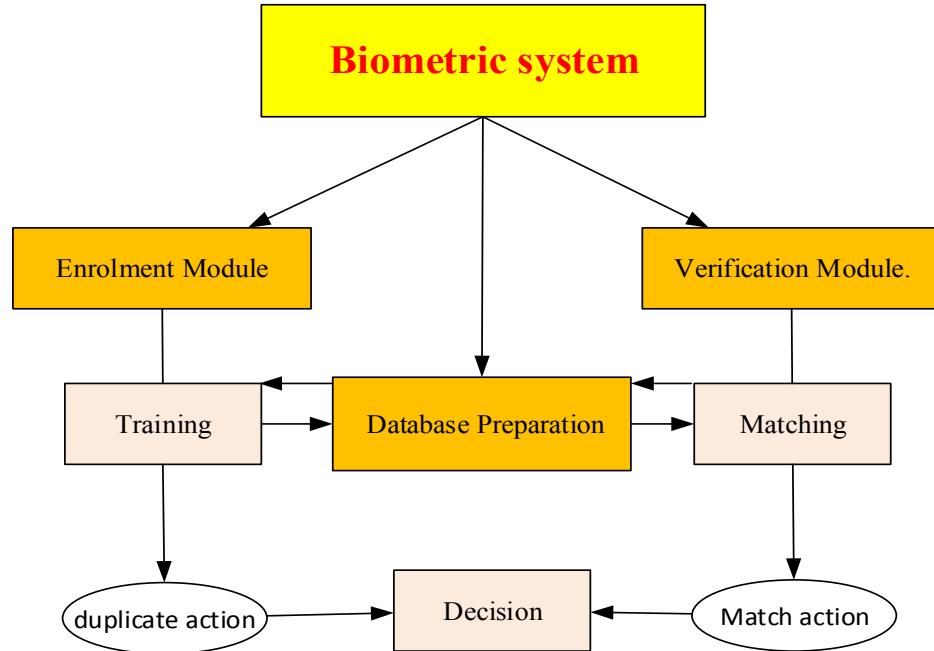


Figure 1.3: the classification of a biometric system.

B) *Models operators*

Logically, from Figure 1.3, a biometrics system for recognition depends on two operating models: verification and identification models. In the verification model, the system authenticates people's identity by matching token biometric data with the original data stored in a database system. The person who requests recognition calls an identity, for example by the use of a password or smart card. The model of decision in the system manages one-to-one comparison to determine whether or not the ideal identity is true or not. The ideality verification is typically used for positive recognition, where the aim is to prevent multiple people from using the same identity [6]. In identification model, the system identifies people by searching the original patterns of the users in the database to find a match. Consequently, the system uses a one-to-many comparison to establish a person's identity without the subject having to claim an identity [7].

C) *Biometric system design*

A biometric system is designed using the following four main modeles:

- Biometric sensor model,
- Feature extraction model,
- Matching model (decision-making model)
- System database model

The sensor model uses biometric data from an individual such as a fingerprint sensor that images the ridge and valley geometrics lines of a user's finger. Meanwhile, feature extraction model, acquires the biometric data to extract a set of relevant features. For instance, local ridge and valley singularities are based on the position and orientation of minutiae points.

Matching scores generated from the features extracted during recognition when the data acquired is compared against the stored templates. The matching model is responsible for decision-making as shown in the component in Figure 3, and this is done when the user who has claimed an identity is either accepted or rejected based on the matching score generated by the matching model.

Finally, the system database model is used by the biometrics system to store the biometric templates of registered users. During the enrolment model, the biometrics characteristics of an individual are first scanned to produce a raw digital of the features [8].

1.2 BIOMETRIC FACIAL RECOGNITION

Computer vision and image processing have become capable of recognising individuals based on the recognition of human faces. The difference between computer vision and image processing in this thesis is that computer vision is involved mainly with feature extraction, and image processing is used to enhance the image extraction information. Hence, face recognition uses an image of a face in a photograph or video, which is converted into features that describe the face's characteristics. Therefore, recognition algorithms use simple geometric models and the science of sophisticated mathematical representations to process the matching of individuals. Moreover, face recognition is used for both verification and identification (open-set and closed-set) purposes. Facial biometrics technology has encountered a number of challenges which need to be overcome. The technology of face recognition works with the most public individual identifiers without the need to use physical action. For instance, the hand or the finger can be placed on a reader but it can take a long time to process data for a group of people so there is no intrusion or delay, and it can also be a source of the transfer of germs. Similarly, people may be required to precisely position their eye in front of a scanner for face recognition systems which invisibly take pictures of candidate's faces as they enter a defined area. In face recognition, people may be entirely unaware of the process, and they do not feel under scrutiny or a violation of their privacy. On the other hand, problems in facial recognition may occur when the face is fully or partly covered, and differences in external factors such as illumination and people's bodily movements may cause problems. Hence, the exact position of the fingerprint or iris can be more precise in recognition systems than face recognition. To overcome these issues, advances in technologies such as a high-resolution camera, can be used in different research that focuses on accuracy and facial movements. Mobile devices and the Internet are increasing the popularity of facial recognition, and the performance of such systems is improving. Many applications have recently been developed to maintain the progress in the field of computer vision, such as in biometric authentication, human-computer interfaces, commercial security, law enforcement, image and video surveillance, immigration, and national and international Identification (ID) systems [9].

1.2.1 *Geometrical Facial Measurement*

The complete facial structure is measured by face recognition systems, as well as the distance between the eyes, nose, mouth, and jaw edges. These measurements are saved in a database and recalled for comparison to identify the person. Human faces have numerous unique landmarks and different peaks with valleys that make up facial features. Moreover, the human face has about 80 nodal points. The important nodal points on the face that have been used in facial recognition technology using image analysis and computer vision are as follows [10]:

- Distance between the eyes
- Width of the nose
- The shape of the cheekbones
- The length of the jaw line

The measurements from these points identify people in terms of numerical code under the title ‘faceprint’, which means that the geometry of each face has unique distances and can be measured in actual life using the available tools to do this work. These distances can be real numbers that measure the distances in terms of inches, millimetres and centimetres.

1.2.2 *Approaches of Face Recognition*

Most current face recognition approaches use one of three methods of algorithmic representation: holistic approaches, local feature approaches, and advanced approaches called geometrical approaches that use the holistic and local approaches which will be discussed in Chapter 5.

A) Holistic approaches

The traditional way to identify people in the past was based on facial geometry. In a surveillance system, if an unknown face appears more than one time then it is stored in a database for further recognition. Meanwhile, face recognition techniques can be divided into two groups based on whether the representation of the face uses holistic texture features and is applied to either the whole face or specific regions, using the facial features of the mouth, eyes, brows, cheeks, and the geometric relationships between them.

In face recognition, the tested images of the face regions are compared with others in the database based on biometric features that are constant throughout the life of an individual regardless of age and environmental conditions. These techniques can be used to identify faces despite variations in perspective. One of the best examples of holistic methods is Eigenface of PCA [11, 12] and LDA [13]. Likewise, Fisher discriminant analysis [14] has been applied to expressly provide discrimination among classes, when multiple training data per class are available. Through the training process, the ratio of between-class difference to within-class difference is maximised to find a basic set of vectors that best discriminate among the classes. Holistic approaches use only limited points of interest which do not destroy any of the information in the image and this is one of the main advantages of these approaches [15, 16]. The Eigenface approach appears to be a fast, simple, and practical method, and has become the most widely used face recognition technique. However, it does not provide variance over changes in poses and scales. The technology used in these approaches are considered to be expensive and require a high degree of correlation between the test and training images, and they also do not perform effectively under large variations in pose, scale and illumination [17, 18].

B) Local approaches

Local matching approaches have been published which accomplish impressive results in face recognition [19-26]. These methods are based on locating several facial features and then classifying the faces by comparing and combining the corresponding local statistics. Recent surveys devote considerable attention to local matching methods [16, 19]. Research into local approaches has developed the robustness of face recognition from holistic approaches, such as modular PCA (MPCA), as a substitute for holistic approaches for entire images. MPCA establishes multiple-eigenspaces around facial components such as the eyes, nose, and mouth to form eigenfeature as shown in Figure 1.4.

Likewise, holistic methods can become modular, such as in modular FDA, with similar gains and losses. These approaches can only alleviate the problems of pose variation to a certain extent, because with local facial features, some image distortions brought about by pose variations still exist. The benefit of localising image matching is also gained at the cost of the extra requirements of feature detection [27]. However, different types of methods can be used in these approaches, such as feature-based (physical) methods. Data for local features such as eyes, nose, and mouth are gathered by extracting their locations

and local statistics (geometric and appearance data), which are then fed into a structural classifier.

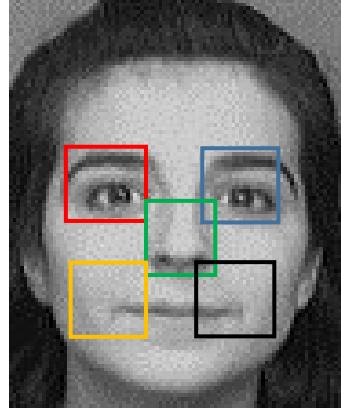


Figure 1.4: PCA eigenfeatures for eyes, nose, and the edge of the mouth [25].

A big challenge for feature extraction methods is feature restoration. The same when the system tries to retrieve features that are invisible due to significant variations, such as changes in the position of the head when matching a frontal image with a profile image [6].

1.3 THE CHALLENGES IN FACE RECOGNITION AND THIS WORK

Previous studies have assumed that face recognition is a specific and challenging case of object recognition. One difficulty arises from the common issue of recognising objects from 2-D images from a side view of the objects where faces appear to be roughly alike, and differences between them are quite subtle [16]. Thus, frontal face view is very compact in image space, which makes it fundamentally impossible for traditional pattern recognition techniques to discriminate among them with a sharp movement of the pose face [28]. Furthermore, the human face does not have regular geometrical shapes that make it inflexible for testing and comprising in terms of finding unique features connecting for the same person. Sources of variation in facial appearance can be categorised into two groups: intrinsic and extrinsic factors [28]. Camera calibration from the proposed system involves extracting the intrinsic and extrinsic information of the camera from a 2-D image. The intrinsic factor of the face recognition requires the information that is measured from the calibration because it is pure to the physical nature and independent of the observer. Whereas, extrinsic factors in face recognition are based

on the appearance of the face changing according to light conditions and illumination, scale and imaging parameters.

This work focuses on face recognition across pose, and the intrinsic parameters of the camera calibration can be used to recognise the intrinsic factor of face recognition that is applied on face recognition from fundamental factors. Evaluations of state-of-the-art recognition techniques conducted during recent years, such as the FERET evaluations [29, 30] are used in this study to demonstrate the results and compare it with different methods.

1.4 WORK MOTIVATION

The motivation for this work connected the geometrical and mathematical aspects of camera calibration methods. As well as, suitable application of face recognition to solve the problems caused by various movements across pose.

1.4.1 *Geometric parameters*

The techniques of face recognition and detection widely use the measurements of regular features of the face. Thus, the main approaches for face recognition can be classified into two categories: appearance based and feature based [31]. Both of these methods are based on extracting geometrical parameters to measure the facial parts regions such as nose, eyes. (Explained further in Section 1.2.1)

The geometrical parameters for the face cannot be changed when the head is moving at different angles. In other words, this work assumes that human faces in real life have the same distances between features and different horizontal poses. Therefore, the changes will effect the object view in the video frames and pictures; for example, the distance between the eyes can look shorter for a side view compared to the front view. Therefore, another motivating factor for this work is to measure the geometrical facial distances from 2D face images in actual distance for the real world, or equivalent to it. The measurements from videos and pictures are based on two types of camera parameters: intrinsic and extrinsic. The main focus is to infer the parameters of the camera from a picture without known information about the camera. Then, from these parameters, a mathematical technique is applied to measure real dimensions from a 2-D image.

These two types of parameters give information about the camera, and thus information for each frame or captured picture. Therefore, these parameters are used to determine the depth and orientations of objects in the scene, such as in 3-D approaches. Therefore, camera calibration method is used to gather this information in most applications and techniques, for example using stereo and mono cameras. Hence, the first motivation of this research is to implement and improve a unique method of camera calibration for different (square, rectangular and triangular) geometrical shapes. The method is based on a geometrical mathematical approach to perspective projection.

In this work, camera calibration technique is used to infer information from low-resolution images and to use information from one view of an image to solve problems in a different position as in cross-pose face recognition. The system of camera calibration builds on a linear system to gain nonlinear information such as in focal length parameter as depth in camera calibration.

1.4.2 *Face recognition across poses*

Most of face recognition systems deal with the output of the camera as an image or video frame. The process performs the identification or verification of the subject or subjects that appear in the camera output picture. The face recognition system is based on the three steps shown in Figure 1.5 [32].



Figure 1.5: Face recognition system

The face detection and feature extraction phases could run simultaneously depending on the different methods and purposes of face recognition applications in biometrics, such as video surveillance, robotics and control of man-machines. Nowadays, difficulties in face recognition techniques arise in handling different poses when the recognition of faces is subjective in-depth rotations. Therefore, differences in face image caused by rotation are often larger than the inter-person differences used to distinguish between identities [27]. Each of the three components that are shown in Figure 1.5 involves challenges that motivate the researchers to find the solution to the problems pose variations and to get better results. Therefore, the main motivation of this thesis is to develop a new technique

for face recognition based on solving the problems caused by variations in pose. The technique proposed requires fewer less complex methods, which tackle the difficulties of face tracking from different orientations. Moreover, the feature extraction technique that includes detecting the position of the pupils of the eyes and measuring the distance between them.

1.4.3 *Quadtree decomposition*

The previous two aspects of the motivation for this work are based on the effect of camera parameters on the captured pictures. Whereas, processing an extensive database of images is affected by the size and dimensions of the images, which can lead to a lengthy process which gives less promising results.

On the other hand, decomposition of an image is a level congruency map of the image that constructs a feature structure for the image. The typical way used to compress the feature structure is to apply a threshold, reducing an image representation to a simple binary structure. Thresholding is subjective and can eliminate much of the valuable information in the picture. In this thesis, the improvement from the previous motivations encourages the study to find another enhancement for the results. Thus, quadtree decomposition is a method that divides a square image into four equal sized blocks. Each block is tested to see if it meets some criterion of homogeneity. The test criterion is applied to the blocks that are not divided any further. Therefore, the proposed use of quadtree depends on determining regions of interest for intensity analysis in images. Hence, quadtree decomposition is used to enhance the tracking of face images across poses, as will be discussed in Chapter 5. This kind of decomposition has been used before in face recognition in different ways. The quadtree decomposition is introduced based on observations in the literature [33, 34] where there is high overlap between various parts of facial regions, and the relationships between local regions play a major role in classification. Zhang et al. used random quadtree decomposition [35] to split templates into small non-overlapping regions, while quadtree decomposition templates into overlapping areas.

According to the results of decomposition, face images in the original training data are reorganised to generate new samples. Quadtree decomposition has mostly been used to control illumination and reduce the dimensional information of images. It has also been used in segmentation and for different enhancements in image analysis [36]. Hence, in

this work it is used in face tracking with various types of movement that are based on previous enhanced information.

1.5 CONTRIBUTIONS

This thesis presents three novel methods representing significant improvements in performance regarding face recognition and computer vision. The contributions of this thesis can be outlined in five areas:

1. Implementing and developing the camera calibration system

The first novel contribution of this work has been to implement the mathematic approach of prospective projection to computer vision camera calibration system. The new method of camera calibration is distinguished from the rest of the methods in low cost requirements for image and camera with high accuracy results. The proposed work introduces a linear camera calibration technique, which does not need prior information about the camera. The technique is a mathematical approach to perspective projection used to infer the main intrinsic and extrinsic parameters of the camera. The parameter of the camera has to process to deal with next step of measurement technique. Consequently, the proposed technique of camera calibration can use low grade of pictures from any camera. Relatively, the results tested different pictures that are captured from different distances and resolution.

2. Camera measurement technique (CMT)

The novel algorithm developed in this thesis relates to the camera measurement technique (CMT) that calculates the depth from 2D images to cover fixed dimensions when the object is moving, as is the problem of across pose face recognition. The CMT can measure regular objects such as free shapes, and irregular ones such as geometrical shapes. It is novel because of the flexibility to determine the difficult (sharp) angles in free shapes that are selected manually.

The geometrical shapes are determined to work fully automatic, and the accuracy of the proposed work is based on the selected points to determine the distance between the points.

3. Face detection

The new algorithm developed in this thesis enhances the face detection to track face views in a various posses. The new method of face detection meets with the aim of the system to recognise face views in various poses. The main algorithm uses computer vision between two methods: a basic face tracking system and an image decomposition system in face quadtree decomposition (FQD). Each method is implemented and developed to overcome the problems with popular face detection and tracking approaches that encounter difficulties in tracking when a face is in wide angle views.

4. Features selection

Another contribution of the novel algorithm in this thesis is a step responsible for selecting the landmark on the face such as eyes, nose and mouth. The face recognition system in this study focuses on two points of interest on the face, which are the centres of the eyes. Thus, the selected landmark points will have a challenge against changes in facial poses. Hence, the methodology of feature extraction developed in this thesis represents a new technique to extract local features to be used in a geometric face recognition approach. The technique developed is named FQD, which includes the face quadtree decomposition method and detection features.

5. Face recognition across pose

This work proposes a new solution for face recognition with variations in pose using perspective projection from novel camera calibration method. The contributions are based on each other and share the information to feed to the next level of the system. This stage of contribution shows the last part of the system that is connected to all the previous contributions to get the expected results of face recognition across the pose. The proposed novel method is based on CMT and FQD, and is called face recognition camera measurement technique (FCMT). The recognition system applies the system for each picture, and the picture for each person is compared with gallery (front view) picture across horizontal view. Training and testing sets were used in identification mode to measure original matching scores.

1.6 THESIS STRUCTURE

This research combines different methods from image processing and computer vision. The work achieves an enhancement in face recognition by proposing a novel system of biometric face recognition. The system is developed to solve the problems of recognising a face in various poses. The contributions proposed in this work cover camera calibration, measurement of dimensions from 2-D images (2-5D from 2-D), feature extraction, face detection, and face recognition across pose. The work of this study is summarised and discusses the new methods in Chapter 3, Chapter 4 and Chapter 5.

Chapter 1 then introduces the first step of the research in terms of implementing and developing the camera calibration system. An overview is given of the proposed method of calibration and its core, and then the characteristics are described after the main approach of regular perspective projection is explained also.

After introducing the contributions in the first chapter, Chapter 2 describes the literature of the main contributions of camera calibration and face recognition, with a brief review of the problems to be solved in each part. Likewise, the classification of the camera calibration related by used approach are explaind. Meanwhile, this chapter classifies the face recognition and discuss the issues of this study.

Chapter 3 introduces the first step of the work of implementing and developing the camera calibration system. This chapter explains the methodology and the base system of the thesis. Starting from the overview of the proposed method of the calibration, then care characteristic, after the mean approach of geometrical perspective projection use. To approve and support the proposed work, at the end of Chapter 3 shows the test of the system and the results in different environments to prove and support the proposed work that is obtained in three stages: camera parameters, focal length and application.

Chapter 4 demonstrates a novel algorithm measurement from camera calibration technique, the system of the measurement was tested with the results to support the theory of the research system. The statistical approach used by the proposed algorithm based on perspective projection is explained. This is followed by a test of system calibration for each angle of rotation. After this, dimensions are infered from an image using automatic calibration, and the types of shapes of the calibration are explaind. The results of the tests

are described of the proposed method using pictures captured from the different resolutions of mobile camera.

In Chapter 5, the face recognition system is proposed using the results described in previous chapters. It also describes the different approaches to face recognition by comparison, and the benefits of the proposed approach are demonstrated. Moreover, this chapter proposes a novel method for face recognition from various poses called the Face Camera Measurement Technique (FCMT). The results of the experiments are then reported comparing different methods that used the same environment and database to deal with the problem of variations in pose.

Finally, Chapter 6 concludes this thesis by summarising the main points of each chapter, discussing the results and suggestions for future work.

Chapter 2 : CAMERA CALIBRATION AND FACE RECOGNITION

2.1 CHAPTER OUTLINE

Face recognition technology has been a fascinating area for the last 40 years, with considerable research attention since 1990 [37]. Meanwhile, cameras have become an everyday part of our life. Camera and applications are widely used in several areas based on face recognition, such as access control, surveillance, driving licenses and passports, image analysis and home security. The face recognition from the various poses and geometrical camera calibration are connected, and can complement each other. This study aims to detect the essential features of the face from 2-D data using and perspective projection of camera calibration.

The triangulation principle of measuring distance used to be a common practice among engineers in ancient times, as seen in the pyramids. These methods are still applicable today. A distance-measuring device of this type has been reported, and a modern commercial gauge based on this principle has been described [22]. Hence, the proposed work of measuring distance technique used geometrical mathematics in camera calibration. The technique applied on the face features for recognition purposes, such as ear-to-ear distance was used to identify subjects and recognise the individuals [23]. Triangulation has a significant advantage over interference techniques, as the absolute distance by optical interferometry to the target is measured.

This chapter links the methodology of camera calibration and face recognition algorithms and the common principles between these approaches is included as part of this thesis. Therefore, the main methods of camera calibration are reviewed with the recent face recognition technology that has received significant attention within this field of study. Biometric systems and camera calibration applications are introduced with particular reference to the main problems encountered with their use.

Then the key technical challenges in solving such problems are outlined. The last part of this chapter focuses on image decomposition techniques relevant to the methods used in face recognition systems.

2.2 INTRODUCTION

Cameras were invented a long time ago; Johann Zahn designed the first camera in 1685, while Joseph Nicephore Niepce took the first photograph in 1814 [38]. When cameras were first introduced, they were expensive, and large amounts of money were needed to own one. However, the pinhole cameras was then introduced in the late 19th century. These cameras were cheap and they became commonly used in everyday life. Pinhole cameras have constant distortion that has to be corrected, and from this point camera calibration comes into research area.

Camera technologies have improved rapidly, and the research into cameras has recently expanded into several areas such as image processing and computer vision. One of the most interesting research areas in camera calibration is how to fix the relationship between the camera image's original units (pixels) and real world units, for example, millimetres or inches. Therefore, it is a significant step to achieve an accurate representation of reality from 2-D images. In actual life, people have a marvellous capability to identify objects regardless of a variety of features dimensions geometries, viewpoints, poses and lighting conditions, and these skills are the same in principle as those in face recognition [39, 40]. Nevertheless, our brains do not have sufficient ability in some respects, for instance in relating and dealing with large amounts of data and the aptitude to accomplish multiple recognition tasks rapidly. Hence, the examination and understanding of how people achieve facial identification is an essential research topic for numerous scientists. These scientists have been trying to overcome the weaknesses of the abilities of the human brain.

However, face recognition has been regarded as an efficient method for subject identification and tracking using information such as camera position, orientation, and focal length. This information can be obtained from parameters of the camera that is used in face recognition, for example when the external parameters selects the orientation of the objects in space and the internal information gives the focal length, when the focal length is popular in 3-D reconstruction and object orientations the face recognition across pose. Therefore, the study system provides knowledge of the scene that can be used for object detection or tracking, which makes it suitable for various surveillance systems. There are two main types of approach used in face identification; holistic and local approaches. Furthermore, another new method in the field of face recognition is called the geometrical approach.

This method combines the two types of technique to achieve better results in a different way. Hence, the research is extended to include different technique that feed the geometrical approach of face recognition from different poses.



Figure 2.1: Axis PTZ IP [41]

However, camera calibration began to be used widely in pan, tilt and zoom (PTZ) cameras [42], as shown in Figure 2.1. These cameras have a significant role in face tracking, and recognition at a distance as an active part of visual surveillance systems. Tracking systems using a single monocular camera have been extensively applied in the past. Monocular systems use two kinds of camera: static and PTZ cameras.

Static camera tracking systems often use background subtraction methods to obtain information about objects in the foreground [43]. Meanwhile, PTZ cameras allow the monitoring system [44, 45] to rotate and zoom in and out to maintain an object of interest within the field of view. The combination of camera calibration and face recognition methods is relatively new and is used to solve problems related to cost, real dimensions, three dimensionality and distance. Face images have low resolution when they are captured at a distance, larger than 5 metres, thus degrading the performance of face matching. Park et al. [14] addressed this problem by proposing an imaging system consisting of a single view and PTZ cameras to acquire high-resolution face images up to a distance of 12 metres. The proposed system included a new coaxial concentric camera configuration of static and PTZ cameras to achieve the distance invariance using camera calibration.

The proposed work uses the standard methodology and uses the parameters of pan, tilt and zoom that are used in PTZ cameras, the names of angles have been changed as some parameters have not been applied, such as zoom parameter. The proposed system enables the estimation of the distance without the need of a mechanical camera that is challenging the high cost of

cameras such as PTZ. These two issues require further improvement and have motivated the proposition to use camera calibration in face recognition. Moreover, in the last three years, several researchers have tried to improve on this work and use the major concept of the link between camera calibration and face recognition for specific purposes [46, 47].

The advantage of the static camera in tracking face systems is that it is easy to operate. Meanwhile, PTZ camera monitoring systems can cover a wide field of view and have a large scale, which are advantages of these systems. However a significant disadvantage of static systems is that it is hugely difficult to acquire high-resolution images of objects interest which would be very useful for behavioural analysis and the detection of abnormal behaviour [48]. Furthermore, these systems lose the panoramic information concerning motion because it has limited view and fixed positions, and it is difficult to obtain information about the object's location in the scene directly. Another problem is the difficulty of foreground extraction because of the non-stationary background caused by camera movement [49-51]. These systems connecting the camera calibration with face recognition use information from standard camera calibration of pan, tilt and zoom cameras, and all recognition is based on real-time processing. Also, different types of calibration can use various parameters of the image and can deal with large databases without a standard camera.

2.3 Classification of camera calibration

Camera calibration is a major step in computational geometry within computer vision techniques. While some information regarding the measurement of images can be obtained using un-calibrated cameras [52], camera calibration is necessary when metric information distances are required. Camera calibration is divided into two phases:

- Camera modelling deals with the mathematical approximation of the physical and optical behaviour of the sensor using a set of parameters.
- The second phase of camera calibration deals with the use of direct or iterative methods to estimate the values of parameters.

Two kinds of parameters are involved. The fundamental parameters relate to the internal geometry and optical characteristics of the image sensor. These basic parameters determine how light is projected through the lens onto the image plane of the sensor, such as the camera pinhole.

Extrinsic parameters measure the position and orientation of the camera in a world coordinate system, which provides metric information for the user in virtual coordinate system instead of the camera coordinate system.

Camera calibration can be classified according to numerous different measurements:

1. Linear and non-linear camera calibration [53, 54].
2. Intrinsic and extrinsic camera calibration. Intrinsic calibration is related only to obtaining the physical and optical measurements of the camera [52, 55, 56]. Extrinsic calibration concerns the measurement of the position and orientation of the camera in the scene [57, 58]
3. Implicit and explicit camera calibration [59]. Implicit calibration is the process of calibrating a camera without explicitly computing its physical parameters. While implicit calibration can be used in 3-D measurement and image reconstruction.
4. The geometrical camera calibration [53, 60, 61] is usually performed by imaging a calibration object whose geometrical properties are known. Different methods have been applied for this kind of calibration using 3-D points, or even a reduced set of 3-D points [62].

The proposed work deals with geometrical camera parameters using two vanish points projections to extract the lines and get the intrinsic and extrinsic parameters that are used in geometrics face recognition approach. The method of vanish point in camera calibration has become popular in face recognition and usually requires expensive tools, images from un-calibrated cameras, and complex algorithms.

2.3.1 *Projective geometry and camera calibration*

Projective geometry is an essential tool for modelling and analysis in computer vision. The interesting idea of using projection points comes from the imaginary aspect of the perspective projection; when these points are in an infinity projective space, there is a possibility to meet other points that have the same characteristics, and then these present a new shape in image plan.

Using vanish points projection has a combinations of results; in terms of these points at infinity in a projective space possibly meeting in the same way with other points, and presenting a new shape in the image plan.

Projective geometry allows short linear algebraic descriptions to be made of the geometrical entities and relationships that are faced with vision problems. A general n -dimensional projective space is defined by $2+n$ basis points (points of interest or face features in this study), and a point in space is represented by a dimensional homogeneous vector defined to scale factor. However, projective transformations of the space are represented by $(n+1) \times (n+1)$ homogeneous matrices, similarly defined up to scale [63]. There are specific representations of geometric entities in one, two and three dimensions and different techniques and approaches are used, starting from dividing the projection methods to linear and non-linear (Figure 2.2). The following section discusses linear perspectives and projective geometry as used in the understanding of computer vision.

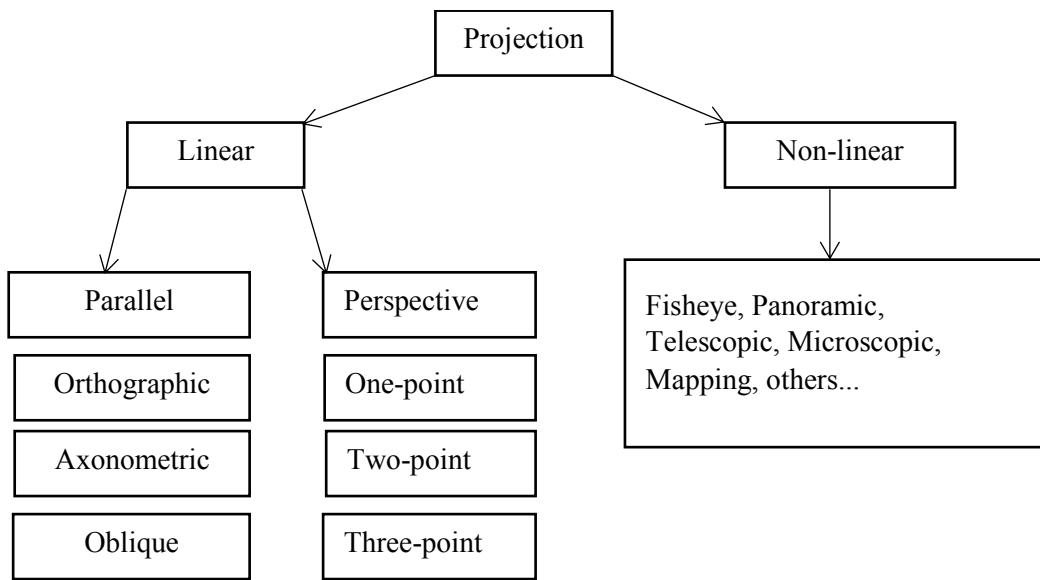


Figure 2.2: Classification of geometric projection

2.3.1.1 *The strengths of perspective projection*

The development of linear perspective in art lead to the development of projective geometry as a mathematical discipline, starting in the Italian Renaissance of the 15th century [64]. The theory of linear perspective was popular between artists, but there remained no standard techniques. The individual artist applied his personal technique to represent depth and space as well as he could until the beginning of the 15th century when Filippo Brunelleschi invented the linear perspective [65]. Thus the artists that followed adopted the concept of linear perspective and continued to modify their techniques. For example, Masaccio [66] adopted a single

vanishing point to portray perspective and space, as can be seen in Figure 2.3. The main point was how to produce an impression of 3-D depth in architectural painting.

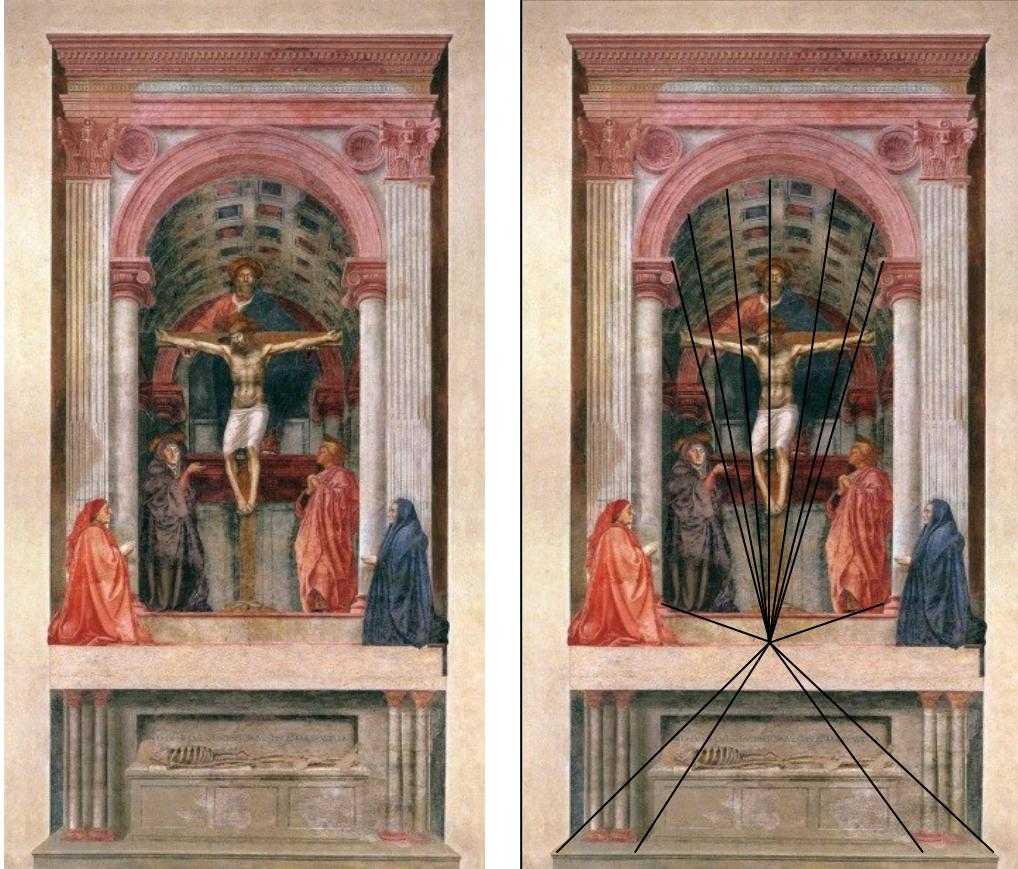


Figure 2.3: The Holy Trinity by Masaccio[67]

Artists have used vanishing points to derive geometrical coordinates from constructions. For instance, the projected square can be split into four equal sub-squares or finding the projection in a rectangle can be found if two of its sides are known. In Figure 2.4, the perspective projection prepared in this basic scene is used to simplify the method of projection that shows the landscape and the horizon. Here, the boundaries of the road are set in parallel as lines in space, while the others appears as retreat to the horizon direction. The line of the horizon is formed by the vanishing directions of the ground plane. The vanishing directions meet at a vanishing point when the horizontal parallel lines cross the corresponding horizon, and the lines are stable even if they lie at different heights above the ground plane. As well as the directions of vanish points from two vertical planes appear to come together in the distance and intersect the line at infinity. Moreover, the meeting points of the horizon lines stay constant as the viewer is changing at the infinity scene. The road always seems to vanish in the same direction on the horizon, and the

stars stay fixed for the viewer during stable movement. Therefore, the lines of vision to infinitely distant points are always parallel, because they meet at infinity.

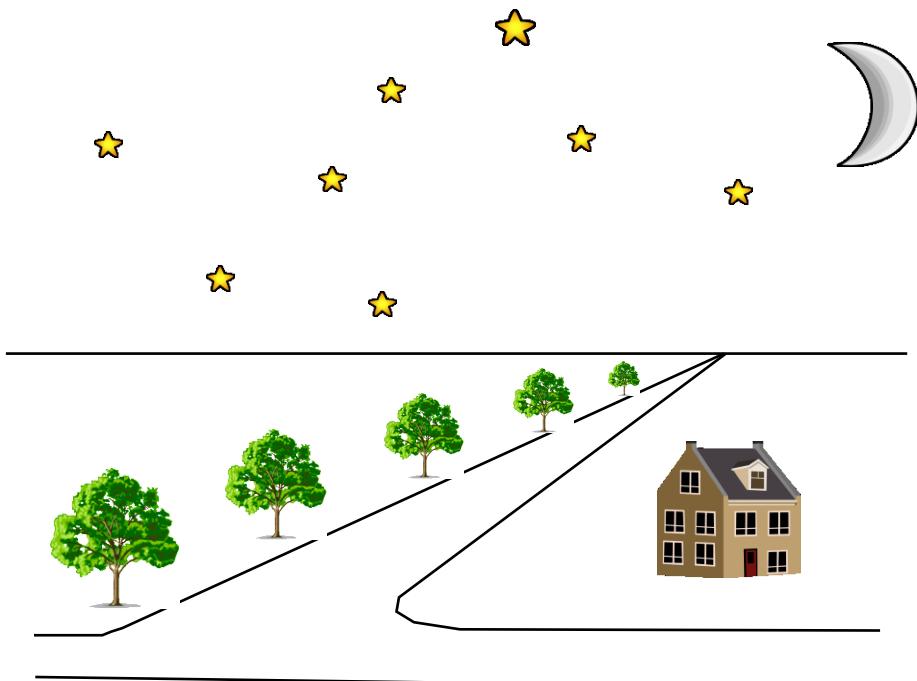


Figure 2.4: Landscape and the horizon

The concept of determining the geometrical dimensions has to be extended to handle phenomena at the infinity meeting with the conception of proposed work. In turn, through further studies within the field of projective geometry, mathematical investigations have been able to explain the principles underlying the technique of linear perspective. Considerable use is made of vanishing points and several practically useful geometric constructions can be derived; for example, to split a projected square into four equal sub-squares, or to find the projection of a parallelogram when two of its sides are known.

2.2.1 Perspective projection in camera

The transformation of perspective projection between the paintings and 2D capture pictures shows new definitions in perspective projection defined as the following:

- Centre of projection
- Focal length
- Principal point
- Principal axis

The position of the centre of the projection for the camera is at the origin O of the 3-D reference edge of the space. From Figure 2.5, the image plane (Π) is parallel to the Y-axis and X-axis plane and explains the distance of f (focal length) alongside with Z-axis from O. Also we can see that a 3D point $P=(X,Y,Z)$ is projected on the camera's image plane at coordinate $P_c = (u, v)$. However, the principal point (O) is the orthogonal projection of the centre of the projection against Π . When the X-axis corresponds to this projection line, then it is the principal axis.

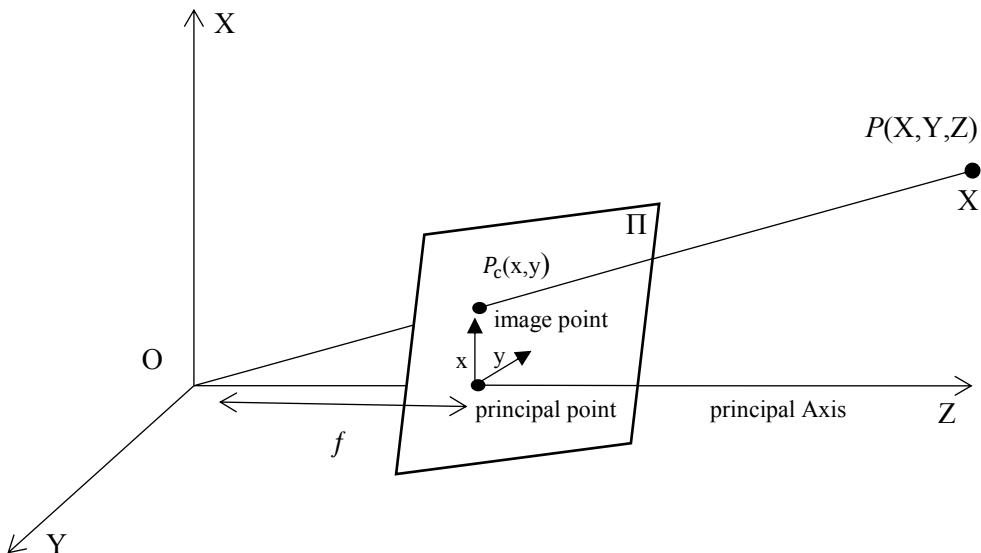


Figure 2.5: Standard perspective projection in camera scene

Therefore, these points of perspective projection can be used to infer the information concerning camera calibration from 2-D [68]. Let (x, y) be the 2-D coordinates of P_c , and (X, Y, Z) are the 3-D coordinates of P . A direct application of Thale's theorem [69], as shown in equations 2.1 and 2.2 calculate x and y :

$$x = \frac{fx}{Z} \quad (2.1)$$

$$y = \frac{fy}{Z} \quad (2.2)$$

Where the optical ray, the line joining X and O, meets the image plane. Suppose that f (focal length)=1 is the scale different of the image and include a full camera calibration into the model.

In homogeneous coordinates [70], represent a 2D point (x, y) by a 3D point $(\dot{x}, \dot{y}, \dot{z})$ by adding a “fictitious” third coordinate. By convention, we specify that given $(\dot{x}, \dot{y}, \dot{z})$ we can recover the 2D point (x, y) as

$$x = \frac{\dot{x}}{\dot{z}}$$

$$y = \frac{\dot{y}}{\dot{z}}$$

In Camera Coordinates, equations 2.1 and 2.2 are transformed to equation 2.3 as follows:

$$\begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & f & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (2.3)$$

The centre of projection of the image coordinates is not the principal point in the 2-D image and scaling along each image axis is different. Thus, the image coordinates undertake further transformation described by matrix K . Correspondingly, the 3-D coordinate system does not usually match with the perspective reference edges, and so the 3-D coordinates are subject to a Euclidean motion defined by some matrix M , that multiplying the affine inhomogeneous coordinates of a point by a 2×2 matrix M is equivalent to multiplying its homogeneous coordinates, giving a new matrix as in equation (2.4):

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \sim K = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & f & 0 \end{pmatrix} M \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (2.4)$$

Camera calibration uses M to localise the 3-D position and poses and the matrix M is the perspective transformation matrix, which relates 3D world coordinates and 2D image coordinates. Hence, it will have 6 degrees of freedom, which characterise the extrinsic camera parameters. Meanwhile, K is independent of camera position and contains the intrinsic parameters of the camera and is represented in a triangular matrix as in the following equation 2.5 [71]:

$$K = \begin{pmatrix} S_x & S_\theta & u_o \\ 0 & S_y & v_0 \\ 0 & 0 & 1 \end{pmatrix} \quad (2.5)$$

Here S_x and S_y stand for the scaling along X and Y axes of the image plane, while S_θ represents the non-orthogonality between the axes (usually $S_\theta \approx 0$). The intersection of the principal axis and the image plane gives the principal point, and (u_o, v_0) are its coordinates of the principal point the intersection of the principal axis and the image plane (x, y). In conclusion, the perspective projection model in a homogeneous coordinate system is designated using a linear equation, which is useful for the accuracy of the system. Most of the material in this section is taken from a tutorial on projective geometry for image analysis [71] in order to briefly introduce the working tools employed in the next chapter. The next section describes the pinhole model-based camera, and its use in projective reconstruction from image correspondences without calibration and an approach to subsequent metric reconstruction.

2.2.2 Pinhole camera model

One of the most significant uses of a perspective projection is for a perspective model, which is the major work of camera geometry or as known, the pinhole camera model. The construction of the pinhole camera as shown in Figure 2.6 allows light rays from in front of the camera to be focused at the pinhole and to be projected on top of the image plane through the camera. The image plane is often considered to lie in front of the camera at the same distance from the pinhole, which helps to get the un-reversed image [72]. The distance from the focal point to the image plane is the focal length.

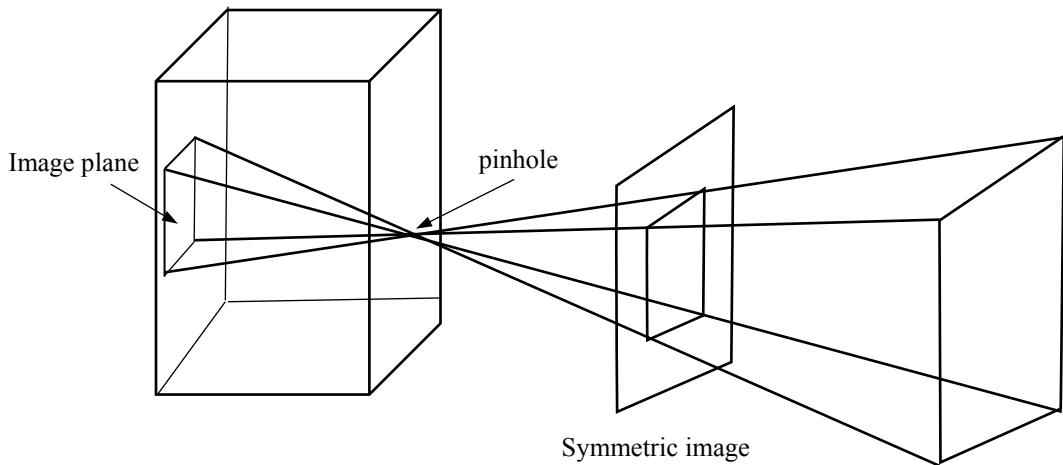


Figure 2.6: Pinhole camera structure

The principle of current proposed work is based on the sutural of pinhole camera model. Each point on the object projected in space from the straight line through the origin centre that is located at (X_0, Y_0, Z_0) into the image plane, and the z-axis of the camera frame is perpendicular to the image plane [73]. Overall, pinhole camera operation shows the mapping of a 3-D scene in a 2-D image and can be summed up by the following two conditions:

- Close objects are projected to become bigger
- Distant objects are projected onto the same region.

These conditions are used for camera calibration to separate between the Euclidean structure of the camera scene, and the camera's zooming features [74]. The geometrical methods of camera calibration take a sampling of image scene to solve the parameters in the linear interpolation method. Haralick [75] was the first to use the main idea of a pinhole camera; He showed how to use the 2-D perspective projection of a rectangle of unknown size, and position it in 3-D space to determine the camera visual angles and parameters on the planes of the rectangle. The obvious differences between approaches concern the form of the interpolation functions, and the mathematical techniques used to derive the parameters, and the main intention of most calibration work has been the solution of the back projection from the internal and external parameters of the camera.

2.3 CAMERA AND PROJECTIVE PARAMETERS

Camera calibration in photogrammetry arranges the coordinates of three 3-D reference points and the corresponding positions of their perspective projection, which helps in allocating the position of the camera as well as its aspect direction. Photogrammetric methods need the internal orientation data in order to describe the metric characteristics of the camera. The internal and external camera parameters are inferred from one or more images of a scene. Hence, this gives information about three-dimensional points in the world coordinate system. The process that is responsible for this transformation is called reconstruction, which uses matched points between one or more images to determine a reconstruction up to a projective transformation of the valid world points. The calibration of cameras involves computing this projective transformation to rectify the reconstruction.

One of the most important tasks in computer vision and photogrammetry is based on a problem called space resection [76]. In this area, many studies showed clarifications from image views such as Koenderink et al. [77] who first showed un-calibrated views of a scene captured with cameras allowing for or preserving parallel relationships. The camera calibration determining these relationships in the projective transformation of the virtual world points is used to rectify the reconstruction of 3D objects.

However, the estimation of intrinsic and extrinsic camera parameters is based on a camera model using a one camera calibration technique. Image analysis and reconstruction of objects from their images have been an active area of research to solve some problems of multiple view technologies, such as the cost rate, accuracy-complexity and relation to the limited angular field of view (FOV) of some cameras [78, 79]. In this thesis, the system calibration of uses a calibration technique that drives camera information parameters from one view. In general, this technique takes a terminal point as a fixed point and rotates a 1-D object around it.

However, the geometric model of the pinhole camera still accurately describes the main functioning principles of many modern cameras. A simple approach for appropriate parameter initialization within this model was proposed in 1999 by Zhang [80] who provided a flexible new technique to calibrate a camera easily. The technique required the camera to observe a planar pattern shown at least two different orientations. Later, Zhang [81] extended his work to propose a camera calibration method that uses a 1-D object that is considered to be simpler than that of the 2-D model plane in structure. Zhang's calibration approach has been proved successful in extensive subsequent research. Moreover, 2-D plane-based calibration techniques require the observation of a planar pattern shown in a few different orientations [53, 82]. As opposed to Tsai's technique [22], knowledge of the plane's motion is not necessary. Because almost anyone can design such a calibration pattern by him/herself, as this set-up is easier for camera calibration [53]. More recently, Zhang [83] has taken a newer technology into account to develop his early camera calibration system. He developed a recalibration technique using a special card works with the Kinect. If users find that the Kinect sensor is not responding accurately to their actions, then it possible to recalibrate it by showing it the card [83]. The Kinect calibration card provides a solution when the Kinect sensor, due to the calibration between the IR projector and the IR camera becoming invalid, produces inaccurate depth values.

The task of camera calibration is conducted to determine the parameters of the transformation between an object in 3-D space, and the 2-D image observed by the camera from visual information (images). This transformation involves extrinsic parameters that are responsible for the orientation (rotation) and location of the camera, while the intrinsic parameters are characteristics of the camera. These parameters inferred from camera calibration are used in camera surveillance when static and PTZ cameras need to be calibrated into a common coordinate system to communicate with each other [84]. The calibration is performed between the pixel coordinates of the static camera, the pan and tilt values from the camera calibration and these internal parameters.

Moreover, calibration is a major step in the reconstruction and recognition of 3-D face images in face recognition systems [85]. The next section covers the main points that are related to camera calibration with face recognition.

2.4 Classification of face recognition techniques

Face recognition is a significant task in biometric systems developed for surveillance purposes, such as searching for wanted criminals, suspected terrorists, and for home or company monitoring. It is a wide area for the researchers to investigate and develop new applications to accomplish a progressive step in systems of security. The term face recognition refers to the identification, using computational algorithms of an unknown face image. This operation can be conducted by comparing the unknown face with the faces stored in a database or one pose from different poses. The stage of face recognition can be divided into: face location; feature extraction and facial image classification.

Various face recognition algorithms exist and each has different characteristics depending on the purpose of their use in recognition system. Thus, they have advantages and limitations. Therefore, it is an open area for research work, and much research has been published on face recognition.

The traditional operation of face recognition algorithms is based on identifying facial features, by extracting landmarks or features from an image of a subject's face. For example, the position, size, and shape of the eyes, nose, cheekbones, and jaws [16]. However, many studies use synthesised facial images which are not directly used for face recognition but instead an identity subspace is learned from the AAM parameters to distinguish between different identities [86].

Blanz et al. [87] extended 2-D AAMs to 3-D morphable models with densified shape vertices. More recently, Mostafa et al. [12] proposed a similar work in which active shape models (ASMs) [13] were utilised for facial landmark detection. Promising results were achieved in pose-invariant face recognition with estimated model parameters. However, the computational cost of fitting such a dense model is high, this prevents this method from being suitable for real-time applications. Further studies have been undertaken in this field providing an illumination model [88] that can extend the proposed study for future analysis. Moreover, the illumination model in recognition is more efficient for pose invariant face representation.

However, dimensionality reduction is significant in facial recognition algorithms so as to avoid the overfitting of data, and reduce computational costs due to less complexity. [89]. Dimension reduction approaches may use pattern recognition as a process of reducing the number of random features under consideration [90]. The process of dimension reduction can be classified based on the algorithm used as follows [91]:

1. Global or local dimensionality reduction

In global approaches the algorithms are based on preserving the global structure of data samples, while local methods preserve the local structure.

2. Supervised or unsupervised methods.

Unsupervised dimensionality reduction methods do not use any class information, whereas the supervised methods can take advantage of any class information available in the data, and may sometimes need further information regarding classes. Al-Shiha et al. [3] have recently proposed supervised and unsupervised multi-linear neighbourhood preserving projection (MNPP) methods for face recognition to remedy this problem. Their method used discriminative feature extraction by extending the original NPP to its multi-linear form, which is suitable for use in face recognition and biometric security applications. The method effectively circumvents the problems of small sample size and the dimensionality dilemma in face recognition tasks compared to other methods such as PCA or LDA.

3. Linear, multi-linear and nonlinear algorithms

Linear and multi-linear algorithms are based on the space of data used while nonlinear algorithms depend on a mapping transformation function.

4. Feature extraction and feature selection.

Features are extracted by applying a mapping function to the linear or multi-linear dimensional data space. Therefore, a subset of the original features are found and selected. However, these changes in the size of the initial data space into a smaller space reduce a few dimensions from the image. Meanwhile, some data analysis methods such as classification, regression, recognition and image retrieval can be conducted in the reduced space more precisely than in the original space. However, facial images can be easily obtained with inexpensive fixed cameras. Thus, good face recognition algorithms and appropriate reprocessing of the images can compensate for noise and slight variations in orientation, scale and illumination.

2.4.1 *Face recognition across poses.*

One of the key challenges encountered in modern face recognition techniques is the complication of supervision arising from varying poses that change in depth information during rotation. These changes caused by rotation are often larger than the inter-person differences used to distinguish between personal identities. Hence, these techniques based on a fixed number of samples per person are supposed to entail less effort in collecting data, and lower cost for storing and processing it.

In current surveys of face recognition techniques [92, 93], pose variation was acknowledged as one of the major unsolved problems in research into face recognition. Thus it has attracted the attention of many researchers in computer vision and pattern recognition. A recent review of face recognition across poses [27] has classified the techniques used in three categories: general algorithms, 2-D techniques, and 3-D approaches. In this thesis, the method of face recognition algorithms that have been used is a combination 2-D techniques and 3-D approaches as they share similar information. In Table 2.1, the classification of 2-D face recognition shows another class through different methods. It also classifies the face recognition across pose for these algorithms that are based on 3-D methods.

Table 2.1: 2-D techniques for face recognition across pose

Classification	Method
Real view-based matching	Beymer's method [94], panoramic view [95]
Pose transformation in image space	Parallel deformation [96], pose parameter manipulation [97], Active appearance models [98, 99], linear shape model [100], Eigen light-field [101].
Pose transformation in feature space	Kernel methods (kernel PCA [102], kernel FDA [103, 104], Expert fusion [105], correlation filters [106], local linear regression [107], tied factor analysis [108].

Table 2.2: Face recognition across pose with assistance of 3-D models

Classification	Method
Generic shape-based methods	Cylindrical 3-D pose recovery [109], Probabilistic geometry-assisted face recognition [110], Automatic texture synthesis [111].
Feature-based 3-D reconstruction	Composite deformable model [112], Jiang's method [113], multi-level quadratic variation minimisation [114].
Image-based 3-D reconstruction	Morphable model [115, 116], illumination cone model [117, 118] , stereo matching [119].

There are two main developments in face recognition techniques:

1. Improving the ability of general face recognition algorithms so that image variation can be tolerated.
2. Designing mechanisms that can eliminate or at least compensate the difficulties brought about by image changes such as pose variations according to original characteristics, such as through 2-D transformations or 3-D reconstructions (Table 2.2).

Recent advances toward automatic head-pose-invariant methods have focused more on facial expression recognition. These use expressions to select points of interest on the face and to recognise it. The methods of facial expression using image expression into the expression subspace can then synthesise new expression models [120]. Using the manufactured images for individual samples, more accurate estimations can be obtained of within-individual variability, which contributes to significant improvements in recognition. However, it can be classified as face-shape-based approaches and face-shape-free approaches. Face-shape-based approaches use either 2-D or 3-D face shape models to decouple rigid movements due to changes in head position and free movements owing to changes in facial expressions. Methods have been proposed which use 3-D facial geometry deformation to recognise facial expressions in 3-D images [121, 122]. These methods require high-quality images and 3-D geometrical data, and thus will not be widely applicable due to expensive and complex hardware requirements [123]. Meanwhile, other methods assume that the 3-D head pose is independent of free facial movements [124, 125] .

2.4.2 The one sample per person problem

The use of one sample for face recognition can be traced back to when the geometric-based methods were popular, where various configurations of features such as the distance between the two eyes were manually extracted from a single face image and stored as templates for later recognition [93, 126]. Furthermore, these face recognition techniques assume at least two samples of the same person are always available. However, many applications in law enforcement involve a significant volumes of face images that need more intelligent and less complex ways to process face information. Most of these applications are based on appearance techniques. Furthermore, the basis of appearance methods for large training is set by learning mechanisms to generate a good performance, thus, partly due to the high-dimensional representation of face images [126]. The solution for this problem is for dimensionality

reduction techniques to be employed. The most prominent subspace approach in the feature extraction process in face recognition is PCA [37] and Fisher's LDA [127]. Fisherface produces discriminant features, whereas PCA (Eigenface) represents the face in expression features of low dimensional subspace with maximum data variance. When each n -dimensional face image x can be represented as a linearly weighted sum of a set of orthogonal basis u_i ($i = 1 \dots n$), then:

$$x = \sum_{i=1}^n a_i u_i \approx \sum_{i=1}^m a_i u_i (\text{when } m \ll n) \quad (2.6)$$

However, the LDA-based algorithm is commonly used to be superior to PCA-based ones, since LDA gives a low-dimensional representation of the optimised objects whereas PCA offers simply object reconstruction [14, 128]. Much research has been published in this field reporting several approaches, and different methods as part of the solution for the cross poses problem, such as features expression in the previous section and more for the persons features themselves. Therefore, the general training set with multiple samples per person [129, 130] is applied to extract discriminative features which are subsequently used to identify the people who are each enrolled with a single specimen. Meanwhile, another research in this field is addressing the problem of one sample being used per person; the problem of considering this as an extreme case of one sample per person commonly happens in real scenarios [131]. Hence, this issue needs to be addressed so as to make face recognition techniques be appropriate in those circumstances. However, it is likely that, for most face recognition technologies, storing only one sample per person in the database will have several advantages. All these studies can be influenced by the noise and the resolution of the image, and thus the next section reviews another part of the proposed system, which is called the decomposition method.

2.5 IMAGE QUADTREE DECOMPOSITION IN FACE RECOGNITION

Face locations, detection and feature extraction are three critical stages in face recognition as mentioned in section 2.4. Therefore, they are important in face analysis applications, and always receive close attention in the design of computer vision and pattern recognition systems. However, to define the correct spatial relevance of physiognomic features and the location such as face tracking, this is still seen as an excessive challenge. Many studies have been published based on improving the accuracy of the selection of the information in both stages by using a method for decomposing images. This improvement cannot be made for the face recognition approach as it can enhance the information of the picture and assists in getting better results

from the stander face recognition methods. However, the problems of PCA and LDA have been addressed [132] and both are considered to have a weak ability to cope with local variations caused by facial expressions. Multi-sub region fusion methods, which divide the face into a set of sub-regions aimed at this issue, reported a better performance. A local discrimination driven face partition method was proposed based on coarse-to-fine quadtree decomposition. Amaral et al. [36] have also proposed to reduce the dimensionality of LBP descriptors to improve classification rates. This work is based on quadtree image decomposition algorithms to apply task-related spatial map segmentation. Moreover, another kind of decomposition method has been developed for face representation and recognition, called the wavelet packet. This was first proposed in 1997 [133] and then used by Garcia et al. [134] for face recognition where each face image is described by a subset of band-filtered images containing wavelet coefficients. An optical image recognition system based on the mechanism of crystal volume holography. Wavelets and wavelet packets have recently emerged as powerful tools for signal compression, particularly in the areas of image and video compression. Wavelet and wavelet packet-based compression algorithms for one-dimensional signals are presented. Along with the results of a compressing method that has been proposed and constructed by Ding et al. [135], who considers that volume holographic associative storage in a photorefractive crystal has some special properties, as shown in Figure 2.7.

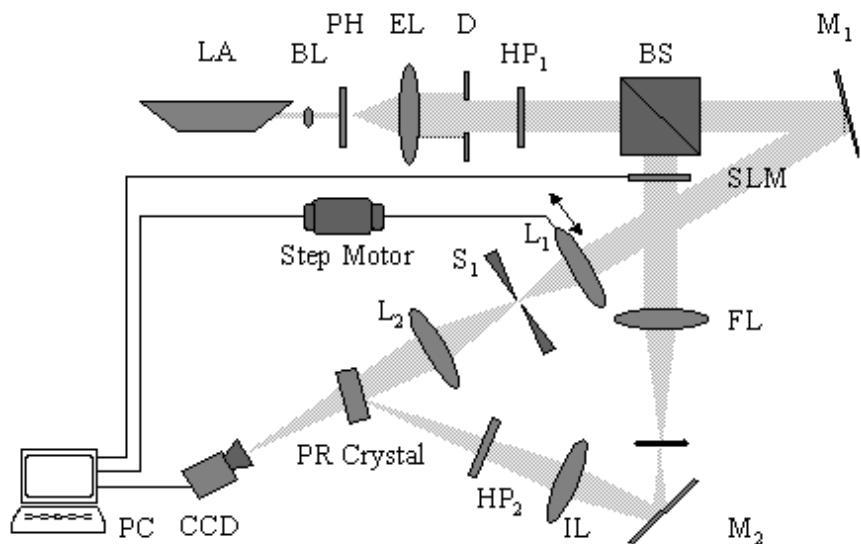


Figure 2.7: Volume holographic image recognition system [135]

The scheme is generally based on an information theory approach that decomposes face images into a small set of particular feature images called ‘Eigenfaces’, which are the principal components of the initial training set of face images. Therefore, developing this method is in cooperation with a previous section of face recognition that can extend into future work of proposed system.

2.6 SUMMARY

This chapter has presented brief descriptions of the main approaches considered research concerning: camera calibration, face recognition and image decomposition. Each section started by explaining the basic principles of the processes used.

The camera calibration is an important part of computer vision and photometric systems, and has an important role in camera surveillance and face recognition. This chapter focused on the perspective projection approaches and discussed the latest research that is used in face recognition that is related to the proposed work. Meanwhile, face recognition needs data for internal parameters of camera calibration for use in many applications such as PTZ cameras. Furthermore, popular technology in face recognition was discussed and the possibility of using it in the proposed method of camera calibration for face recognition across pose was considered. Problems with each technique were listed and possible solutions suggested. Finally, the main components of systems used to improve the accuracy of face tracking and feature selection in the proposed method were described. The first approaches of image decomposition in face recognition and the latest publications on this subject were also discussed.

Chapter 3 :CAMERA CALIBRATION BASED ON PERSPECTIVE PROJECTION

The method of camera calibration used in this study is based on the perspective projection of geometrical shapes. This chapter gives the mathematical details of the existing method. First, an introduction explains the calibration matrix for the camera and demonstrates the parameters of camera calibration that the system needs to determine the depth and the length of the dimensions of the object. The proposed methods used to calculate these parameters using vanishing points are then described and tested, before illustrating how this approach is utilised. The results of the experiment will be discussed in section 3.4. Finally, the findings of this work are summarised.

3.1 INTRODUCTION

Camera calibration is a very important method in computer vision and image processing. Many applications and algorithms use camera calibration as an indispensable step in order to obtain the best results. In 3-D reconstruction, camera calibration is used to determine a set of camera parameters that describe the mapping between 3-D references and 2-D images. There are several research publications [53, 136-138] that refer to camera calibration in control point extraction from images, model fitting, and errors originating as what the proposed system use at this stage.

The parameters for camera calibration from the images give information of third dimension for the objects, which is considered as a background process of the production of the image. Therefore, the processing of camera parameters can be used in measuring the distances of the objects from 2-D images in the actual real world. Thus, it can be widely used in many applications that are based on geometrical perspective such as architectural and medical imaging. Face recognition is one of the many fields that uses camera calibration, and numerous used principle face points to recognise the human face based on distances between facial features. The calibration methods are different depending on the purpose of the field. The best results are based on high requirements of the data image, which is related to the cost of the cameras and the complexity of the method. The two views of object and stereo camera have been used in purpose of reconstruction to get the depth information of the objects [53, 139]. Other approaches use more than two cameras [140] to get more accurate results.

In this chapter, information about depth is gained using one camera (without prior information) for testing, and one for imaging. Moreover, a new system of camera calibration is implemented which is based on perspective projections. The novelty of this work concerns the different geometrical shapes of camera calibrations with high accuracy of information. Hence, in this chapter the mathematics and general concepts are explained concerning the use of camera calibration for rectangular, triangle and square shapes to get the intrinsic camera parameters. The next chapter will then include further analysis of all geometrical shapes utilising the proposed algorithm aiming to give the best results in face recognition.

3.1.1 Camera Calibration Matrix

Before describing the characteristics of camera calibration, the camera calibration matrix needs to be understood. The parameters of a pinhole camera are represented in a 4-by-3 matrix called the camera matrix. This matrix maps represent the 3-D world scene into the image plane [141]. The extrinsic parameters have to be calculated from the matrix to represent the 3-D coordinates of the camera in space; the intrinsic parameters represent the optical centre and focal length of the camera [142]. The world points are transformed into camera coordinates using the extrinsic parameters. The camera coordinates are mapped into the image plane using the intrinsic parameters. The following equations show the relationship between the characteristics and the coordinates of the camera. The 1×3 matrix of the 2-D image points are multiplied by a scale factor W in equation 3.1, where is the same as world points matrix for 3-D multiply by P that is presented as camera matrix as in equation 3.2.

$$W[X \ Y \ 1] = [X \ Y \ Z \ 1]P \quad (3.1)$$

Scale factor Image points world points

$$P = \begin{bmatrix} R \\ t \end{bmatrix} K \quad (3.2)$$

Camera matrix Intrinsic rotation and translation Intrinsic matrix

3.2 CAMERA CALIBRATION CHARACTERISTICS

In computer machine vision, camera calibration is the process of determining the internal camera geometric and optical characteristics (intrinsic parameters) and the 3-D position and orientation of the camera frame within a certain world coordinate system (extrinsic parameters) [61]. These parameters are used to calculate the camera matrix. The external parameters transfer the information about an object from a 3-D world coordinate system to the 3-D camera coordinate system (Figure 3.1). On the other hand, the internal parameters characteristics project the points of the objects from the 3-D camera coordinates into 2-D image coordinates.

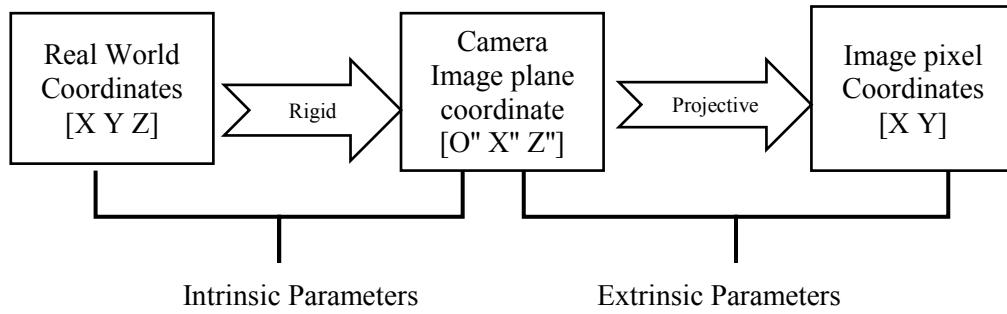


Figure 3.1: Camera calibration characteristic.

In the present study, the system uses these parameters to infer 3-D information from a 2-D image and measure objects. The camera projection captures 2-D images of 3-D objects in the real world. Then, in a reverse process of backwards projection, the 3-D scene structure is recovered from images to obtain the real world dimensions. These dimensions are difficult to infer when the focal length of the camera is unknown. A camera image generates 3-D information from the 2-D image. Hence, the aim is to calibrate the camera from known geometric shapes in a 2-D image, and then to infer the real dimensions for all objects in an image. The estimation of 3-D from 2-D images [143] uses a sheet of A4 paper as a reference to calculate the camera parameters. The following section shows how the information for each characteristic from Figure 3.1 is recovered.

3.2.1 Extrinsic parameters

Extrinsic parameters define the location and orientation of the camera reference frame. When measurements are made by a particular coordinate system, this is called the frame of reference for the representation of objects for example using points, lines and surfaces in the real world [144]. These parameters uniquely identify the transformation from an unknown camera reference frame and the known world reference frame. These parameters determine the translation vector between the relative positions of the origins of the two reference frames. Another benefit of using the extrinsic parameters is finding the rotation matrix that brings the corresponding axes of the two frames into alignment with each other.

The parameters responsible for the transformations of the coordinate system from 3-D world coordinates to 3-D camera coordinates are T and R , as shown in Figure 3.2, where T is the position of the origin in the world coordinate system and R is the rotation matrix.

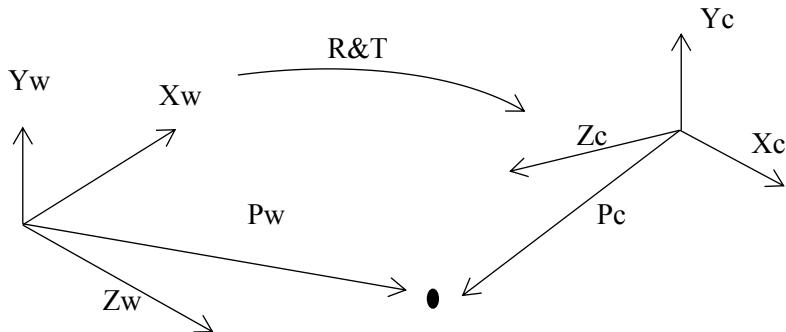


Figure 3.2: Intrinsic parameter transformation

However, the pinhole camera model is based on the principle of collinearity, where each point in the object space is projected by a straight line through the projection centre into the image plane. The origin of the camera coordinate system is at the projection centre at the location (X_0, Y_0, Z_0) in the object coordinate system, and the z-axis of the camera frame is perpendicular to the image plane.

Rotation is represented using the Euler angles ω , ϕ , and κ that define a sequence of three simple rotations around the X , Y and Z axes respectively. The rotations are performed clockwise, first around the X -axis, then the Y -axis that has already been rotated once, and finally around the Z -axis that was rotated twice during the previous stages.

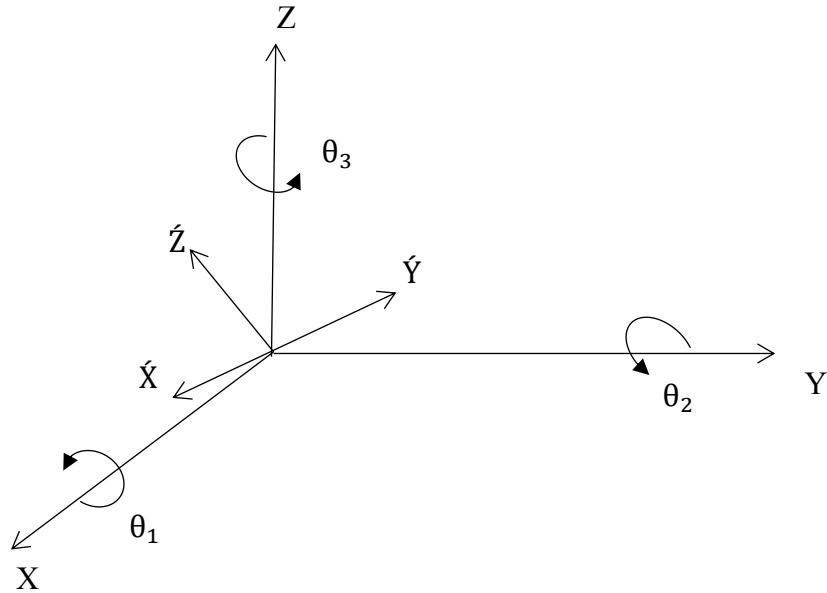


Figure 3.3: Elements of rotation

In this study the algorithm employs the main concept of projection with a different aspect of novelty that connects with various points in the geometrical shapes from 2-D image. Moreover, the elements of rotation for θ , ϕ , and ξ are changed to be θ_1 , θ_2 and θ_3 as shown in Figure 3.3. The new elementary rotations around the X, Y, and Z axes represent the angles of rotation from the corner plane reference.

When θ_1 , is responsible for the rotation of the Y-axis, which is rotated around the Z-axis by angle call the pan angle in 0° . Thus, θ_2 the abscissa axis of the image plane, is the rotated X-axis. Hence, counterclockwise rotates around the X-axis that call the tilt angle. Finally, θ_3 , the ordinate axis of the image, is the rotated X-axis while counterclockwise rotates about the y-axis that call the swing angle.

3.2.2 Intrinsic parameter

The 3-D in computer vision and 2-D images have a pixel coordinating system that correlates the following physical parameters:

- Focal length
- Size of the pixels
- Position of the main points
- Camera orientation

The new researches concentrated on the problem of self-calibration of a camera, when the focal length from intrinsic parameters is known [145]. The algorithms used two images taken from different zoom or views [146, 147]. In other research, a stereo camera was used to get two views instead of zooming [148-150]. Denzler et al. [151] used two cameras to get information about objects in real time. Meanwhile, there are researchers looking for sequences pictures from video frames to get information about the focal length [152, 153].

In this study, the system uses intrinsic parameters by using the theoretical part of the perspective projection of a pinhole camera and estimates the focal length. Since the camera parameters started to involve detecting and tracking objects [154], the research area in computer vision has become a popular and challenging area. The research involved in recognition system of a medical image has encouraged the study to make advance steps in the biometric field. The proposed system is based on geometrical parameters such as window face tracking and different orientation of the fixed distances between the features. Thus this study implements the camera calibration based on perspective projection and a unique method is used to estimate the focal length to get decent intrinsic parameters. The proposed method uses one picture captured from a mono-camera from one view.

The camera model most often used in computer vision is the pinhole camera with a perspective projection where image rays pass through a single camera point. Therefore, lines in the scene are projected as lines onto the image plane. An interesting property of the projective space is that parallel lines intersect at a point on the image, unlike in the simple case of Euclidean space where parallel lines never cross. It is convenient to say that the point of intersection of parallel lines is placed at infinity and its projection onto the image plane is the vanishing point (VP) [155, 156].

The literature on the extraction of vanishing points dates back to the late 1970s [157] and straddles the fields of photogrammetry, computer vision, and robotics. These points give important information on the 3-D structure of the scene and is considered as an important concept in computer vision and camera calibration [158]. For example, in feature matching of the vanishing point can reduce the search space information [159]. Also, another feature includes the matching viewpoint of images by using parallel vanish points to recover the image in the 3-D scene [160].

An interesting part to use in the proposed method of camera calibration and measurement is the vanishing point, that is used in detection algorithms when the methods of clustering line intersections are based on the density of line intersections in a local region [161]. These methods are incapable of discriminating between true vanishing points and points that naturally arise as the mutual intersection of many lines in the image. Hence, the proposed study found the formula that can estimate the intrinsic parameters of the camera calibration to get the real world information from the 2-D image. Different kinds of algorithms are used to obtain the vanishing points, such as those explained in the next sections.

3.2.2.1 The central vanishing points

The perspective projection for the principal vanishing point designates a situation where only one direction point is needed to draw an entire form or scene. The centre of vision that projected from one point of perspective is called a central viewpoint. However, the direction point needs a one-point of perspective because the sides of all objects are moving away from the image plane in the same direction as seen in Figure 3.4.



Figure 3.4: Vanishing point from one direction.

In art, where perspective was initially developed, artists used only one direction point for everything within the drawing or painting [162]. It was not for several more centuries that the understanding and use of more than one direction point entered into the operation for most artists. One-point perspective is an active method used in computer vision to obtain the parameters of the camera and to determine the 3-D coordinates, as it is easy to

calibrate between the camera and scene. Hence, studies have used this method to calibrate a camera using 1D object points or lines [163-165].

Augeas et al. [166] introduced the concept of the self-calibration of a 1D projective camera from point correspondences, and described a method for defining the two intrinsic parameters of a 1D that a camera requires by the trifocal tensor of three images. The central vanishing point is one of the linear systems that is very simple to formulate and solve, which is an advantage of using it. Likewise, it does have disadvantages and the main problems faced includes the following:

- it does not provide information about the camera parameters.
- it cannot be used to model radial distortion[167].
- it is hard to impose constraints such as a known focal length.
- the error function will not be minimised.

For these reasons, non-linear methods are preferred to define non-linear error parameters of intrinsic and extrinsic radial distortion between projected 3-D points and image positions. Moreover, the camera parameters are minimised using non-linear optimisation techniques to give a higher accuracy of measurement

3.2.2.2 Two-point perspective

The second type of perspective uses two points that have the same concept as the previous type of one-point perspective, but with two different directions for the two points. The point of view of 2-D is based on two vanishing points connected by a horizontal line. This kind of perspective is useful when inferring 3-D information from 2-D. Figure 3.5 shows the two-point perspectives where it can be observed that two directions of points are pointing away from the image plane. V1 is the first vanish point and V2 is the second vanish point.

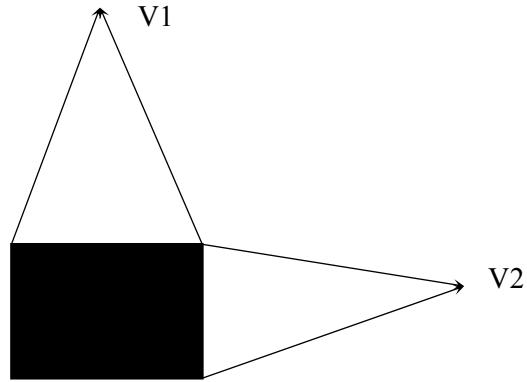


Figure 3.5: Two-vanishing point perspective from image plane.

The direction of the two projection lines which meet at the vanishing points, are looking into the distance between the viewpoint and the object at the same corresponding position; thus, there will be no difference between the direction point and the viewpoint of that object. For example, each window face detection in the proposed algorithm illustrated two-point direction from each angle of each corner. In simpler terms, with the direction point on the left, we see that part of the object from the left. Guillen et al. [69] proposed the first approach that uses only two vanishing points, which was used to build structural scene models. Therefore, Grammatikopoulos et al. [168] used the relationship between two vanishing points of orthogonal directions and the camera parameters to find the intrinsic parameters using the calibration sphere obtained from several images containing two VPs. Lee et al. [169] also used a method to calibrate a camera using two orthogonal vanishing points from image streams without assuming that the principal point is known.

The presented work used two vanishing points from four points that were observed from the corner of geometrical shape; each parallel corners give one vanishing point in orthogonal direction to another. The connection of these points helps in estimating the focal length of the camera, which is then used to infer the rest of the intrinsic parameters in camera calibration.

3.3 PERSPECTIVE PROJECTION

Early research in computer vision to solve the problem of perspective projection from 2-D was conducted by Haralick [75], where the method of projection used assumed that the camera lens is the origin and that the lens views down the y-axis. The image plane is a known distance f in front of the lens and is orthogonal to the lens optical axis that was recovered from the coordinate system based on the four corners of a rectangle.

The perspective projection (x^*, z^*) of a 3-D point $(x, y, \text{ and } z)$ is given by equation 3.3:

$$\begin{pmatrix} x^* \\ z^* \end{pmatrix} = \begin{pmatrix} \cos \theta_3 \sin \varepsilon \\ -\sin \theta_3 \cos \varepsilon \end{pmatrix} \begin{pmatrix} \dot{x} \\ \dot{z} \end{pmatrix} \quad (3.3)$$

Where \dot{x} is represented in equation 3.4 as follows:

$$\dot{x} = f \frac{x \cos \theta_1 + y \sin \theta_1}{-x \cos \varphi \sin \theta + y \cos \varphi \cos \theta + z \sin \varphi} \quad (3.4)$$

and \dot{z} is shown in equation (3.5):

$$\dot{z} = f \frac{x \sin \varphi \sin \theta + y \sin \varphi \cos \theta + z \sin \varphi}{-x \cos \varphi \sin \theta + y \cos \varphi \cos \theta + z \sin \varphi} \quad (3.5)$$

If the perspective projection (\dot{x}, \dot{z}) is known, then the array of 3-D points having (x^*, z^*) for its perspective projection can be determined [170].

The system of the project has implemented the method of Haralick and used it after making a major change to the purpose and the parameters that were used in the original system. One of the main improvement of the system considers from the estimate the focal length and uses it to draw the 3-D coordinate system of geometrical shapes as input, and L and W are the length and width of the shapes. Hence, a, b, c and d are the parameters of the recovered plane which is $ax+by+cz+d=0$, R_θ , R_α and R_σ the three angles of rotation are recovered.

3.3.1 Two Vanishing Points Used to Infer the Intrinsic Parameters of the Camera

As a part of camera calibration based on perspective projection, this section will explain the major steps required in finding the vanishing points so as to connect them and locate the main parameter of the camera, which is the focal length. The system is based on a coordinate system that will enable accurate estimates to be made of the objects in the images from the 2-D surface of the object [170, 171]. From the four corners of the reference triangular, rectangular and square geometrical shapes are used, there are three angles of rotation of the camera which are calculated as in Figure 3.6, considering OXYZ as the real world coordinate system and O''X''Z'' as the image plane coordinates (Figure 3.6). The human eye is assumed to be the centre of the coordinate system at O. Any point (A) on a plane in space will have its projection on an image plane at B, and θ_1 , θ_2 and θ_3 are the three angles of rotation around X, Y and Z.

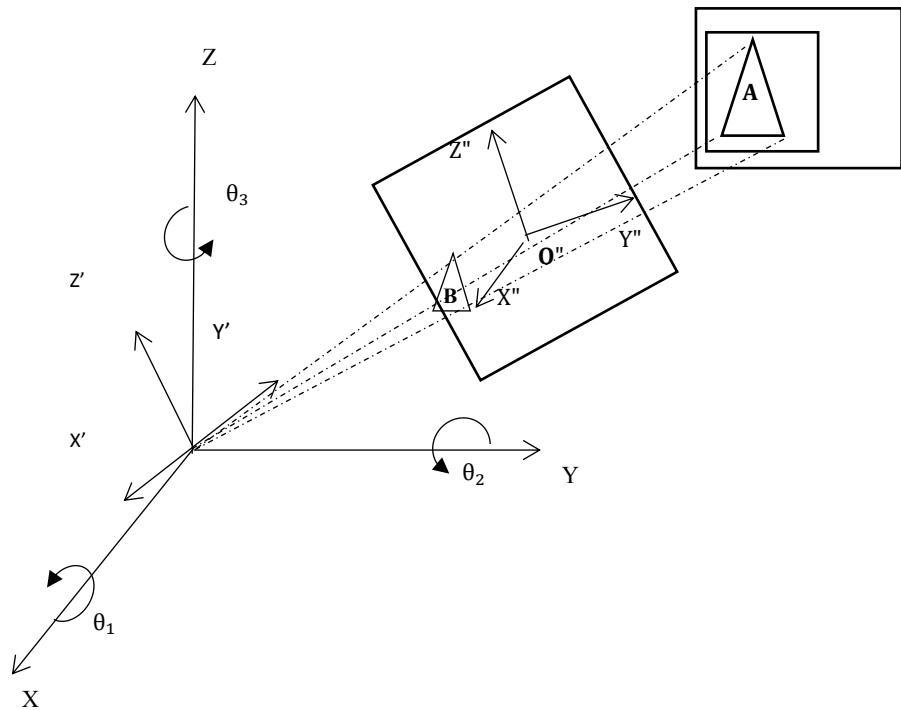


Figure 3.6: Camera perspective showing projection of a point in space

The image plane located between the distance front of the lens and orthogonal to the lens optical axis as represented in equation 3.6. Here $\mathbf{a} = (a_1, a_2, a_3)$ is the coordinate vector of A with respect to OXYZ and $\mathbf{b}'' = (b_1'', b_2'')$ is the projection on the middle plane of the transformations as shown in plane projected on O''X''Z''. \mathbf{n} is the direction vector of the image plane on OXYZ [170, 171].

$$\mathbf{b}'' = \left(\frac{f}{(n \cdot a)} \right) \times R \cdot \mathbf{a} \quad (3.6)$$

where R is the rotation matrix obtained after rotating the axes about Z , X , and Y by θ_1 , θ_2 and θ_3 and \mathbf{n}'' is the direction vector of the image plane on $O''X''Z''$ and is given as $\mathbf{n}'' = (0, 1, 0)$. b'' is repeated for all the four corners of the geometrical shape and, from these, the three angles of rotation θ_1 , θ_2 and θ_3 can be derived [170, 171].

3.3.2 Focal Length from Vanishing Points

Based on Guillou *et. al* [69], two parallel lines in the 3-D actual world will converge to one point on the image. The two axes of the geometrical reference shape projected from the four angles of the geometrical shape has two parallel lines which intersect with each other to give $V1$ and $V2$ as shown in Figure 3.7. Then from the two vanishing points line, the coordinate system can be used to calculate the nearest point to the centre.

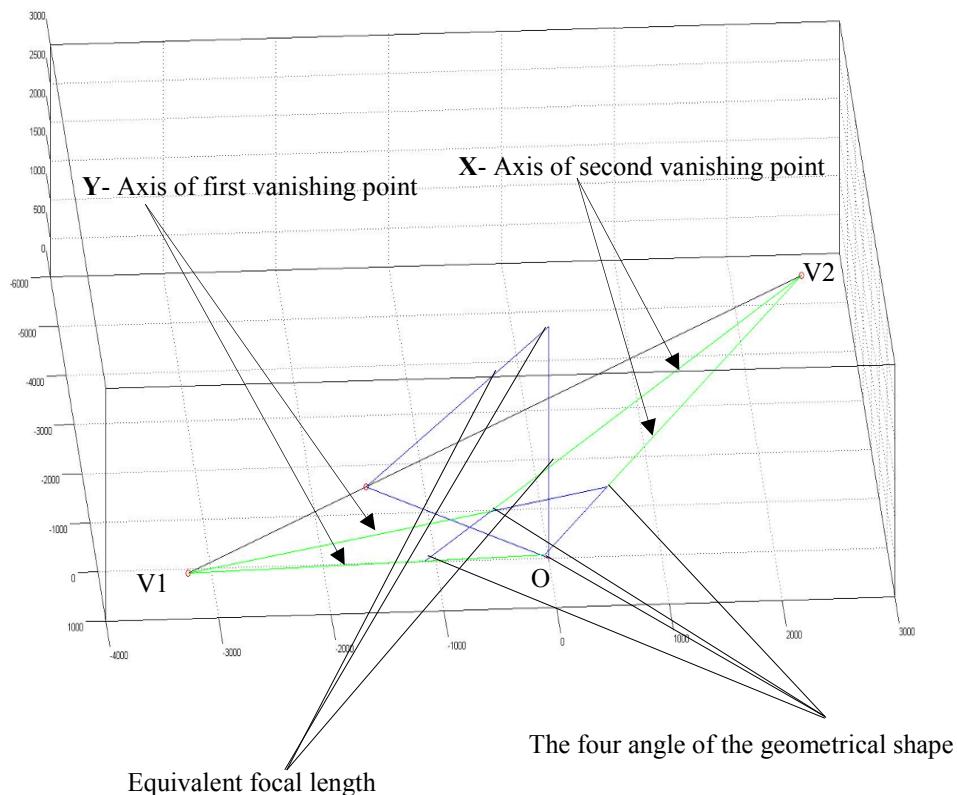


Figure 3.7: Coordinate system of the distance between the centres of two vanishing points and the centre of the lens

Therefore, the distance from the centres of the origin points (O) of ‘X’ and ‘Y’ to the vanishing points (V₁, V₂) are calculated from equations 3.7 and 3.8 as follows:

$$O_x = \frac{\frac{V_{x1}(V_{y2} - V_{y1})^2}{(V_{x2} - V_{x1}) - V_y(V_{y2} - V_{y1})}}{\frac{(V_{y2} - V_{y1})^2}{(V_{x2} - V_{x1}) + (V_{x2} - V_{x1})}} \quad (3.7)$$

$$O_y = \frac{(V_{y2} - V_{y1})(A_x - V_{x1}) + V_{y1}}{V_{x2} - V_{x1}} \quad (3.8)$$

The coordinator system in Figure 3.8 classifies the position of the O_x and O_y between the two vanishing points and centre of the original point.

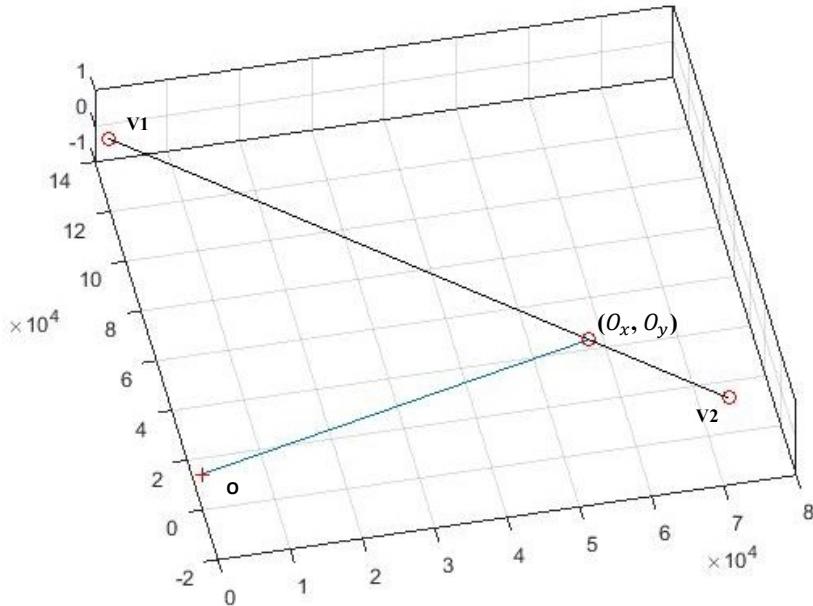


Figure 3.8: The nearest point on the vanish line to the original point

Then from the nearest point, the distance from the centre of the image plane to the centre of the image plane projection can be calculated **PPo** using equation 3.9:

$$|\mathbf{PPo}| = \sqrt{(O_x)^2 + (O_y)^2} \quad (3.9)$$

After this, to determine the focal length, one point for each vanishing point needs to be recognised in equations 3.10 and 3.11 as follows:

V1 is the point of the first vanishing point.

$$|V1| = \sqrt{(O_x - V_{x1})^2 + (O_y - V_{y1})^2} \quad (3.10)$$

V2 is the point of the second vanish point.

$$|V2| = \sqrt{(O_x - V_{x2})^2 + (O_y - V_{y2})^2} \quad (3.11)$$

Hence, the algorithm aims to recover the distance of focal length in front of the lens that is orthogonal to the optical lens axis as in Figure 3.8. In face recognition, from the same previous formulations, the algorithm system enables the determination of the same focal length for each person cross pose from the images in equation 3.12.

$$f = \sqrt{(V1)^2 + (V2)^2 - \mathbf{P}\mathbf{P}\mathbf{o}} \quad (3.12)$$

From the focal length and the image coordinates of the corners of geometrical shapes, the three angles of rotation can be calculated. Considering $\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3$ and \mathbf{P}_4 the image corners coordinates of the shape (Figure 3.9).

Likewise,

$$\mathbf{P}_1 = \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix}, \mathbf{P}_2 = \begin{pmatrix} x_{1+w} \\ y_1 \\ z_1 \end{pmatrix}, \mathbf{P}_3 = \begin{pmatrix} x_1 \\ y_1 + L \\ z_1 \end{pmatrix}, \mathbf{P}_4 = \begin{pmatrix} x_1 + w \\ y_1 + L \\ z_1 \end{pmatrix}$$

Then the corresponding perspective can be determined from these corners a followings:

$$\mathbf{P}'_1 = \begin{pmatrix} x'_1 \\ z'_1 \end{pmatrix}, \mathbf{P}'_2 = \begin{pmatrix} x'_2 \\ z'_2 \end{pmatrix}, \mathbf{P}'_3 = \begin{pmatrix} x'_3 \\ z'_3 \end{pmatrix}, \mathbf{P}'_4 = \begin{pmatrix} x'_4 \\ z'_4 \end{pmatrix}$$

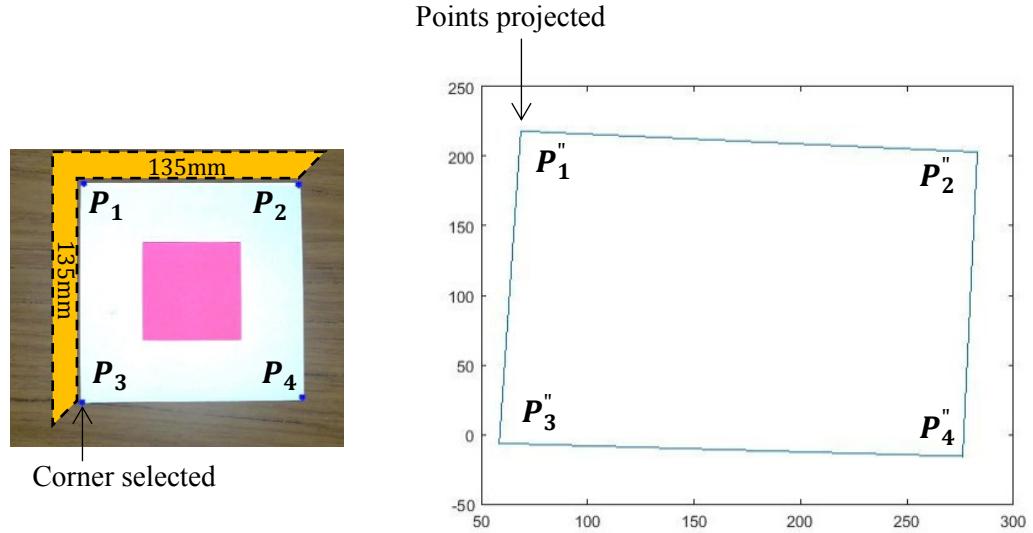


Figure 3.9: Recovered projection points from the corner shape

The two axes of the geometrical reference shape projected from the four angles

Formally, the swing angles are:

$$\theta_2 = \tan^{-1} \frac{(Px_1''Pz_2'' - Px_2''Pz_1'')(Pz_3'' - Pz_4'') - (Px_3''Pz_4'' - Px_4''Pz_3'')(Pz_1'' - Pz_2'')}{f((Px_3'' - Px_4'')(Pz_1'' - Pz_2'') - (Px_1'' - Px_2'')(Pz_3'' - Pz_4''))} \quad (3.13)$$

$$\theta_1 = \tan^{-1} \frac{\cos \theta_2 (Px_1''Pz_3'' - Px_3''Pz_1'')(Px_2'' - Px_4'') - (Px_2''Pz_4'' - Px_4''Pz_2'')(Px_1'' - Px_3'')}{f((Px_1'' - Px_3'')(Pz_2'' - Pz_4'') - (Px_2'' - Px_4'')(Pz_1'' - Px_3''))} \quad (3.14)$$

$$\theta_3 = \tan^{-1} \frac{A(Pz_3'' - Pz_1'')B(Pz_4'' - Pz_2'') - C(Pz_4'' - Pz_3'') + (Pz_2'' - Pz_1'')}{A(Px_1'' - Px_3'') - B(Px_2'' - Px_4'') - C(Px_1'' - Px_2'')D(Px_3'' - Px_4'')} \quad (3.15)$$

Where the constants A, B, C and D are:

$$A = \frac{(Px_2''Pz_4'') - (Px_4''Pz_2'')}{(Px_1'' - Px_3'')(Pz_2'' - Pz_4'') - (Px_2'' - Px_4'')(Pz_1'' - Pz_3'')} \quad (3.16)$$

$$B = \frac{(Px_1''Pz_3'') - (Px_3''Pz_1'')}{(Px_1'' - Px_3'')(Pz_2'' - Pz_4'') - (Px_2'' - Px_4'')(Pz_1'' - Pz_3'')} \quad (3.17)$$

$$D = \frac{(Px_3''Pz_4'') - (Px_4''Pz_3'')}{(Px_1'' - Px_3'')(Pz_2'' - Pz_4'') - (Px_2'' - Px_4'')(Pz_1'' - Pz_3'')} \quad (3.18)$$

$$C = \frac{(Px_1''Pz_2'') - (Px_3''Pz_1'')}{(Px_3'' - Px_4'')(Pz_1'' - Pz_2'') - (Px_1'' - Px_2'')(Pz_3'' - Pz_4'')} \quad (3.19)$$

Once the focal length and camera angles are obtained using equation 3.12 the image coordinates of the four corner points can be converted into their corresponding 3-D coordinates using reverse transformation. Hence, the corresponding 3-D coordinates identify the position of the piece of paper in space, and thus the equation for the plane of paper can be derived, which will be of the form $z = k$ (constant). It will be the same as the equation for flat surface of the object, since the paper lies on it. Now as we know the equation for the flat surface, any measurement can be taken on the same plane by finding the distance between a pair of points using the distance formula is explained further in the next chapter

3.4 RESULTS AND EXPERIMENT

In this section, the proposed method of camera calibration experiments is summarised. The results show the observed information of camera parameters and the analysis carried out to prove the accuracy of the parameters from the proposed method.

3.4.1 Camera Parameters

The most important parameters of the camera that are employed in the system are the intrinsic parameters represented by Σ , θ , and Φ . These parameters are determined from selected angles of rotation that can be observed with fixed and robust dimensions. After calculating the estimated focal length, it is used to fix the unknown distance between the camera and the object. The first experiment calculated the mean internal parameters of the camera from selective images that were affected by Gaussian noise as shown in Table

3.1. The selecting of rectangular, square and triangular corner in the 2-D image, each point is projected in space as explained in section 3.3. The principal points of calibration depend on the accuracy of the selection of each corner of the image selected manually. For this experiment, picture a captured from mobile camera GT-I9300 SAMSUNG Galaxy III, 2448×3264 pixels was used. The camera was set up 80 cm away from the calibration photo object shape. The results show slight differences in the values of the intrinsic parameters caused by variations in the manual selection of the corners of the calibration shapes. The images were affected by Gaussian noise for seven different values until they reached 0.750 when the noise covered all the images and did not show any shape to calibrate the system.

However, the automatically fixed points for the corners give the system the same parameters without any effect of noise. Table 3.2 shows the results of the experiment with fixed automatic selection.

Table 3.1: Intrinsic Camera Parameters from Manual Point Selection

Noise in dB	00.0	0.25	0.125	0.250	0.500	0.750
Real data						
θ	1.0558	0.9053	1.3751	1.2793	0.7869	1.2547
Σ	0.4975	0.6339	0.1628	0.2910	0.7659	0.2960
Φ	1.3228	1.3598	1.1227	1.6025	1.2232	0.9824
F	3.5438	4.2283	3.0660	4.2283	4.1154	4.5984

Table 3.2: Intrinsic Camera Parameters from Automatic Point Selection

Noise in dB	00.0	0.25	0.125	0.250	0.500	0.750
Real data						
θ	0.1551	0.1551	0.1551	0.1551	0.1551	0.1551
Σ	1.4129	1.4129	1.4129	1.4129	1.4129	1.4129
Φ	1.5883	1.5883	1.5883	1.5883	1.5883	1.5883
F	4.1321	4.1321	4.1321	4.1321		4.1321

3.4.2 Focal Length and Application

Based on the pinhole camera model (Figure 3.10) there is a relationship between the size of the object and distance. The object becomes bigger as the camera becomes closer, which means that the internal focal between the lens and the image on the sensor becomes smaller.

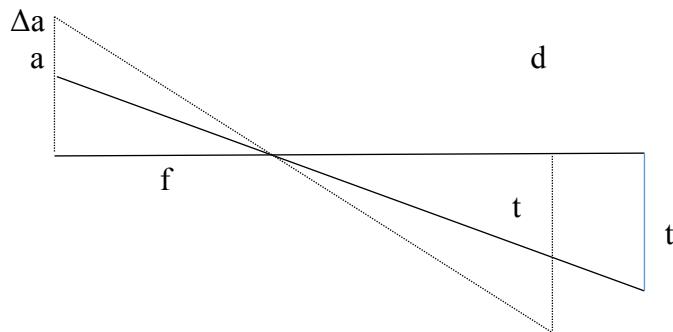


Figure 3.10: The pinhole camera

When (a) is the original image size of the actual object. Δa is the change in image due to the movement of the camera forward or backwards, where d is the original distance from the optical centre of the camera to the actual object and Δd is the change in position due to the movement of the camera forward or backwards. Here, f is the focal length of the camera, and t is the size of the actual object. From the pinhole imaging principle and the mathematics of lenses represented in equation 3.20:

$$\frac{a + \Delta a}{f} = \frac{t}{d + \Delta d} \quad (3.20)$$

Equation 3.21 then substitutes the known distance instead of the focal length to determine the unknown distance from the second image. The lens equation expresses the quantitative relationship between the objects distance in the real world (d_o), the image distance from the camera (d_i), and the focal length (f). The equation is stated as follows:

$$\frac{1}{f} = \frac{1}{d_o} + \frac{1}{d_i} \quad (3.22)$$

pictures from four different distances between the camera and the plane of the object as shown in Figure 3.6 were captured by the Samsung Galaxy SIII mobile camera with a resolution of 8 MP. The manufacturer's focal length for the camera is 4 mm; the first results compared the pixels in millimetres based on the original focal length, which gives 1mm=993 pixels.

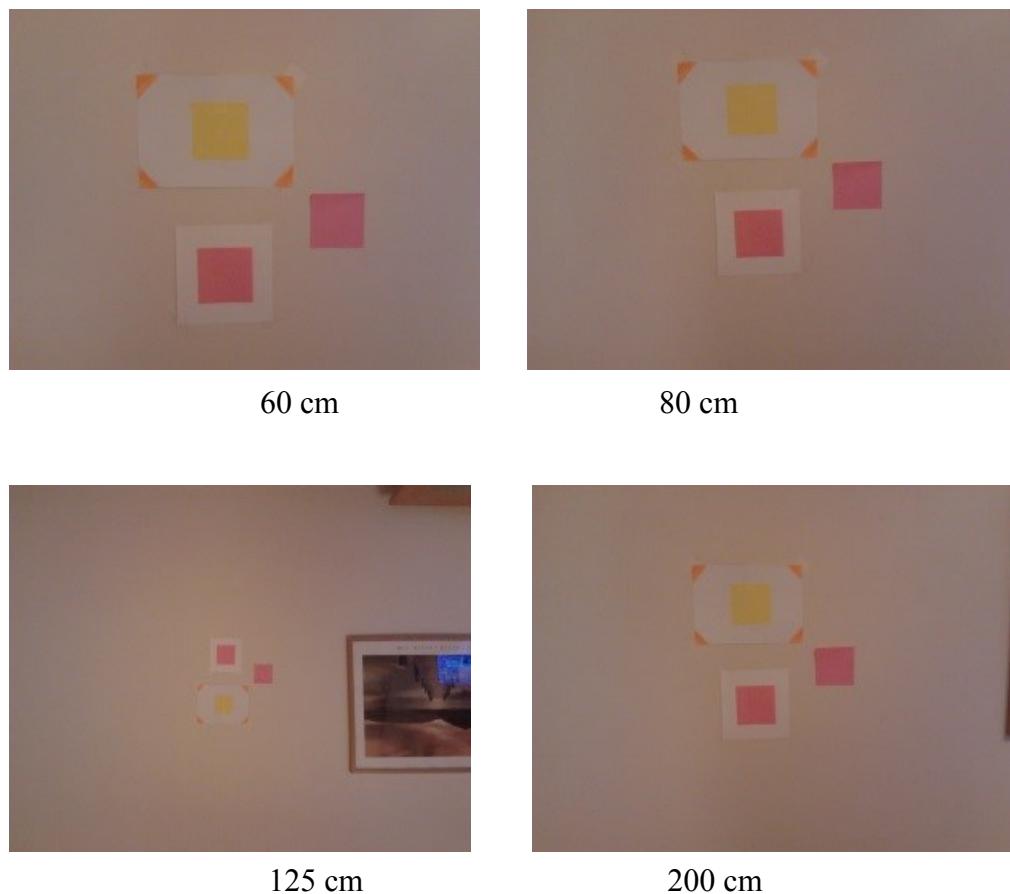


Figure 3.11: Four pictures captured from different distances in cms

The results shown in Table 3.3 show that in practice the focal length decreases when the depth of field increases. When, dc : Distance between the lens of the camera to the plane of the object and fc : Internal focal length of the camera. Section 3.3 has explained the method used for calibration and how the two vanishing points are used to give focal length,

Table 3.3: The internal focal length of the camera from four distances

Distance				
dc	60cm	80cm	125cm	200cm
fc	4mm	3.3 mm	3mm	2.9mm

which is an important parameter in the proposed method of measurement technique in the next chapter. In the following, the practical tests show the coordinate system of the focal length for one image from 4 different distances.

Each distance has two views of the focal length that shows the two vanishing points and the line that connects them with nearest point to the center of the view as explained in section 3.3 and shown in Figure 3.12.

3.4.2.1 *Test of Focal Length Accuracy*

The accuracy of determining the focal length is evaluated using selective database as four pictures captured from various distances, as in the previous experiment. The camera was calibrated to the corners of the object's shape (square and rectangle), and the internal parameter of focal length was calculated from a real picture based on the reference measurement. The distance of focal length that was affected by noise was compared with real known values to calculate the error for each level of noise. Moreover, the mean error and standard deviation were calculated for each level of noise. The results are shown in Figure 3.13 where it can be seen that there is a gradual curve between the level of noise between 0 and 3, and from main error from 1 to 14, with a fixed standard deviation for each level of noise.

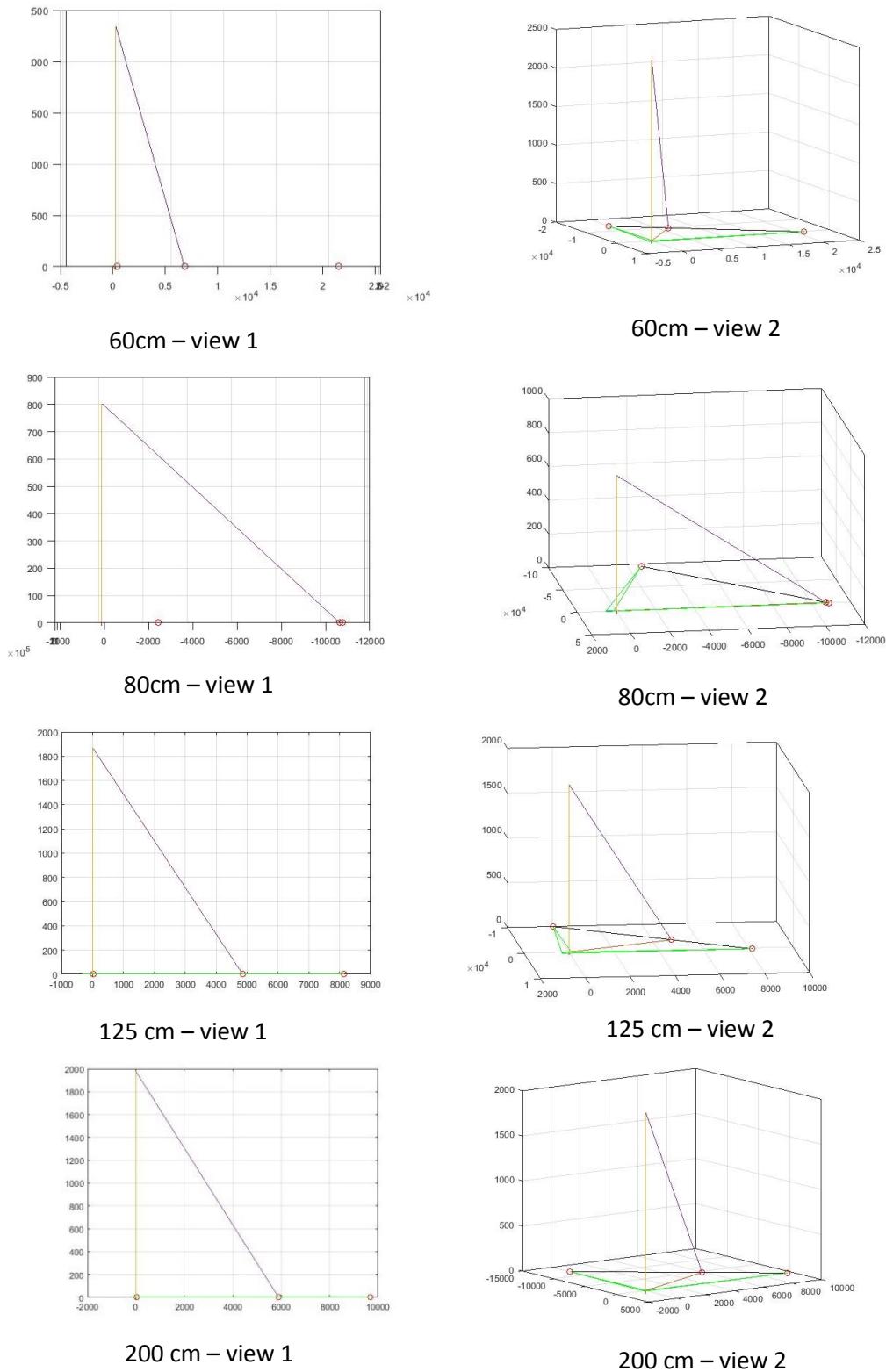


Figure 3.12: Coordinate system of focal length

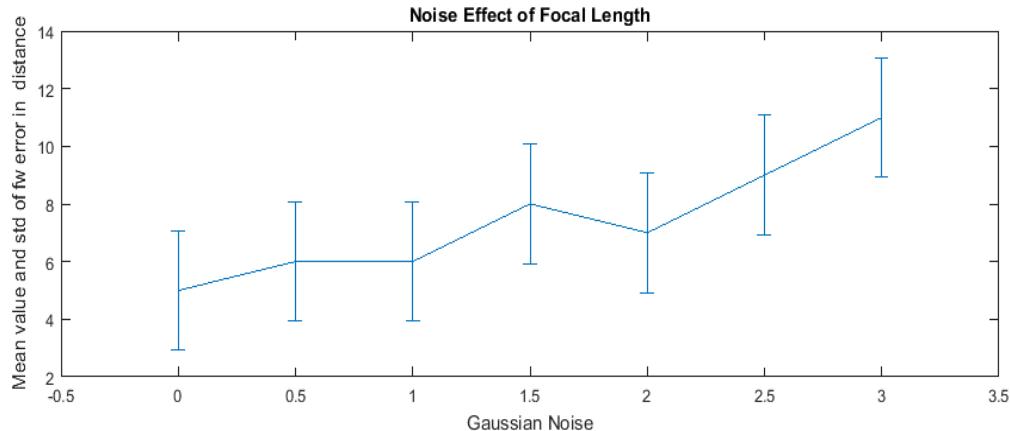


Figure 3.13: Effect of noise (dB) on focal length (mm)

The x-axis shows seven levels of Gaussian noise from 0 to 0.75 db. The y-axis shows the mean value and standard deviation of the error in focal length measurement calculated based on the difference between the distances of focal length for the four variant distances (60, 80, 120, 200cm). The error was at 5mm at zero level noise, and 1mm errors were observed more or less for all the four distances. The rate of error increased slightly until 11mm at the maximum noise 0.75 db. The maximum level of noise made the picture extremely blurry and difficult to recognise any objects. The range of the error supports the system to deal with pictures from different quality levels, as there is no significant error.

3.4.2.2 Focal Length in Distance

The proposed method of camera calibration uses a model pinhole camera for two pictures captured for the same object plane of reference calibration. One of the images used as a reference to calibrate the camera, where the distance between the camera lens and the object plane is known. The proposed method of auto-calibration will give the parameter of focal length for the second picture as what we expect in the proposed study and shown in Figure 3.14. Moreover, the second picture will not have any dimensions or information known about the camera that can determine from the same reference. Hence, the proposed method of camera calibration determining the unknown distance of any pictures captured from a different angle of the camera, as represented in equation 3.22.

$$\frac{Do}{Dx} = \frac{X1}{X} \quad (3.23)$$

D_o is the real distance for the original reference (2-D image). D_x is the distance between the camera and the different image that is calculated from the system. X_1 is one dimension of the reference that is calculated from the system, X is the dimension of real reference.

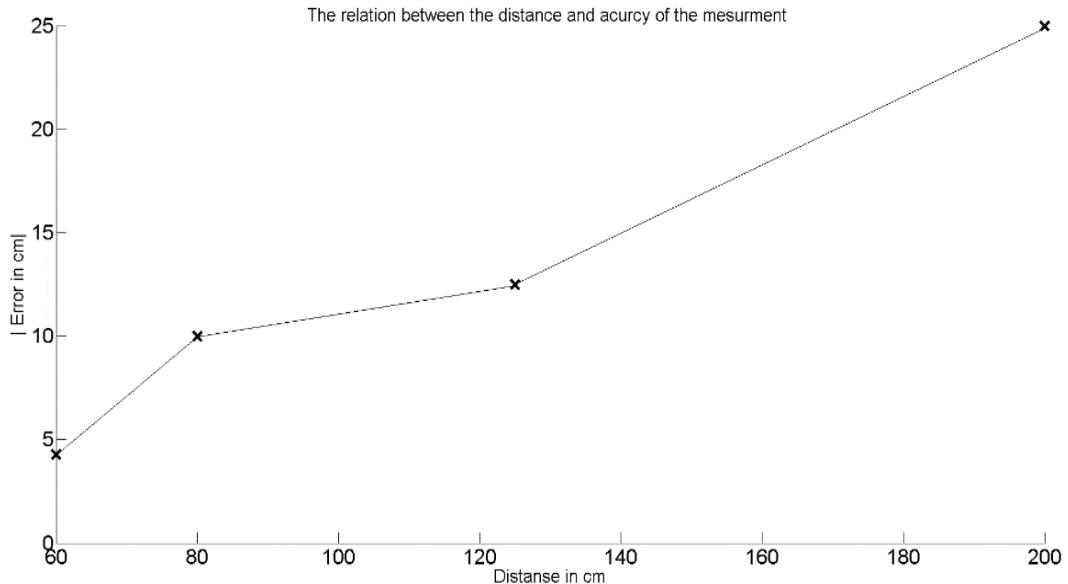


Figure 3.14: The error rate of focal length

The error for the first distance of 60mm was just 4mm but reached 25mm when the camera was 2 metres away. The experiment conducted with manual selection using a free-hand mobile camera, which might be more accurate using the automatic method of calibration. When pictures have been captured using free-hand movement from mobile cameras, this can give rise to different information for each picture, which increases the possibility of getting more errors in applications. Therefore, ensuring the fixed movement of the camera would significantly reduce the errors of the pictures.

3.4.3 Discussion

The proposed method extracts the intrinsic parameters of the camera calibration from the 2-D images as the first step in proposed system. Each of Σ , θ , Φ and f have the effect for the angles rotation from the image plane into space. Therefore, the parameters will have the same distances for the dimension affected by camera movement to the image plane. The results were tested also under the effect of the Gaussian noise from 0 to 0.75 db as shown in Figure 3.13 where it can be seen there were very close error values between the variants Gaussian noise as shown in Table 3.3. The results were taken from semi-automatic selection of the corner object in the image (Table 3.1). Meanwhile, Table 3.2 shows constant values when the calibration was full-automatic for the geometrical object

from the image. Thus, accuracy of the results assumed to be based on selection for the four corresponding points of the calibration, which is made slightly different from the effect of the noise.

However, the focal length considers as a main parameter that can solve the problem of finding the distance measurement from 2D image [172-175]. Thus, the focal length was calculated with the camera parameters shown in Table 3.1 and Table 3.2. Moreover, the focal length parameter can be used in applications to determine the accuracy of measurement and dimensions reconstruction. The first experiment used focal length to calculate the distance between the lens of the camera and the object plane. Based on the pinhole camera, the distance between the camera and the object became smaller when the object is far to the camera and the results in Table 3.3 shows that the focal length of the camera decreased gradually when the distance to the image plane decreases. The distance between the lens and the object started at 60 cm when the internal focal length was 4 mm, and finished with a distance of 2 metres with focal length measurement of 2.9 mm. The intrinsic value of the focal length used in equations 3.4 and 3.5 was used to determine the actual distance between the camera and the object. The accuracy of focal length determination has been tested by using the distance of focal length in application. The mean error and standard deviation of focal length was calculated under the effect level of Gaussian noise.

As shown in Figure 3.13 the plotted results give a gradual increased curve with a fixed level of standard deviation for each level of noise. The results indicate an error of 2mm in focal length for the first level of noise for the four different positions of the camera. The curve of Figure 3.13 is growing slightly until the peak 12mm of the error, when the internal differences of the error between the distances are constant into 2mm. These results are based on semi-automatic selection of the focal length distances for each position of the camera. Therefore, the internal error shows small difference between the camera distances 60, 80, 120, 200 cm, compared with the original dimension that gives us 0.6% of internal. Meanwhile, the percentage error of the furthest distances gave 4.4% in overall error. The error was calculated based on the comparison between the real informations from the objects and the calculated informations from the proposed system.

The next experiment used the focal length in application of distances after testing the accuracy of one side of the dimension. The accuracy of focal length in distance from the geometrical shape in the picture shown in Figure 3.14, the result shows the error in millimetres compared with the real world distance. As shown in Figure 3.11 the relationship between the distance between the camera and object and the error in measurement are correspondingly proportional. The error was less than 5mm for the first position of the camera at 60 cm and it considered as a minimum error of the experiments. Meanwhile, the error was 25 mm when the camera was far away from the object at 2 metres. In addition, the focal length was tested from different distances and provided higher accuracy compared to the real world. The finding of the parameters from the proposed method of camera calibration supported the previous theory of the proposed study.

3.4 SUMMARY

This chapter presents the first work in my thesis and based on these results there are further methods that I will incorporate in order to obtain the final results of face recognition. The camera calibration method is based on perspective projection and determining the parameters of the camera is the first step. The physical and logical background of camera calibration was discussed before using the proposed method. This method of camera calibration uses the pinhole camera model to infer the camera parameters. The experiments and results were tested and compared with different scenarios of the camera changing its position. The method is flexible for use with different geometrical shapes as a reference of calibration and allows the object in the image to be calibrated without the need to capture it by the user of the system. Merely knowing the dimensions of any object in the picture is sufficient to be able to calculate the dimensions of the rest of the objects in the image plane. Also, the use of vanishing points helps in inferring the intrinsic parameters such as focal length.

The next chapter uses the internal parameters of the camera to measure the dimensions of objects with regular and irregular shapes.

Chapter 4 : 3-D INFORMATION MEASUREMENT FROM 2-D IMAGE

Virtual life gives a 3-D view that arises from human stereovision that includes the perception of depth, which is a complex process. The eyes connect to the brain estimating the measurements of the objects around us by comparing the objects with each other that are big, and small, long or short. In recent years the communications network and camera mobility have quickly involved, and these developments have motivated the investigation of new methods to gain information about object dimensions from a 2-D image.

This chapter demonstrates the new camera measurement technique (CMT) based on camera calibration. The method is supported by the test system and describes the experiments conducted using it. It starts with a brief review of popular methods that use input images in a high dimensional subspace as part of the proposed method and refers to the challenges and dimensionality problems. Furthermore, the statistical approach of the proposed algorithm based on perspective projection is explained in Chapter 3. The system of calibration is then tested for each angle of rotation and object dimensions are subsequently inferred from an image using automatic calibration. The types of shapes of the calibration are explained in Section 4. The results of using the proposed method are described using pictures captured from different resolution of mobile cameras that are presented and discussed. Finally, the conclusions of the chapter are provided.

4.1 INTRODUCTION

Measuring the shape and appearance of physical objects remains one of the most challenging problems in computer vision and graphic fields. Methods for measuring the shape of real-world objects can be categorised by two geometric approaches when identifying a common point seen from different lines of sight, or by photometric approaches using multiple light positions.

The resolution of airborne instruments [176] to the sub-meter level have been increased due to recent advances in LIDAR technologies, that opens up the possibility of creating detailed maps over a large area. A better framework for handling line-of-sight calculations has been developed by using voxelized approach (A voxel representation of a scene has spatial data as opposed to the conventional rasterization view) to LIDAR processing that allows us to retain detail in overlapping structures more than the current approaches that exist [177-179]. Photometric methods for estimating shape have proved useful for this task because they directly measure normal surfaces, which are effective for rendering applications or recovering reflectance [180]. There has been significant progress toward developing general-purpose techniques, but existing approaches still rely on strong assumptions about the surface properties of the target object that are not always true in practice [181]. Geometrical cues and constraints provide valuable information as to how certain image features should be interpreted. For instance, in many human-made scenes there exist some straight lines, which are also parallel in 3-D. Once these lines are identified, it can infer 3-D structures from 2-D features, and this constrains the search for other structures [182]. On the other hand, camera calibration is an important step inferring the internal and external parameters of the camera [80]. Intrinsic parameters such as focal length, image sensor format and principal point are crucial for applications that require accurate metric information, such as depth and dimensional measurements proposed in this study from 2-D images. The algorithm focuses on the distance between the sensor of the camera and the image plane coordinates, from which the focal length can be calculated. Each camera has a value of focal length given by the manufacturer, whereas the proposed approach estimates focal length without the need to get information from the camera itself. As with existing approaches, the input to the proposed algorithm consists of a single image captured by a mono-camera. The camera mobility was used to feed the input of the system from different kinds of camera resolutions. The output of the proposed system should then include a measure of selective distances in the image.

The shapes that were used to measure are regular and irregular, which is coming as an output of the proposed new camera measurement technique CMT. The new technique is used to infer the dimensions of a real object using an image via the automatic calibration of the camera. This method utilises different geometric shapes to make comparisons in the calibration, and thus to decrease the degree of error. Various square, rectangular and triangular shapes are used to calibrate the camera. Since the position from one of the geometrical shapes is known, thus, measurements can be inferred for the objects or plane from the image. The technique is based on perspective projection of the parallel lines of two vanishing points and explained in Chapter 3. The difference between the parameters of the camera manufacturer can affect the accuracy of measurement, such as the resolution of the lens. The proposed method achieved high accuracy in distance measurement from 2-D images, and even in the worst-case scenario such as when the camera has noise, is not in a good position/difficult angle or hand vibrations, can be affected by 5% error when the camera resolution was 2MP, and the picture was captured from a free hand mobile camera.

The main contributions of this chapter include the calibration of the camera for three different geometrical shapes. A new geometric technique for measuring shapes in an actual scene from a 2-D image is then investigated.

4.1.1 *The Challenges*

In general, camera calibration is considered to be an important part of surface reconstruction methods. It depends upon the precise values of the image pixels that require cameras to be radiometrically calibrated [183] that require high resolution for each pixel value and expensive cameras. Furthermore, the complexity of the algorithms, such as when determining the camera transfer function by collecting a series of images of the same scene at different exposure levels [184-186] was seen as a significant challenge. In addition, the use of stereo-camera to obtain depth information from a scene is another problem faced when using a mono-camera. Moreover, triangulation methods of measurement uses limits of the objects for measurement and more with regular shapes. Meanwhile, there are many different approaches that can be more efficient in measurement based on free shapes with more curves of incline or decline. The pictures from these approaches showed the object lying flat, which is difficult to guess the size or depth from different side views.

The most challenging in the proposed approach is how to calculate the 2.5-D from the 2-D image that is lying in space and determine the distance of the depth for the object at low cost and good accuracy. For example, the measurement of cells of irregular shapes and different sizes for medical purposes. On the other hand, the building reconstruction have different regular and trigger shapes and curves, in this area there are different tools for measurement that are expensive and have the different subject of studies such as laser or ultrasounds.

4.2 NUMERICAL METHODS

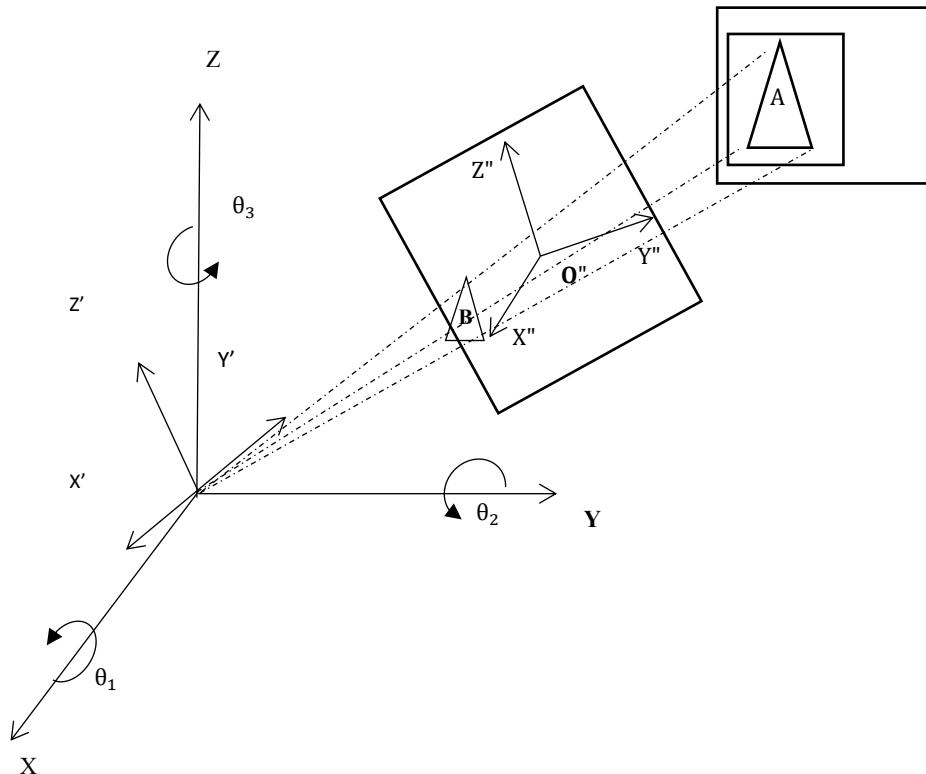


Figure 4.1: Camera perspective showing projection of a point in space.

The new method proposed a technique to measure the distance and dimensions for the objects surfaces from 2-D images. The geometrical approach of perspective projection based on camera calibration system has been implemented and used in this method. The methodology of the calibration measurement system is based on generating a coordinate system which enables accurate estimates to be made of objects in the images. The system automatically detects points in space and uses them for the calibration of different shapes. For ease of use, a square piece of note paper (square), a rectangular sheet of A5 paper and a triangular ruler are used. The shapes are placed on a flat surface, and then from the four

corners of the reference shapes all three angles of rotation of the camera are calculated. The camera projection is as explained below, in which OXYZ is the real world coordinate system and O"X"Z" are the image plane coordinates (Figure 4.1) The human eye is assumed to be at the centre at point O. Any point A lying on a plane in space will have its projection on an image plane at B, where θ_1, θ_2 and θ_3 are the three angles of rotation.

Then, the swing angles in θ_1, θ_2 and θ_3 as achieved using equations 3.13, 3.14 and 3.15 from Chapter 3. The distant measurement equation recover from the image plan points when is the plane equation is given by: $ax+by+cz+d=0$ the three points are:

$$a = y_1(z_2 - z_3) + y_2(z_3 - z_1) + y_3(z_1 - z_2) \quad (4.1)$$

$$b = z_1(x_2 - x_3) + z_2(x_3 - x_1) + z_3(x_1 - x_2); \quad (4.2)$$

$$c = x_1(y_2 - y_3) + x_2(y_3 - y_1) + x_3(y_1 - y_2); \quad (4.3)$$

$$d = -(x_1(y_2 \times z_3 - y_3 \times z_2) + x_2(y_3 \times z_1 - y_1 \times z_3) + x_3(y_1 \times z_2 - y_2 \times z_1)); \quad (4.4)$$

After obtaining the four points from the corner calibrated shapes, the system enables you to recover the 3D coordinates (xd, yd, zd) of a 2D point (xp, yp) located on the target plane After been known a, b, c and d from 4.1, 4.2, 4.3 and 4.4 respectively. The system recovered the coordinate system, with the four corner of the rectangle as input and the parameter of geometrical shapes reference then feed into the geometrical projection to measure the distance.

$$A = a(x \times \cos(\theta_3) - f \times \sin(\theta_3) \times \cos(\theta_2) + z \times \sin(\theta_3) \times \sin(\theta_2)) \quad (4.5)$$

$$B = b(x \times \sin(\theta_3) + f \times \sin(\theta_3) \times \cos(\theta_2) - z \times \cos(\theta_3) \times \sin(\theta_2)) \quad (4.6)$$

$$C = c(f \times \sin(\theta_2) + z \times \cos(\theta_2)) \quad 4.7)$$

When the A, B and C the parameters to find the distance of original point, from the summation of the three points from the distance of the centre point can be calculated and fed into Lambda (equation 4.8) to solve angles of rotations.

$$\lambda = -\frac{d}{A + B + C} \quad 4.8)$$

$$xd = \lambda(x \times \cos(\theta_3) - f \times \sin(\theta_3) \times \cos(\theta_1) \times \sin(\theta_2)) \quad 4.9)$$

$$yd = \lambda(x \times \sin(\theta_3) + f \times \cos(\theta_3) \times \cos(\theta_1) - \cos(\theta_3) \times \sin(\theta_2)) \quad 4.10)$$

$$yd = \lambda(f \times \sin(\theta_2) + z \times \cos(\theta_2)) \quad 4.11)$$

When, the *Ratio* is the distance of the width for the reference over the distance to origin point, the distance can calculate as following:

$$Ratio = \frac{W}{d_o} \quad 4.12)$$

$$d_o = \sqrt{(xd1 - xd2)^2 + (yd1 - yd2)^2 + (zd1 - zd2)^2} \quad 4.13)$$

And the distance measurement equations in real world is:

$$Dis = \sqrt{(Ratio \times d_o)} \quad 4.14)$$

After capturing the image, the first step is to identify the pixel locations of the corners of the geometric shape in the computer image. The semi-automatic process is used which enables the corners of the reference shape to be chosen. There should be a contrast between the shape and the background plane of the object so that the corners are selected as accurately as possible. After the equation for the flat surface is obtained, the user can select, again manually, the corners of the object and the distances between them are calculated.

4.3 TEST OF THE SYSTEM

The previous section has explained that the CMT depends on three angles of rotations defined as θ_1 the tilt angle θ_2 the pan angle and θ_3 the swing angle. Therefore, it needs to ascertain that changes in these angles do not affect the system different angles of capturing the pictures. A full test was conducted for the position of camera when it captured the pictures. When, swing and pan angles tested from 0° to 180° and the tilt tested until 2 meter distance away from the camera to the references as the in Figure 4.2.

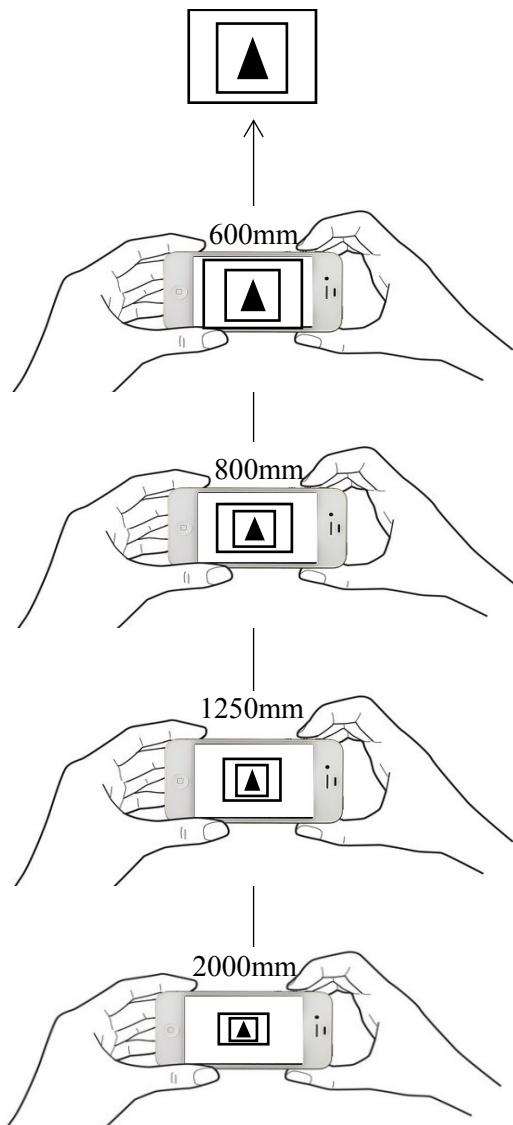


Figure 4.2: Distance between camera and references varied from 500 mm to 2m

4.3.1 *Tilt Angle Variation*

The early experiments of tilt angle has been prepared to verify the system before moving to measurement purpose as it will be used in next part of this section. As shown in Figure 4.1, the system considers θ_1 , θ_2 and θ_3 the tilt angles rotations that are changing depending on the camera movement. The first point of rotation θ_1 shows the effect of the tilt for the camera when it changes as the distance between the camera and object varies. The consequence of tilt angle on the accuracy has been tested by varying the distance between the camera and three square triangular and rectangular geometrical shapes from 600 mm to 2000 mm. The other two angles, θ_2 and θ_3 were kept at 0°angle. The corresponding changes in measurements from real dimensions are shown in Figure 4.2, depending on the differences between the actual length and measured distance by the technique. The relationship between the points of the objects corners of the calibrations is inversely proportional to image plan in the sensor of the camera, which makes the objects become smaller when the camera comes closer to the objects and bigger when it moves further away from the objects. The reasons for this relationship are as follows:

Firstly, when the angle is decreased, the calibrated shape from the image is converted to be vertical on the image sensor plane of the camera. The practical experiment on the tilt angle found that the projected object from the image became difficult to determine when the camera came closer. The reason is the projected image plane of the sensor become bigger and then out of the bounders of the camera sensor. Thus, it makes it hard to select the corners of the shapes precisely, which offers possible failure to determine the object from 2-D images.

Secondly, to minimise the size of the objects from the image, each pixel of image that has been converted in the picture shows the real world in different measurements and the same in mirror. Thus, the semi-automatic selection of the measurement from images has a risk for human use that affected the inaccuracies in the final distance measurement. Hence, the mean error between the real measurements of depth and length and the inferred measurements from the 2-D image in the proposed method has a small difference in tilt angle.

The averages were measured between the two dimensions to record the error for each shape of calibration, as shown in Figure 4.3. The same results were shown for the rectangular shape up to the distance of 1400 mm, and then we see a reduction in error at

1800 mm and 2000 mm. Meanwhile, triangular shape has the highest error as compared with other shapes, but it showed that it is affected by distance as well. The measurement error was increased slightly with distance, which is expected because the corners of the objects become unclear. Also, the accuracy of the selected points of dimensions can be another reason to achieve this error. This error can be reduced when the selection of the corners for the shapes calibration change from semi-automatic to fully automatic.

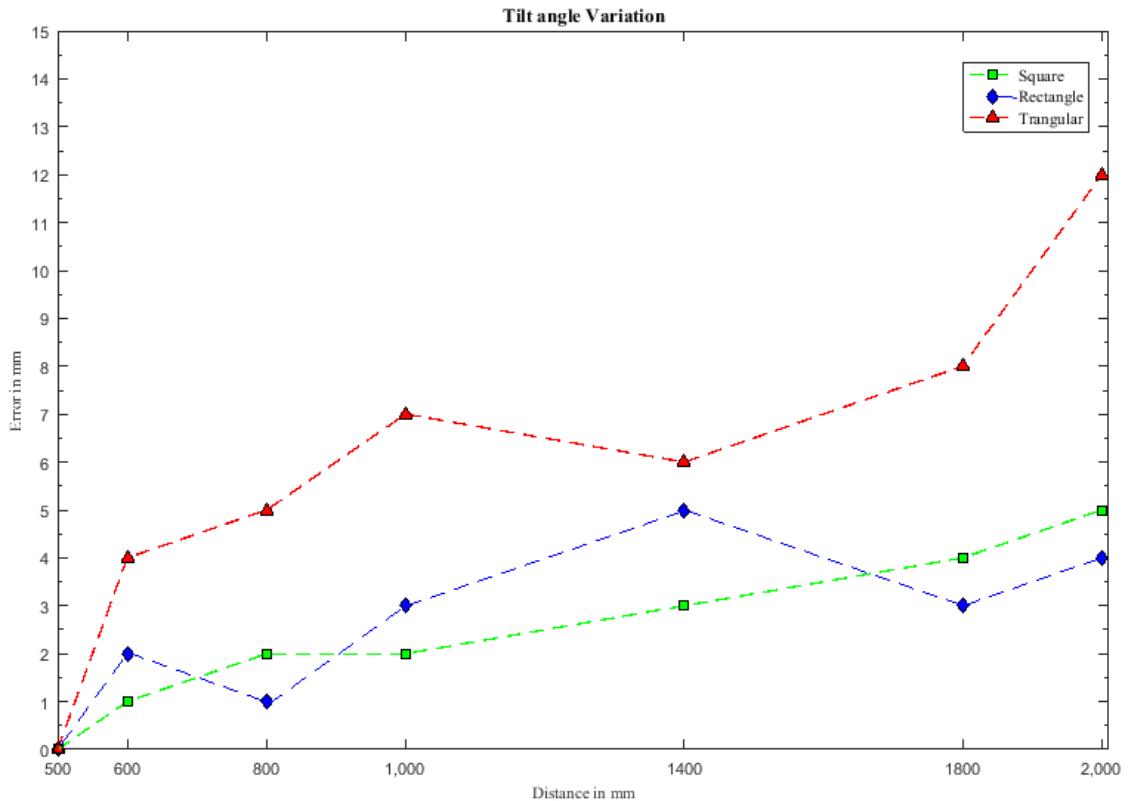


Figure 4.3: The average error of the depth and length in tail variation for three different geometrical shapes of calibration.

4.3.2 Cross Pose Angle Variation

Another point that was tested is point θ_2 that is the response to the pan angle of camera rotation. The difference inaccuracy was measured at angles of 0 and 90 degrees, and the distance between the camera and the image plane was measured at 1 metre. Point θ_3 is held at zero and $\theta_1 = 45$ degrees. The change of position with respect to the object comes by changing the camera angles of mobile camera that captured the pictures that affects the camera perspective of the objects. In Figure 4.4, the practical changing of pan angle shows that the camera cannot get the view of the object from angle 0 and 180. Hence, the

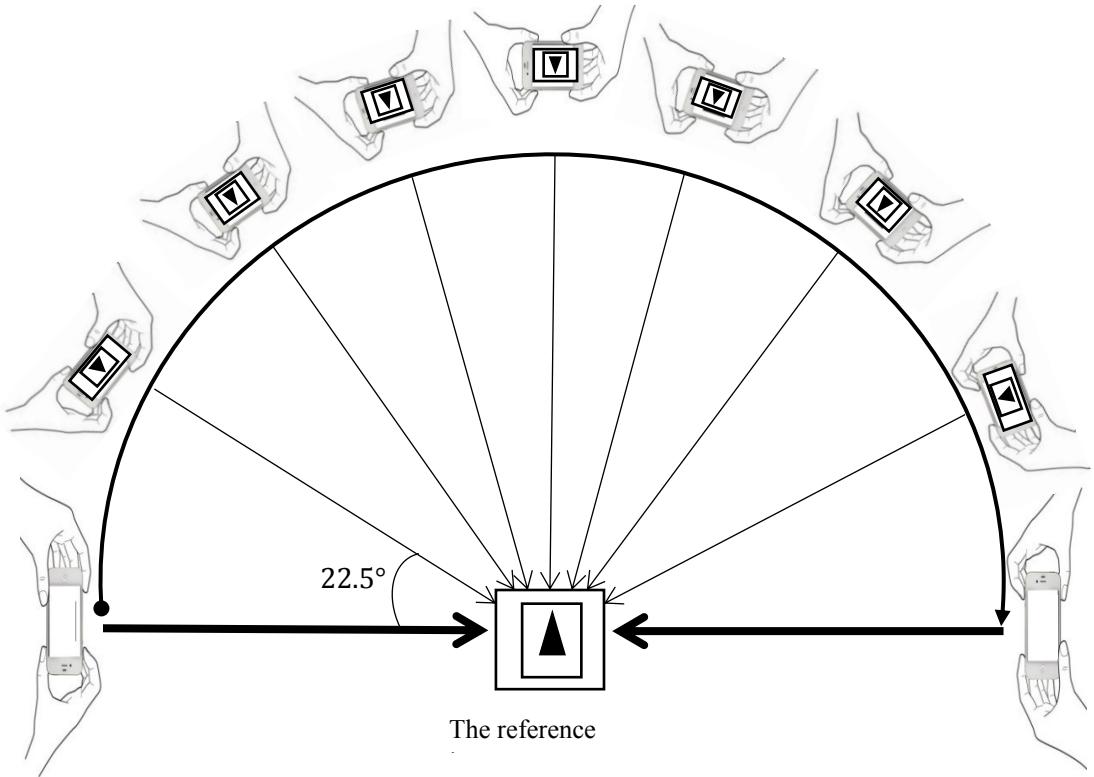


Figure 4.4: Changing the pan angle from 0 to 180° degrees (left to right),

$$\theta_1 = 45 \text{ degrees}, \theta_3 = 0^\circ$$

system of calibration needs to select the corners of the objects that are influencing the accuracy of the results. Therefore, the proposed method used clear angles that allowed to the camera to capture the object in the scene then get in the process without any effect on the object dimensions on the image. Figure 4.5 shows the deviation of inferred measurements from the original sizes. The variation in pan angle did not significantly affect accuracy since there was no difficulty in identifying the sheet corners. The study and practical work on pan angle found the normal distribution of average error for width and depth in difficult camera view angles is 0.8 mm, which did not affect the results. Hence, average results for width and depth were used to show the differences with positions that enabled the system to apply for face as image across pose in further work. Evidently, the camera cannot capture the object when the pan angle is between the camera image plane 0° and 180°, because the camera view is perpendicular on the geometric shape reference in flat space, which makes the angle of vision zero in both angles as shown in Figure 4.5. The most interesting part of the test was shown between the angles of 40° to 140° degrees when the results show the same average error.

These were early practical work that will move the project to the next step and support the main system of face recognition across pose in the next chapter.

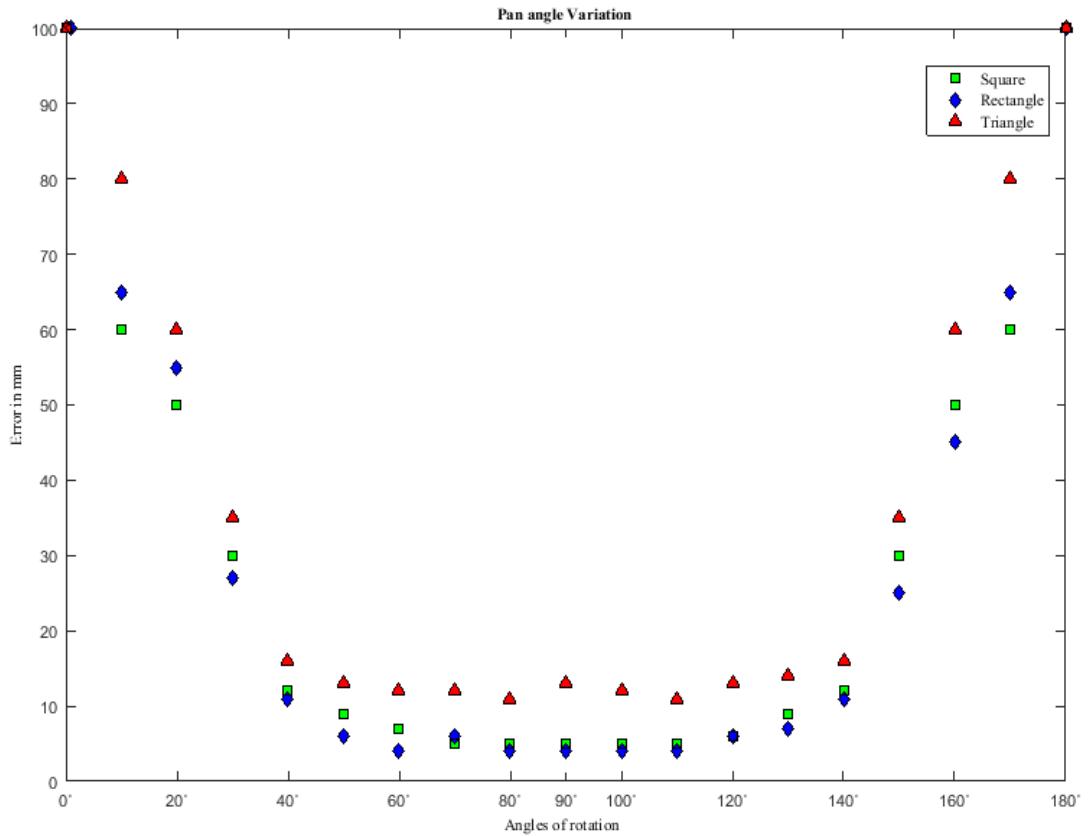


Figure 4.5: Average error for the depth and length with pan angle from 0 to 180°, for three geometrical shapes of calibration.

4.3.3 Swing Angle Variation

Variations in error with changes in swing angle, θ_3 , were also studied, and the results are shown below. The pan and tilt angles were fixed at 0 and 90 degrees respectively with a distance of 1.5 m between the camera and the image plane. The camera was rotated from the same starting position in front of the shapes as shown in Figure 4.6.

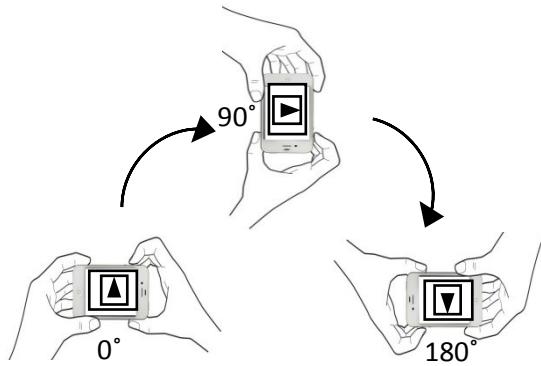


Figure 4.6: Swinging for the camera around 180° , when $\theta_1 = 90^\circ$ and $\theta_2 = 0^\circ$.

However, the views of the image were changed by modifying the swing angle, without any significant affect on the accuracy measurement, as it was in the target of the camera to capture it. Meanwhile, the measurements on the selected corners on the objects by swing angles will change the positions of the points without changing the corners shape in the image plane. The deviations were small for both dimensions of depth and length, so the average between dimensions was measured and compared with orignal ones to show the error in millimetre as shown in Figure 4.7. The average error between the depth and the length of the square was constant in each swing angle except for the angle of 110° , when it was the same as in rectangular shape with an error of 1mm more or less. The difference between the averages gained by automatic selection and human selection of the calibration points, is minimal. Furthermore, the triangle has more changes and variants in measurements, roughly 1mm more than other shapes. The camera orientation while taking a photograph and the positioning of the paper on the object plane does not affect the accuracy.

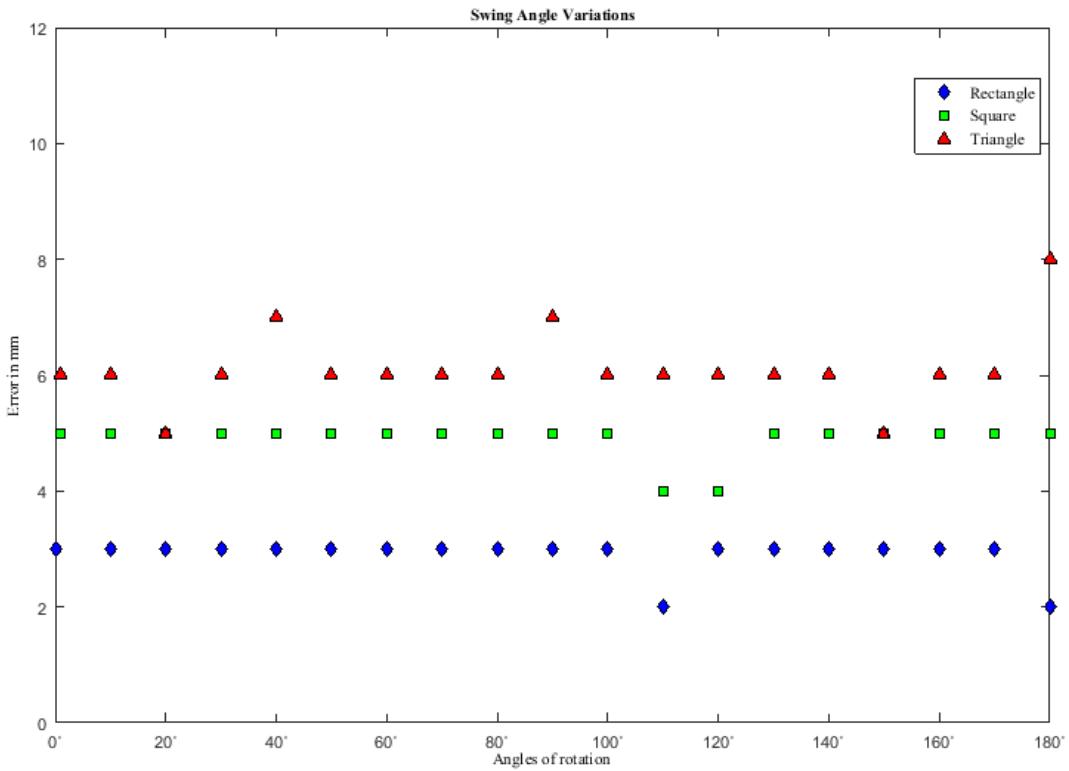


Figure 4.7: Average error in the depth and length with swing angle varying from 0 to 180°, for three geometrical shapes.

4.4 INFERRING DIMENSIONS FROM AN IMAGE USING AUTOMATIC CALIBRATION

Different applications may be used to infer 3-D information from 2-D. For example, applications concerning the site of the object, and 3-D information concerning the position and orientation of moving cameras. Both intrinsic and extrinsic parameters as explained in Chapter 3 need to be calibrated. In this project, the proposed system uses the measurement technique and face tracking in the next chapter. Moreover, the camera calibration needs be accomplished only once. There are different kinds of information used in 3-D applications that is concentrated on the position and orientation of the camera to get the object information, such as a camera when held by a robot about the target world coordinate system [61]. However, there are disadvantages of these methods regarding accuracy, as required by higher resolution images. Other methods use different views of an object to get 3-D from 2-D information images [47]. Again there are problems associated with such methods, as they require more than one image that depends on the contrast to create a 3-D face image. Extensive research has been undertaken [187, 188] to find certain visual inputs that act as reference objects.

The first step is to implement the Haralick algorithm [75] that operates on relative angles from different geometric shapes as shown in Figure 4.8. Further, we will need to determine an accurate estimation of the focal length from the vanishing points and camera models. Because of its important role in 3-D reconstruction and calibration, the detection of vanishing points in a scene has to be active, especially when no human intervention is required. One proposed method [189] is based on a geometrical approach in which all finite and infinite vanishing points are estimated for an image of a human-made environment. The solution is centred on the grouping of line segments that are identified in the picture; these represent data on the projective space. The calculation of the third vanishing point is not feasible in this work as the possibility of a third pair of parallel lines being present in the image is not certain. An approach to estimate the focal length using two vanishing points from the projection of a rectangle has also been proposed in [190]. This method allows for the calculation of the focal length without the need for multiple views of the object.

The most important step in any 3-D system is the generation of 3-D content. Many state-of-the-art camera models are specifically designed to produce 3-D models directly such as the stereoscopic dual-camera. These cameras utilise the same plane for configuration for mono-cameras. Each acquires a single eye's view, and depth information is computed using binocular inequality [191]. A depth range camera is another example. This is a conventional video camera enhanced with an add-on laser element, which captures a normal 2-D RGB image and a corresponding depth map. However, the main method for the conversion of data is to produce a depth map for each frame of the 2-D image [191]. In another study [190], a camera is modelled as a perspective camera method on the surface to make the calibration process easier and not requiring the utilisation of any precisely machined device with known geometry, such as a calibration grid.

4.4.1 *Regular Shape*

In order to overcome the challenges of camera calibration and vanishing points, the present study proposed an algorithm that uses a simple method to infer the dimensions of objects in an image captured from a camera with unknown parameters. The calibration uses one regular shape triangular, rectangular or square with a known size of the reference shapes as shown in Figure 4.8. This it can then be used to infer the sizes of objects in the image plane. In this work, a novel algorithm is used which is based on perspective projection.

The implementation of this scientific basis creates a novel technique that enables the dimensions of objects to be determined from the real world by using any image taken by any camera.

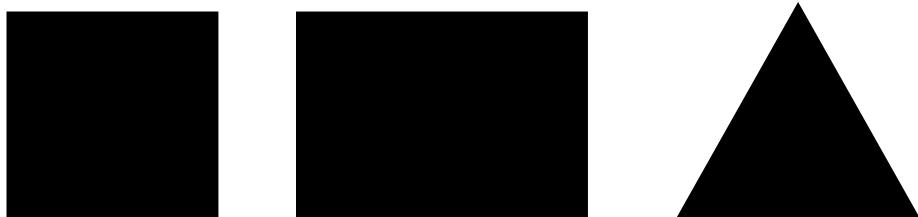


Figure 4.8: Three different geometrical shapes.

4.4.1.1 *Rectangle*

This part of the project uses geometrical shape to calibrate the camera and analyse the information from 2-D images. Haralick had started to use perspective projection in scene according to previous work [75], his calibration was related to three points of perspective for different orientations. Haralick algorithm provided information from rectangle shape to camera examining route. Experiments conducted to determine the exact 3-D coordinates from a rectangle of known size as it is widely known in photogrammetry. A sheet of A5 paper (Figure 4.9) was chosen for use as a rectangular shape in these experiments. First of all, the system used the perspective projection of the coordinates of three 3-D points to obtain the corresponding positions of the object plan and then it was possible to compute the position of the camera. The triangle point system of calibration extended from Fischler and Bollesof to find 3-D points [192].

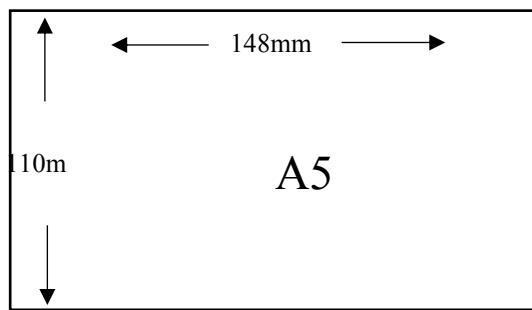


Figure 4.9: shows size reference paper $110\text{mm} \times 148\text{mm}$.

On the other hand, the angles of a rectangle should be computable from the standard 2-D dimensions which, from these angles, the projection drew in space image plan of 2-D to show the 3-D scene. Hence, the advantage of using the known width and length to create a new novel algorithm to determine the objects seen from the rectangular to the actual

dimension. The measured dimensions could be in units of mm or metres. The technique uses the first reference to compare the distance of the objects and measure it. The distance that was required for the rectangular is 279 mm for the width and 210 mm for the length. Chapter 3 explained the equations and algebraic coordinates used in this chapter, but it is not complicated because this project simplifies the algebra to the visual scene of the mathematics.

4.4.1.2 *Square*

To maximise the benefit from camera calibration to infer the intrinsic parameters, the measurement technique used geometrical square shape. One dimension of the 2-D surface can be used to derive the other dimension. The same idea of the previous step of using the rectangular technique, but with more use in real life. Many of the objects are different in shape but can always have standard dimensions geometrically; the four angles are selected from the meeting of the length and width in the same distance to draw the 3-D scene in space. As a known reference to the further dimension, the system used 2-D note paper $135\text{mm} \times 135\text{mm}$ as shown in Figure 4.10.

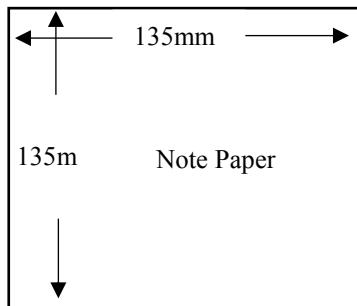


Figure 4.10: Size of the reference paper $135\text{mm} \times 135\text{mm}$.

4.4.1.3 *Triangle*

The triangle is the third reference shape used to calibrate the camera and to project the angles from 2-D to 3-D. The advantage of using the triangle shape in the proposed method is the additional benefit of another geometric shape that has different angles and prospect in the projection system. Hence, the new shape gives the proposed system a variety of measuring different surfaces of the objects and 3-D reconstruction. However, the use of the triangle can be a risk for the purpose of accurate measurement due to the analytical methodology used. The four angles of the shape compute to three points in space during the projection transformation. While, in the case of the triangular shape there are not enough (four) angles to achieve the accurate measurement. Thus, the system uses

estimating position on the other side of the third corner to complete the fourth angle. In the experiment it used two triangular meters of equal side triangle and the isosceles triangle (Figure 4.11). The experiment showed differences 15mm less or more in low resolution of the camera in the worse scenario which does not make a difference in the reconstruction of the shapes in 3-D, but it can make an effect in the field of measurement.

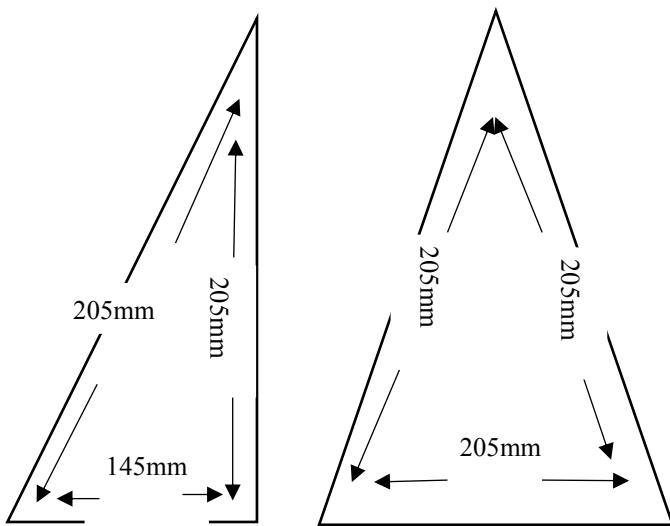


Figure 4.11: Isosceles triangle and equilateral triangle references.

The proposed study is assumed to be carefully accurate and the future work is believed to work in reconstruction building. Therefore, the work has improved the results of the calibration for this shape, which has not been used before in published work. The triangle shape can be investigated further in future work in different fields of research but not as an active part of a proposed project. Also, the practical work has not used the triangle shape in the experiment of face recognition.

4.4.2 Irregular Shape

To be able to understand the forms of the world and to make this understanding useful in the service of humanity, we need to be able to measure and have knowledge of the dimensions of shapes; in other words, the areas related to space and the position, size and shapes of things in it. When we know how to apply and understand the relationships between shapes and sizes we will be better prepared to use these in our everyday lives. Geometry assists in doing that because it provides knowledge of how to deal with measurements and the relationships between lines, angles, surfaces and solids.

The method proposed in this study uses the calibration of the camera based on a geometrical approach which employs perspective projection to measure the surfaces of objects. Our life is surrounded by objects of different shapes, the practical experiments on this subject found that irregular shapes are most convenient to use in different fields nowadays. Therefore, this study has developed a system of measurement to include irregular shapes. The camera calibration takes the previous regular shapes to determine the intrinsic camera parameter to feed the CMT and measure the points of selections from the image. Previous research has been interested in shapes and sizes, as well as in others field of visual abilities. The proposed method of study has included the both of interest that concentrate in imaginary space using geometry such as the shapes and size and visual information from 2-D images. The diversity in the method of calibration and the flexibility of CMT for 2-D for irregular shapes has many advantages in several approaches, such as medical uses. The CMT used different free shapes in the experiment to explain the ability of covering different areas as shown in Figure 4.12.

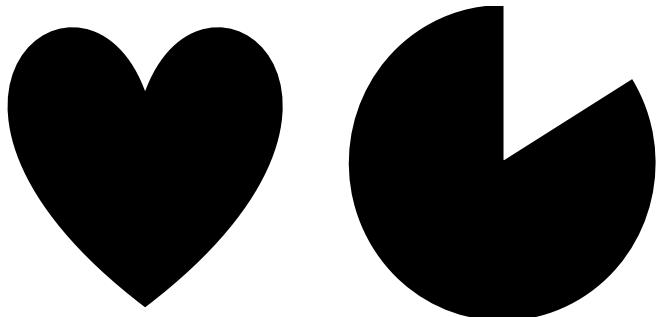


Figure 4.12: Two different irregular shapes to be measured.

4.5 RESULTS AND EXPERIMENTS

4.5.1 *Measurement of Dimensions for Regular Shapes*

The method has been developed through computer simulations based on the geometrical mathematics used in Chapter 2 to determine real world dimensions from objects in images. It has been observed that the accuracy of the technique may depend on camera orientation. The deviation of inferred measurements from real dimensions depends on the camera angle used. When the camera angle changes, the perspective projection of the reference sheet onto the image plane changes as well and affects the accuracy of measurements taken. The following experiment measures the dimensions of surfaces by

placing different sheets of papers with the selected geometric shapes and a triangular ruler at 90°. The real dimensions of all of the shapes are shown in Figure 4.3 and the images were captured using a Samsung mobile Galaxy SIII (GT-i9300), which has a resolution of 8 MP. The results are described in detail in Figure 4.13 and

In another experiment using the calibration of the camera for one shape and the measurement of the others, the camera was calibrated for a sheet of A5 paper and then the dimensions of the rest of the shapes were inferred (Table 4.1).

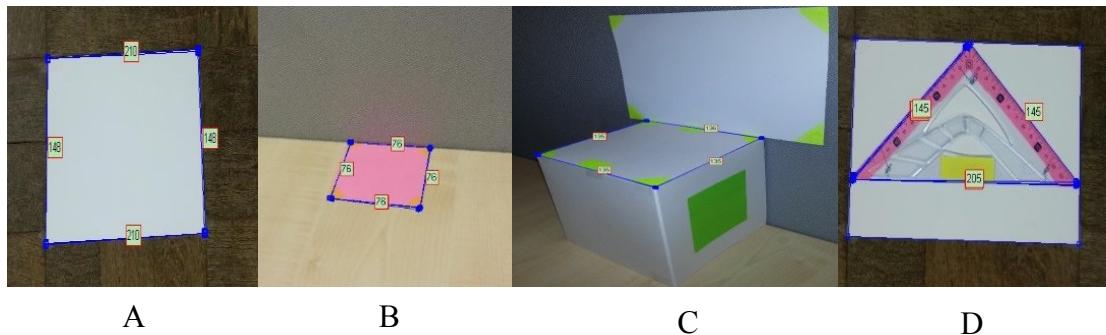


Figure 4.13: The results of CMT for the standard size of shapes.

In order to measure the accuracy of the inferred dimensions given, three different kinds of cameras were used a Nokia Xpress Music 5220 that supports a 2-MP, a Samsung Galaxy Mini GT-S5570 with a 3-MP camera, and a Samsung Galaxy SIII with camera 8 megapixels.

Table 4.1: The result of measurement for the geometrical shapes

Geometric shapes		Parameters of the image	
		Real dimensions (mm)	CMT Dimension (mm) of image
A5 paper	A	210×148	210×148
Note-paper	B	76×76	76×76
Card blank	C	135×135	135×135
Triangular ruler	D	145×205	145×205

From four different distances between the camera and the plane, the error between the real dimensions and those obtained with the cameras were measured. Figure 4.14 shows the variations in the accuracy of measurements for the four pictures captured by the 2 MP camera, taken from four different distances between the camera and the real plane.



Figure 4.14: Pictures captured from different distances by 2 MP Camera.

The errors in measurement shows outstanding results in that the accuracy of measurement, which was within 1 mm when the picture was at a distance of 600 mm from the plane (Figure 4.15).

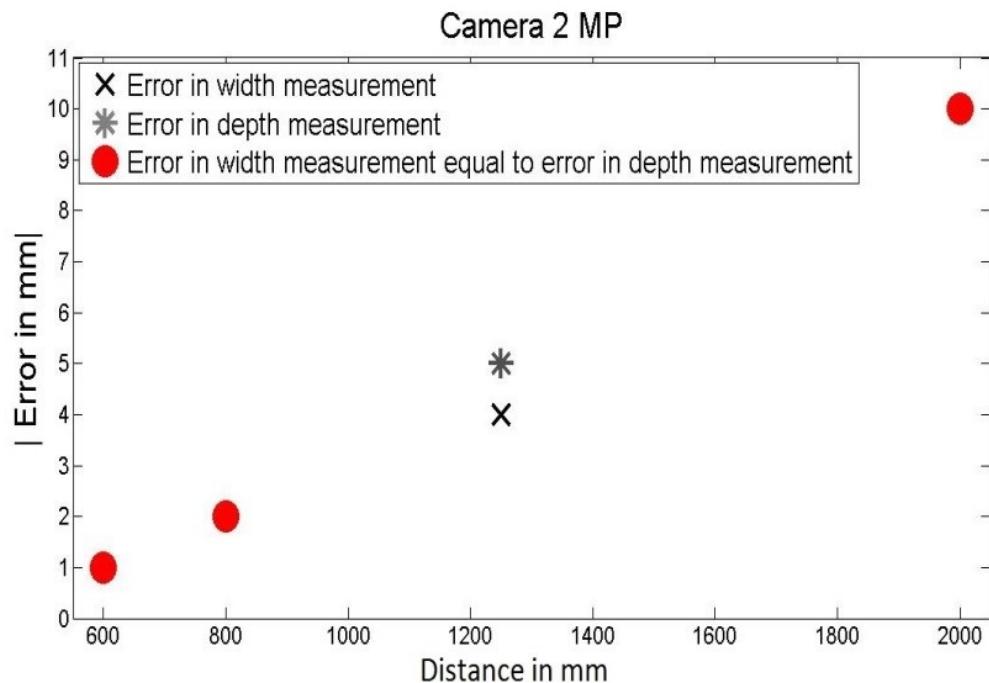


Figure 4.15: Alteration in error with distance -2 MP.

In another experiment, the pictures have been captured by 3 MP camera from different distances. Whereas, Figure 4.16 shows the minimum and maximum errors in inferred dimensions regarding distance. The accuracy shows variation of the width in 7 mm, which is considered as the lowest point of accuracy obtained with this camera. The final experiment and results are expressed in Figure 4.16.

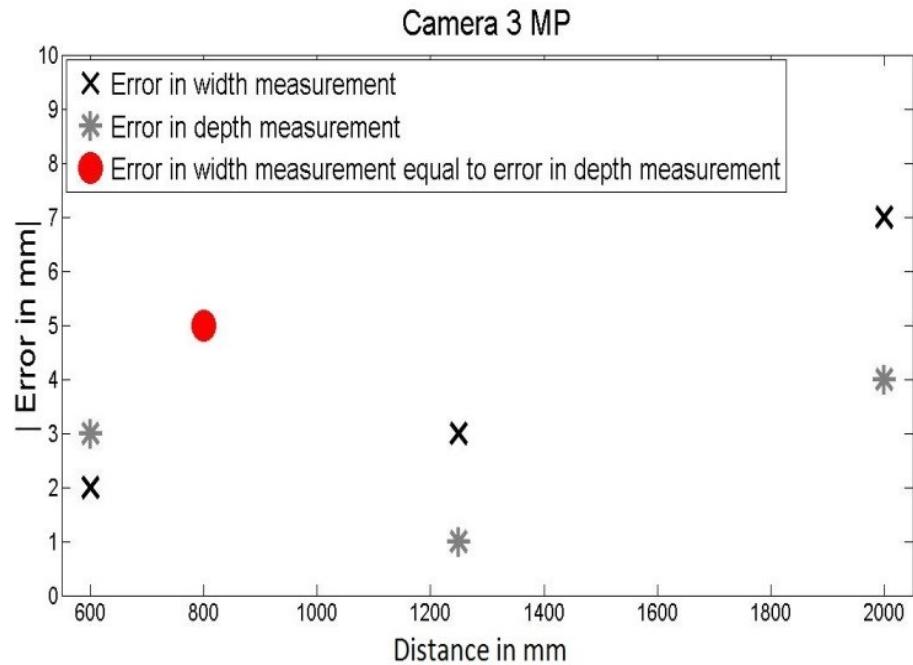


Figure 4.16: Alteration in error with distance -3 MP.

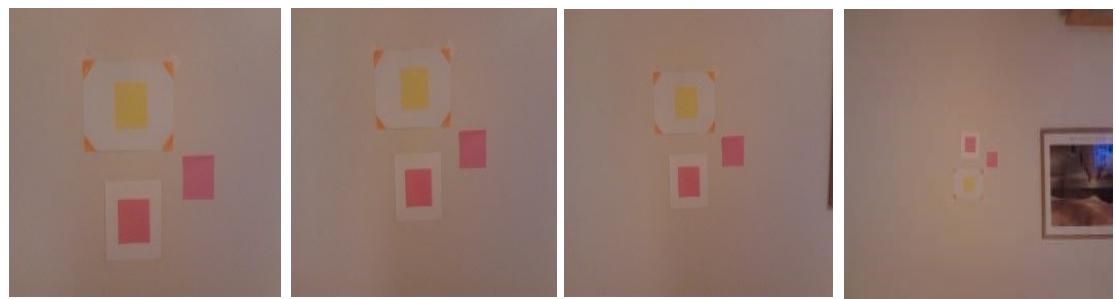


Figure 4.17: Pictures captured from different distances by 8 MP Camera.

Figure 4.17 shows the highest camera resolution used in the experiments, which is 8 MP. The practical experiments ensured the accuracy of the dimensions, which can obtain a measurement of 5 mm for both width and depth (Figure 4.18).

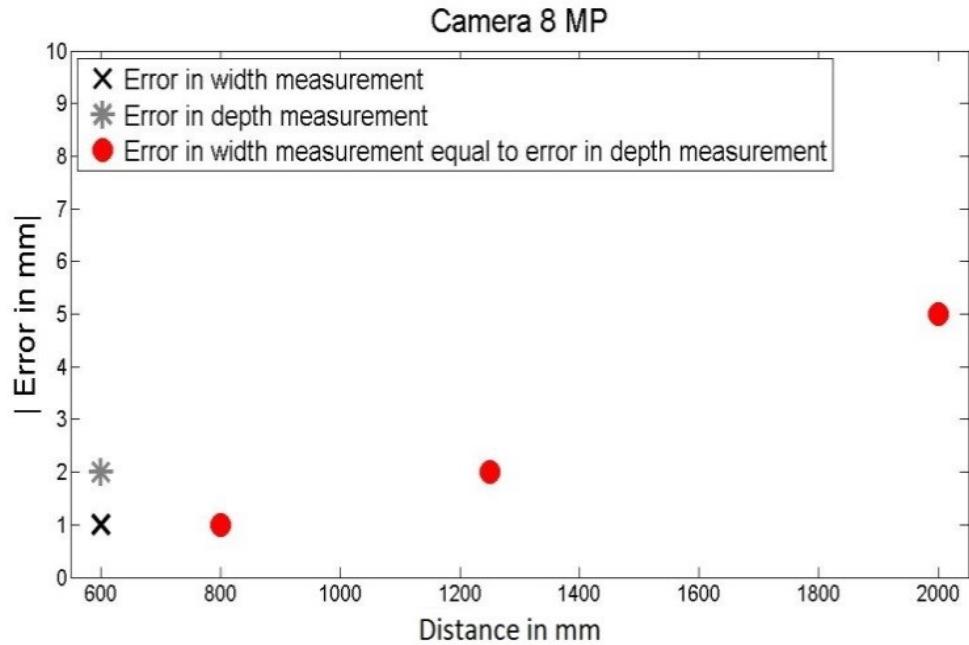
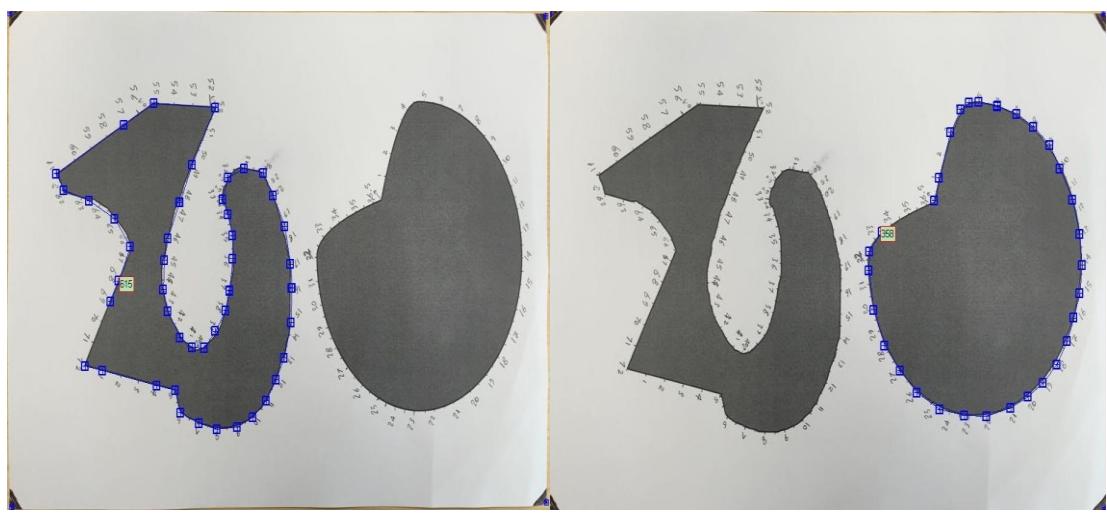


Figure 4.18: Alteration in error with distance -8 MP.

4.5.2 Measurement of Dimension of Irregular Shapes

In a further experiment to prove the effectiveness of the method; irregular shapes were measured from 2-D images, which gives an excellent opportunity to develop and use in a different field. Figure 4.19 shows the measurement for half of a heart as an irregular shape. The corner of an A4 paper reference is used to calibrate the camera from known dimensions that enables the CMT to calculate the boundaries of the half of the heart.



B

A

Figure 4.19: Measurement of an irregular shape.

The results inferred 280 mm for the half of the heart from the real measurement of 275 mm. Figure 4.20 demonstrates the measurement of dimensions of two irregular shapes. The objects were created using the core painting programme in Windows 7. The results from applying CMT on the two images A and B from Figure 4.20 supported the accuracy and robustness of the proposed system because the two images have many curves and angles. Hence, the measurement of image B shows 358 mm from the real measurement 335 mm, that shows 6.6% of error and 1% more in image A.

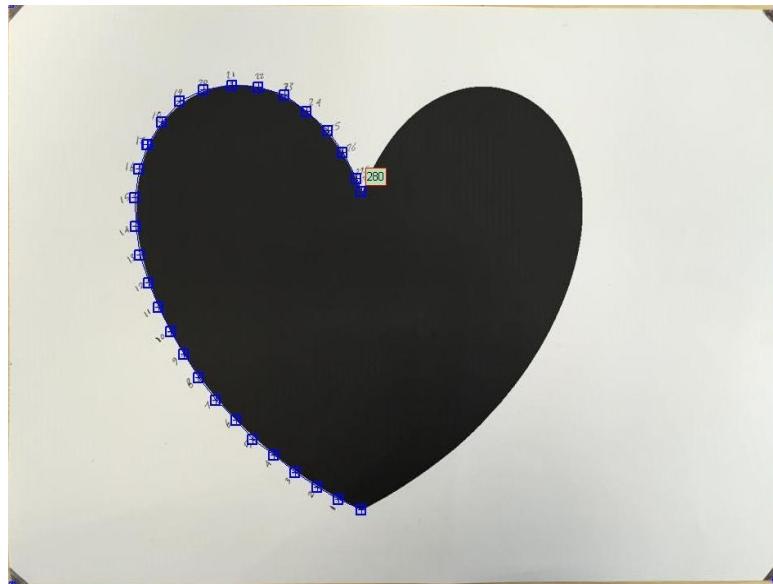


Figure 4.20: Measurement of two different irregular shapes.

4.6 DISCUSSION

The experiments used 2-D images captured from three different cameras. This variation does not affect the results of the measurements, because the perspective projection uses the object plane of calibration to give the value of focal length used to measure the depth of the object. The results of experiments for regular geometrical shape show slight differences in accuracy which vary between the geometrical shapes. The rectangle and the square give almost the same distance from the reality and it has been used for the proposed method. The error in measurement was highest when the camera was farthest from the object plane at 2 metres. The error was 10mm between the real object dimensions and the measurement using the proposed technique from a low-resolution 2MP camera . This error reduced as the experiment changed the camera resolution within different mobile cameras, if 3 MP camera was used the error was 7mm, and 5mm was achieved when the camera resolution was 8 MP.

On the another hand, the results of irregular shapes measurement shows the error in distances between the real dimensions in 2-D and the distances that measured by CMT. The percentange of error in measurement was 1.8%, when the calibrated shape has one or two direction of the curve as in Figure 4.12. Whereas, the other shapes of measurement had 6.6% and 7.6% of error when the curves had random boundaries.

Finally to summarise the results of irregular shapes from the experiments, the Table 4.2 shows the initial measurement of the 2-D image shapes that were chosen manually.

Table 4.2: Summary of measurement error

Figures	Real dimensions	Inferred dimension	Error in mm	Error Percentage MSE
Figure 4.19	280	275	5	1.8%
Figure 4.20 (A)	358	335	23	6.6%
Figure 4.20 (B)	615	690	75	7.6%

4.7 SUMMARY

In this chapter, a new flexible automatic technique has been developed to calibrate a camera to be used for easily inferring the dimensions of the objects. The technique only requires the camera to observe a planar measurement from different geometric shapes. A rectangle, square and triangle are used as flat objects to determine the camera parameters that enable dimensions of the real objects to be inferred. The system of calibration was tested for three angles of a rotation before testing it in the experiments. Tilt, pan and swing rotation gave different angles of object rotation and the test showed the accuracy of inferred measurements with only small levels of error that did not affect the main purpose of measurement. However, the weakest triangle shape in calibration and measurement were selected based on the practical test of the shape that shows limitation of use in measurement. Moreover, the camera view angles were tested and selected the range of possible view. For example, angles of 0° and 180° in camera pan give no views of the object, which prevented the calibration and then measurement of dimensions. As it planned from the research to use the calibration system in face recognition cross pose, the previous experiment focused on the tilt and pan angles, and the swing that have not shown significant problems or large difference in distances.

The calibration depends on the constant pixels of the image, which is influenced by changing the resolution of the picture. Many new algorithms and techniques use a high-resolution stereo camera to infer the depth of 3-D dimensions from 2-D images. Thus, the use of an algorithm faces many challenges related to image resolution, changes in distance and shape, and the cost of the technique. The experimental results show that the inferred dimensions were highly accurate. The technique allows the use of varying resolutions and the results show that the algorithm works with low-grade image quality such as 2MP resolution and with good quality, such as 8MP, which provides excellent accuracy in inferring real dimensions. The 2MP, 3MP and 8MP cameras used are available at low cost for any user. Furthermore, three different geometrical shapes were used to measure regular and irregular shapes to give a wide opportunity to accommodate the largest possible number of formats that are measured around us. Furthermore, the results tested the variant distance of the measurement up to 3 meters to allow the technique to work widely in-door for different purposes of measurement. This technique can be applied to any image captured from any camera.

The triangle shape was not used in these experiments as it showed the maximum error of the measurement in test part compared with the other geometrical shapes of calibration. Also, the project will need the rectangle and square shapes that are used in the window for face detection to face recognition. In the next Chapter 5, the research will present the new system for face recognition based on the inferring information from intrinsic parameters of camera calibration that was achieved in this chapter.

Chapter 5 : NOVEL METHOD OF FACE RECOGNITION FROM VARIOUS POSES

Camera surveillance and face recognition are useful tools to identify criminal activity and avoid crime. The various face poses is a grand problem that is a challenge in face recognition and computer vision. In this chapter, a novel method of face recognition from various pose is proposed based on a geometrical approach of measurement that is efficient in handling the range of pose variations within $\pm 60^\circ$ of rotation. The method is called the Face Camera Measurement Technique (FCMT).

A unique approach of camera calibration from a 2-D image is demonstrated, using fixed and robust facial landmarks to ensure a reliable estimation of real dimensions. In the first part of the system, the images need to be decomposed into grayscale before feature extraction and selection. The original image is then transformed into a full-face pose in order to accurately estimate the distance between the eyes. Finally, recognition can be made more stable against the pose differences when the distance between the eyes in frontal picture is compared to all pose pictures. Extensive and systematic experimentation using the FERET database showed that the proposed method consistently outperforms single-task based baselines, as well as state-of-the-art methods for the pose problem.

Different approaches to face recognition are discussed and the benefits of the proposed approach are identified. The proposed system of face recognition started with quadtree decomposition technique to end with face detection. Comparing different methods and different database supports the results, and the system components have been tested for each level of progress.

5.1 INTRODUCTION

Recently there have been many advances and developments in communication technologies and cameras to estimate real dimensions from 2-D images. This link between the technologies make such tasks easier and more accurate due to recent research in computer vision and camera calibration [53]. Therefore, face recognition from still images and videos are involved as part of this development, based on computer vision methods. The verification of a person's identity is an active research for different purposes, and the most common applications used are in the field of surveillance [193]. Moreover, many appearance-based learning methods are used for facial recognition. Traditional linear subspace learning techniques for dimensionality reduction such as proposed by Haijun et al. [194] represent input data as vectors or 2-D matrices and a solution is found by optimal linear mapping to a lower dimensional space. However, Haijun et al. technique is inadequate when dealing with sensorial data as it can cause a loss in the natural structure identified in the original sensorial data. The technique is also associated with high computational costs and memory demands.

The camera calibration method used in this work measures and projects the image plane into 3-D. Nevertheless the matching faces with different poses is challenging since it represents a change in 3-D space, but there is only the information of 2-D appearances in images of the face. One way to recognise a face in various poses is to reconstruct the 3-D shape of an input face. Geometric approaches derive the head pose from the geometric configuration of facial features depending on how the face is positioned about the silhouette of the head, or how much the nose deviates from bilateral symmetry [195]

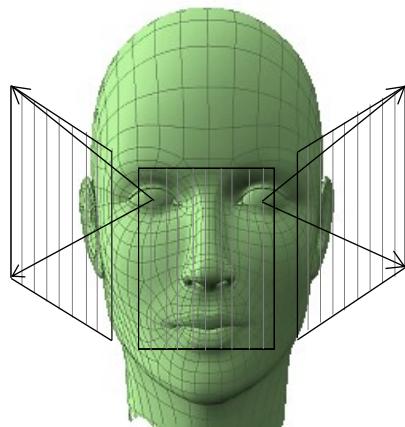


Figure 5.1: The theory of the proposed method of 3D artificial human face.

One group of approaches uses the outer corners of the eyes, the outer corners of the mouth and the position of the nose [195]. Thus, face images can be matched by identifying various poses.

Another technique is a linear subspace-based approach for face recognition where a face image can be represented in training as a line as a combination of images. Figure 5.1 shows that the geometrical shape of a face is always constant and can move with image reference planes without variation. In this chapter, images from the FERET database [29] were used with different face images captured from different viewpoints of each person between $\pm 60^\circ$. This work on face recognition with various poses using coefficients estimated from 2-D face images. However, the 2-D face images only reflect the incomplete parts of the 3-D face, and pictures in different poses correspond to various parts of the 3-D face. The constant measurement for the face during the rotation can be formed as a recognition problem. Consequently, the measurement of the distance between the eyes was taken as a parameter to recognise the person from the images.

5.1.1 *Face Recognition Approaches*

Three main tasks of face recognition may be named: holistic, local and geometry approaches. Such systems compare the image of a tested person with photos of people who have access permission to joint used object. The proposed study tested several methods for the above-mentioned tasks to conclude it in the geometric approach.

Knowledge of the internal parameters of the camera is a necessary step in 3-D system, as well as photogrammetric applications that infer knowledge about faces and locations from images. Inferring the dimensions of objects from 2-D images is considered a useful method for camera projection to define the internal parameters of the camera [171, 196]. In the same development, approaches between camera projection and face recognition have been used recently to identify the distance [44, 197-199]. However, much of the analysis used for this area depends on illumination, rotation (pin, tail) and different ages for the same person. Camera calibration and applications have received significant attention [200, 201]. Patch correspondences based on 2-D affine transformation enhanced by Ashraf et al. [202] provide parameters to facilitate face images across pose to give 3-D geometric information. Research to identify novel approaches that directly measure the similarity between patches with the basic idea, ridge regression and lasso regression are explored by Annan Li et al. [203] using coupling bias and variance in the regression for different poses. There are three main approaches to face recognition within this field of research, which challenge the various facial orientations; a holistic, local and geometric approach as it has been explained in Chapter 1.

5.1.1.1 Geometrical Approaches

This study proposed a new approach to face recognition, which is geometrical. It is a combination of both holistic and feature extraction approaches. 3-D Images are used in geometrical methods when the image of a person's face is caught in 3-D. The information of the face enable the proposed method to follow the curves of the eye sockets and select the main interesting points such as the pupils. Even a face in profile would serve in depth information and feature measurements, which gives it sufficient information to construct a full face [204, 205]. Usually, the 3-D system in face recognition uses many steps to achieve the recognition such as detection, position, measurement, representation, and matching. The proposed approach used similar information that used a 3D system without the need to reconstruct the faces, with minimum face view requirements. The novelty of this work is that the calibration of the camera from 2-D images is used to measure the distances between the facial features, with changes to the depth being managed by the system.

On the other hand, these approaches are practical in facial expression recognition [206-209] that is based on the unique points in the special region on the face. For instance, Figure 5.2 shows the first experiments that have been done in this work in real time, based on the distance between three points ear to ear. Conversely, matching faces from different poses in 2-D images is challenging since a pose varies in 3-D space but only information from the 2-D appearance is given of images of the face. One-way to recognise a face across poses is to reconstruct the 3-D shape of an input face. Geometric approaches calculate the information of the face when it moves across pose from the geometric configuration of facial features when the face is positioned about the profile of the head, and the shape of the nose deviates from bilateral symmetry [195].

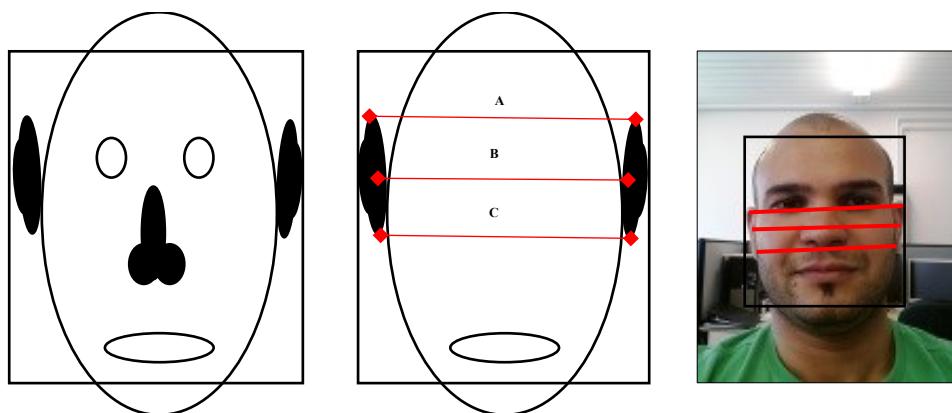


Figure 5.2: Harvesting dimensions ear to ear from three difference points.

Another group of approaches uses the outer corners of the eyes, the outer corners of the mouth and the position of the nose [210]. The most similar method for the proposed method was presented by Annan Li et al. [211] and was further improved using face representation by coupling bias and variance in the regression for different poses. Therefore, the proposed method has been tested and compared with Blanz and Vetter [116], Ashraf et al. and Annan Li et al. This works based on various pose face recognition using coefficients estimated from 2-D face images as can be seen in Figure 5.3. However, the 2-D face images only reflect incomplete parts of the 2.5D face and images in different poses correspond to various parts of the 2.5D face. The camera calibration in proposed approaches measures and converts the dimensions from a 2-D plane to 2.5D.



Figure 5.3: Concept of the proposed method with front, left and right side references of various poses from FERET database.

5.2 FACE TRACKING AND DECOMPOSITION

The proposed system of face recognition is built from three reliable algorithms that are used to obtain accurate information for facial points. These three algorithms are:

- camera calibration,
- dimension measurement,
- face tracking and recognition.

Previous chapters have explained the first two algorithms related to the camera calibration and the technique for measurement. This section, introduces the first part of the system, which is called face quadtree detection FQD. The method is a fusion method between face detection and quadtree decomposition image. The algorithm of face detection is based on quadtree areas of the face to get the best results compared with other methods such as Viola-Jones [212-214].

5.2.1 Quadtree Decomposition

Quadtree decomposition is a method that divides a square image into four equal sized blocks. Each block is tested to see if it meets some criterion for homogeneity. The test criterion is applied to the blocks that are not divided any further. Therefore, the proposed use of quadtree depends on determining regions of interest for intensity analysis in images [215-217]. In Figure 5.4, the early results of using decomposition shows one sample of the ORL database, when the algorithm was applied, then the CMT showed the measurement between the eyes for two different images. For example, when $S(k,m)$ is nonzero, then (k,m) is the upper-left corner of a block in the decomposition, and the size of the block is given by $S(k,m)$. By default, quadtree decomposition splits a block unless all elements in the block are equal.

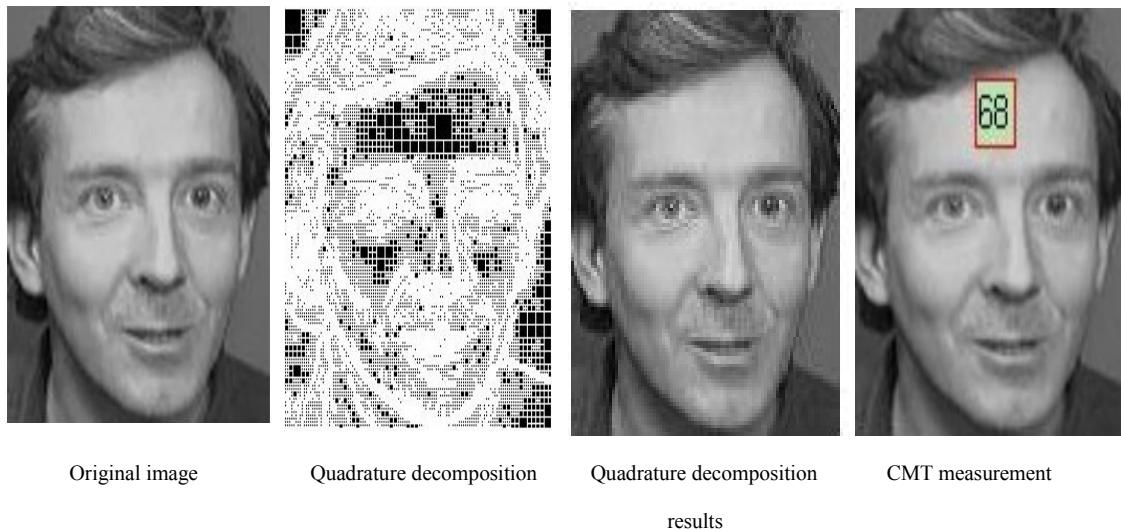


Figure 5.4: The process of features measurement applied with Quadrature decomposition and retrieval information image to the original.

The results showed that quadtree decomposition for images reduces the error of tracking by half, compared to the results obtained from the Viola-Jones algorithm. However, tracking the face is the initial part in recognition problems and is a mechanism to calibrate the camera. The testing of the quadtree decomposition algorithm on still images gives high performance for face tracking and the selection of landscape points on the face. As shown in Figure 5.4, the ORL database [218] was initially used for 40 people, each having ten different images with different

expressions, poses, or features. Furthermore, from the 400 face images was compared the effect of quadtree decomposition on the face tracking and the selection of faces. When face tracking by the Viola-Jones algorithm was analysed, there was an unsuccessful result where the system could not detect 60 pictures out of 400. Whereas, only 2 pictures could not be detected with our fusion proposed methods of FQD. To ensure the comparison was fair, the results were tested with 100 subjects from the FERET database with each analysis having nine images of the face with different poses $\pm 60^\circ$. The benefits of the fusion of these algorithms include giving lower error in the tracking results and more accurate selections of landscape points. The FQD method is based on the accuracy of the dimensions gained in CMT.

5.2.2 Face Tracking

The proposed method of face recognition considered the Viola-Jones algorithm as a useful method for face tracking [212-214]. The proposed system has implemented, developed, and enhanced the Viola-Jones algorithm further so that the facial landmarks are selected automatically. The classification of proposed method of tracking is qualified by extracting features from 2D images. These extracted features are then fed into the face to detect it. The proposed method of face tracking used viola et al. [219] and cascade classifiers used to train the classifiers defines the smallest region containing the object. When the detector incrementally scales, it is the input image that is used to locate target objects. At each scale increment, a sliding window, whose size is the same as the training image size, scans the scaled image to locate objects. The algorithm is flexible and adapts to facial poses with different angles for the tracking in $\pm 60^\circ$. Moreover, the algorithm was enhanced to use any size of the image with colour and grayscale images in sharp angle of view. The results tracking for each view of the face is dependent on the geometrical mathematics.

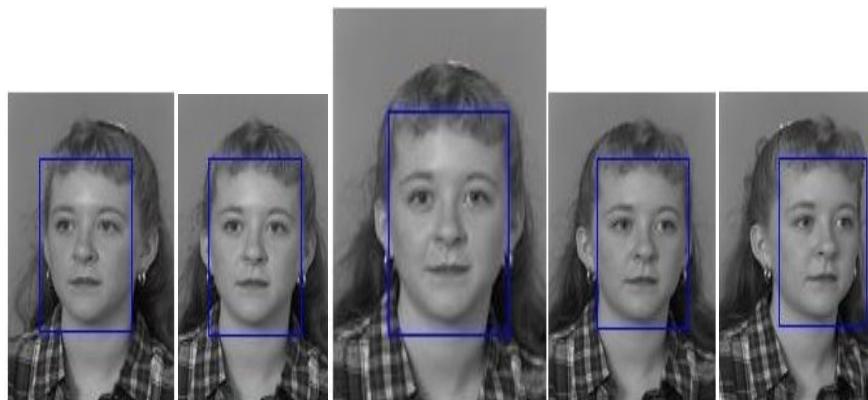


Figure 5.5: Example of face tracking.

Furthermore, the human face have different inclines and declines areas that can affect the measurement plane between two points, thus the proposed system of CMT has addressed this problem and proved it by measured 3D shape. The results were compared to the original dimensional distances, and the concept was demonstrated in the cube shape as shown in Figure 5.6. Meanwhile, the early example from the results in Figure 5.5 shows the accurate tracking for 6 images across poses for one object taken from ORL database.

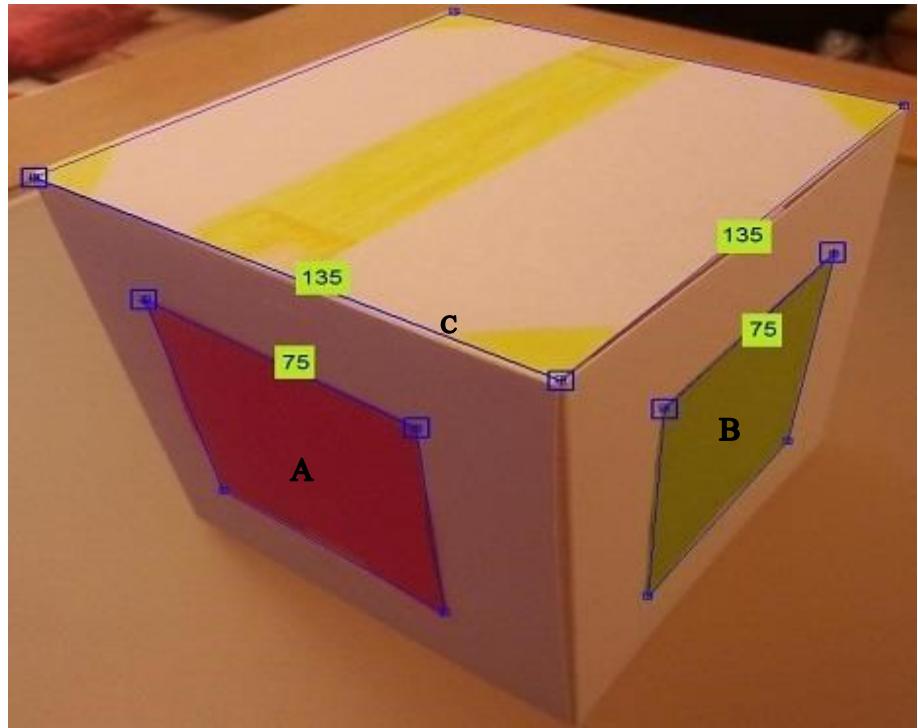


Figure 5.6: Three side of calibration for cube (A, B, C) shows the frontal and two sides of view.

5.2.3 FQD

The propose method of face recognition is based on a problem of face recognition across poses. Therefore, from the previous implementation of the quadtree image decomposition algorithm and face tracking that developed from algorithm viola et al. [219] that is based on four stages: Haar Feature Selection, Creating an Integral Image, adaboost training and cascading classifiers. The proposed algorithm used both algorithms combined into one algorithm, called Face Quadtree Decomposition (FQD). The automatic measurement of CMT based on the selected points of the face to improve the feature points on the face, as the proposed system can affect the accuracy of finding the same point when the face is rotated using quadtree decomposition, which is explained in Figure 5.4. The FQD selects the points of interest that the system uses in face recognition such as the eyes, nose and mouth. Whereas, the FQD covered all the detection

for the face features and used the pupils as the main selection points to build the system of recognition as shown in Figure 5.7.

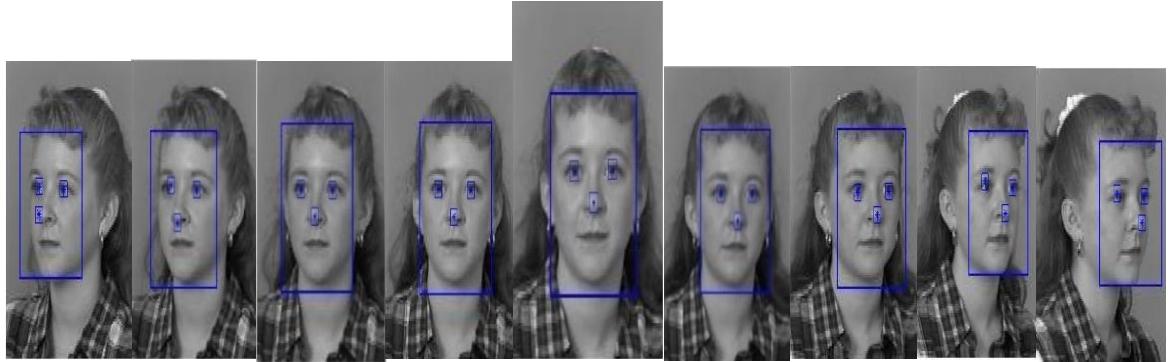


Figure 5.7: The output of FQD on one object from FRET database.

The performance of the Viola-Jones algorithm after the face quadtree decomposition algorithm achieved high accuracy of the box face tracking, and decreased the error of tracking to 1%. This part was successfully used as a base for our calibration to fix the overall face dimensions because some images gave an error of tracking. Also, the auto selection of the landscape face points for pupils was more robust with the change of the face orientation. The best tracking performance was achieved for the more challenging angles of tracking at +60° and -60°. Previous tests for the CMT showed that the accuracy in the worst case scenario from a long distance to 3m, for the low-resolution camera, 2 MP, was 94% which is accurate for our purposes [196]. Our new method uses a novel technique in camera calibration CMT that measures the 2.5D from 2-D images [196] and facial tracking to recognise the face under changing orientations cross angles ±60°. The method used camera calibration with face tracking while it crossed facial orientation to get the actual dimensions, and the measurement was from two landmarks of fixed points- the pupils of the eyes.

5.3 FCMT

After tracking the face and selecting the main facial points on of interest accurately, these features are used with the CMT technique for each image and object across poses. The new system of Face Camera Measurement Technique FCMT is based on the measurement of the distance between a person's pupils across pose angles between -60 to +60. These distances are measured in mm [220]. The window of faces measured based on the actual life in researchers of inter-pupillary distance (IPD), the distance between the centres of the pupils of the two eyes. The Table 5.1 shows the average between pupiles distance for male and female personnel in the USA army[220, 221].

Table 5.1: IPD values (mm) [219]

Gender	Sample size	Mean
Male	1771	64.7
Female	2205	62.3

In practice, any image chosen from the database has to be firstly pre-processed with FQD to track the face and select the corners of a window of the face. Afterwards, FQD automatically selects the two points of the pupils of the eye. The output image received from this is then used by the CMT to determine the internal camera parameters for camera calibration. The inter-papillary distance can be measured by CMT to give a unique identification for every person. The window of face detection for each individual is used with different orientations, and Figure 5.8 shows the steps of the FCMT system.

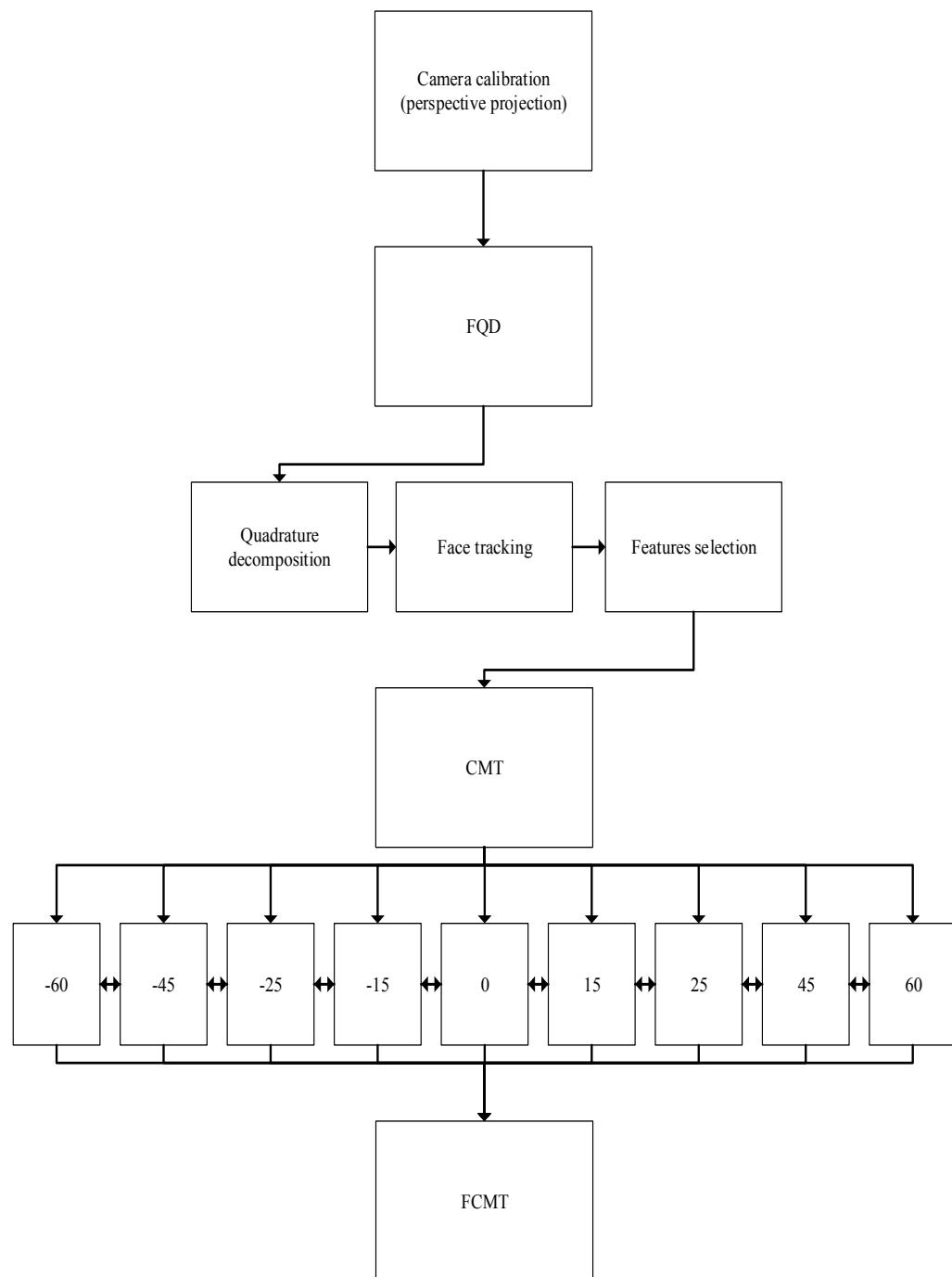


Figure 5.8: The FCMT system.

5.3.1 Physical and Mathematical Model of FCMT

The camera calibration is a necessary step in 3-D computer vision in order to extract metric information from 2-D images [222]. Furthermore, automatic face recognition systems compute a similarity score calibration between an official enquiry [197]. Hence, this paper presents a novel method to track and recognise the faces of an individual from a single still image across poses. The first step of the FCMT is to generate a coordinate system of perspective projection from Haralick algorithm [75], that operates on the corner of geometric shapes from the 2-D images. The geometric face box is used to calibrate the camera, as shown in Figure 5.9. The Viola-Jones tracking algorithm [28-30] is applied after enhancement to fit the system as explained in the previous section. The general idea is to generate a coordinate system that will enable accurate estimates to be made of the cropped face box from various poses in images. The system detects a point in space automatically by calibrating the geometric face box. Then, from the four corners of the box, three angles of rotation in the plane are projected. The camera projection is explained below, in which OXYZ is the real coordinate system and O''X''Z'' are the image plane coordinates as shown in Figure 5.9. The human eye is assumed to be the centre of the coordinate system at O.

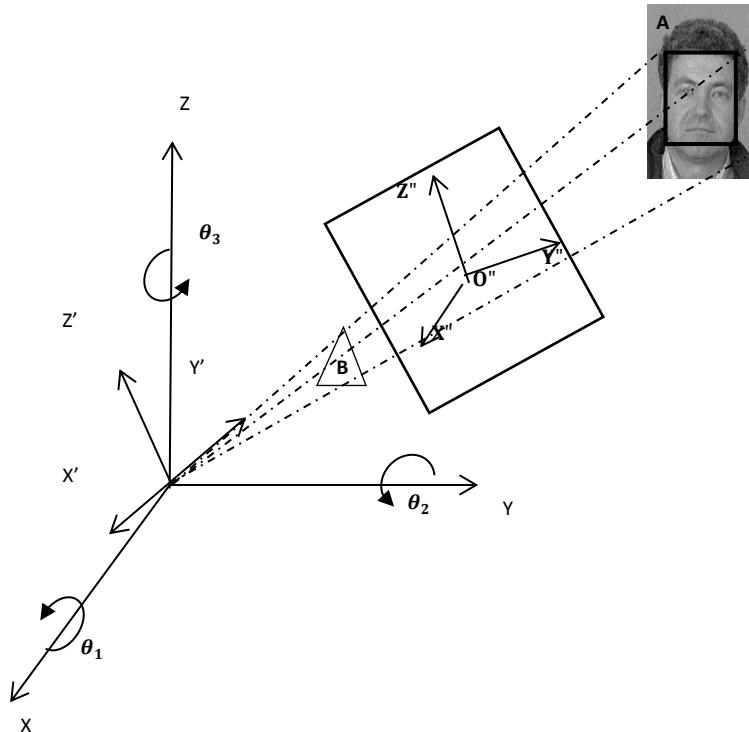


Figure 5.9: Camera perspective showing projection of a point in space.

Any point A lying on a plane in space will have its projection on an image plane at B and θ_1 , θ_2 and θ_3 are the three angles of rotation. This method allows the calculation of focal length without the need for multiple views of the object. From the basic system explained in Chapter 3, the focal length and image coordinates of the corners of the geometrical face box, the three angles of rotation can be calculated [18].

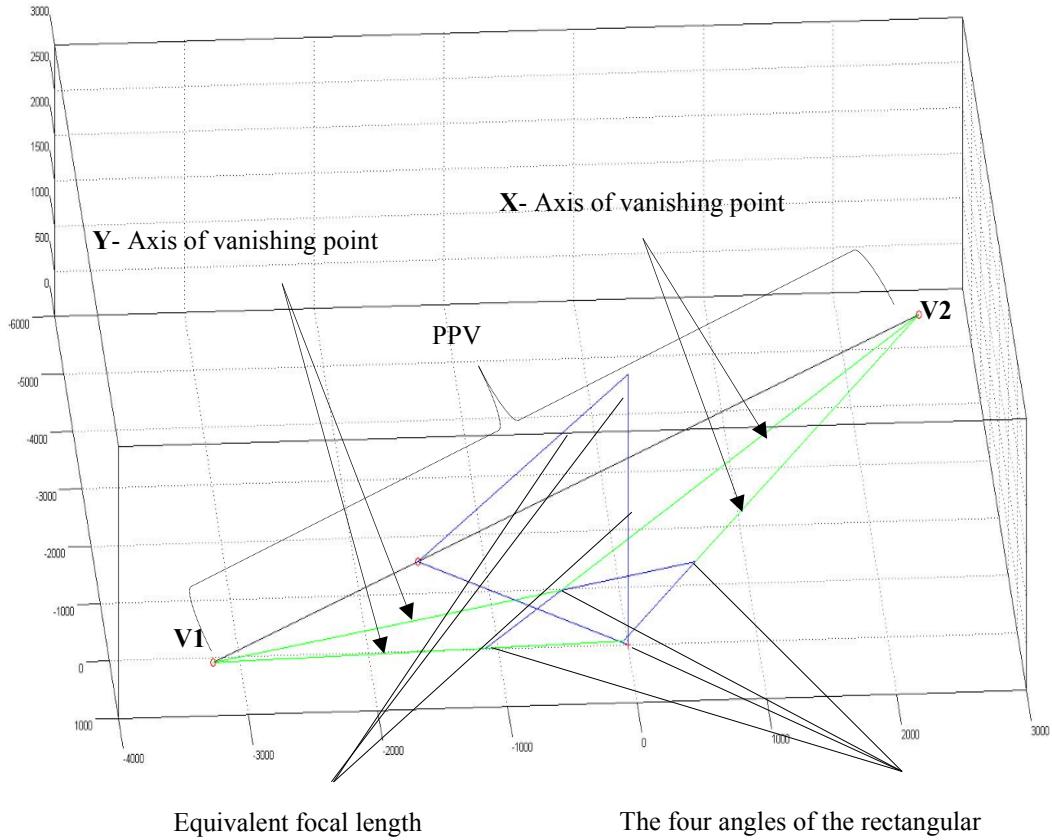


Figure 5.10: FCMT coordinator system.

5.3.2 Coordinate System of Camera Perspective

The mathematical modelling was conducted using Matlab. Figure 5.10 shows the coordinate system for the CMT system that is based on two vanishing points in space. The vanishing points come from the Y-axis to meet at the first vanishing point. From the side of the axis comes the second vanishing point from the X-axis. The connecting line between the two vanishing points is used to find the point closest to the centre of the image plane **PPv** from equation 5.1:

$$\mathbf{PPv} = \sqrt{(Ax)^2 + (Ay)^2} \quad (5.1)$$

Where \mathbf{PPv} is the distance from the centre of the image plane to the centre of the image plane projection. Then from \mathbf{PPv} and the two-point axes we can find the equivalent focal length of the face box. Accordingly, we can determine the same focal length for each person across pose images.

5.4 FACIAL DATABASE

Two standard facial databases, the ORL [223] and FERET [30, 224], were used to evaluate the effectiveness of the proposed FCMT method for face recognition across poses. Details of the facial databases and the pre-processing method used have been described in Sections 5.2 and 5.3. Experimental studies of performance in face identification were applied on the ORL database and with further progress with FERET. Only experiments with FERET images have been used for the cross pose test, these and experiments were carried out to show the efficiency and effectiveness of the proposed method. All of the experiments were software undertaken using an Intel(R) Core TM i5 CPU processor with 3.20GHz and 8.00 GB memory and MATLAB [225]. In the experiments, all of the training and experimental external data were fixed into same dimensional tensors via the learned subspaces when the dimensions of each image sample varied depending on which database has been set without any specific dimension. The ORL and FERET databases contain only 2-D data. All facial images from the two databases were automatically cropped and normalised to ensure that all images were of the same size.

For face tracking, the images were applied directly to 256x256 greyscale matrices for the 2-D methods by using FQD. There were some tests to illustrate the results for each database using a certain number of training samples per subject in each different experiment. All of the images in both databases have been utilised after training for testing and identification. The training and testing sets were used in identification mode to measure genuine matching scores. Therefore, there were three subsets for each database: for modulation (decomposition), training, and testing as one step and the last step of recognition (matching). The training and testing sets were used to evaluate the performance of the FCMT in identifying and recognising each image in the next step of the system.

5.5 RESULTS AND EXPERIMENTS

5.5.1 Real-time Face Recognition

The first experiment on face recognition used CMT for individual people in real time. Manual selection from the computer vision system was employed to determine the point-to-point distance between the two ears.

It has been observed that the accuracy of the technique may depend on the camera orientation. The deviation of inferred measurements from real dimensions depends on three points in ears because, as the camera angle changes, the perspective projection of the reference sheet onto the image plane changes and this affects the accuracy of the measurements taken. The following experiment measures the dimensions between ears for different people in ages, colour and gender by calibrating the camera on a piece of a note-paper as a reference.

The real dimension of all the shapes is shown in Table 5.2: *Results* from the images captured using a Samsung mobile Galaxy SIII (GT-i9300) camera that has a resolution of 8 MP. As shown in **Error! Reference source not found.**, the experiments conducted with images of males and female between 28 to 42 years old from different ethnicities.



Figure 5.11: The experiment of real-time face recognition for different subjects

From the four corners of the geometric reference shape the camera calibration and CMT were applied to calculate the dimensions in actual life based on the manual use of FQD points selected. In this case, the system used front face images taken with a real camera to obtain the dimensions of the faces and to determine the point distance between the ears.

Each ear has 3 points as shown in Figure 5.12:

- Point A: The distance between the helices of two ears (Top of the ear)
- Point B: The distance between the antihelices of the two ears (Middle of the ear)
- Point C: The distance between the lobe of the two ears (End of the ear)

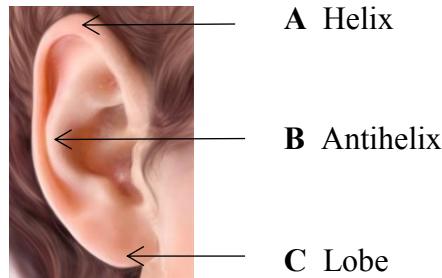


Figure 5.12: Names of ear points.

The measured distances between the three points across the face shows the difference between each point and the variants points to give the accurate recognition for faces. The results are described in detail in Figure 5.13 and Table 5.2.

Table 5.2: Results of the actual distances and CMT

Subject	The real distances			Distances from CMT		
	Point A	Point B	Point C	Point A	Point B	Point C
1	174	163	147	172	161	142
2	145	140	130	143	143	126
3	195	185	177	190	180	173
4	180	170	140	176	176	146

Figure 5.13, shows the difference between the estimated results and the true distances between the pairs of the three points shown in Table 5.2. The accuracy of the three variables points A, B, C, and the results show high levels of recognition.

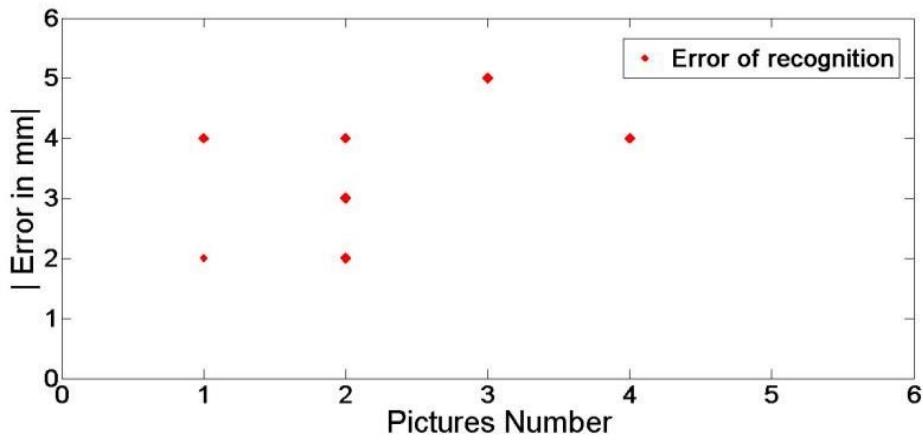


Figure 5.13: Accuracy of face recognition based on CMT

The dimensions inferred by the proposed method are accurate within 5mm. This approach shows that the different features of the human ear can be used for biometric recognition. The results of the distance between the three points on the ear (A, B, C) indicates that they are unique from one person to another. Whereas, point B was more prominent than the rest of the results of measured distances on the ear. Also, there is little difference in subject number 2 between the points A and B. In summary, the accuracy of face recognition is 95% and the images were captured from a portable camera. These results were from an early experiment that has been carried out in face recognition using CMT. The acceptable error of the proposed system is the maximum error that is 5mm.

5.5.2 Face Tracking

5.5.2.1 FQD with the ORL Database

The results of the experiments on quadtree image decomposition in FQD showed that the images reduce the error of tracking to half of that obtained by the results from the Viola-Jones algorithm. Tracking the face is the initial stage of the task recognition and is used as a mechanism to calibrate the camera. The testing of the quadtree decomposition algorithm with still images gives high performance for face tracking and the selection of landscape points on the face. Figure 5.14 shows images from the ORL [218] database for 40 persons, each with ten different images with different expressions, poses, or features. From the 400 face images the effect of the FQD on the face tracking for each person was compared.

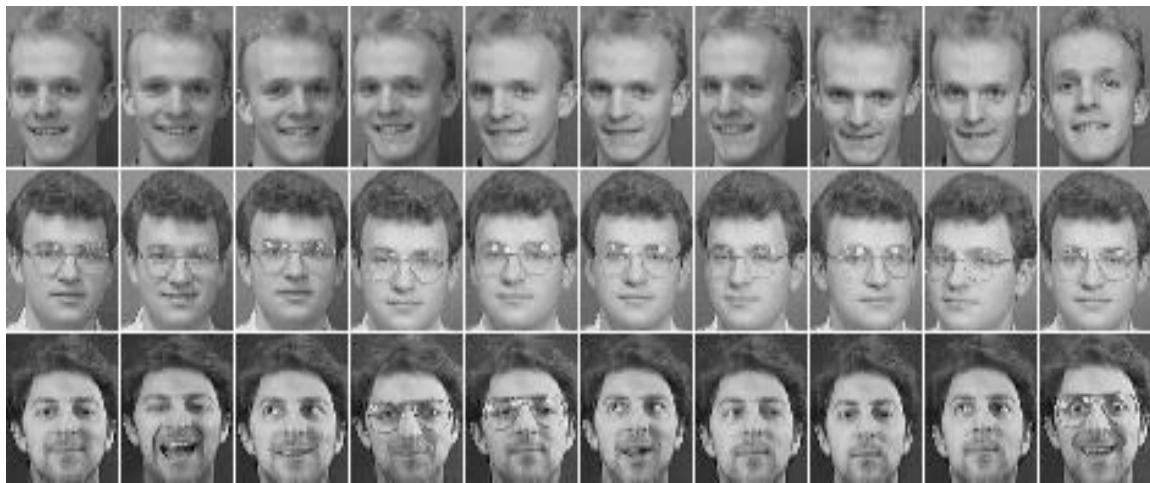


Figure 5.14: Sample of ORL database facial image.

The performance of face tracking from Figure 5.15 shows that it is the false face tracking by Viola-Jones algorithm showed a reduction of 60 pictures from 400.

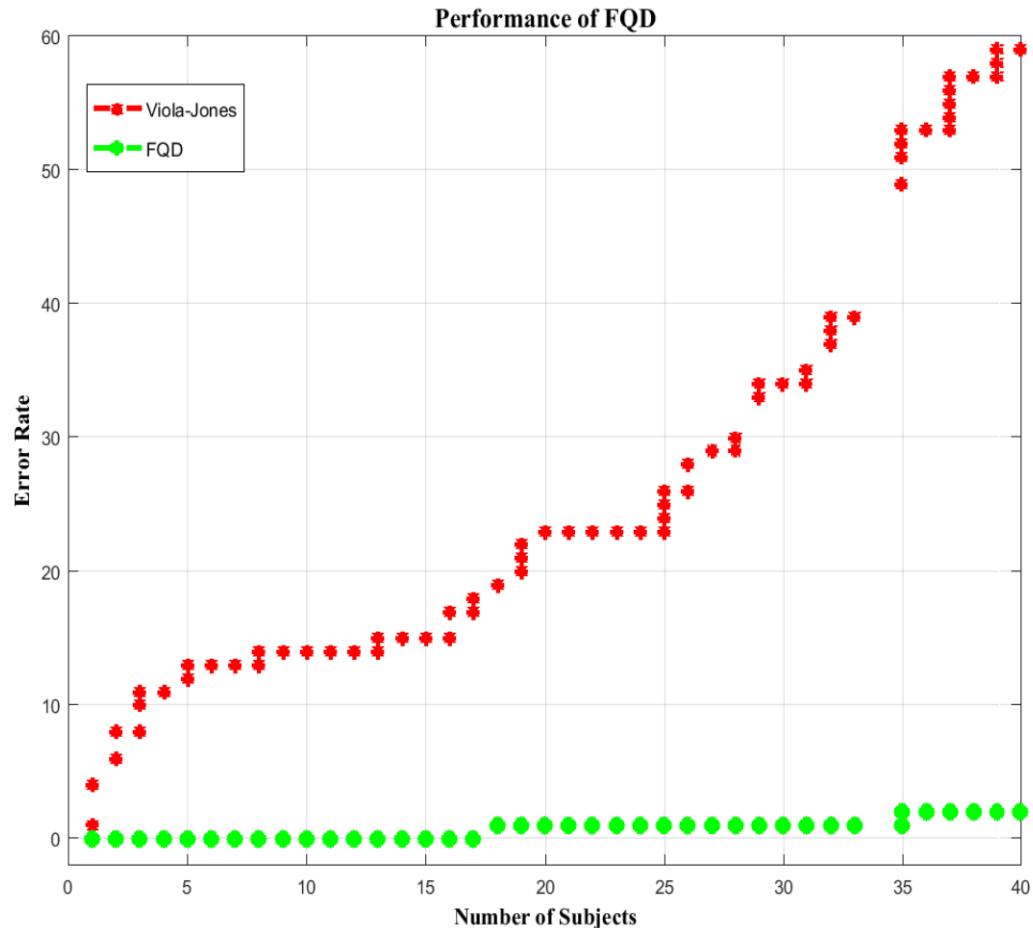


Figure 5.15: Performance in tracking faces using images from the ORL database for 40 people where for each person there are 10 different pictures.

While the proposed method of development FQD was reducing only two pictures from the entire database. To provide a more optimal scenario regarding computational efficiency, the images were fitted to 256x256 pixels. This experiment was achieved using 40 people, for both the training and test faces.

5.5.2.2 FQD with the FERET database

To make sure that the performance of FQD was unbiased for more than one database; FQD was tested with 100 different subjects from the FERET (Figure 5.16) database with each subject having nine images of the face across poses $\pm 60^\circ$.



Figure 5.16: Sample of FERET database facial images.

Therefore, as shown in Figure 5.17 the red plot shows the error of tracking faces for the Viola-Jones algorithm where there are more than 100 false recognitions from 900 images. Meanwhile, the green line represents the FQD results showing only one falsely recognised image from the 900 images. The advantage of the fusion of these algorithms is to give less error in tracking results and a more accurate selection of landscape points. The proposed work is based on the accuracy of the dimensions that infer the camera projection method.

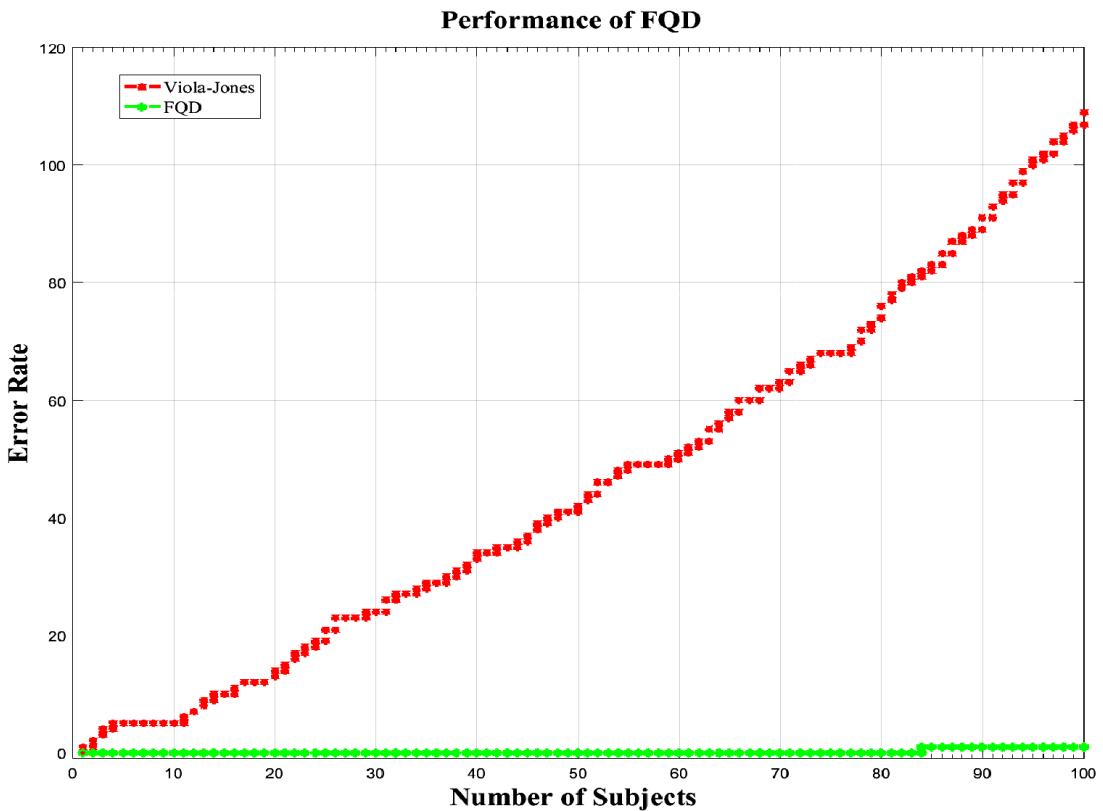


Figure 5.17: Performance in tracking face images from the FERET database, of 100 people where person has 9 pictures in different poses.

5.5.3 Face Recognition

The second achievement of the results of this step for face recognition comes after face tracking and interesting point selection on the face from FQD. The face recognition in the last step of the system presented in FCMT is applied to each image from the database to measure it, and find the unique distances of the features for each person.

5.5.3.1 Face Recognition Error Rate

The results of FCMT experiments were compared with those of Annan Li et al. [211], who used the same environment as our system without any effect of illumination or expressions. Also, Annan Li achieved best results compared with the rest of the methods such as Blanz and Vetter[116] and Ashraf et al. The algorithm of FCMT tested 100 subjects, out of the 200 subjects in the FERET database each of whom has nine images. The FERET pose database used the yaw angle of 0° , 60° , 40° , 25° , 15° , -15° , -25° , -40° , and -60° . Meanwhile, a randomly selected subject was used to test the FCMT algorithm and prove the results. Figure 5.18 shows the recognition error rates for each pose angle for comparison. The green curve shows the present results, while the red curve shows those of Annan Li et al.

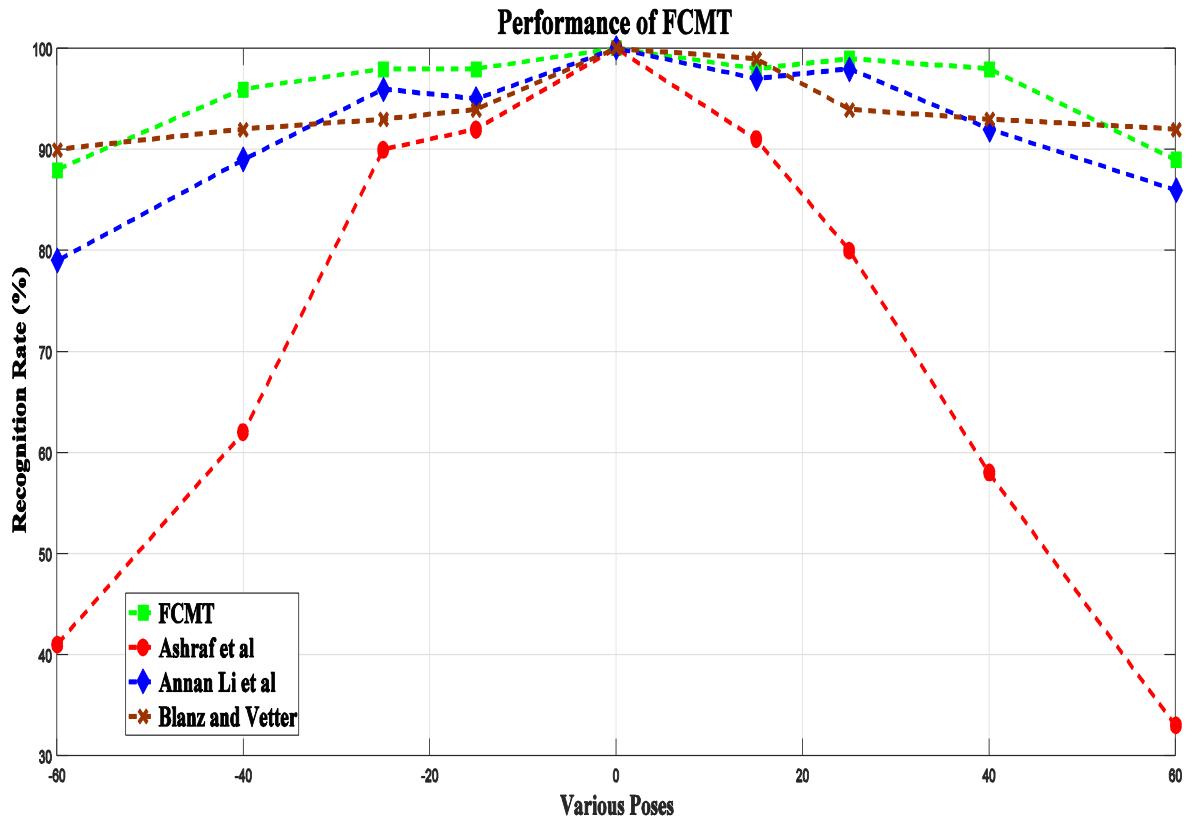


Figure 5.18: Performance of FCMT compared to typical cross-pose face recognition methods on the FERET database.

The recognition rate for the FCMT method gradually increases from an angle of -60° until peaking at an angle of 0° (front face). Annan Li's results begin at a lower recognition rate of 86% and peak at 100% recognition at the front face. From 0° angle to $+60^\circ$, the green curve of FCMT gradually decreases, while Annan Li's curve decreases and then rises again before decreasing sharply to a lower rate of recognition. The closest results to the proposed method is compared with Blanz and Vetter, but it lacks stability in both side of the angle, as we can see in $\pm 20^\circ$ angles in Table 5.3. In general, the proposed method gives results with a symmetrical curve indicating equal results for plus and minus angles of rotation, while the previous method gives very different results for rotation around the $\pm 60^\circ$ poses. Moreover, the average between the angles $\pm 60^\circ$ was calculated to explain the performance of our system as the mean average, and the results are shown at mean squared error (MSE) as following:

Table 5.3: The error percentage in recognition between $\pm 60^\circ$ angles

Methods	Average Recognition rate (%)				MSE
	$\pm 60^\circ$	$\pm 40^\circ$	$\pm 20^\circ$	$\pm 15^\circ$	
FCMT	88.5	97.5	98	98.5	95.62
Annan Li <i>et al.</i>	82.5	90.5	97.5	96.5	91.75
Ashraf <i>et al.</i>	37	60	85	91.5	68.37
Blanz and Vetter	91	91.5	93.5	96.5	93.12

The results shown as main error scores significantly outperforms the proposed FCMT in terms of accurate recognition at all angles. The proposed method gives more consistent performance for the same rotation with the two databases tested (FERET and ORL). Finally, the proposed system integrates the two CMT and FQD algorithms. The performance comparison analysis shows that the FCMT is a more efficient and robust method because it gives the highest performance in various poses. The reason for this may be that it is more robust to changes in general conditions with regard to the image capture procedure. It can be observed by comparing the physical theory of our system, as it measures the same distance for faces in different jaw angles that are considered fixed in real life.

5.5.3.2 Confusion Matrix of Recognition Rate Based on Pupils Distance

Having applied FCMT and comparing the results for face recognition with those of another method, the pose is assumed to be known from the previous experiment.

The algorithm tested 100 subjects out of 200 subjects in the FERET database each with nine images. Meanwhile, a randomly selected subject was used to test the algorithm. The gallery pose is the frontal of face pose that is compared with the rest of the poses. The rest of the pose is non-frontal that includes the poses around the gallery.

When, there are nine possible poses in total of the head cores pose in horizontal $\pm 60^\circ$. The experimental results of FCMT and feature poses representation are plotted into 9×9 confusion matrices as shown in Table 5.4. In this table, “EP” denotes the estimated pose poses, and “RP” denotes the real poses. In the experiments, the facial landmarks are also automatically labelled by FQD. As can be seen from Table 5.4, the recognition performance is also affected the difference between gallery and real pose. Each image from the database was tested with it and

then with the rest of the pictures based on the unique number of the measurement that was calculated by FCMT. The main range of the distance for the face features is measured by millimetres between 50 and 65 mm; this range is too short and might result in over similarity between the objects.

Therefore, the system used to measure by decimals numbers 1.000 between 0.50 extra for the similar images to make it unique for the entire object across pose and with the rest of the images. There will be a vast number of probabilities that can be used in a large database of measurements in millimetres. These numbers are divided into decimals for more accurate results for each pose. The approximate mean error percentage was calculated to show minor similarities in the database. The error shows little similarity between the pictures and the same error for both sides of the matrix that support the results, when the results were compared horizontally and vertically.

5.5.3.3 *Face Recognition rate per person across poses*

The confusion matrix tested the recognition results per each pose starting with the same pose and then with the rest of the pose. The results were taken before from the first measurement by CMT in the range between 50.0 and 65.0 mm before adding any decimal numbers (0.000) to reduce the similarity between the subjects.

Table 5.5 shows the results in the confusion matrix for each person across 9 poses at angles from -60° to 60° . When views are compared for different poses, it gives 100% recognition of the same reproducible measurement between the features. Comparing with the rest of the poses gives a slight difference in percentage error across pose. Each image was tested with itself and then was tested with the rest of the pose. The test algorithm based on yes or no, scoring 0 or 1 respectively. When it is not the same number, the number was marked as number one, and the same marked as zero.

Table 5.4: Confusion matrix of recognition rate and relative mean error for each vector, Horizontal (H) and Vertical (V).

RP EP \	-60	-40	-25	-15	0	5	25	40	60	Main Error (V)
-60	94.38	94.65	95.88	96.25	97.39	96.04	96.05	94.63	94.48	4.5
-40	94.54	92.68	94.08	94.77	94.59	94.42	93.82	93.92	95.68	5.5
-25	95.86	94.08	92.68	93.29	93.05	93.21	92.99	92.87	97.26	6.0
-15	96.23	94.77	93.29	92.08	92.73	93.21	92.6	93.81	98.04	6.0
0	96.37	94.67	93.05	92.73	92.14	92.79	92.75	93.42	97.72	6.0
15	96.01	93.91	93.12	93.3	92.83	91.58	92.42	93.88	97.97	6.0
25	96.02	94.42	92.99	92.6	92.67	92.42	92.06	93.74	97.91	6.0
40	94.44	93.92	92.71	93.71	93.42	93.88	93.74	91.34	95.9	6.0
60	94.4	95.86	97.31	97.95	97.85	97.97	97.91	95.89	93.66	3.5
Main Error (H)	4.5	5.5	6.0	5.0	6.0	6.0	6.0	6.0	3.5	

After all the images were measured, each pose calculated the percentage of the recognition rate and show it in Table 5.5.

Table 5.5: Face recognition rate per person across poses

EP \ RP	-60	-40	-25	-15	0	15	25	40	60
-60	100	92	91	95	74	91	96	93	95
-40	92	100	91	93	92	97	92	90	97
-25	91	91	100	85	87	92	87	91	98
-15	95	93	86	100	81	87	85	89	100
0	74	92	85	81	100	80	83	91	96
15	91	97	92	87	81	100	83	91	97
25	96	92	87	85	83	83	100	89	95
40	93	90	91	89	91	91	89	100	97
60	95	97	98	100	96	97	95	97	100

5.6 SUMMARY

In this chapter, a new solution is proposed for face recognition in various poses using geometrical approach for camera calibration. The novel method is based on the combination of CMT and FQD, and is called the FCMT.

First integration and enhancement were used to prepare the images to apply the measurement technique. The method used for tracking faces was enhanced and the features of the face across poses were automatically selected with a new algorithm entitled FQD. The results of experiment show that this method is more accurate and successful when using FQD as in section 5.2.3. As common face feature used for recognition tasks, the distance between the eyes was used to identify people across poses. Also, the results of projection measurements that were transformed from 2-D images showed a fixed distance for different poses. The second part of the proposed method presented in this chapter to complete the system is the FCMT technique. After images have been pre-processed with FQD, the CMT uses the output image to determine the internal camera parameters for camera calibration. Furthermore, the two landscape points of the eyes can be measured by CMT to give a unique identification for every person. The window face detection for each individual is used with different orientations. Experiments were applied to the real-time face in section 5.5.1, with other features of the face (ear to ear) the points selected manually and the camera setup to automatic. The system used captured images to do the experiment, and the experiment successfully met with the technical part of the system. The study is aware that the database should be reliable and have many subjects, but in this case, the system was tested with the real-time picture in order to support the next experiment.

The results of the experiments conducted show that the proposed approach significantly improves recognition performance when the testing face pose angles are wide. Therefore, the proposed system is more efficient and robust and showed the highest performance with images from the FERET database compared to a previously reported method. Therefore, the results indicate that the proposed method is close to that of the state-of-the-art methods.

Chapter 6 : CONCLUSION AND FUTURE WORK

This chapter summarises the research presented in the thesis. The contributions made of facial recognition and computer vision by the proposed system are briefly discussed and summarised. However, there are further questions to be addressed in the future work, and recommendations are made concerning potential research directions with the aim of achieving more efficient feature fusion techniques in different fields, such as building reconstruction and medical science.

1.1 CONCLUSION

Computer vision and face recognition have been researched for decades, and while humans can easily recognise sounds and images, no machine vision system has yet been designed that can operate accurately without actual measurement and perspective of viewing distance or image noise. This thesis presents research conducted to improve on pattern and face recognition using camera calibration. However, research to design more robust automatic facial recognition systems is challenging due to a significant appearance variability of human faces, insufficient data and the complexity of pattern distribution. This section briefly outlines the major findings achieved and contributions made in this thesis.

The work in this thesis has satisfied the aims of the research set out in Chapter 1. Chapter 2 represented a literature review of the main contributions that were mentioned in Chapter 1 to classify the camera calibration, including the main basics of geometric approaches. The main method used perspective projection, which was historically used in painting and more recently, in the use of image and video. Amongst the methods of face recognition, the thesis focuses on the main problem that is related to face recognition across pose. The proposed method also provides a solution to the one sample per person face problem. The decomposition image will be described at the end of this chapter as the final contribution, highlighted with previous studies and the solution that is provided in the field of the study.

The first contribution of this thesis described in Chapter 3 is a flexible, new, automatic technique that has been developed easily to calibrate a camera manually and automatically. The technique is based on the perspective projection of two vanishing points from the object in the 2-D image. The technique only requires a captured picture from any camera, then it can observe the information of the image. Rectangles, squares and triangles were used as flat objects to determine the camera's intrinsic and extrinsic parameters. The previous work carried out on camera calibration was described to explain the proposed work of camera calibration. The

experiments and results were tested and compared with different scenarios of a camera changing in position.

The internal parameters of the calibrations were then used to measure the dimensions of objects with regular and irregular shapes. In Chapter 4 a novel technique was developed to infer accurate object dimensions from a 2-D image. The previous information of the camera calibration enables the measurement technique (CMT) to determine the actual 2.5D from the image. The system of calibration was tested for three, tilt, pan and swing angles of rotation before applying it in the experiments. The results showed high level of accuracy of the measurements with minimal error that did not affect the measurement purposes. The technique is considered to be flexible to measure free shapes, and regular geometrical shapes have been tested to suit the facial requirements in measurement. The technique allows the use of varying camera resolutions; and test results indicate that the algorithm works with low-level image quality of 2MP resolution and with good quality 8MP resolution, both providing excellent accuracy in cameras used inferring the real object dimensions. The 2MP, 3MP and 8MP cameras are available at low cost for any user. Furthermore, the algorithm uses three different geometrical shapes to measure the regular shapes and irregular shapes, which gives a wide opportunity to accommodate the largest possible number of formats measured surrounding us.

Chapter 5 is the main part of this thesis that links computer vision with biometrical facial recognition. Three approaches devided by the research were described in this chapter including face tracking, image decomposition and FCMT.

After applying the quadtree image decomposition to enhance the image, the FQD tracked the faces automatically and selected the feature face points. The main proposed method is achieved to complete the system in FCMT technique. The images were pre-processed with FQD and the output image from the CMT technique in Chapter 4 was used to harvest the internal parameters of camera calibration. Furthermore, the two landscapes of the eyes can be measured by CMT to give a unique identification for every face on different views across pose. The results show the recognition error percentage for every single angle for comparison. From -60° to 60° angle of face rotation, the main error recognition rate shows the significant improvement in our proposed method compared with different methods that used the same database and more complex methods. Moreover, the results take the confusion matrix of recognition rate and the estimated mean error for each vector, Horizontal (H) and Vertical (V). The error shows little similarity between the pictures and the same error for both sides of the matrix, which supports the results when compared horizontally and vertically.

In conclusion, this thesis has proposed a solution to the problem of pose variations in face recognition by using the novel method of camera calibration. As a current landscape face point was used to recognise the face, the work focused on the distance between the eyes compared with the face window that was detected from the face tracking method to identify people. The result of camera calibration for the image enables the projection of the picture plane into 2.5-D. The perspective projection allows the holistic dimensions of the face to be fixed in different orientations. Experiments show that this approach significantly improves recognition performance when the pose angles are large. The analysis of the proposed system demonstrates its higher effectiveness as well as robust performance with images from the FERET database, as well as the use of ORL to improve the enhancement in face tracking from FQD. The results indicate that the proposed method gives performance close to that of state-of-the-art methods.

6.2 FUTURE WORK

Based on the work presented in this thesis, there are several possible finder directions that can be investigated. Each proposed contributions has intrusting future work rather than the main proposed work, as the following:

- The new method of camera calibration used one view of image from mono camera which proved as an acceptable solution for the same purposes and low cost compared with expensive cameras such as stereo cameras. Moreover, camera calibration is an important part of computer vision and 3-D reconstruction. Therefore, the information extracted from various views helps in the reconstruction of 3-D shapes, the views can capture by camera or even from frame video can be used to reconstruct the depth as shown in Figure 6.1 . The focal length and other camera parameters are applied already in different algorithms in 3-D reconstruction systems. The purpose of the 3-D reconstruction have wide use and keep improving side by side with the development of the technology of portable camera. The most appropriate field of this technique will be reconstruction building that uses in architecture and objects reconstruction.

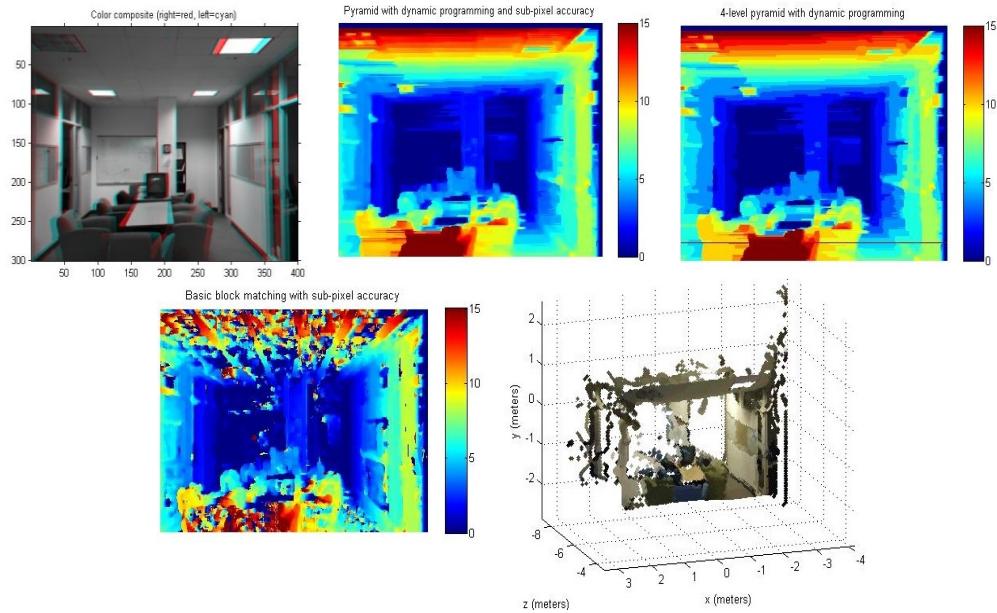


Figure 6.1: 3D reconstruction from 2 views image (right and left)

Meanwhile, face recognition across poses from the 3-D face recognition is an attractive approach in face recognition. Thus, it will be unique and more robust when using it with proposed method. Figure 6.2 shows experiment on face reconstruction.

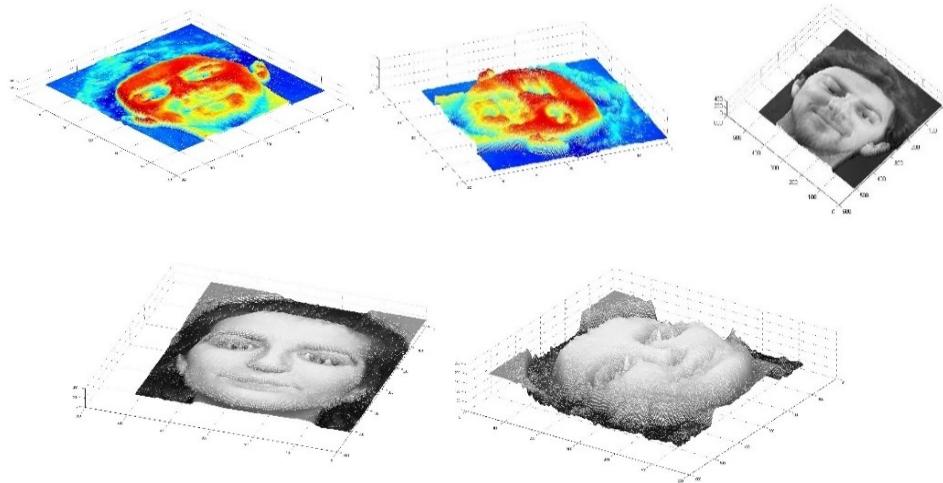


Figure 6.2: the 3D reconstruction of face recognition from one view using ORL database.

The second contribution of the proposed system which could be more interesting and unique in future work is applying the CMT technique on any captured image to determine the measurements which is giving exact dimension of real word.



(a)



(b)

Figure 6.3: MCT future work.

Figure 6.3 shows two pictures: a) civil engineer carrying out a survey to measure the distances, b) indoor CMT technique to measure the distances. The accuracy of the system was determined in chapter 4 under the effect of noise and low quality of the camera and was about 95% accurate. Moreover, CMT can be applied in a medical image as it has wide attention areas to measure the cancer cells and bones length from x-ray ray images.

- The proposed method of FCMT can be developed to minimise the error in accuracy and to provide a robust system that can be applied in different applications. The recommended future work for the proposed system is to use more than two facile features when the landmark points are increased.

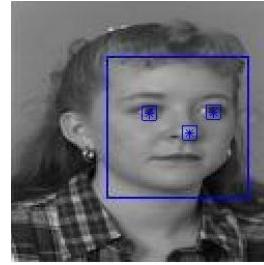


Figure 6.4: FQD applied for three points.

Some of the future work in face recognition has been started with early experiments. Hence, the FQD selected more than 6 points per face such as pupils, nose and mouth as shown in Figure 6.4. As well as, the free cross face rotation considers the horizontal cross pose face recognition, the future work in this enhancement was applied by FQD and FCMT as in Figure 6.5 and will be an open area for the researcher to improve it and use it for different purposes from head pose image database [226].



Figure 6.5: FQD free poses points detection.

The detection features of FQD technique were enhanced from the Viola-Jones algorithm and has a specific method that can deal with more points with further accuracy. The face recognition expression has been vastly improved recently, and it has the basics that can apply for this work including the landscape points, and the measurements for the expressions such as happiness or sadness, which can lead to a complete system in the future for face recognition. Finally, the first work carried out on face recognition FCMT that uses the ear-to-ear distance was unique as it used in the real camera measurement for faces. In the future, the algorithm can be applied to a real video camera

REFERENCE

- [1] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, pp. 4-20, 2004.
- [2] M. Burge and W. Burger, "Personal Identification in Networked Society," ed: Kluwer Academic Publishers, 2005.
- [3] M. Gobi and D. Kannan, "A Secured Public Key Cryptosystem for Biometric Encryption," *International Journal of Computer Science and Network Security (IJCSNS)*, vol. 15, p. 49-56, 2015.
- [4] R. Bhatia, "Biometrics and face recognition techniques," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, 2013.
- [5] M. Paulini, C. Rathgeb, A. Nautsch, H. Reichau, H. Reininger, and C. Busch, "Multi-Bit Allocation: Preparing Voice Biometrics for Template Protection," *Odyssey 2016*, pp. 291-296, 2016.
- [6] J. L. Wayman, "Fundamentals of biometric authentication technologies," *International Journal of Image and Graphics*, vol. 1, pp. 93-113, 2001.
- [7] S. Pankanti, R. M. Bolle, and A. Jain, "Biometrics: The future of identification [guest editors' introduction]," *Computer*, vol. 33, pp. 46-49, 2000.
- [8] C. A. Shoniregun and S. Crosier, *Securing biometrics applications*: Springer, 2008.
- [9] R. Chellappa, A. K. Roy-Chowdhury, and S. K. Zhou, "Recognition of humans and their activities using video," *Synthesis Lectures on Image, Video & Multimedia Processing*, vol. 1, pp. 1-173, 2005.
- [10] R. Newman, *Security and access control using biometric technologies*: Cengage Learning, 2009.
- [11] V. P. Kshirsagar, M. R. Baviskar, and M. E. Gaikwad, "Face recognition using Eigenfaces," in *Computer Research and Development (ICCRD), 2011 3rd International Conference on*, 2011, pp. 302-306.
- [12] H. Zhang, W. Liu, L. Dong, and Y. Wang, "Sparse eigenfaces analysis for recognition," in *2014 12th International Conference on Signal Processing (ICSP)*, 2014, pp. 887-890.
- [13] H. Moon and P. J. Phillips, "Computational and performance aspects of PCA-based face-recognition algorithms," *Perception*, vol. 30, pp. 303-321, 2001.
- [14] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 711-720, 1997.

- [15] T. S. Jebara, "3D pose estimation and normalization for face recognition," McGill University, 1995.
- [16] R. Jafri and H. R. Arabnia, "A survey of face recognition techniques," 2009.
- [17] Y. Gao and M. K. H. Leung, "Face recognition using line edge map," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 764-779, 2002.
- [18] Y. Gao and Y. Qi, "Robust visual similarity retrieval in single model face databases," *Pattern Recognition*, vol. 38, pp. 1009-1020, 2005.
- [19] J. Zou, Q. Ji, and G. Nagy, "A comparative study of local matching approach for face recognition," *IEEE Transactions on image processing*, vol. 16, pp. 2617-2628, 2007.
- [20] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 2037-2041, 2006.
- [21] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in *European conference on computer vision*, 2004, pp. 469-481.
- [22] X. Geng and Z.-H. Zhou, "Image region selection and ensemble for face recognition," *Journal of Computer Science and Technology*, vol. 21, pp. 116-125, 2006.
- [23] T.-K. Kim, H. Kim, W. Hwang, and J. Kittler, "Component-based LDA face description for image retrieval and MPEG-7 standardisation," *Image and Vision Computing*, vol. 23, pp. 631-642, 2005.
- [24] A. M. Martínez, "Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 748-763, 2002.
- [25] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on*, 1994, pp. 84-91.
- [26] K. Tan and S. Chen, "Adaptively weighted sub-pattern PCA for face recognition," *Neurocomputing*, vol. 64, pp. 505-511, 2005.
- [27] X. Zhang and Y. Gao, "Face recognition across pose: A review," *Pattern Recognition*, vol. 42, pp. 2876-2896, 2009.
- [28] C. Nastar and M. Mitschke, "Real-time face recognition using feature combination," in *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, 1998, pp. 312-317.
- [29] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1090-1104, 2000.

- [30] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," *Image and Vision Computing*, vol. 16, pp. 295-306, 1998.
- [31] R. Brunelli and T. Poggio, "Face recognition: Features versus templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, pp. 1042-1052, 1993.
- [32] P. F. De Carrera, "Face recognition algorithms," 2010.
- [33] J. Lu, Y.-P. Tan, and G. Wang, "Discriminative multimanifold analysis for face recognition from a single training sample per person," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, pp. 39-51, 2013.
- [34] T. Windeatt and G. Ardeshir, "Decision tree simplification for classifier ensembles," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 18, pp. 749-776, 2004.
- [35] C. Zhang, X. Liang, and T. Matsuyama, "Small sample size face recognition using random quad-tree based ensemble algorithm," 2013.
- [36] V. Amaral, G. A. Giraldi, and C. E. Thomaz, "A Statistical Quadtree Decomposition to Improve Face Analysis," 2016.
- [37] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, pp. 71-86, 1991.
- [38] R. Leggat, "A History of photography."1995.
- [39] A. Haro, M. Flickner, and I. Essa, "Detecting and tracking eyes by using their physiological properties, dynamics, and appearance," in *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, 2000, pp. 163-168.
- [40] R. Huber, H. Ramoser, K. Mayer, H. Penz, and M. Rubik, "Classification of coins using an eigenspace approach," *Pattern Recognition Letters*, vol. 26, pp. 61-75, 2005.
- [41] <https://www.axis.com/gb/en/>.
- [42] A. Del Bimbo, F. Dini, A. Grifoni, and F. Pernici, "Pan-tilt-zoom camera networks," *Multi-Camera Networks: Principles and Applications*, p. 189-211, 2009.
- [43] Z. Cui, A. Li, G. Feng, and K. Jiang, "Cooperative object tracking using dual-pan–tilt–zoom cameras based on planar ground assumption," *IET Computer Vision*, vol. 9, pp. 149-161.
- [44] U. Park, H.-C. Choi, A. K. Jain, and S.-W. Lee, "Face tracking and recognition at a distance: A coaxial and concentric PTZ camera system," *IEEE Transactions on Information Forensics and Security*, vol. 8, pp. 1665-1677, 2013.

- [45] M. Al Haj, A. D. Bagdanov, J. Gonzalez, and F. X. Roca, "Reactive object tracking with a single PTZ camera," in *Pattern Recognition (ICPR), 2010 20th International Conference on*, 2010, pp. 1690-1693.
- [46] J. C. Neves, J. C. Moreno, S. Barra, and H. Proen  a, "Acquiring high-resolution face images in outdoor environments: a master-slave calibration algorithm," in *Biometrics Theory, Applications and Systems (BTAS), 2015 IEEE 7th International Conference on*, 2015, pp. 1-8.
- [47] W. L. Woo and S. S. Dlay, "3D shape restoration using sparse representation and separation of illumination effects," *Signal Processing*, 2014.
- [48] M. Al Haj, C. Fern  andez, Z. Xiong, I. Huerta, J. Gonz  lez, and X. Roca, "Beyond the static camera: Issues and trends in active vision," in *Visual Analysis of Humans*, ed: Springer, 2011, pp. 11-30.
- [49] J. K. Suhr, H. G. Jung, G. Li, S.-I. Noh, and J. Kim, "Background compensation for pan-tilt-zoom cameras using 1-D feature matching and outlier rejection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, pp. 371-377, 2011.
- [50] S. W. Kim, K. Yun, K. M. Yi, S. J. Kim, and J. Y. Choi, "Detection of moving objects with a moving camera using non-panoramic background model," *Machine Vision and Applications*, vol. 24, pp. 1015-1028, 2013.
- [51] C. Micheloni and G. L. Foresti, "Real-time image processing for active monitoring of wide areas," *Journal of Visual Communication and Image Representation*, vol. 17, pp. 589-604, 2006.
- [52] J. Salvi, X. Armangu  , and J. Batlle, "A comparative review of camera calibrating methods with accuracy evaluation," *Pattern Recognition*, vol. 35, pp. 1617-1635, 2002.
- [53] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1330-1334, 2000.
- [54] R. Horaud, R. Mohr, and B. Lorecki, "Linear camera calibration," in *Robotics and Automation, 1992. Proceedings., 1992 IEEE International Conference on*, 1992, pp. 1539-1544.
- [55] R. K. Lenz and R. Y. Tsai, "Techniques for calibration of the scale factor and image center for high accuracy 3-D machine vision metrology," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, pp. 713-720, 1988.
- [56] M. A. Penna, "Camera calibration: A quick and easy way to determine the scale factor," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, pp. 1240-1245, 1991.

- [57] Y. Liu, T. S. Huang, and O. D. Faugeras, "Determination of camera location from 2-D to 3-D line and point correspondences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 28-37, 1990.
- [58] C. C. Wang, "Extrinsic calibration of a vision sensor mounted on a robot," *IEEE Transactions on Robotics and Automation*, vol. 8, pp. 161-175, 1992.
- [59] G.-Q. Wei and S. De Ma, "Implicit and explicit camera calibration: Theory and experiments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, pp. 469-480, 1994.
- [60] O. D. Faugeras and G. Toscani, "The calibration problem for stereo," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1986, pp. 15-20.
- [61] R. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," *IEEE Journal on Robotics and Automation*, vol. 3, pp. 323-344, 1987.
- [62] Z.-Q. Hong and J.-Y. Yang, "An algorithm for camera calibration using a three-dimensional reference point," *Pattern Recognition*, vol. 26, pp. 1655-1660, 1993.
- [63] D. Liebowitz, "Camera calibration and reconstruction of geometry from images," University of Oxford, 2001.
- [64] A. Criminisi, I. Reid, and A. Zisserman, "Single view metrology," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, 1999, pp. 434-441.
- [65] H. Smallwood, "Projective Geometry: Perspectives from Art and Mathematics," Fort Lewis College, 2009.
- [66] D. Raynaud, "Knowledge and Beliefs Regarding Linear Perspective," in *Studies on Binocular Vision*, ed: Springer, 2016, pp. 15-35.
- [67] J. A. Aiken, "The Perspective Construction of Masaccio's" Trinity" Fresco and Medieval Astronomical Graphics," *Artibus et historiae*, pp. 171-187, 1995.
- [68] C. Brauer-Burchardt and K. Voss, "Robust vanishing point determination in noisy images," in *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, pp. 559-562 vol.1. 2000.
- [69] E. Guillou, D. Meneveaux, E. Maisel, and K. Bouatouch, "Using vanishing points for camera calibration and coarse 3D reconstruction from a single image," *The Visual Computer*, vol. 16, pp. 396-410, 2000.
- [70] J. Bloomenthal and J. Rokne, "Homogeneous coordinates," *The Visual Computer*, vol. 11, pp. 15-26, 1994.

- [71] R. Mohr and B. Triggs, "Projective geometry for image analysis," in *XVIIth International Symposium on Photogrammetry & Remote Sensing (ISPRS'96)*, 1996.
- [72] K. D. Gremban, C. E. Thorpe, and T. Kanade, "Geometric camera calibration using systems of linear equations," in *Proceedings. 1988 IEEE International Conference on Robotics and Automation*, 1988, pp. 562-567 vol.1.
- [73] J. Heikkila and O. Silven, "A four-step camera calibration procedure with implicit image correction," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1997, pp. 1106-1112.
- [74] S. M. Seitz, "Image-based transformation of viewpoint and scene appearance," Citeseer, 1997.
- [75] R. M. Haralick, "Determining camera parameters from the perspective projection of a rectangle," *Pattern Recognition*, vol. 22, pp. 225-230, 1989.
- [76] T. Schenk, "Introduction to photogrammetry."
- [77] J. J. Koenderink and A. J. Van Doorn, "Affine structure from motion," *JOSA A*, vol. 8, pp. 377-385, 1991.
- [78] K. H. Strobl, W. Sepp, and G. Hirzinger, "On the issue of camera calibration with narrow angular field of view," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2009, pp. 309-315.
- [79] K. H. Strobl and G. Hirzinger, "More accurate camera and hand-eye calibrations with unknown grid pattern dimensions," in *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, 2008, pp. 1398-1405.
- [80] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, 1999, pp. 666-673.
- [81] Z. Zhengyou, "Camera calibration with one-dimensional objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 892-899, 2004.
- [82] P. F. Sturm and S. J. Maybank, "On plane-based camera calibration: A general algorithm, singularities, applications," in *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, 1999.
- [83] Z. Zhang, "Microsoft Kinect Sensor and Its Effect," *IEEE MultiMedia*, vol. 19, pp. 4-10, 2012.
- [84] U. Park, "Face Recognition: face in video, age invariance, and facial marks," Michigan State University, 2009.

- [85] G. Medioni and R. Waupotitsch, "Face modeling and recognition in 3-D," in *Analysis and Modeling of Faces and Gestures, 2003. AMFG 2003. IEEE International Workshop on*, 2003, pp. 232-233.
- [86] G. J. Edwards, T. F. Cootes, and C. J. Taylor, "Face recognition using active appearance models," in *Computer Vision—ECCV'98*, ed: Springer, 1998, pp. 581-595.
- [87] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, pp. 1063-1074, 2003.
- [88] L. Zhang and D. Samaras, "Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, pp. 351-363, 2006.
- [89] U. Fayyad and R. Uthurusamy, "Evolving data into mining solutions for insights," *Communications of the ACM*, vol. 45, pp. 28-31, 2002.
- [90] A. Jain and D. Zongker, "Feature selection: Evaluation, application, and small sample performance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 153-158, 1997.
- [91] A. A. M. Al-Shiha, "Biometric face recognition using multilinear projection and artificial intelligence," *PhD diss., University of Newcastle Upon Tyne*, 2013.
- [92] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino, "2D and 3D face recognition: A survey," *Pattern Recognition Letters*, vol. 28, pp. 1885-1906, 2007.
- [93] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang, "Face recognition from a single image per person: A survey," *Pattern recognition*, vol. 39, pp. 1725-1745, 2006.
- [94] D. J. Beymer, "Face recognition under varying pose," in *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on*, 1994, pp. 756-761.
- [95] R. Singh, M. Vatsa, A. Ross, and A. Noore, "A mosaicing scheme for pose-invariant face recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 37, pp. 1212-1225, 2007.
- [96] D. Beymer and T. Poggio, "Face recognition from one example view," in *Computer Vision, 1995. Proceedings., Fifth International Conference on*, 1995, pp. 500-507.
- [97] D. González-Jiménez and J. L. Alba-Castro, "Toward pose-invariant 2-d face recognition through point distribution models and facial symmetry," *IEEE Transactions on Information Forensics and Security*, vol. 2, pp. 413-429, 2007.

- [98] F. Kahraman, B. Kurt, and M. Gokmen, "Robust face alignment for illumination and pose invariant face recognition," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1-7.
- [99] T. F. Cootes, G. V. Wheeler, K. N. Walker, and C. J. Taylor, "View-based active appearance models," *Image and vision computing*, vol. 20, pp. 657-664, 2002.
- [100] I. A. Kakadiaris, G. Passalis, G. Toderici, M. N. Murtuza, Y. Lu, N. Karampatziakis, *et al.*, "Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 640-649, 2007.
- [101] R. Gross, I. Matthews, and S. Baker, "Appearance-based face recognition and light-fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 449-465, 2004.
- [102] C. Liu, "Gabor-based kernel PCA with fractional power polynomial models for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 572-581, 2004.
- [103] J. Huang, P. C. Yuen, W.-S. Chen, and J. H. Lai, "Choosing parameters of kernel subspace LDA for recognition of face images under pose and illumination variations," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 37, pp. 847-862, 2007.
- [104] J. Yang, A. F. Frangi, J.-y. Yang, D. Zhang, and Z. Jin, "KPCA plus LDA: a complete kernel Fisher discriminant framework for feature extraction and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 230-244, 2005.
- [105] T. K. Kim and J. Kittler, "Design and fusion of pose-invariant face-identification experts," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, pp. 1096-1106, 2006.
- [106] M. D. Levine and Y. Yu, "Face recognition subject to variations in facial expression, illumination and pose using correlation filters," *Computer Vision and Image Understanding*, vol. 104, pp. 1-15, 2006.
- [107] X. Chai, S. Shan, X. Chen, and W. Gao, "Locally linear regression for pose-invariant face recognition," *IEEE Transactions on Image Processing*, vol. 16, pp. 1716-1725, 2007.
- [108] S. J. D. Prince, J. H. Elder, J. Warrell, and F. M. Felisberti, "Tied factor analysis for face recognition across large pose differences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 970-984, 2008.

- [109] Y. Gao, M. K. H. Leung, W. Wang, and S. C. Hui, "Fast face identification under varying pose from a single 2-D model view," *IEE Proceedings-Vision, Image and Signal Processing*, vol. 148, pp. 248-253, 2001.
- [110] X. Liu and T. Chen, "Pose-robust face recognition using geometry assisted probabilistic modeling," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005, pp. 502-509.
- [111] X. Zhang, Y. Gao, and M. K. H. Leung, "Automatic texture synthesis for face recognition from single views," in *18th International Conference on Pattern Recognition (ICPR'06)*, 2006, pp. 1151-1154.
- [112] M. W. Lee and S. Ranganath, "Pose-invariant face recognition using a 3D deformable model," *Pattern recognition*, vol. 36, pp. 1835-1846, 2003.
- [113] D. Jiang, Y. Hu, S. Yan, L. Zhang, H. Zhang, and W. Gao, "Efficient 3D reconstruction for face recognition," *Pattern Recognition*, vol. 38, pp. 787-798, 2005.
- [114] X. Zhang, Y. Gao, and M. K. H. Leung, "Recognizing rotated faces from frontal and side views: An approach toward effective use of mugshot databases," *IEEE Transactions on Information Forensics and Security*, vol. 3, pp. 684-697, 2008.
- [115] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," in *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, 1999, pp. 187-194.
- [116] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 1063-1074, 2003.
- [117] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Generative models for recognition under variable pose and illumination," in *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, 2000, pp. 277-284.
- [118] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 643-660, 2001.
- [119] C. D. Castillo and D. W. Jacobs, "Using stereo matching for 2-d face recognition across pose," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1-8.
- [120] I. Kotsia and I. Pitas, "Facial expression recognition in image sequences using geometric deformation features and support vector machines," *IEEE Transactions on Image Processing*, vol. 16, pp. 172-187, 2007.

- [121] Y. Sun and L. Yin, "Facial expression recognition based on 3D dynamic range model sequences," in *European Conference on Computer Vision*, 2008, pp. 58-71.
- [122] L. A. Jeni, A. Lörincz, T. Nagy, Z. Palotai, J. Sebők, Z. Szabó, *et al.*, "3D shape estimation in video sequences provides high precision evaluation of facial expressions," *Image and Vision Computing*, vol. 30, pp. 785-795, 2012.
- [123] O. Rudovic, M. Pantic, and I. Patras, "Coupled Gaussian processes for pose-invariant facial expression recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, pp. 1357-1369, 2013.
- [124] Y. Zhang and Q. Ji, "Active and dynamic information fusion for facial expression understanding from image sequences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 699-714, 2005.
- [125] R. El Kalouby and P. Robinson, "Real-time inference of complex mental states from facial expressions and head gestures," in *Real-time vision for human-computer interaction*, ed: Springer, 2005, pp. 181-200.
- [126] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," *Josa A*, vol. 4, pp. 519-524, 1987.
- [127] S. Chen, J. Liu, and Z.-H. Zhou, "Making FLDA applicable to face recognition with one sample per person," *Pattern recognition*, vol. 37, pp. 1553-1555, 2004.
- [128] F. Hafiz, A. A. Shafie, and Y. M. Mustafah, "Face recognition from single sample per person by learning of generic discriminant vectors," *Procedia Engineering*, vol. 41, pp. 465-472, 2012.
- [129] M. Yang, L. Van Gool, and L. Zhang, "Sparse variation dictionary learning for face recognition with a single training sample per person," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 689-696.
- [130] W. Deng, J. Hu, and J. Guo, "Extended SRC: Undersampled face recognition via intraclass variant dictionary," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, pp. 1864-1870, 2012.
- [131] H.-K. Ji, Q.-S. Sun, Z.-X. Ji, Y.-H. Yuan, and G.-Q. Zhang, "Collaborative probabilistic labels for face recognition from single sample per person," *Pattern Recognition*, vol. 62, pp. 125-134, 2017.
- [132] C. Zhang, X. Liang, and T. Matsuyama, "Multi-subregion face recognition using coarse-to-fine Quad-tree decomposition," in *Pattern Recognition (ICPR), 2012 21st International Conference on*, 2012, pp. 1004-1007.

- [133] A. Z. Kouzani, F. He, and K. Sammut, "Wavelet packet face representation and recognition," in *1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation*, 1997, pp. 1614-1619 vol.2.
- [134] C. Garcia, G. Zikos, and G. Tziritas, "A wavelet-based framework for face recognition," in *Int. Workshop on Advances in Facial Image Anal. Recognition Technology, 5th European Conf. Computer Vision*, 1998.
- [135] L. Ding, Y. Yan, Q. Xue, and G. Jin, "Wavelet packet compression for volume holographic image recognition," *Optics communications*, vol. 216, pp. 105-113, 2003.
- [136] J. Kannala and S. S. Brandt, "A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 1335-1340, 2006.
- [137] K. Nakano, M. Okutomi, and Y. Hasegawa, "Camera calibration with precise extraction of feature points using projective transformation," in *Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No.02CH37292)*, 2002, pp. 2532-2538.
- [138] J. Heikkila and O. Silven, "A four-step camera calibration procedure with implicit image correction," in *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, 1997, pp. 1106-1112.
- [139] C. Shanshan, Z. Wuheng, and Z. Lijun, "Camera calibration via stereo vision using Tsai's method," in *2009 First International Workshop on Education Technology and Computer Science*.
- [140] B. Li, L. Heng, K. Koser, and M. Pollefeys, "A multiple-camera system calibration toolbox using a feature descriptor-based calibration pattern," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013, pp. 1301-1307.
- [141] Z. Zhang, *Camera calibration*: Springer, 2014.
- [142] J. Heikkila and O. Silvén, "A four-step camera calibration procedure with implicit image correction," in *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, 1997, pp. 1106-1112.
- [143] S. Sreekaladevi, W. Wai Lok, S. Dlay, P. Gore, G. Johnson, and D. Swailes, "Inferring 3D dimensions of flat objects from a single 2D image," in *Intelligent Computer Communication and Processing (ICCP), 2013 IEEE International Conference on*, 2013, pp. 123-126.
- [144] Zhang, Ruo, Ping-Sing Tsai, James Edwin Cryer, and Mubarak Shah. "Shape-from-shading: a survey." *IEEE transactions on pattern analysis and machine intelligence* 21, no. 8 (1999): 690-706.

- [145] P. Sturm, "On focal length calibration from two views," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, 2001, pp. II-145-II-150 vol.2.
- [146] P. Sturm, "On focal length calibration from two views," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, 2001, pp. II-145-II-150 vol. 2.
- [147] R. Gherardi and A. Fusiello, "Practical autocalibration," in *European Conference on Computer Vision*, 2010, pp. 790-801.
- [148] N. Alvertos, "A zoom-based stereo camera model," in *Southeastcon'89. Proceedings. Energy and Information Technologies in the Southeast.*, IEEE, 1989, pp. 482-485.
- [149] E. Krotkov, K. Henriksen, and R. Kories, "Stereo ranging with verging cameras," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 1200-1205, 1990.
- [150] M. S. S. Kumar and N. Avinash, "Stereo Image Rectification Using Focal Length Adjustment," in *Signal and Image Processing (ICSIP), 2014 Fifth International Conference on*, 2014, pp. 241-249.
- [151] J. Denzler, M. Zobel, and H. Niemann, "Information theoretic focal length selection for real-time active 3D object tracking," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, 2003, pp. 400-407 vol.1.
- [152] Jin, Hailin. "Initialization for robust video-based structure from motion." U.S. Patent 8,934,677, issued January 13, 2015.
- [153] Z. Zhou, H. Jin, and Y. Ma, "Robust plane-based structure from motion," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012, pp. 1482-1489.
- [154] Moeslund, Thomas B., Adrian Hilton, and Volker Krüger. "A survey of advances in vision-based human motion capture and analysis." *Computer vision and image understanding* 104, no. 2 (2006): 90-126.
- [155] R. Orghidan, J. Salvi, M. Gordan, and B. Orza, "Camera calibration using two or three vanishing points," in *Computer Science and Information Systems (FedCSIS), 2012 Federated Conference on*, 2012, pp. 123-130.
- [156] H. Kogan, R. Maurer, and R. Keshet, "Vanishing points estimation by self-similarity," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 755-761.

- [157] M. Horn, x00E, x010D, ek, and S. Maierhofer, "Extracting vanishing points across multiple views," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, 2011, pp. 953-960.
- [158] A. Criminisi, I. Reid, and A. Zisserman, "Single view metrology," *International Journal of Computer Vision*, vol. 40, pp. 123-148, 2000.
- [159] L. Quan and R. Mohr, "Matching perspective images using geometric constraints and perceptual grouping," in *2nd International Conference on Computer Vision (ICCV'88)*, 1988, pp. 679--683.
- [160] R. Cipolla, T. Drummond, and D. P. Robertson, "Camera Calibration from Vanishing Points in Image of Architectural Scenes."
- [161] J. C. H. Leung and G. F. McLean, "Vanishing point matching," in *Image Processing, 1996. Proceedings., International Conference on*, 1996, pp. 305-308 vol.2.
- [162] D. G. Stork, "Computer vision, image analysis, and master art: part 1," *IEEE MultiMedia*, vol. 13, pp. 16-20, 2006.
- [163] Z. Zhang, "Camera calibration with one-dimensional objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 892-899, 2004.
- [164] Z. Zhang, "Camera calibration with one-dimensional objects," in *European Conference on Computer Vision*, 2002, pp. 161-174.
- [165] X. Cao and H. Foroosh, "Camera calibration without metric information using 1D objects," in *Image Processing, 2004. ICIP'04. 2004 International Conference on*, 2004, pp. 1349-1352.
- [166] O. Faugeras, Q. Long, and P. Sturm, "Self-calibration of a 1D projective camera and its application to the self-calibration of a 2D projective camera," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1179-1185, 2000.
- [167] V. Chari and A. Veeraraghavan, "Lens Distortion, Radial Distortion," in *Computer Vision*, ed: Springer, 2014, pp. 443-445.
- [168] L. Grammatikopoulos, G. Karras, and E. Petsa, "An automatic approach for camera calibration from vanishing points," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 62, pp. 64-76, 2007.
- [169] D. H. Lee, K. H. Jang, and S. K. Jung, "Intrinsic Camera Calibration Based on Radical Center Estimation," in *CISST*, 2004, pp. 7-13.
- [170] R. M. Haralick, "Using perspective transformations in scene analysis," *Computer Graphics and Image Processing*, vol. 13, pp. 191-221, 1980.
- [171] S. Sreekaladevi, W. L. Woo, S. Dlay, P. Gore, G. Johnson, and D. Swailes, "Inferring 3D dimensions of flat objects from a single 2D image," in *Intelligent Computer*

Communication and Processing (ICCP), 2013 IEEE International Conference on, 2013, pp. 123-126.

- [172] D. Bequé, C. Vanhove, A. Andreyev, J. Nuyts, and M. Defrise, "Correction for imperfect camera motion and resolution recovery in pinhole SPECT," in *Nuclear Science Symposium Conference Record, 2004 IEEE*, 2004, pp. 2507-2510.
- [173] Y. Shih, B. Guenter, and N. Joshi, "Image enhancement using calibrated lens simulations," in *European Conference on Computer Vision*, 2012, pp. 42-56.
- [174] G. Berkovic and E. Shafir, "Optical methods for distance and displacement measurements," *Advances in Optics and Photonics*, vol. 4, pp. 441-471, 2012.
- [175] A. Flores, E. Christiansen, D. Kriegman, and S. Belongie, "Camera distance from face images," in *International Symposium on Visual Computing*, 2013, pp. 513-522.
- [176] D. Baumgardner, J. L. Brenguier, A. Bucholtz, H. Coe, P. DeMott, T. J. Garrett, *et al.*, "Airborne instruments to measure atmospheric aerosol particles, clouds and radiation: A cook's tour of mature and emerging technology," *Atmospheric Research*, vol. 102, pp. 10-29, 2011.
- [177] S. Hagstrom and D. Messinger, "Line-of-sight measurement in large urban areas using voxelized lidar," in *SPIE Defense, Security, and Sensing*, 2012, pp. 837907-837907-12.
- [178] Hagstrom, Shea T. "Voxel-Based LIDAR Analysis and Applications." PhD diss., ROCHESTER INSTITUTE OF TECHNOLOGY, 2014.
- [179] U. Pyysalo, J. Oksanen, and T. Sarjakoski, "Viewshed analysis and visualization of landscape voxel models," in *24th International Cartographic Conference, Santiago, Chile*, 2009.
- [180] T. Weyrich, H. Pfister, and M. Gross, "Rendering deformable surface reflectance fields," *IEEE Transactions on Visualization and Computer Graphics*, vol. 11, pp. 48-58, 2005.
- [181] M. Holroyd, J. Lawrence, G. Humphreys, and T. Zickler, "A photometric approach for estimating normals and tangents," in *ACM Transactions on Graphics (TOG)*, 2008, p. 133.
- [182] A. Tai, J. Kittler, M. Petrou, and T. Windeatt, "Vanishing point detection," in *BMVC92*, ed: Springer, 1992, pp. 109-118.
- [183] T. B. van Driel, S. Herrmann, G. Carini, M. M. Nielsen, and H. T. Lemke, "Correction of complex nonlinear signal response from a pixel array detector," *Journal of Synchrotron Radiation*, vol. 22, pp. 584-591, 2015.

- [184] E. D. Paul and M. Jitendra, "Recovering high dynamic range radiance maps from photographs," presented at the ACM SIGGRAPH 2008 classes, Los Angeles, California, 2008.
- [185] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *ACM SIGGRAPH 2008 classes*, pp. 31-41, 2008.
- [186] T. Mitsunaga and S. K. Nayar, "Radiometric self calibration," in *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, 1999.
- [187] R. Sundareswara and P. R. Schrater, "Bayesian modelling of camera calibration and reconstruction," in *3-D Digital Imaging and Modeling, 2005. 3DIM 2005. Fifth International Conference on*, 2005, pp. 394-401.
- [188] R. Y. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," *Robotics and Automation, IEEE Journal of*, vol. 3, pp. 323-344, 1987.
- [189] Andaló, Fernanda A., Gabriel Taubin, and Siome Goldenstein. "Detecting vanishing points by segment clustering on the projective plane for single-view photogrammetry." In *Information Forensics and Security (WIFS), 2010 IEEE International Workshop on*, pp. 1-6. IEEE, 2010.
- [190] K. Y. K. Wong, P. R. S. Mendonca, and R. Cipolla, "Camera calibration from surfaces of revolution," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, pp. 147-161, 2003.
- [191] L. Seung-Hyun, L. Jae-Hong, and K. Min Young, "Dual camera based wide-view imaging system and real-time image registration algorithm," in *Control, Automation and Systems (ICCAS), 2011 11th International Conference on*, 2011, pp. 1766-1770.
- [192] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, pp. 381-395, 1981.
- [193] E. Murphy-Chutorian and M. M. Trivedi, "Head pose estimation in computer vision: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 607-626, 2009.
- [194] Zhang, Haijun, QM Jonathan Wu, Tommy WS Chow, and Mingbo Zhao. "A two-dimensional Neighborhood Preserving Projection for appearance-based face recognition." *Pattern Recognition* 45, no. 5 (2012): 1866-1876.
- [195] G. Fanelli, J. Gall, and L. Van Gool, "Real time head pose estimation with random regression forests," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, 2011, pp. 617-624.

- [196] M. M. Fakhir, W. L. Woo, and S. S. Dlay, "Inferring dimensions from an image using automatic calibration," in *Communication Systems, Networks & Digital Signal Processing (CSNDSP), 2014 9th International Symposium on*, 2014, pp. 776-780.
- [197] M. I. Mandasari, M. Günther, R. Wallace, R. Saeidi, S. Marcel, and D. A. van Leeuwen, "Score calibration in face recognition," *IET Biometrics*, vol. 3, pp. 246-256, 2014.
- [198] H.-C. Choi, U. Park, and A. K. Jain, "Ptz camera assisted face acquisition, tracking & recognition," in *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, 2010, pp. 1-6.
- [199] W. Jin, G. Geng, K. Li, and Y. Han, "Parameter estimation for perspective projection based on camera calibration in skull-face overlay," in *Virtual Reality and Visualization (ICVRV), 2013 International Conference on*, 2013, pp. 317-320.
- [200] D. S. Pierce, T. S. Ng, and B. R. Morrison, "A novel laser triangulation technique for high precision distance measurement," in *Industry Applications Society Annual Meeting, 1992., Conference Record of the 1992 IEEE*, 1992, pp. 1762-1769.
- [201] M. M. Fakhir, W. L. Woo, and S. S. Dlay, "Face Recognition Based on Features Measurement Technique," in *Modelling Symposium (EMS), 2014 European*, 2014, pp. 158-162.
- [202] A. B. Ashraf, S. Lucey, and T. Chen, "Learning patch correspondences for improved viewpoint invariant face recognition," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, 2008, pp. 1-8.
- [203] A. Li, S. Shan, and W. Gao, "Coupled bias-variance tradeoff for cross-pose face recognition," *IEEE Transactions on Image Processing*, vol. 21, pp. 305-315, 2012.
- [204] A. M. Bronstein, M. M. Bronstein, R. Kimmel, and A. Spira, "3D face recognition without facial surface reconstruction," *Technion-Computer Science Department Technical Report*, 2003.
- [205] V. D. Kaushik, A. Budhwar, A. Dubey, R. Agrawal, S. Gupta, V. K. Pathak, *et al.*, "An efficient 3D face recognition algorithm," in *New Technologies, Mobility and Security (NTMS), 2009 3rd International Conference on*, 2009, pp. 1-5.
- [206] J. Chen, Z. Chen, Z. Chi, and H. Fu, "Facial Expression Recognition in Video with Multiple Feature Fusion," *IEEE Transactions on Affective Computing*, vol. PP, pp. 1-1, 2016.
- [207] D. Ghimire and J. Lee, "Geometric feature-based facial expression recognition in image sequences using multi-class adaboost and support vector machines," *Sensors*, vol. 13, pp. 7714-7734, 2013.

- [208] G. Gao, K. Jia, and B. Jiang, "An Automatic Geometric Features Extracting Approach for Facial Expression Recognition Based on Corner Detection," in *2015 International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*, 2015, pp. 302-305.
- [209] Durmuşoğlu, Alptekin, and Yavuz Kahraman. "Facial expression recognition using geometric features." In *Systems, Signals and Image Processing (IWSSIP), 2016 International Conference on*, pp. 1-5. IEEE, 2016.
- [210] A. Gee and R. Cipolla, "Determining the gaze of faces in images," *Image and Vision Computing*, vol. 12, pp. 639-647, 1994.
- [211] A. Li, S. Shan, and W. Gao, "Coupled bias-variance tradeoff for cross-pose face recognition," *Image Processing, IEEE Transactions on*, vol. 21, pp. 305-315, 2012.
- [212] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, 2001, pp. I-511-I-518 vol. 1.
- [213] R. Lienhart and J. Maydt, "An extended set of haar-like features for rapid object detection," in *Image Processing. 2002. Proceedings. 2002 International Conference on*, 2002, pp. I-900-I-903 vol. 1.
- [214] Y.-Q. Wang, "An Analysis of the Viola-Jones face detection algorithm," *Image Processing On Line*, vol. 4, pp. 128-148, 2014.
- [215] E. Shusterman and M. Feder, "Image compression via improved quadtree decomposition algorithms," *Image Processing, IEEE Transactions on*, vol. 3, pp. 207-215, 1994.
- [216] L. N. Trefethen, *Spectral methods in MATLAB*: SIAM, 2000.
- [217] W. Gautschi, "Orthogonal polynomials, quadrature, and approximation: computational methods and software (in matlab)," in *Orthogonal Polynomials and Special Functions*, ed: Springer, 2006, pp. 1-77.
- [218] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *Applications of Computer Vision, 1994., Proceedings of the Second IEEE Workshop on*, 1994, pp. 138-142.
- [219] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, 2001, pp. I-I.
- [220] N. A. Dodgson, "Variation and extrema of human interpupillary distance," in *Electronic imaging 2004*, 2004, pp. 36-46.

- [221] Smith, George, and David A. Atchison. "The Eye and Visual Optical Instruments." *The Eye and Visual Optical Instruments, by George Smith and David A. Atchison*, pp. 830. ISBN 0521478200. Cambridge, UK: Cambridge University Press, February, 1997
- [222] S. Bo, H. Qing, H. Chao, and M. Q. H. Meng, "A new camera calibration method for multi-camera localization," in *Automation and Logistics (ICAL), 2010 IEEE International Conference on*, 2010, pp. 7-12.
- [223] O. Olivetti, "Oracle research laboratory face database of faces," ed.
- [224] P. J. Phillips, M. Hyeonjoon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1090-1104, 2000.
- [225] I. MathWorks, *MATLAB: the language of technical computing. Desktop Tools and Development Environment, version 7* vol. 9: MathWorks, 2005.
- [226] N. Gourier, D. Hall, and J. L. Crowley, "Estimating face orientation from robust detection of salient facial structures," in *FG Net Workshop on Visual Observation of Deictic Gestures*, 2004.