“DATA FUSION IN DEPTH IMAGES: APPLICATION TO FACIAL BIOMETRICS”

THESIS SUBMITTED TO THE GOA UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN

ELECTRONICS

By

Mr. ANIKETH A. GAONKAR, M.Sc.

RESEARCH WORK CARRIED OUT UNDER

THE GUIDANCE AND SUPERVISION

Of

Dr. RAJENDRA S. GAD, Ph.D.

Professor in Electronics
School of Physical and Applied Sciences
Goa University, Taleigao Plateau
Goa- 403 206

JULY, 2021

CERTIFICATE

This is to certify that the thesis titled "**Data Fusion In Depth Images: Application To Facial Biometrics**" being submitted by Aniketh A. Gaonkar to Goa University for the award of the degree of Doctor of Philosophy, is an original research work carried out by him under my supervision. In my opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree.

Dr. Rajendra S. Gad

Professor in Electronics & Research Guide
School of Physical and Applied Sciences
Goa University

July, 2021

Goa University

Taleigoa Plateau, Goa - 403 206

DECLARATION

I, **Aniketh A. Gaonkar**, hereby declare that the work presented in this thesis entitled "**Data Fusion In Depth Images: Application To Facial Biometrics**" is my own original contribution and it has been written by me in its entirety. I have honestly and clearly acknowledged all the works of others whenever reported in this thesis.

Aniketh A. Gaonkar

(Research Scholar in Electronics)
School of Physical and Applied Sciences

July, 2021

Goa University

Taleigao Plateau, Goa - 403 206

Dedicated to

My Family & Friends

for their endless support

ACKNOWLEDGEMENT

Acknowledgment is not merely a formality but a way of expressing deep gratitude towards the help and support offered by near and dear ones. A very well known shloka in Sanskrit

" प्रेरकः सूचकश्चैव वाचको दर्शकस्तथा।
शिक्षको बोधकश्चैव षडेते गुरवः स्मृताः॥"

which means "The one who inspires, one who informs, one who recites, one who guides, one who teaches, and the one who awakens, these are the six Gurus to remember." And I feel that my guide has all these six qualities, and I have witnessed the same over the various phases of research/Ph.D. life. With this, I take an opportunity to express a deep sense of gratitude towards my guide, Dr. Rajendra S. Gad, Professor in Electronics & Vice-Dean, School of Physical and Applied Sciences, Goa University. His scholarly inputs, valuable guidance, suggestions, inspirations, and firm convictions; motivated me to boost my research career. Moreover, his expert guidance and unwavering enthusiasm for research have helped me to bring out a good piece of research.

The outcome of this research also has valuable suggestions, feedback, and critical comments of VC's nominees Dr. M. Kunhanandan, Asst. Professor in Mathematics, School of Physical and Applied Sciences, Goa University, and Dr. Jivan S. Parab, Assoc. Professor in Electronics, School of Physical and Applied Sciences, Goa University. I would like to thank Dr. M. Kunhanandan for his input and suggestions in developing the hole-filling filters w.r.t. mathematical aspects and other valuable comments during the Departmental Research Committee (DRC) meetings. Further, I would like to extend my thank to Dr. Jivan Parab for sharing his research expertise and valuable suggestions throughout. His continuous appreciation made me work more enthusiastically towards achieving the research goals. I would also like to thank Dr. A. Mohapatra (Former VC's nominee) for his valuable comments and suggestions during my initial research period.

I would also like to thank Prof. Gourish Naik (HAG, Former Dean of FNS & Former Head, Department of Electronics, Goa University), a person with superior knowledge, for his

support and guidance. His appreciation, critical comments, and questioning during the DRC meetings have mutually helped me to grow.

I am very thankful to the Ministry of Electronics & Information Technology (MeitY), Government of India, for granting a fellowship under Vishveshwarya Ph.D. Scheme under Electronics System Design & Manufacturing (ESDM) and IT/IT Enabled Services (IT/ITES) sectors.

A special thanks to Dr. Narayan Vetrekar for being a wonderful labmate, helping me to design the data acquisition protocols, resolving errors in the codes, clearing my doubts, and always motivating me. Further, I want to thank Ms. Bhagyada Pai Kane, Ms. Shweta Sawal Desai, and Mr. Saurabh Vernekar (Postgraduate students, Department of Electronics, 2015 batch) for setting up necessary arrangements in the 3-D imaging laboratory and synchronize the volunteers on time for biometric data. Finally, I want to thank all the volunteers for participating willingly in biometric data collection for academic research.

I would like to thank Dr. Panem CharanArur, Dr. Marlon Sequeira, Mr. Caje Pinto, and Mrs.Yogini Prabhu for being wonderful lab members. I would also like to acknowledge Dr. Sulaxana Vernekar, Dr. Ingrid-Anne Nazareth, Dr. Vithal Tilvi, Dr. Supriya Patil, Mr. Sameer Patil, Dr. Vinaya Gad, Dr. Shaila Ghanti, Dr. Niyan Marchon, Dr. Udaysingh Rane, Dr. Reshma Rauth Deasi, Dr. Sandesh Bugade, Mrs.Lina Gad Parab, Mrs. Rama Murkunde and Mrs. Amrita N. Vetrekar for helping me in all possible ways.

A sincere thanks to Mr. William D'Souza, Mr. Vishant Malik, Mrs. Ashwini Velip, Mrs. Pushpa Andrade, Mr. Agnelo Lopes for supporting and helping me with the administrative works regarding the processing of Ph.D. files.

I would also like to express my gratitude to Prof. Anuradha Wagle (Controller of Examinations, Goa University) for showing concern towards my research and for granting me leave whenever needed for my research work. I would also like to thank Mrs. Maya Sawant (former AR-E(PG)), Mr. Madhusudan Lanjewar (former AR-E(PG)), Ms. Qubilah Dsouza (AR-E(UG)), and staff of the Exam-Professional Section for managing the work in my absence during my research period. I would also like to thank the Academic and Administration Division of Goa University for processing the necessary Ph.D. files.

A special thanks to Dr. Beena Verenkar (Assoc. Professor in Chemistry; Govt. College of Arts, Science & Commerce – Khandola), who had immense faith in me during my bachelors and motivated me to pursue research.

Friends always add beautiful colors to the canvas of your life. I am privileged to have an excellent bond of friendship with Sanket and Chandan at the Goa University premises. I am grateful to them for sharing valuable moments together and for always being special friends on this journey. I would also like to acknowledge my friends – Harshada Gauns, Mitesh Gad, Supriya Chari, Tanvi Bukkam, Prajyoti Sawant, Dr. Rahul Kerkar, Pratik Asogekar, Dr. Shambhu Parab, Dr. Diviya Vaigankar, Sajiya Mujawar, Sulochana Shet, Alisha Malik, Dr. Bhakti Salgaocar, Dviti Mapari, Sankrita Naik Gaonkar, Nikita Verenkar, for their support. Finally, deep gratitude to the members of my special groups SNAPS (Shravan, Nishant, Pundalik, Saurabh) and BSPARK (Bipin, Selvia, Prasanna, Roshini, Kalpesh) for always motivating and encouraging me to achieve my goal.

This acknowledgment would not be complete without expressing my gratitude toward my family. Family is the one who stands with you in your ups and downs, but we hardly acknowledge their deeds. Here, I would like to take a moment to thank my father, Mr. Arjun Gaonkar, and my mother, Mrs. Anita Gaonkar, for continuously encouraging me to pursue higher education and standing with me in all my challenges. I am grateful to them for the enormous effort, and hard work they have put in that helped me to achieve my goals. In addition, I am thankful to my younger brother Abhijit Gaonkar for his help and valuable support in all means.

I would like to thank each and everyone who has helped me in some way or the other throughout my research life.

Above all, I owe everything to God for granting me wisdom, strength, willpower, good health, and determination to endure all the obstacles that came in the way of my work and thus, making me accomplish my goal.

- Aniketh A. Gaonkar

ABSTRACT

Facial biometrics has received paramount attraction and has grabbed a unique position in the research area. It is an easily accessible and convenient trait as compared to the other traits of biometrics. One can find substantial research in 2D biometrics; however, its outreach has limitations like illumination variation and pose variation. The 3D faces can overcome the limitations that commonly affect the 2D system as the 3D system has more spatial information than 2D, in the form of depth. The research in the 3D domain was an expensive task until the development of the low-cost 3D Kinect camera.

Here in this thesis, we have generated a Kinect-based GU-RGB-D database having variation in pose (-90^0, -45^0, 0^0, $+45^0$, $+90^0$), expressions (smile, eyes closed), occlusion (paper was covering half part of the face), and captured in two different environmental conditions (Controlled and Uncontrolled). This makes it a perfect building block to study the practical challenges of RGB-D systems. Here, preliminary studies using the score level fusion and pixel-level image fusion have been performed on the generated GU-RGB-D database and on the publicly available EURECOM database.

The Kinect camera being a low resolution leads to data loss and creates holes in the image during the acquisition process due to various factors affecting the reflectance of IR. Filling these holes is an essential task as it degrades the image quality and affects the overall system's performance. This thesis presents kernel-based hole filling filters (LI- Filter, EA-Filter & WA-Filter) for the depth images at a pre-processing stage. To quantify the performance of the proposed filters, the experimental evaluation has been performed on the GU-RGB-D & on the publicly available IIITD RGBD databases using local and global features extractor algorithms such as PCA, HOG, LBP, LPQ, GIST, BSIF, and LogGabor.

Further, we have used a Collaborative Representation Classifier (CRC) and an Image Set Classification approach to study the performance on the RGB-D database. In CRC based approach, the images are first fused using 2- Discrete Wavelet Transform (2DWT) followed by feature extraction using PCA, HOG, LBP, LPQ, GIST, BSIF Log Gabor, and CNN. The obtained features are then classified using CRC. In the Image Set Classification approach, the images are fused using pixel-level image fusion and CNN-based image fusion, and they are

classified using image sets to quantify recognition performance. Here the features are extracted from the depth and fused images independently, and these are used to learn the set classification algorithms like AHISD, CSD, CDL MMD, MDA, SANP).

LIST OF PUBLICATIONS

1. **A. A. Gaonkar**, N.T. Vetrekar, R.S Gad; "Hole Filling And Image Fusion Approach For RGBD Database"; International Journal of Engineering Research and Technology (IJERT), ISSN 0974-3154, Volume 13, Number 12 (2020), pp. 5113-5122

2. **Gaonkar A. A**., Gad M.D., Vetrekar N.T., Tilve V.S., Gad R.S. (2017) Experimental Evaluation of 3D Kinect Face Database. In: Mukherjee S. et al. (eds) Computer Vision, Graphics, and Image Processing. ICVGIP 2016. Lecture Notes in Computer Science, vol 10481. Springer, Cham. https://doi.org/10.1007/978-3-319-68124-5_2

3. **A. A. Gaonkar**, N.T. Vetrekar, R.S Gad; "Collaborative Representation With Hole Filling Techniques For Kinect RGBD Face Recognition", - (Communicated)

4. **A. A. Gaonkar**, N.T. Vetrekar, R.S Gad; "Fusion Based Image Set Classification Approach For RGBD Images"; – (Communicated)

CONFERENCE PUBLICATIONS

1. **A. A. Gaonkar**, MD.Gad, N.T.Vetrekar, R.S.Gad, G. M. Naik, "3D Biometrics For Defense Security: Multimodal Fusion Approach", Fourth International Conference on Electronic Warfare - EWCI 2016. 22 - 25 February 2016, Bangalore, India

2. CharanArur Panem, **A. A. Gaonkar**, U. V. Rane, A. B. Pandit, R. S. Gad, "Sensors Data Fusion Architecture Over MIMO: Case Study of Quadcopter", International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) – 2016, Chennai T.N., India

3. **A.A. Gaonkar**, MD. Gad, N.T. Vetrekar, Vithal Shet Tilve, R.S. Gad, "Experimental Evaluation Of 3D Kinect Face Database" ICVGIP 2016 – 19 Dec 2016 IIT Guwahati

4. NT. Vetrekar, Raghavendra Ramachandra, **A.A. Gaonkar**, G.M. Naik, R.S. Gad, "Extended Multispectral face recognition across two different age groups: An Empirical Study", ICVGIP 2016, 18 - 22 December, IIT Guwahati, India

5. **A.A. Gaonkar**, N.T. Vetrekar, Vithal Shet Tilve, R.S. Gad, "3D Kinect Face Recognition: Patch Filling Technique" VLSI symposium 2017, 28th April 2017, Goa

6. **A.A. Gaonkar**, R.S. Gad, "Deep Learning In Image Processing & Embedded Systems – A Review" VSI symposium 2018, 23th March 2018, Goa University

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF FIGURES

LIST OF FIGURES

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES

LIST OF TABLES

LIST OF TABLES

LIST OF TABLES

LIST OF TABLES

LIST OF ABBREVIATIONS

2.5-D: 2.5 Dimensional

2-D: 2 Dimensional

3-D: 3 Dimensional

AHISD : Affine Hull Based Image Set Distance

BSIF : Binarized Statistical Image Feature

CHISD : Convex Hull Based Image Set Distance

CMC : Cumulative Match Curve

CNN : Convolutional Neural Network

CRC : Collaborative Representation Classifier

DCNN : Deep Convolutional Neural Network

DCS : Discriminant Color Space

DCT : Discrete Cosine Transform

DLQP : Depth Local Quantized Pattern

DWT : Discrete Wavelet Transform

EA - Filter: Exponential averaging based filtered depth images

G-LBP : Gradient Local Binary Pattern

HOG : Histogram of Oriented Gradients

ICA : Independent component analysis

ICP : Iterative closest point

IR : Infrared

KLDA : Kernel Linear Discriminant Analysis

LBP : Local Binary Pattern

LGBP : Local Gabor Binary Pattern

LI- Filter: Linear interpolation-based filtered depth images

LG: LogGabor

LPR : Local Phase Quantization

LIST OF ABBREVIATIONS

MDA : Manifold Discriminant Analysis

CDL : Covariance Discriminative Learning

MLNN : Multi-Layer Neural Network

MMD : Manifold-Manifold Distance

PCA: Principal Component Analysis

PLS : Partial Least Squares

RDF : Random Decision Forest

ReLU Layer: Rectified Linear Unit

RGB : Red Green Blue

RGB-D : Red Green Blue - Depth

ROC : Receiver Operating Curve

SANP: Sparse Approximated Nearest Point.

STFT : Short-Term Fourier Transform

SVM: Support Vector Machine

WA- Filter: Weighted averaging based filtered depth images

WO - Filter: Depth images without filtering

CHAPTER 1:

INTRODUCTION

1.1 Overview

The escalating need for security enhancement in today's world has attracted researchers to the area of biometrics. Traditional authentication systems that are token-based or knowledge-based [1] are not entirely reliable and are inefficient to meet the fast-growing world's high-security needs. The knowledge-based systems, such as passwords, are sometimes difficult to remember and annoying for a user, particularly when alphanumeric or non-dictionary words are used. On the other hand, these passwords can easily be cracked/hacked by various alphanumeric combinations [2]. Similarly, token-based systems like Identity cards can be stolen or lost [3], making it precarious to an individual or an organization or the highly secured areas like the defense base. Due to several such issues with these systems and being independent of an individual's inherent attributes, and it varies for every individual, biometric-based authentication/identification came into the limelight [4]. Alternatively, biometric-based systems offer a stable and reliable solution to recognize individuals by employing automated or semi-automated methods based on their biological characteristics [4].

Biometric authentication systems can be broadly classified into physiological (such as the face, iris, palm, finger, etc.) and behavioral (such as signature, gait, voice, keystrokes, etc.) attributes of an individual depending upon the types of measurements used for computations [5]. Physiological characteristics are based on an individual's physical attributes such as the face, fingerprint, iris, palmprint, etc., while behavioral characteristics are associated with an individual's behavior such as voice, keystrokes, signature, etc. The physiological biometric traits are mostly 2-dimensional signals, and behavioral biometric traits are 1-dimensional, mainly time-domain signals [5]. The biometric authentication or identification systems are preferred over the traditional methods for various reasons, such as: An individual needs to be physically present at the time of authentication. These systems obviate the need to remember the password or to carry a token. These systems are pattern recognition-based, which identifies an individual by determining the authenticity of its specific physiological or behavioral characteristics.

The basic operations of the biometric system (physical or behavioral) can be classified as registration, pre-processing, feature extraction, and matching. These operations are briefed as follows [6][7]:

- Registration/Enrollment: This is the initial process in which the biometric data is acquired and stored in the database as templates.
- Pre-processing: The acquired data also has the collection of unwanted information. This is removed in this stage by applying the various pre-processing techniques to avoid the system's overall performance degradation.
- Feature Extraction: In this stage, the best quality features are obtained from the acquired data using various feature extraction algorithms.
- Matching: Here, the matching scores are computed from the obtained biometric training and testing features, and accordingly, the final output of the system is obtained.

An ideal biometric trait should fulfill the following seven established principles/characteristics to become a part of a biometric application. [1] [8]

- Universality: The trait used in the biometric application should be available with all the individuals/ users accessing the application. If not, it will be a less attractive biometric trait and will have a reduced power of subject discrimination.
- Uniqueness: The trait should have the ability to differentiate amongst the individual identities, i.e., it should be unique for every subject. The identical biometric features may have rare to no applications in biometric systems.
- Permanence: The trait of biometric of an individual should not be a time variant over a period concerning the matching algorithms. However, this principle may not withstand soft biometrics or transient biometrics as some applications don't require long-term biometric recognition.

- Measurability: The biometric trait should be acquired and digitized without causing any inconvenience to an individual. Moreover, the acquired data should be fit for extracting the representative feature sets.
- Acceptability: The individuals using the biometric systems should be comfortable in providing their biometric data. The factors that can affect acceptability are the subject's resistance towards the measurement of a specific biometric trait, the subject's trust in the system, and the privacy of the specific biometric trait.
- Circumvention: This refers to the ease of imitating the trait of an individual using artifacts. A biometric system should be robust to attacks and spoofing.
- Performance: The final performance of the system, i.e., the recognition accuracy and the requirement of resources, should be within the constraints of the application.

However, none of the biometric traits are ideal, i.e., a single biometric trait does not meet all the above requirements, but they are admissible. The significance of a specific biometric trait to some applications depends upon the nature and needs of the system design and the biometric trait properties [1].

1.2 Face Biometrics

The various biometric traits can be deployed to develop identification and recognition systems based on human physiological or behavioral modes. One of the most active biometric traits used over decades for security enhancement is face recognition, as the facial trait is non-intrusive and user-friendly [9] compared to other biometric traits [10]. Further, this technology is independent of human monitoring and easy to deploy and maintain.

Face recognition is an everyday task which an individual performs quickly, accurately, and frequently. However, automated algorithms' implementation to achieve the same task becomes a challenging research problem that has received proportionate attention in the literature.

The non-intrusive properties of face biometric allow capturing face images at varying stand-off distances without the user's co-operation in a covert manner. Further, with superior recognition accuracy and wide-scale utility in various applications ranging from simple access control to highly secure cross-border applications, face biometric have received significant attention in academic as well as in commercial industry sectors [11]. Therefore, even though the other technologies based on the traits like fingerprint and iris, which are more accurate and mature [12], they cannot completely replace the need for face biometrics.

Research in the area of face recognition can be traced back to the 1960s [13]. A computer to recognize a human face was researched by Woodrow W. Bledsoe, Helen Chan, and Charles Bisson in the year 1964 to 1966. Further, Peter Hart continued this research and used a set of images instead of feature points to optimize the results. Later in the 1970s, Goldstein, Harmon, and Lesk developed an automatic human face identification system using 21 specific subjective markers like hair color and lip thickness. The approach had good recognition accuracy but was impractical to apply for many faces. Turk and Pentland [14] in 1991 proposed a principal component analysis (PCA) based method (eigenface algorithm) to handle face data. Thereafter several algorithms were developed inspired by [15–17]. Christoph von der Malsburg [18], in 1997, designed a system that had the ability to identify people from non-clear photos. Followed this work, the research in the area of face recognition diverged into two paths. Face recognition using 3D view is proposed and implemented in systems such as Polar and FaceIt [19].

1.3 Related Work In 3D Face Biometrics

Despite significant progress in the area of face recognition, the number of challenges due to pose and illumination variation, expressions, occlusions remains unsolved in the literature, which affects the system's performance [20–24], and is required to be addressed.

In comparison to the 2D face recognition, 3D has good robustness and high fake resistance. 3D systems employ rich facial geometric information. It is robust to facial pose variations and has less effect of ambient light conditions on it. Thus, it has the potential to overcome the inherent limitations of 2D face recognition. Hence this is used for high-security areas [25]. The 3D faces have more spatial information in the form of depth which is an inherent property associated with 3D face recognition and is robust against the uncontrolled environment compared to 2D biometrics. 3D biometric has presented significant improvisation in 3D face acquisition and 3D face matching [11]. The improved 3D imaging devices and processing algorithms have attracted the researcher community towards developing a reliable face recognition system [23].

In the late 1980s, a small 3D face database was engaged with a curvature-based method by [26] and obtained 100% recognition accuracy. Further, in 1996, the experiments of combining frontal and side view performed by [27] have shown improvement in the recognition accuracy. Later with the development of the 3D scanning devices (specifically laser based and structured light technology based) of high ability, more and more 3D face recognition research has been proposed and also contributed to a large number of 3D databases. The details of the existing laser scanner-based and stereo imaging-based 3D face databases in literature are described in detail in chapter 3 (Table 3.1).

The 3D biometric research was an expensive task [9] until an efficient, low-cost RGB-D Kinect camera was developed, which provides RGB image and depth information [28]. The images captured with the Kinect camera have low resolution and noise. Yet, it has more spatial information in the form of depth, which is a robust inherent property associated with 3D face recognition against the uncontrolled environment. The Kinect-based RGB-D databases available in the literature are discussed in detail in chapter 3 (Table 3.2).

Feature extraction is one of the essential stages in biometrics, as it is necessary to obtain high-quality features from the raw data to have good performance of the system. The research on 3D face recognition is mostly focused on the high-resolution data acquired using high-cost 3D

scanners and under-controlled environments, whereas the work based on low-resolution cameras such as Kinect is minimal compared to scanning devices. Min et al. [29] have proposed a Kinect based real-time 3D face identification system in which the face region is detected and segmented by thresholding the depth values. In the subsequent step, the face images are cropped to a standard resolution. Further, the probe is registered with several intermediate references using EM-ICP [30] algorithm to obtain matching. In another work, Min et al. [31] have generated a 3D database based on the Kinect sensor having 52 subjects over two sessions and has 2D, 2.5D, 3D, and video data. Here recognition rates are calculated for 2D, 2.5D, and 3D-based face data using standard face recognition techniques like PCA, LBP, SIFT, LGBP, ICP, and TPS, and also RGB and Depth images were fused using score-level fusion.

Mantecon et al. [32] have presented the recognition results using only depth information captured from Kinect 2 sensor. In this work, the authors have proposed the Depth Local Quantized Pattern [DLQP] descriptor, a modified version of the original LBP operator. This modification extracts the robust and high distinguishable features between different patterns. Further, the output of the descriptor was engaged in training and testing an SVM classifier. Mantecon et al. [33], in their other work, proposed a face recognition algorithm based on a bag of dense derivative depth patterns (Bag-D3P) which is a highly discriminative image descriptor. This descriptor involves four different stages; dense spatial derivatives are computed in the first stage to encode the 3D local structure and quantized in a face-adaptive fashion in the next stage. A compact vector description is created in the third stage by a multi-bag of words from the quantized derivatives. In the final stage, the global spatial information is added by the spatial block division. After that, the SVM classifier is engaged to obtain the recognition task. Further, the results are compared with the different state-of-the-art approaches: LBP + SVM, SIFT + SVM, PCA + neural networks (NN), LGBP (Gabor and LBP features) + NN, LPQ + SVM, HOG + SVM; a BSIF + SVM, and a DLQP (depth local quantized patterns) + SVM.

Neto et al. [34] have developed a face recognition system using the 3DLBP method for the data captured from the Kinect sensor. The features obtained from 3DLBP are used to train the SVM classifier to compute the final results. The work proposed by Li et al. [9] also presents a Kinect based face recognition study that utilizes both depth and RGB images. The authors have generated the depth map from the point clouds obtained from Symmetric filling in his work. In symmetric filling, the point cloud is mirrored after correcting the pose. Here the Euclidean distance is computed between the point in the mirrored cloud and its closest point on the original image. Depending upon the Euclidean distance, the point is added to the original point cloud if the threshold is smaller. The RGB images are transformed to Discriminant Color Space (DCS) before utilizing. Further, for face recognition, the SRC (Sparse Representation Classifier) algorithm is applied on pre-processed depth and DCS color texture separately. The DCS texture comprises three channels and needs to be stacked into one vector before the application of SRC. The set of two obtained similarity scores are normalized and summed for the final result.

Hazým Kemal Ekenel et al.; (2007) obtained 3-D face recognition approach using the discrete cosine transform (DCT), which is a local appearance-based model at the feature level [5]. Tri Huynh et al. [35] have proposed a new LBP based descriptor, namely Gradient-LBP (G-LBP), for gender recognition task on EURECOM and Texas database. Enrico Vezzetti et al.;(2014) proposed a new 3D face recognition algorithm, whose framework is based on extracting facial landmarks using the geometrical properties of facial shape [7]. Ajmera et al.; (2014) computed CRR based on modified SURF descriptors and image enhancement techniques and filters like adaptive histogram equalization, NLM filter, etc., for their internal database and compared it with EURECOM and Curtinface database and also has performed scored level fusion [9].

Kim et al. [36] have developed a 3D face recognition system based on a deep convolutional neural network (DCNN) and a 3D augmentation technique. This approach makes use of the existing pre-trained version of the VGG-face, which is fine-tuned for the depth data. To train

the CNN network, large data is needed; thus, authors have generated the expressions and occlusions to deal with the shortage of data. A face recognition system developed by Lee et al. [37] is based on a deep learning approach. It utilizes the face images captured using a consumer-level RGB-D camera. Here the recognition process comprises of three parts: depth image recovery, deep learning for feature extraction, and joint classification. The network is first trained with the RGB data and then fine-tuned for the depth data for transfer learning.

Kinect being the low-cost 3D camera has led to the development of few more 3D databases in addition to the conventional scanner-based databases, as mentioned in Table 1.2. This has, in turn, give a strong upward push to 3D research. Some of the Kinect-based work, including their approach, feature extraction methods, classification methods, and fusion strategies, are tabulated in Table 1.3.

1.4 Hole Filling In 3D Face And Related Work

The quality of a captured image plays a very important role in face recognition. The captured data using low-resolution cameras like Kinect, RealSense D435i has low-quality images compared to the high-quality scanners-based images. These cameras capture the RGB and the Depth images. Here, the captured RGB image quality is good, but the captured depth images have missing information in the form of holes [38].

The Kinect camera uses an Infrared (IR) projector and sensor to capture the depth images. The IR projector projects a structured infrared light on an object, which is received by the depth sensor; further, the controller estimates the depth of an object and outputs the depth map image. Due to the intrinsic features of the Kinect and the external environmental parameters changes, the captured depth image has an ineligible quality (noisy and having holes) to use in any computer vision application, and filling these holes becomes challenging work. The reasons for the appearance of holes in the depth images include [39]:

- The horizontal distance between IR projector and the depth sensor is centimeters apart; hence along the edges of objects, some occlusion areas will occur.
- Surfaces reflect the infrared light with mirror reflection and not with diffuse reflection; this causes the depth image to lose the depth information in highlight surfaces.
- The infrared-absorbed surfaces will lead to loss of information in depth images.
- The flickering artifacts may be generated in the depth image as the structured light is randomly projected, and the noise will differ in each depth frame.
- In the presence of high ambient infrared light, IR dots involved in producing the depth images on an object will be indiscernible to the IR camera. Thus, fewer pixels will be correctly captured, thereby deteriorating the image [40].

The holes (missing information)in the depth images affect the overall performance of the system; thus, it needs to be addressed and filled at the pre-processing stage. In literature, one can find the contribution of Yu Mao et al. [41] towards the identification and filling of holes. Here the holes are identified at the initial stage and are filled based on depth histogram and linear interpolation and graph-based interpolation methods. A research of dis-occlusion removal in Depth Image-Based Rendering (DIBR), a hierarchical hole-filling (HHF), and depth adaptive hierarchical hole-filling tactics were used where a pyramid approach is followed from lower resolution estimate of 3D wrapped image to estimate the hole pixels value, is addressed by Mashhour Solh et al. research [42]. A hole-filling algorithm to improve the image quality of DIBR has been proposed by Dan Wang et al. [43, 44]. Here, the order of hole filling is determined by the sum of the priority calculation function and the depth information. Further, the gradient information is used as auxiliary information to find the best matching block. Litong Feng et al.[45] has worked on an adaptive background-biased depth map hole-filling method. Amir Atapour-Abarghouei et al. [46]have addressed hole filling by Fourier transform and Butterworth high/low pass filtering. Here the texture synthesis method is used for high-frequency details, and structural inpainting is used for inpainting the low-frequency information. Further high-frequency depth synthesis has been performed by query

expansion concept with the final output and then recombined in Fourier space. Liang et al. [47] had proposed a segmentation-based approach for inpainting stereo images. Here the constraint was that the missing information in one stereo image might be filled from another image in both color and depth images using depth-assisted texture synthesis. Some other approaches have to be used to fill the holes in the depth images in the absence of stereo or multi-camera views. Breckon et al. [48] had proposed a surface completion technique based on the nonparametric propagation of existing scene information from the known /visible scene areas to the unknown/invisible 3D regions in combination with the preliminary underlying geometric surface completion. Kang Xu et al. [39] has proposed filtering method for small holes and occlusion area in the Kinect based depth images. In this approach the holes are filled using 4-neighbor-pixels interpolating algorithm.

1.5 Motivation

3D face recognition has received blooming attention and interest from the scientific community in recent years due to the development of low-cost Kinect cameras. The literature shows that the 3D face has the ability to overcome the limitations of the 2D face approach, i.e., limitations due to illumination and pose/angular variations, as the 3D has additional information in the form of depth thus it can serve as a robust approach in the areas of high security. Along with security, cost efficiency and processing time also play an important role. The scanner-based 3D systems are expensive and time-consuming as compared to low-cost cameras such as Kinect. However, the Kinect-based databases available for research are very few as compared to the scanner-based databases.

Considering this fact, the generation of the Kinect-based GU-RGB-D database formed the base of this research. The database protocol was designed with expressions, occlusion, angles/pose variations so as to have all possible variations in one single database. The Kinect being the low-resolution camera, the holes get developed (missing information) while

capturing the data. The literature survey directs that to obtain high performance, filling the holes/missing information is an important step. Thus this thesis presents variable kernel-based three hole filling techniques/filters for the depth images. Further to quantify the performance of these, extensive experimental evaluation across the GU-RGB-D and IIIT-D databases is performed on depth images using various state-of-the-art algorithms. The study is also extended for various fusion approaches and advanced classification approaches. The aim of this thesis is to help in solving the difficulties in 3D face recognition by using some simple setups and approaches, which would be of great scientific importance for future research.

1.6 Thesis Contribution

Given the entire literature survey, the thesis entitled "**Data Fusion In Depth Images: Application to Facial Biometrics**" is the compilation of the work listed as follows:

- Generated a Kinect-based GU-RGB-D face database with variations in pose /angles, expressions, occlusions, and illumination, covering all the possible variations under one database. The database is captured in two sessions so as to explore the scope to study the effect of control and uncontrolled environment.

- Presented the two preliminary studies on GU-RGB-D and EURECOM database by using PCA as a feature extractor and having the score level and the pixel-level image fusion approach.

- Designed kernel based hole-filling filters with the contribution from neighborhood pixels for the depth images acquired using Kinect sensor to enhance the performance of the 3D face recognition system. This is experimented on GU-RGB-D database with seven different feature extraction methods such as Principal Component Analysis (PCA), Histogram of Oriented Gradient (HOG), Local Binary Pattern (LBP), Local Phase Quantization (LPQ), GIST, Binarized Statistical Image Features (BSIF), and

LogGabor to demonstrate the significance of employing the hole filling techniques to improve the performance of the state-of-the-art face recognition methods.

- Proposed fusion scheme based on 2D-Discrete wavelet transform and collaborative representation classifier (CRC). The scheme combines the RGB and depth image (after hole filling) using 2D-Discrete wavelet transform, which is followed by a robust collaborative representation classifier (CRC) for RGBD based face recognition. Presented an extensive experimental evaluation based on the proposed scheme and the designed hole filling filters on GU-RGB-D and IIIT-D databases. The verification and recognition rate is computed for eight different feature extraction methods such as Local Phase Quantization (LPQ), Local Binary Pattern (LBP), Histogram of Oriented Gradient (HOG), GIST, LogGabor, Binarized Statistical Image Features (BSIF), Principal Component Analysis (PCA), and deep convolutional neural network features extracted at `conv5' layer to demonstrate the applicability of our approach for improved performance analysis.

- Presented the Image Set Classification study based on various Image Set Classification algorithms i.e., MMD: Manifold-Manifold Distance MDA (Manifold Discriminant Analysis), CDL (Covariance Discriminative Learning), AHISD (Affine Hull Based Image Set Distance); CHISD (Convex Hull Based Image Set Distance), SANP (Sparse Approximated Nearest Point).

- Presented the image set classification study on depth images, fused RGB and depth images using image-level pixel fusion and CNN based image fusion. The experimental results are computed on the GU-RGB-D database using seven different feature extraction algorithms like Principal Component Analysis (PCA), Histogram of Oriente Gradient (HOG), Local Binary Pattern (LBP), Local Phase Quantization (LPQ), GIST, Binarized Statistical Image Features (BSIF), and Convolution Neural Networks (CNN).

1.7 Thesis Outline

The thesis is organized into a total of six chapters as described below:

- **Chapter-1: Introduction;** presents a detailed discussion on 3D biometric face recognition, and related contributions in literature are summarized. Further, motivations and contributions are specified in this chapter about the significant findings of this thesis.
- **Chapter-2: Algorithms, Techniques, And Methods;** presents the algorithms, methods, and techniques corresponding to the state-of-the-art feature extractions, fusion, and classification methods, used in this thesis.
- **Chapter-3: Generation Of GU-RGB-D Database & Preliminary Studies;** presents in detail the GU-RGB-D database generation protocol and the preliminary studies based on score level fusion and pixel-level image fusion.
- **Chapter-4: Pre-processing And Feature Extraction Methods;** presents in detail the designed variable kernel-based hole filling filters and the experimental evaluation of the same by using different state-of-the-art feature extraction algorithms.
- **Chapter-5: Advance Classification Approach For Performance Analysis;** presents the performance analysis using the collaborative representation classifier (CRC) based and Image Set Classification approach.
- **Chapter-6: Summary & Conclusion;** presents the summary and the conclusive remarks based on the research.

Table 1.1: Studies based on Kinect based 3D facial Databases

Databases	Authors	Year	Approach Based on:	Features extraction	Classifier	Fusion
EURECOM	[31]	2014	-	PCA, LBP, SIFT, LGBP, ICP, TPS	-	Score level (weighted sum)
EURECOM, VAP, IIIT-D	[49]	2019	LBP-RGB-D-MSVM algorithm	LBP	Multiclass Support Vector Machines (MSVM)	Feature level (feature concatenation)
EURECOM, Curtin Face	[50]	2016	Similarity values from texture images and depth	HOG	Joint Bayesian algorithm	Score level (weighted sum)
EURECOM, VAP, IIIT-D	[51]	2018	Learning complementary features from multiple modalities and common features between	CNN	AlexNet [52], GoogLeNet-BN [53], VGG-16 [54]	Score level (weighted sum)

			different modalities			
EURECOM, IIIT-D, HRRFaceD, Biwi Kinect Head Pose database	[33]	2016	Face image descriptor: bag of dense derivative depth patterns (Bag-D3P)	Bag-D3P, PCA, LBP, HOG, LGBP, DLQP, SIFT, LPQ, BSIF	Support Vector Machine (SVM) , Neural Network (NN)	-
EURECOM	[35]	2012	Gradient-LBP (G-LBP)	LBP. 3DLBP, GLBP	Support Vector Machine (SVM)	-
Curtin Face, IIIT-D	[55]	2020	Multimodal attention network (feature-map attention and spatial attention)	Convolutional feature extraction	4 fully connected layers serve as a classifier	Two-layer attention mechanism
EURECOM, VAP, IIIT-D	[28]	2014	Entropy Maps and Visual Saliency Map	HOG	Random Decision Forest (RDF)	Match Score Level Fusion (Weighted Sum) Rank Level

						Fusion (Weighted Borda Count)
IIT-K, EURECOM, Curtin Face	[56]	2014	modified SURF descriptors	SURF	-	weighted score fusion
VT-KFER	[57]	2015	Baseline feature sets	LBP, Distance based 3D features	SVM	-
IKFDB	[58]	2021	-	HOG	Support Vector Machine, Multi-Layer Neural Network, Convolutional Neural Network	-

CHAPTER 2:

ALGORITHMS, TECHNIQUES

AND METHODS

CHAPTER 2

This chapter gives a brief description of the various publicly available databases used in our study and the state-of-the-art techniques/algorithms used throughout the experimental evaluation in this thesis. The organization of the chapter is as follows; section 2.1 describes the publicly available EURECOM and IIIT-D databases and their generation protocols. The different local and global feature extraction algorithms such as PCA, HOG, LBP, LPQ, GIST, BSIF, Log Gabor, and CNN, which are used for extracting features in the experimental protocol, are discussed in the section 2.2. Further, section 2.3 describes the different fusion strategies employed in this research work, such as pixel-level image fusion, 2D-discrete wavelet transform-based image fusion, and CNN-based image fusion. The Collaborative Representation Classifier (CRC), a Classification technique used for the RGB-D database, is also surveyed in section 2.4.1. Section 2.4.2 has the description of the Image Set Classification algorithms like Manifold-Manifold Distance (MMD), Manifold Discriminant Analysis (MDA), Covariance Discriminative Learning (CDL), Affine Hull Based Image Set Distance (AHISD), Convex Hull Based Image Set Distance (CHISD), and Sparse Approximated Nearest Point.

2.1 Databases

This section briefly describes the publically available RGB-D face databases used in this study. Along with the GU-RGB-D database (described in the later chapter), we have used EURECOM and IIIT-D databases for performing different analyses in this thesis.

2.1.1 EURECOM Database

EURECOM [31] database is a collection of well-aligned 2-D, 2.5-D, 3-D, and video data, captured using a Kinect sensor. The database has a total of 52 subjects, including 38 males and 14 females from different countries. Here the data is available in two sessions, and each session has four types of data modalities for each subject: 2-D RGB image, 2.5-D depth map, 3-D point cloud, and RGB-D video sequence. In both the sessions, there are nine facial variations: Neutral face, strong illumination, mouth open, smiling, occlusion by hand, occlusion by sunglasses, occlusion by paper, left and right face profile, face profile, and the

same acquisition protocol was maintained over the two sessions. All the images are captured under controlled conditions, comprised of natural light and the LED diffusion light. The sample images from the EURECOM database are shown in figure 2.1.

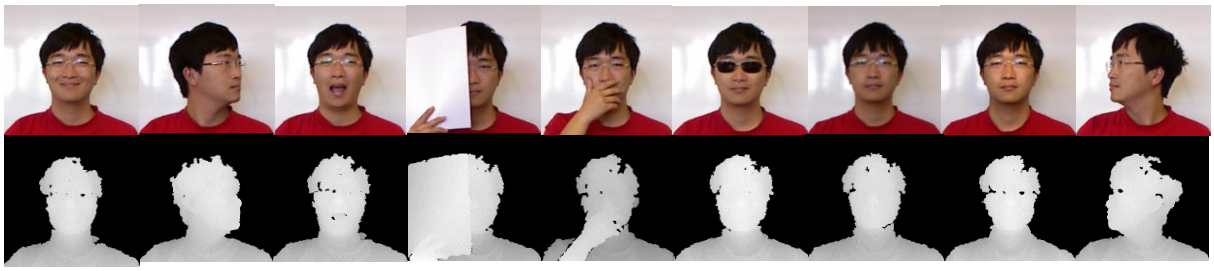


Fig 2. 1: Sample images from EURECOM database

2.1.2 IIIT-D RGB-D Database

IIIT-D RGB-D [59] face database is captured using Microsoft Kinect sensor and comprises of total 106 male and female subjects. The database has multiple RGB-D images of each subject with a minimum count of 11 images and a maximum of 254 images per subject. This database has a total of 4605 images. The sample images from the IIIT-D RGB-D database are shown in figure 2.2.

Fig 2. 2: Sample images from IIIT-D RGB-D database

2.2 Feature Extraction Algorithms

Feature extraction is an essential step in image processing as it extracts the most relevant information from the raw data and represents it in a lower dimensionality space [60]. In the area of facial image processing, facial features can be either extracted globally (face as a

whole) or locally (features like nose, eyes, etc.) from the input images. Global features correspond to the structural anatomy, while local features provide details in facial appearance. In this thesis, seven different types of feature extraction techniques are employed, and the same are described in subsections 2.2.1 – 2.2.8 in detail.

2.2.1 Principle Component Analysis(PCA)

Principle Component Analysis [61][62] is a dimensionality reduction technique used for feature extraction. In image or signal processing, this technique is mainly used to reduce the size of the feature vector, which is in turn used for solving recognition or classification problems. The technique preserves the important information and removes the redundant information from the pattern/image. In the application of face recognition, an eigenface-based approach is used. A face contains a certain set of essential characteristic features called principle components or Eigenfaces and are extracted from the original image with the help of principle component analysis.

The recognition process involves the following operations:

i. Acquire the training set of face images.
ii. Compute the Eigenfaces over each image from the training set, wherein around 90% cumulative higher eigenvalue vectors are considered, which define the face space. The eigenfaces can be recalculated or updated as and when the new faces are augmented.
iii. Calculate the distribution in this new dimension space by projecting the face image of each known subject onto the face space.

This data can be structured to be used in further processing, thereby eliminating the overhead of re-initializing, decreasing the computation time, and improving the system's performance [63].

After initializing the training, the operation involved in the recognition process is as follows:

i. Calculate a set of weights of the input image in terms of eigenfaces by projecting the input image onto each of the Eigenfaces from the training set.
ii. Determine if the image is a face or not by checking if the image is sufficiently close to face space.
iii. On determining a face, classify the weight pattern as known or as an unknown subject.
iv. Update the weights or eigenfaces as known or unknown (if required).
v. If the same unknown subject/face is seen several times, its characteristic weight pattern can be calculated and incorporated into known faces.

2.2.2. Histogram of Oriented Gradient (HOG)

Histogram of Oriented Gradients [64] [65] is a feature descriptor method used to extract features from the images by computing the magnitude of gradient vectors. This technique counts the occurrences of gradient orientation in localized portions of an image which means the entire image is broken into smaller regions, and the gradients and orientation are calculated for each region. This has extensive use in image processing and computer vision. The main steps involved in computing HOG features are summarized as follows:

i. The input image I(x,y) is divided into the blocks corresponding to M × M pixel size and subsequently in smaller and smaller cells. These block sizes and cell sizes can be customized based on the user's requirement.

ii. The magnitude G(x,y) and direction θ(x, y) of the gradients for each pixel location within the cell is computed using x and y directional gradient i.e. $G_x(x, y)$ and $G_y(x, y)$ As in Equation 2.1.

$G(x,y)=\sqrt{G_x(x,y)^2 + G_y(x,y)^2}$;

$$\theta(x, y)= \begin{cases} \tan -1\left(\frac{G_y(x,y)}{G_x(x,y)}\right) - \pi & if\ G_x(x,y) < 0\ and\ G_y(x,y) < 0 \\ \tan -1\left(\frac{G_y(x,y)}{G_x(x,y)}\right) + \pi & if\ G_x(x,y) < 0\ and\ G_y(x,y) > 0 \\ \tan -1\left(\frac{G_y(x,y)}{G_x(x,y)}\right) & otherwise \end{cases} \quad (2.1)$$

Where, $G_x(x,y)$ = I(x + 1, y) − I(x − 1, y) and $G_y(x,y)$ = I(x, y + 1) − I(x, y − 1) depicts the details of horizontal and vertical gradients, respectively, at the given pixel location.

iii. The magnitude of gradients vector computed for each pixel within the cells is placed in either one of the orientation bins according to the gradient angle.

iv. Finally, all the histograms corresponding to each of the blocks are concatenated to obtain the final HOG descriptor.

2.2.3 Local Binary Pattern (LBP)

Local Binary Pattern (LBP) [66] is an effective texture descriptor for analyzing the images. This technique threshold the neighboring pixels based on the value of the current pixel [67] and efficiently captures the local spatial patterns and the grayscale contrast in an image. The steps involved in the computation of the LBP descriptor from an image are explained below[68].

i. In an image, I(x,y), for every pixel (x,y), choose 'N' neighboring pixels at a radius of R.

ii. For the current pixel (x,y), calculate the intensity difference with its N neighboring pixels.

iii. Threshold the central pixel with the neighboring pixels, and all the negative differences are assigned 0, and all the positive differences are assigned 1. This

results in a binary pattern. The equation of the basic version of LBP can be given by equation 2.2

$$LBP(x,y) = \sum_{N=0}^{N-1} S(g_N - g_c)\, 2^N \quad (2.2)$$

Where g_c is a central pixel value positioned at (x,y), g_N is one of the neighboring pixel value within radius R. N is the total neighborhood pixel number, and the function S(x) is defined such that

$$S(x) = \begin{cases} 1 & if\ x > 0 \\ 0 & otherwise \end{cases} \quad (2.3)$$

iv. Next, the histogram is computed to determine the frequency value of the binary patterns. Then, depending upon the number of pixels involved in the LBP computation, histogram bins are determined.

v. All the normalized histograms are concatenated, resulting in a feature vector.

2.2.4 Local Phase Quantization (LPQ)

Local Phase Quantization (LPQ) is a texture descriptor method based on the blur invariance property of the Fourier phase spectrum [69]. It has been used as a feature descriptor to recognize blurred face images in biometrics [70, 71]

. This method employs 2-dimensional Short-Term Fourier Transform (STFT) across the rectangular neighborhood of each pixel position to extract the local phase information.

For a given image I(x,y), STFT on the local neighborhood $N_{x,y}$ at each pixel position x,y of the image can be computed from Equation 2.4

$$STFT(u,(x,y)) = \sum_{(x_1,y_1)\epsilon N_{x,y}} I(x, y - x_1, y_1) e^{-2\pi j u^T (x_1,y_1)} \quad (2.4)$$

Further, only four frequency components corresponding to u = [u_1,u_2,u_3,u_4] are considered in LPQ, where, $u_1 = [\nu,0]^T$, $u_2 = [0,\nu]^T$, $u_3 = [\nu,\nu]^T$ $u_4 = [\nu,-\nu]^T$ and ν is the scalar frequency parameters. Out of the total eight output coefficients obtained for the input image, four coefficients correspond to the real part while the other four corresponds to the imaginary part, resulting in eight binary coefficients. These binary coefficients, when represented in the decimal pattern it generated output I_{LPQ} consist of phase information. Finally, the LPQ feature descriptor can be expressed as the histogram of these decimal values.

2.2.5 GIST

GIST [72, 73] is used to recognize the scene from the image by extracting local and global semantic information. For extracting features, the image is initially divided into several blocks, and the Gabor filters of different scales and different orientations are applied to these blocks to compute the required features.

The steps involved in computing the GIST features for a given image are as follows:

i. Apply the Gabor filters [74] (Equation 2.5) corresponding to 4 different scales and 8 different orientations, resulting in 32 Gabor filters.

$$G(x,y,\theta_w,\sigma_x,\sigma_y) = \exp\left\{-\frac{1}{2}\left[\frac{x_1^2}{\sigma_x^2}+\frac{y_1^2}{\sigma_y^2}\right]\right\}\exp\left[\frac{2\pi i x_1}{\rho}\right] \quad (2.5)$$

Where $x_1 = x\cos\theta_w + y\sin\theta_w$ and $y_1 = -x\sin\theta_w + y\cos\theta_w$; here ρ and θ_w are the wavelength and orientation of the sinusoidal plane wave respectively. σ_x and σ_y is the standard deviation of the Gaussian envelop across x−direction and y−direction, respectively, and w = 1,2,...,m; which represents the number of orientations, here m=8.

ii. To obtain the feature map, perform a convolution operation between the input image and 32 Gabor filters, resulting in 32 feature maps.

iii. Fragment each feature map into 16 portions of 4×4 grids and average the feature values within each of the feature maps.

iv. Further, concatenated the 16 averaged values computed from 32 feature maps to obtain the final GIST of 512 (16 average value x 32 feature map) dimensions.

2.2.6 Binarized Statistical Image Feature (BSIF)

Binarized Statistical Image Feature (BSIF) [75] is a statistical learning-based descriptor that computes a binary code string by thresholding. The pixel's code value is considered as the local descriptor of the image intensity pattern surrounding the pixel. Independent component analysis (ICA) is employed to extract the independent vectors from random samples of the training image set. To generate the BSIF features, the input image is convolved with the predefined textural filters formulated using a natural image database.

The BSIF process can be summarized as follows:

i. Conversion of the input image to a grayscale image

ii. Patch selection from the grayscale image

iii. Subtraction of the mean value from all the components

iv. Patch whitening process

v. Estimation of ICA components

The filter response for the input image $I(x,y)$ and linear filter v_i of size m x m is obtained by convolution as in equation (2.6).

$$Z_i = v_i * I(x,y) \tag{2.6}$$

where Z_i is the response of the i-th filter, and the binarized features b_i is obtained by assigning b_i=1 if $Z_i > 0$ and $b_i = 0$ otherwise.

2.2.7. Log-Gabor

Log-Gabor [76, 77] is a feature descriptor in line with GIST, which uses banks of Gabor filter of different scales and orientations. The filter response of the Log-Gabor filter is Gaussian in nature on a logarithmic frequency scale, allowing it to capture more significant characteristics features in the higher frequency region. The equation (2.7) gives the transfer function of Log Gabor on a linear frequency scale.

$$G_{lg}(\omega) = \exp\left\{\frac{-\log\left(\frac{\omega}{\omega_0}\right)^2}{2\left(log\frac{m}{\omega_0}\right)^2}\right\} \quad (2.7)$$

where ω_0 is the central frequency component of the filter, and *m* is the bandwidth scaling factor. Further, to maintain the consistency in filter shape, the ratio $\frac{m}{\omega_0}$ must be kept constant.

2.2.8 Convolution Neural Network (CNN)

Convolutional neural networks (CNN) [78, 79] is a feed-forward Artificial Neural Network widely used for image processing and computer vision applications. In CNN, interneuron connections are inspired by the biological animal visual cortex (a visual mechanism) [80]. CNN recognizes the visible pattern directly from the pixel images with minimal pre-processing. Like other neural networks, CNN comprises an input layer, multiple hidden layers, and an output layer. The hidden layer comprises of Convolution layers, Pooling Layers, ReLU, and fully connected layers, thus helps to learn higher-order features in data. Each of the hidden layers learns to extract complex features of any given set of input images. The processing capacity of this network depends on the number of hidden layers; thus, it can be refined by varying depth and breadth. The detailed description of architectural layers of CNN and working is as given below:

i. Convolution Layers: These are the basic building blocks of CNNs that extract the features from the input image. The first convolution layer extracts the low-level features like edges, lines, corners, etc. This layer convolves the input image with a set

of kernels corresponding to the features to be extracted (learnable filters) to produce feature maps. The kernel moves pixel by pixel (one pixel at a time) from left to right of the input image, starting from the top left corner. On reaching the top right corner of the image, it moves down to the first element of the next row and again moves from left to right. This process continues till it reaches the bottom right. Small sets of pixels are used for computations depending upon the size of the kernel, thus preserving the spatial relation between the pixels. The size of the feature map is controlled by three parameters depth, stride, and zero paddings.

ii. Pooling Layer: This layer acts as a nonlinear down sampling layer. Max pooling is the common down sampling technique used in these layers. Here each input image is divided into non-overlapping sub-regions (two-dimensional spaces), and the maximum value of each is recorded [81]. The pooling layer gradually reduces the number of parameters and computations (controls overfitting) by reducing the size of the representation. Other pooling techniques used are average, sum, etc.

iii. ReLU Layer: Rectified Linear Unit is an element-wise nonlinear operation applied to every pixel, and the negative pixel values in the feature map are replaced by zero. This layer increases the nonlinear properties of the decision function and the entire network without affecting the receptive field of the convolution layer.

iv. Fully connected layer: The term fully connected implies the connection of every neuron from the previous layer to the next layer. The output of multiple convolutions and max-pool layers is a high-level feature of the input image. This layer classifies the input image into various trained classes depend on the extracted features. Finally, a classifier like softmax or some other classifying technique is used to classify the inputs.

2.3 Image Fusion

Image fusion techniques integrate the complementary data from the multiple input images generated, usually using various modalities. As a result, the resultant image will be more informative and complete than any of the input images [82]. Furthermore, image fusion helps improve geometric corrections, sharpen the image, replace the defective data, enhances invisible features, and provide a better dataset for decision-making [83]. Image fusion can be divided into two groups; spatial domain fusion, where the input image pixels are directly taken into consideration, and Transform domain fusion, where the frequency domain of the input image is considered [82]. We have employed image fusion techniques over two modalities to fuse RGB and depth images to improve the performance of the face recognition algorithms. The employed fusion techniques, i.e., Pixel-level image fusion, 2D-discrete wavelet transform-based image fusion, and CNN-based image fusion, are briefly described in this section.

2.3.1 Pixel-Level Image Fusion

In pixel-level average image fusion, the fused image $I_3(x,y)$ is obtained by averaging the pixel intensities of both the input images $I_1(x,y)$ and $I_2(x,y)$and can be formulated as equation (2.8)

$$I_3(x,y) = \frac{I_1(x,y)+I_2(x,y)}{2} \quad \forall\ (x,y) \tag{2.8}$$

2.3.2 Discrete Wavelet Transform (DWT) Based Image Fusion

In Discrete Wavelet Transform, the input image is decomposed into approximate and informative components that convert the image from spatial domain to frequency domain. Image fusion using DWT can be generalized as follows [82];

i. DWT is performed on both the input images to be fused to obtain wavelet lower decomposition.
ii. These decomposition levels are further fused by implementing different fusion rules, using suitable operators.

iii. Finally, Inverse Discrete Wavelet Transform is performed on the fused decomposed level to reconstruct and obtain the final fused image.

We have employed a 2-Level Discrete Wavelet Transform carried out using the Haar mother wavelet function in our work. The result of wavelet decomposition obtains the wavelet coefficients in seven sub-band images [84] that correspond to one approximation, two horizontal, two vertical, and two diagonal coefficient details.

2.3.3 Convolutional Neural Network (CNN) Based Image Fusion

Convolutional Neural Network (CNN) is basically a deep learning model which learns the hierarchical feature representation for image/signal data with different levels of abstraction. As the basic operation of the CNN is convolution, thus it is feasible to apply CNN for image fusion [153]. The fusion approach can be described in the following steps:

i. Focus detection: Here, the pre-trained CNN model is fed with the input image to generate the score map. The score map consists of focus information, i.e., each coefficient indicates the focus property. Further, the focus map of the same size as the input images of the two or more modalities are generated by averaging the overlapping patches from the score map.

ii. Initial segmentation: The generated fused focus map is segmented into a binary map.

iii. Consistency verification: The popular consistency verification approaches, i.e., small region removal and guided image filtering, are employed to refine the binary segmented map and to generate the decision map 'Z' further.

iv. Fusion: The obtained decision map Z is used to calculate the final fused image T using pixel-wise weighted-average rule and can be given as in equation (2.9).

$$T(x,y) = Z(x,y)I_1(x,y) + \big(1 - Z(x,y)\big)I_2(x,y) \quad (2.9)$$

where, I_1 and I_2 are the input images from the two modalities.

2.4. Feature Classification Techniques

Feature classification is a prime step in image processing as it measures the level of similarity or dis-similarity within the feature vectors obtained from the images. This section briefly described the feature classification technique employed in this thesis apart from simple distance metrics.

2.4.1 Collaborative Representation Classifier (CRC)

Collaborative Representation Classifier (CRC) [85] has emerged as one of the robust feature classification methods in the face recognition domain. It is an extended version of the Sparse Representation Classifier (SRC) and computes the maximum likelihood ratio between the test sample image and the other classes in a joint manner. In order to perform the final feature classification, the maximum likelihood of the test sample is computed against the other classes from the training set. Let the equation (2.10) represent the feature vector (m dimensions) of each image;

$$Z = \{Z_1, Z_2, \ldots, Z_b\} \in \mathbb{R}^{m\,x\,N} \tag{2.10}$$

where b is the total number of classes and N is the total number of images across the classes.

The expression for CRC can be represented by a general modal as follows:

$$\alpha' = \arg\,\min_{\alpha}(\| I - Z\alpha \|_2^2 + \mu \| \alpha \|_2^2) \tag{2.11}$$

where $\alpha = \alpha_1 \ldots. \alpha_b$ is the coefficient vector, μ is the regularization parameter, and 'I ' is an input test image is given by $I \epsilon \mathbb{R}^m$.

2.4.2 Image Set Classification Algorithms

Image set classification is a technique where a set of images represents each class as compared to the traditional recognition or classification problems, where a single image is involved in the learning process [86, 87]. Here a set of test images are assigned to the label of

the nearest training set using distance criteria. The image set classification procedure can be briefly described as follows [86]:

i. Images belonging to the same class are considered as an image set.
ii. The most representative sample images are extracted from the set.
iii. The intrinsic property of this set is learned through a proper probability distribution.

Basically, image set classification is the measure of similarity between the most similar two sets of the same class and having the most similar output responses of the probability distribution models. Even though the images belong to the same class, there may be a remarkable difference due to various conditions such as lighting, posture, etc. thus, the image set classification is used to draw the intrinsic property from a set of images. We have employed six different image set classification techniques; namely, Affine Hull Based Image Set Distance (AHISD), Covariance Discriminative Learning (CDL), Convex Hull Based Image Set Distance (CHISD), Manifold-Manifold Distance (MMD), Manifold Discriminant Analysis (MDA) and Sparse Approximated Nearest Point (SANP). The details related to each of the technique are explained in the section 2.4.2.1 to 2.4.2.6 in the following section.

2.4.2.1 Affine Hull Based Image Set Distance (AHISD) & Convex Hull Based Image Set Distance (CHISD)

These techniques treat a set of images in linear space and characterize each set by the affine or convex hull (a convex geometric region) spanned by its feature points. Here to compare the different sets, the distance of closet approach (geometric distance) is employed between the convex models.

Let the set of samples be $Z_{ci} = \{ Z_{c1}, Z_{c2}, Z_{c3,\dots}, Z_{cm}\} \in \mathbb{R}^d$ corresponds to the class 'c', and each sample image can be expressed in 'd' dimensional feature vector. For the given set, the affine or Convex hull is formulated as follows [87, 88]:

$$A_c^{aff} = \{Z = \sum_{k=1}^{m} \beta_{ck} Z_{ck} \mid \sum_{k=1}^{m} \beta_{ck} = 1\} \quad (2.12)$$

Further, the affine hull can be parametrized by choosing some point μ_c as a reference point in affine space where

$$\mu_c = \frac{1}{m}\sum_{k=1}^{m} Z_{ck} \quad (2.13)$$

and the equation (2.12) can be rewritten as

$$A_c^{aff} = \{Z = \mu_c + U_c v_c \mid v_c \in \mathbb{R}^q\} \quad (2.14)$$

Where U_C is the feature vector in the affine subspace and v_c is a vector of free parameters expressed with respect to U_c of reduced dimensions within the subspace.

By introducing the upper bound 'U' and the lower bound 'L' on the allowable coefficient β, the equation (2.15) can be modified as the equation for the convex hull approximation.

$$A_c^{caff} = \{Z = \sum_{k=1}^{m} \beta_{ck} Z_{ck} \mid \sum_{k=1}^{m} \beta_{ck} = 1, L \leq \beta_{ck} \leq U\} \quad (2.15)$$

Where L = 0 and U ≥ 1

2.4.2.2 Covariance Discriminative Learning (CDL)

In Covariance Discriminative Learning (CDL) [89] approach, the image set $Z=\{Z_1, Z_2, Z_3, ...Z_m\} \in \mathbb{R}^d$ is represented by its natural second-order statistic (covariance matrix, *C*). The covariance matrix of each set is computed using equation (2.16)

$$C = \frac{1}{m-1}\sum_{k=1}^{m}(Z_k - \bar{Z})(Z_k - \bar{Z})^T \quad (2.16)$$

Where m is the number of samples in the image set, Z_k denotes the k-th image and $\bar{Z}$ is the mean of the image samples. Further, the Log-Euclidean distance approach is adopted to map the covariance matrix from the Riemannian manifold to a Euclidean space. This approach is

adopted as the classic learning algorithm is operated in the vector spaces associated with the Euclidean matrix and cannot take the direct input point from the manifold. Kernel Linear Discriminant Analysis (KLDA) [90] and Partial Least Squares (PLS) [91] are the two classification algorithms used in this algorithm

2.4.2.3 Manifold-Manifold Distance (MMD)

The Image set containing images pertaining to the same subject and having large variations is modeled as a manifold, and the distance between the two manifolds is computed in Manifold-Manifold Distance (MMD) [87, 92, 93] approach. Manifold can be considered as an extended subspace and can be modeled with the collection of local linear models due to the fact that the local linearity holds across the global nonlinear manifold. Consider two manifolds, M_1 and M_2 :

$M_1 = \{C_u : u = 1, 2, 3, \dots, m\}$

$M_2 = \{C'_v : v = 1, 2, 3, \dots, n\}$

Where u and v are the u-th and the v-th component subspaces of manifold M_1 by C_u and M_2 by C'_v, respectively, and m and n are the number of a local linear subspace in M_1 and M_2, respectively. The distance between the two manifolds M_1 and M_2 is computed using equation (2.17)

$$d(M_1, M_2) = \sum_{u=1}^{m} \sum_{v=1}^{n} f_{uv} d(C_u, C'_v),$$

$$\text{such that } \sum_{u=1}^{m} \sum_{v=1}^{n} f_{uv} = 1, f_{uv} \geq 0 \qquad (2.17)$$

From the above equation, it can be seen that the MMD computation has three essential components:

i. Local linear model construction C_u and C'_v (component subspaces)
ii. Subspace to Subspace Distance measurement $d(C_u, C'_v)$ (local model distance) ;
iii. Global integration of local distances; f_{uv} (choice of weights).

2.4.2.4 Manifold Discriminant Analysis (MDA)

Manifold Discriminant Analysis (MDA) [87, 94] is a discriminative learning method that enhances the compactness of local data within each manifold and maximizes the margin of manifolds. The two key points involved in the computation of MDA are as follows:

i. Local Linear Model: An effective clustering method is employed to extract the cluster set for each manifold, and each cluster is a local liner model. This local liner model characterizes the margin between the global nonlinear manifolds.

ii. Discriminative learning: MDA maps the multiclass manifolds into an embedding space by learning a linear discriminant function.

The main aim of obtaining the maximum separability between-class and to enhance the compactness within-class can be obtained from the following equations.

$$S_w = \sum_{p,q} \| u^T x_p - u^T x_q \|^2 \, w_{p,q} = 2u^T X(D - W) X^T u \qquad (2.18)$$

$$S_b = \sum_{p,q} \| u^T x_p - u^T x_q \|^2 \, w'_{p,q} = 2u^T X(D' - W') X^T u \qquad (2.19)$$

$$\text{Maximize } J(u) = \frac{|S_b|}{|S_w|} = \frac{u^T X L_b X^T u}{u^T X L_w X^T u} \qquad (2.20)$$

Where $x_p \in C_{i,k}$ and $x_q \in C_{j,k}$ the local models, D and D′ are the diagonal matrices with $d_{p,p} = \sum_q w_{p,q}$ and $d'_{p,p} = \sum_q w'_{p,q}$ diagonal elements. L_w=D-W and L_b=D'-W' are the Laplacian matrices corresponding to S_w and S_b, respectively.

2.4.2.5 Sparse Approximated Nearest Point (SANP)

The Sparse Approximated Nearest Point (SANP) to compute the between–set distance can be defined as a pair of nearest points on the image sets, which can be sparsely approximated by the sample images of the respective individual set. SNAPs points should satisfy the following constraints:

i. The two points should have a small Euclidean distance between them.

ii. Each of the two points should be able to be approximated by a sparse combination of sample images in the corresponding image set.

To obtain the SANPs of the two image sets, consider the affine hull models of two image sets (μ_m,U_m) and (μ_n,U_n) corresponding to the data matrices X_m and X_n. The process is formulated as follows;

$$E_{v_m,v_n} = \| (\mu_m + U_m v_m) - (\mu_n + U_n v_n) \|_2^2 ,$$

$$P_{v_m,\alpha} = \| (\mu_m + U_m v_m) - X_m \propto \|_2^2 ,$$

$$R_{v_n,\beta} = \| (\mu_n + U_n v_n) - X_n \beta \|_2^2 \quad (2.21)$$

$$\min_{v_m,v_n,\alpha,\beta} E_{v_m,v_n} + \lambda_1 \left(P_{v_m,\alpha} + R_{v_n,\beta}\right) + \lambda_2 | \propto |_1 + \lambda_3 |\beta|_1 \quad (2.22)$$

where E_{v_m,v_n} is the distance between images *m* and *n*, the individual fidelities between these two points, and their sample approximations are preserved by $P_{v_m,\alpha}$ and $R_{v_n,\beta}$ respectively, $| \propto |_1$and $|\beta|_1$ are used to enforce the approximations to be sparse.

CHAPTER 3:

GENERATION OF GU-RGB-D DATABASE

&

PRELIMINARY STUDIES

Face recognition [23] is one of the prominent areas of research as the acquisition of the facial trait is a non-intrusive, easily obtainable, and convenient biometric trait compared to the other biometric traits like iris, voice, gait, etc. The 2D facial images have well-defined roots in the world of biometric research due to the low cost of its acquisition system and wide availability. But, the 2D face recognition system faces its limitations when it comes to mostly illumination and pose variation [95]. In order to overcome these shortcomings of 2D recognition, the 3D recognition system captured the market due to high-security concerns from the local to the defense level. However, research in 3D biometric was an expensive task as the expense of system requirements for acquiring 3D images was very high and time-consuming [9] until the development of an efficient, low-cost RGB-D Kinect camera. This system provides 2D RGB images and depth information, i.e., distance from each object pixel to the sensor [28]. The essential requirement of any research problem is a set of data needed to train a system (such as a machine learning model) and perform various analyses to quantify the system's performance.

In biometrics, databases are essential in developing reliable and robust recognition systems. In addition, they serve as the standard platform for evaluating the various state–of–the–art recognition algorithms [96]. Considering these facts, a large number of face databases are developed to serve the two primary purposes [31]:

i. To test the robustness of the face recognition algorithms for single or multiple variations (Yale Face Database B [97] is one example of this where the database has 405 viewing conditions, i.e., 9 pose and 45 illumination conditions).
ii. To help in the development of face recognition algorithms for a particular data modality (for example, the Honda/UCSD [98, 99] video frame database was developed for video-based face recognition problems.

Some of the face databases used for different analysis and study purpose in the literature includes: FERT database [100], FRGC database [101], AR database [102], Olivetti Research Lab (ORL) database [103], pose-illumination-expression (PIE) database [104], labeled face in wild (LFW) [105], surveillance cameras face (SCface) database [106], MOBIO (mobile

biometry) database [107], etc. A more detailed list of the face databases and their descriptions can be found in [108].

There is a wide collection of 2D face databases, whereas the number of 3D face databases is relatively less in literature. Most of the existing 3D face databases like ND-2006, FRGC, BJUT-3-D, GavabDB, UMB-DB, etc., have used a high-quality laser scanner to acquire the data. These high-quality laser scanners provide very accurate details of the face, but they have a high acquisition time and hence need careful cooperation from the subject. On the other hand, the databases captured using high-quality stereo imaging systems such as BU-3-DFE, Texas 3-DFRD, XM2VTSDB, etc., also give similar performance accuracy as that of the laser scanners. A brief description of some of the 3D face datasets captured using laser scanners, and stereo imaging systems in the literature are as follows:

- ND-2006 [109] is a collection of 13450 images captured from 888 subjects. A maximum of 63 images was scanned per subject, and it has 6 different types of expressions, i.e., neutral, happiness, sadness, surprise, disgust, and other.

- FGRC [101] database is one of the widely accepted standard database to evaluate the 3D face recognition algorithms. This database is divided into FRGC v 1.0, where the training set has 943 images obtained from 273 subjects, and FRGC v 2.0, where the training set contains 4007 images of 466 subjects and additional expressions.

- BJUT-3D [110] database is a Chinese database comprises of 500 subjects (an equal number of male and female subjects). This database is generated using CyberWare 3030 RGB/PS laser scanner, which stores texture information along with the shape.

- GavabDB database [111, 112] has a total of 549 three-dimensional images/mesh surfaces corresponding to 61 subjects. Each subject has 9 variations, i.e., 2 frontal, 4 rotated images without expressions, and 3 frontal with different expressions. This

database has captured intrinsic variations like pose, expressions, and extrinsic variations in changing background, light effects, etc.

- UMB-DB database [113] is a collection of 1473 2D, and 3D front images scanned from 143 subjects. There are minimum 9 acquisitions per subject, including 3 neutral expressions, 3 expressions, i.e.smile, bored and angry, and 3 face occlusion with different objects like hat, scarf, and hands. This database also has occlusion such as eyeglasses, holding phones, partially occluded by the hair, and other miscellaneous objects.

- BU-3DFE database [114] has a total of 2500 3D scans and 2D texture information of 100 subjects obtained using the stereo photography technique. This database has 6 variations of expressions, i.e., happiness, anger, sadness, surprise, fear, and disgust.

- Texas 3DFRD, i.e., Texas 3D Face Recognition Database [115], is a collection of 1149 pairs of face texture information and scanned images of 118 subjects. The said database is generated using an MU-2 stereo imaging system at a high spatial resolution of 0.32 mm.

More details about the 3D face database acquired using 3D scanners, or stereo photographic techniques are provided in Table 3.1. As mentioned earlier, obtaining the 3D database was an expensive and time-consuming task until the development of the Kinect camera. The details regarding the Kinect camera based 3D databases are organized in Table 3.2. A brief description of the Kinect camera based databases and the feature extractions are discussed as follows:

- Rui Min et al. [31] have generated a 3D database based on the Kinect sensor having 52 subjects captured over two sessions for 2D, 2.5D, 3D, and video. Here recognition rates are calculated for 2D, 2.5D, and 3D-based face data using standard face

recognition techniques like PCA, LBP, SIFT, LGBP, ICP, and TPS, and also RGB and Depth images were fused using score-level fusion.

- Ajmera et al. [56] have computed CRR based on modified SURF descriptors and image enhancement techniques and filters like adaptive histogram equalization, NLM filter, etc., for their internal database. The internal database (IIT-K) is a collection of 100 male and female subjects captured at 0º, 15º, 30º, 45º, 60º, 75º, and 90º pose angles. These results are compared with Eurecom and Curtin face database. The authors have also performed the study based on scored level fusion.

- R.I. Hg et al. [116] have developed an RGB-D Face database (VAP database) of 31 subjects containing 1581 RGB images (and their depth images). The said database has 17 different pose variations and facial expressions captured using Kinect sensor. The capturing process was repeated three times per subject. Here the authors have developed a face detection protocol using the curvature analysis technique and reported its performance on the VAP database.

- Gaurav Goswami et al. [59] had generated an IIIT-D RGB-D face database of 106 subjects with multiple images per subject. The number of images per subject varies from 11 to 254 images making the total count as 4605 images. The authors have also proposed an algorithm for 3D face recognition, which involves computation of entropy map and visual saliency map followed by HOG descriptor for feature extraction and the use of Random Decision Forest (RDF) classifier for establishing identity. The algorithm was tested for IIIT-D and EURECOM databases.

- Merget, Daniel et al. have generated a Face-Grabber database [117] based on the facial expressions of 40 subjects. The database is captured using Kinect v2 and consists of 67,159 frames of color and depth images. It has six emotion variations, i.e., sadness, disgust, fear, happiness, anger, and surprise. It also consists of a

sequence of six-second head scan with a neutral expression and random facial expression frames of ten seconds.

- RAP3DF [118] database is generated by Rafael Alexandre Piemontez et al. and has a collection of 267 samples obtained from 64 subjects. This database has the collection of two groups of images of the participants. The first group has captured images of frontal position without any expression, while the second group has 6 different facial expressions, i.e., surprise, happiness, fear, anger, sadness, and disgust. Further, each sample in the database has three types of images, i.e., an infrared image, a visible image, and a depth image.

- Seyed Muhammad Hossein Mousavi et al. [58] have generated a Kinect v.2 based IKFDB, i.e., Iranian Kinect face database consisting of 40 subjects. The database has a collection of more than 100000 recorded color and depth frames. Seven main expressions are captured between frames 150-250. In addition, pitch and yaw action has been considered in the databases to resolve the recognition problem from any angle. The authors have used HOG descriptor for feature extraction followed by Support Vector Machine (SVM) [119], Multi-Layer Neural Network (MLNN) [120], and Convolutional Neural Network (CNN) [120] algorithms for classification purpose.

Table 3. 1: Existing Laser scanner based and stereo imaging-based 3D facial databases [31, 121–123]

Databases	Scanner	Number of subjects	Total Images	Pose Variations	Occlusion	Expressions	Reference
3DRMA	Structural Light-based 3D face scanner	120	360	Frontal, up/down, limited left/right	-	-	[124]
FUS	Minolta Vivid 700	37	222	-	-	Neutral, smile, scared, angry, squint, frown	[125]
GavabDB	Minolta Vi-700 laser range scanner	61	549	Frontal, left profile, right profile, looking up, looking down	-	Neutral, smile, accentuated laugh, random gesture	[111]
FRGC v 1.0	Minolta Vivid 3D scanner	273	943	-	-	-	[101]
FRGC v 2.0	Minolta Vivid 900/910 3D scanner	466	4007	-	-	Neutral, surprise, happy, puffy cheeks, anger,	[101]

						frown	
BU3D-FE	Stereo photography, 3DMD digitizer	100	2500	-	-	Neutral, angry, fear, sadness, disgust, happiness, surprise	[114]
CASIA	Minolta Vivid 910 range scanner	123	4059	Frontal, tilt left and right 20◦−30◦, up and down, 20◦−30◦, left and right 20◦−30◦, left and right, 50◦−60◦, left and right 80◦−90◦	-	Neutral, smile, eyes closed, anger, laugh, surprise	[126]
FRAV3D	Minolta Vivid 700 red laser light scanner	105	1696	Frontal looking up and down in X-axis direction, 25◦ Y-axis right turn, 5◦ Y-axis left turn, small and severe Z-	-	Neutral, open mouth, smile, and gesture	[127]

				axis right turn			
ND 2006	Minolta Vivid 910 range scanner	888	13450	-	-	Neutral, surprise, sadness, disgust, happiness, undetermined	[109]
MSU	Minolta Vivid 910 range scanner	90	533	-	-	Neutral , Smile	[128]
ZJU-3DFED	InSpeck 3D MEGA Capturor DF	40	360	-	-	Neutral, smile, surprise, sad	[129]
Bosphorus	The Inspeck Mega Capturor II 3D scanner	105	4652	13 yaw, pitch & cross rotation	Hair, mouth, eye, eyeglasses	34 expressions	[123]
University of York	Stereo vision 3D camera	350	5250	Frontal, up, down	-	Neutral, eyes closed, eyebrows	[130]

						raised, happy, anger	
BJUT -3D	CyberWare 3030RGB/PS laser scanner	500	-	-	-	-	[114]
Texas 3-D	MU-2 stereo imaging system	118	1149	-	-	Neutral, smile/talk with open/closed eyes and/or open/closed mouth	[115]
UMB-DB	Minolta Vivid 900 laser scanner	143	1473	-	Scarf, hat, hands in random positions, eyeglasses, hair, miscellaneous	Neutral, smile, angry, bored	[113]
3D TEC	Minolta scanner	214	428	-	-	Neutral, smile	[131]

Table 3. 2: Existing Kinect based 3D facial databases

Database	No. of Subjects	Sessions	Variations				Reference
			Angles/Poses	Occlusion	Expressions	Illumination	
EURECOM	52	2	Neutral face, Right, Left	Paper & Hand on face, Sunglasses	Smiling, Mouth open	Single pose	[31]
VAP	31	3	combination of 17 vertical and horizontal face poses	-	Smile, Sad, Yawn, Anger	-	[116]
Curtin Face	52	1	various poses	Sunglasses	Various expressions	Yes	[9]
IIIT-D	106	1	various poses	-	Various expressions	Yes	[59]
IIT-K	100	1	0°,15°,30°,45°,60°,75°, 90°	-	-	Yes	[56]

VT-KFER	32	1	Frontal, Right, Left	-	6 random facial expressions	-	[57]
Face-Grabber	40	1	-	-	Sadness, Disgust, Fear, Happiness, Anger, and Surprise; Neutral & Random facial expression	-	[117]
KaspAROV RGB-D video face database	108	1	Yes	-	Yes	Yes	[132]
RAP3DF	64	1	Arbitrary poses	-	Happiness, Surprise, Fear, Sadness, Anger, Disgust.	-	[118]
IKFDB	40	1	Pitch and Yaw action	-	7 Facial expressions	-	[58]

3.1 Contributions

With the literature background of the various 3D facial databases, we have generated the Kinect-based GU-RGB-D database having variations in pose/angles, expression, and occlusion to cover all possible variations under one database. This database also has the scope to study the effect of controlled and uncontrolled environments on the RGB-D database. Further, the preliminary study on the GU-RGB-D and EURECOM database has been performed using Principle Component Analysis (PCA) based feature extraction algorithm. The contributions can be summarized as follows:

- Generation of Kinect-based GU-RGB-D database having variation in pose /angles, expressions, occlusions, illumination and captured under controlled and uncontrolled environmental conditions.
- Preliminary study on GU-RGB-D and EURECOM database using Principle Component Analysis (PCA) algorithm as a feature extraction method.
- Presented study based on score level fusion for RGB and depth images of both the databases.
- Presented study based on Pixel level image fusion of RGB and depth images.

The rest of the chapter is distributed as follows; section 3.2 gives the entire GU-RGB-D database generation process under which sub-section 3.2.1 describes the 3D imaging setup, sub-section 3.2.2 provides the basic working principle of the Kinect camera, and subsection 3.2.3 describes the data acquisition protocol. Next, section 3.3 describes the preliminary study on the GU-RGB-D and EURECOM database with two different methodologies (based on score level fusion & Pixel level image fusion). Then, sub-section 3.3.1 provides the study based on score level fusion (methodology 1), and the results w.r.t the GU-RGB-D and EURECOM databases are given in the subsections of 3.3.2. Finally, the study based on Pixel level image fusion is described in subsection 3.3.3, and its results are discussed in a subsection of 3.3.4.

3.2 Generation Of GU-RGB-D database Using Kinect Camera

This section gives the details of the GU-RGB-D database generation process, including the image setup used for capturing images under the controlled and uncontrolled environment, the basic principle of Kinect camera, and the data acquisition protocol of the GU-RGB-D database.

Fig 3. 1: Pictorial view of the 3-D imaging set up for a generation of GU-RGB-D database

3.2.1 3D Imaging Setup

The 3D biometric imaging laboratory has been set up at our workplace in a dark room and is equipped with the Xbox 360 Kinect depth camera from Microsoft, QTH light sources, and a computer system. The laboratory is facilitated with controlled and uncontrolled environmental conditions in order to capture the images in the same manner. The Xbox 360

Kinect depth camera consists of two parts; the RGB camera used to capture 2D image information and an infrared projector combined with a monochrome CMOS sensor, which acquires depth information, which is the distance between subject and sensor, i.e., depth. The pictorial view of the 3-D imaging setup for the generation of the GU-RGB-D database is given in Figure 3.1.

Fig 3. 2: Kinect camera-based 3D imaging laboratory: Controlled condition

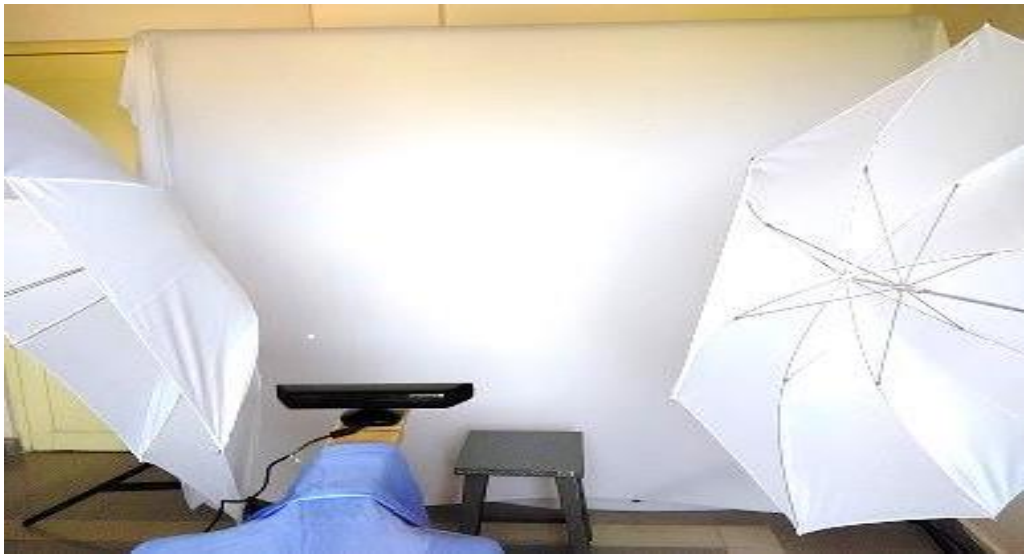


Fig 3. 3: Kinect camera-based 3D imaging laboratory: Uncontrolled condition (window open)

The Kinect sensor is mounted parallel to the ground at the height of 1.5 meters and approximately at a distance of 1.25 meters from the subject. A white muslin cloth backdrop is placed behind the subject to maintain a uniform background and equal illumination on all sides. To capture the data under the controlled environmental condition (Figure 3.2), the two QTH light sources of 600 watts are placed at an angle 45^0 normal to the subject's position. The direct projection of light on the subject is avoided with the white muslin cloth umbrellas mounted in front of light sources. The uncontrolled environmental conditions (Figure 3.3) were obtained by exposing the subject to the ambient light by opening the windows while capturing the images.

3.2.2 Basic Principle of Kinect Camera

Microsoft Kinect camera is a light-coded range camera that can estimate the 3D geometry of an acquired seen. It consists of an RGB camera, an infrared (IR) emitter/projector, an infrared (IR) camera/sensor, and a multi-array microphone. RGB images are captured by the RGB camera directly, whereas the IR projector and IR sensor act together to capture a depth image. The IR projector projects an IR light pattern (predesign pattern of spots) on the scene at the wavelength of 830 nm, and the reflected pattern is captured by the IR camera sensor working on the same wavelength. This captured pattern is further compared with the known pattern to produce the disparity map I_D having disparity value '*d*' at each point. This disparity map is further used to compute the depth map via the active triangulation method, which serves as the basic principle of the system [31][133]. The detailed process, along with the mathematical formulations, is available in [134] and [135].

3.2.3 3D Image Acquisition Protocol for GU-RGB-D Database Generation

After proper calibration of the camera, the image acquisition was performed to maintain and confirm the protocol of the experiment. The highest resolution for Kinect color sensor (1280x960) and Kinect depth sensor (640x480) has been selected among the resolution parameters while capturing the database. The GU-RGB-D Database is collected in two

sessions under controlled and uncontrolled environmental conditions for our institution students and staff. The image capturing protocol is designed as shown in Figure 3.4. In the database, we have introduced a total of eight variations per subject in the image acquisition process, having variation in pose (-90^0, -45^0, 0^0, $+45^0$, $+90^0$), variation in expressions (smile, eyes closed), and occlusion (paper was used to cover the vertical half part of the face).

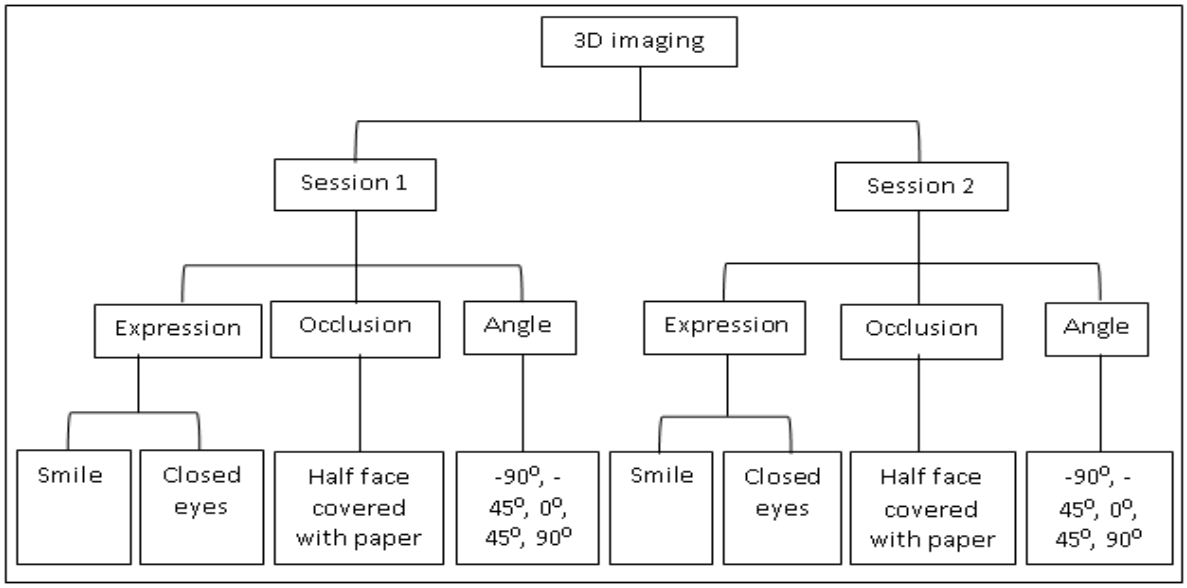


Fig 3. 4: 3D image acquisition protocol

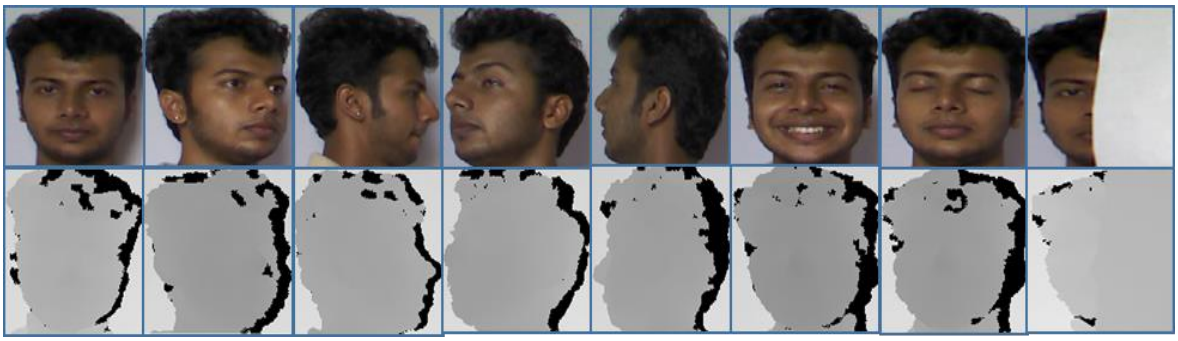


Fig 3. 5: Images from GU-RGB-D database

The GU-RGBD database is collected for 64 subjects, out of which 15 are females and 49 are males from various age groups captured in two different sessions. Session 1 as controlled environmental conditions and session 2 as un-controlled environmental conditions. A total of 16 images were captured for every subject, i.e., eight RGB and eight depth images, in every session. The sample images of the database are shown in Figure 3.5. The database has a total of 2048 images, i.e., 64(subjects) x 32(images per subject) = 2048. The details of the GU-RGBD database as per the designed acquisition protocol are given in Table 3.3.

Table 3. 3: Details of GU-RGB-D database

Facial Variants		RGB Images				Depth Images			
		Subjects	**Session**	**Samples**	**Total Images**	**Subjects**	**Session**	**Samples**	**Total Images**
Front	0º	64	2	1	128	64	2	1	128
Pose/Angles	+45º	64	2	1	128	64	2	1	128
	-45 º	64	2	1	128	64	2	1	128
	+90 º	64	2	1	128	64	2	1	128
	-90 º	64	2	1	128	64	2	1	128
Expression	Smile	64	2	1	128	64	2	1	128
	Eyes Closed	64	2	1	128	64	2	1	128
Occlusion	Paper on face	64	2	1	128	64	2	1	128

3.3 Preliminary Study On RGB-D Databases

We have experimented on the in-house generated GU-RGB-D database and the publicly available EURECOM database for preliminary study and analysis. The said databases have the collection of the multimodal facial images captured using a Kinect camera of 64 subjects (15 females, 49 males) and 52 subjects (14 females, 38 males), respectively. The databases are captured in two sessions and have facial images with different facial expressions, different lighting conditions, and occlusions: neutral, open mouth, smile, left profile, right profile, occlusion mouth, occlusion eyes, occlusion paper, and light on condition.

The fusion of various biometric traits is one of the security enhancement schemes in which two or more biometric modalities/traits are fused [136]. Kinect-based imaging setup captures 2D RGB images as well as depth images. The fusion of these two modes would ultimately

have the upper hand in the performance of biometric systems leading to higher security. There are four basic fusion schemes: sensor level, feature extraction level, matching similarity score level, and decision level fusion [136]. Fusion at the sensor level is done by combining the raw output of the sensors themselves and can be applied under very limited conditions. Fusion at the feature extraction level is quite restricted because if features are homogenous, then they can be combined into a single feature vector, but if inputs are not homogenous, then it becomes challenging to combine them. Fusion in the most multimodal biometric system is often implemented at the matching score level, as it is simple to access and combine the scores generated by the matching module. Also, decision-level fusion can be implemented, but here a small amount of information is available, i.e., merging of multiple accept/reject output into a single decision; therefore, it is not a very accurate fusion strategy alone. However, decision-level fusion is more effective when it is combined with other fusion techniques. Here in this study, we have implemented matching score level fusion of the scores generated using PCA [61][62].

Using the PCA algorithm, one can express the large 1-D vector of pixels from a 2-D image into a reduced principle component of the feature space called eigenspace projection. Eigenspace is calculated by identifying the eigenvectors of the covariance matrix derived from a set of facial images (vectors). The eigenvectors corresponding to nonzero eigenvalues of the covariance matrix produce an orthonormal basis for the subspace within which most image data can be represented with a small amount of error. The eigenvectors are sorted from high to low according to their corresponding eigenvalues. The eigenvector associated with the largest eigenvalues reflects the greatest variance in the image, and the smallest eigenvalues are associated with the least variance.

3.3.1 Methodology 1: Study Based On Score Level Fusion

The framework of methodology 1 can be seen in Figure 3.6. The two databases were cropped manually to obtain the region of interest in the preprocessing stage. This is followed by feature extraction using the Eigen face-based Principle Component Analysis (PCA) algorithm, one of the well-known dimension reduction techniques.

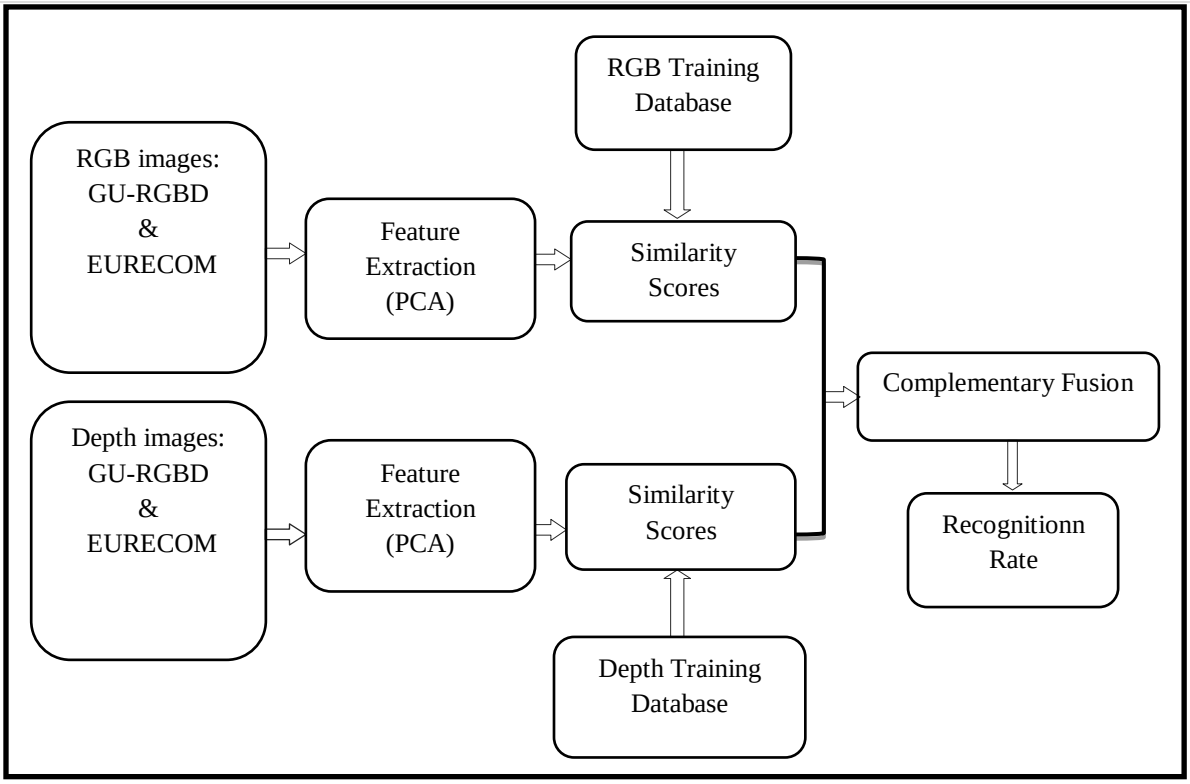


Fig 3. 6: Framework of implemented methodology 1 and fusion system

The RGB and Depth datasets have different variations such as pose variation, smile, occlusion, and various angles. The features from each variation are extracted using the PCA algorithm, and the similarity scores are obtained by testing them against the RGB and depth training dataset of a fixed variation, respectively. Finally, the similarity scores are fused together using complementary fusion in order to obtain the enhancement in recognition rate. In the complementary fusion approach, the weightages of the filter for RGB and depth images are varied based on the value of 'α' as given in Equation.3.1.

$$F_{score} = (1- \alpha)*RGB_{score} + \alpha*Depth_{score} \quad (3.1)$$

The complementary fusion is selected in this study as it implicitly implements the sum/mean fusion at $\alpha = 0.5$ and weighted fusion otherwise.

3.3.2 Results and Discussion

We have considered the data of only 25 subjects from the EURECOM database out of available 52 subjects for computing the preliminary study and the analysis. Similarly, data of 25 subjects from the GU-RGBD database out of available 64 subjects are engaged separately for processing the results. The images from both databases were cropped and downscaled to 96 x 96 dimensions to reduce computational time. The set of neutral images (front face) from session 1 of the respective databases are used as the training set. And the rest of the database, including all the variations from session 1 and session 2, are used for testing. The protocol detail can de be seen in Table 3.4.

The features are extracted from the training and the testing datasets using Principle Component Analysis (PCA) to compute similarity scores. The computed scores are used to calculate the recognition rates of different variations against the respective training set. The complementary fusion is implimented as per Equation (3.1) for five different values of $'\alpha'$ (α = 0.3, 0.4, 0.5, 0.6, 0.7), i.e., by varying the weightage of the RGB or Depth scores. When α = 0.5, equal weightages are given to RGB and Depth Scores. For α = 0.3 & 0.4, the scores of the RGB component is at the higher weightage as compared to the depth and its other way round when α = 0.6 & 0.7. The computed recognition rates for RGB, Depth, and Fusion at Rank - 5 are tabulated in Tables 3.5 & 3.6.

It can be seen that engaged fusion methodology has enhanced the recognition rate as compared to the recognition rate obtained for RGB and depth independently. It can be further noted that by using complementary fusion, i.e., by changing the weightage of either RGB or Depth, the system performance can be improved to a higher level of security. This can be considered as a method to compensate for the system's performance degradation due to the low quality of either depth or RGB images.

Table 3. 4: Evaluation protocol

Pose/Variation		Session	Total Subjects
GU-RGB-D	**EURECOM**		
Training Dataset			
0° pose/Front	Neutral	Session I	25
Testing Dataset			
0° pose/Front	Neutral	Session I	25
45° pose	Illumination	Session I & II	25 each
90° pose	Occlusion by Sunglasses	Session I & II	25 each
(-45)° pose	Occlusion by Hand	Session I & II	25 each
(-90)° pose	Smiling	Session I & II	25 each
Paper on face	Open mouth	Session I & II	25 each
Eyes closed	Occlusion by Paper	Session I & II	25 each
Smile	Left	Session I & II	25 each
-	Right	Session I & II	25 each

3.3.2.1 Evaluation of EURECOM Database

The recognition rates computed for RGB, depth, and fusion using the described methodology for the EURECOM database are tabulated in Table 3.5. Some of the major observations from Table 3.5 are as follows:

- When the 'occlusion by hand' variation of session 1 is tested against the training set, it gives a higher recognition rate for fusion, i.e., 96% for the combination of depth and RGB scores for $\alpha = 0.3$ & 0.4. A similar trend can be seen in the case of the variation 'smiling' of session 1 and session 2 and the variation 'mouth open' of session 2.

- Further, 'occlusion by sunglasses' gives better performance 96% when it is having equal weightage for both RGB and depth. The variation 'neutral' of session 2 is, when tested against the training set, performs 84% for 0.4, 0.5, and 0.6 values of alpha.

- Individual recognition rates in the case of 'left' and 'right' variation in both the sessions are quite lower as only the partial face triangle is available for the computation. However, complementary fusion has improved its performance to some extent.

Table 3. 5: Recognition Rates computed at Rank 5 for EURECOM database using a complementary fusion approach

Variations	RGB	Depth	Fusion				
			(0.3)	(0.4)	(0.5)	(0.6)	(0.7)
Session 1							
Neutral	-	-	-	-	-	-	-
Illumination	96	36	**96**	**96**	**96**	80	76
Occlusion by Sunglasses	80	56	**88**	**88**	**92**	76	72
Occlusion by Hand	88	52	**96**	**96**	**88**	84	76
Smile	100	68	**100**	**100**	96	96	92
Open mouth	96	92	**100**	**100**	**100**	**100**	**100**
Occlusion by Paper	36	24	**44**	**52**	**48**	**40**	**40**
Left	44	24	**44**	**48**	40	28	28
Right	32	20	**32**	**32**	28	28	**32**
Session 2							
Neutral	72	64	**80**	**84**	**84**	**84**	**76**
Illumination	64	64	**80**	**80**	**76**	**76**	**72**
Occlusion by Sunglasses	56	64	**68**	**68**	**72**	**72**	**76**
Occlusion by Hand	60	56	**60**	**68**	**64**	**64**	**72**
Smile	76	64	**80**	**80**	**76**	**76**	72
Mouth Open	56	52	**72**	**72**	**68**	**60**	52
Occlusion by Paper	36	36	**48**	**40**	**40**	**48**	**56**
Left	20	16	**32**	**28**	**24**	**24**	**28**
Right	28	28	**32**	**32**	**32**	**32**	**40**

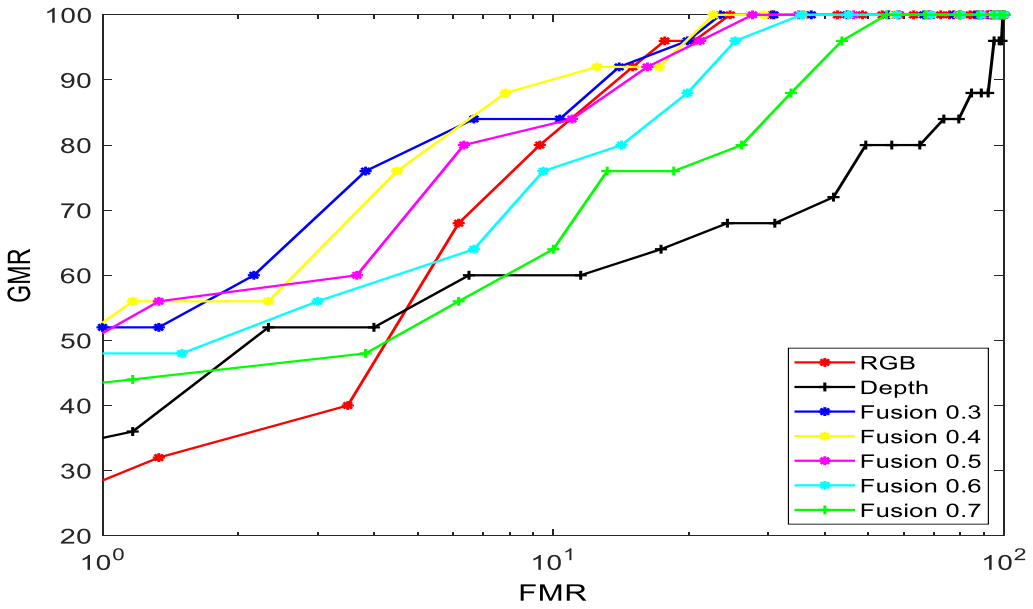


Fig 3. 7: Receiver Operating Curve **(**ROC) plot demonstrating the performance on variation 'smile' (session 1) using complementary fusion for α = 0.3 & 0.7

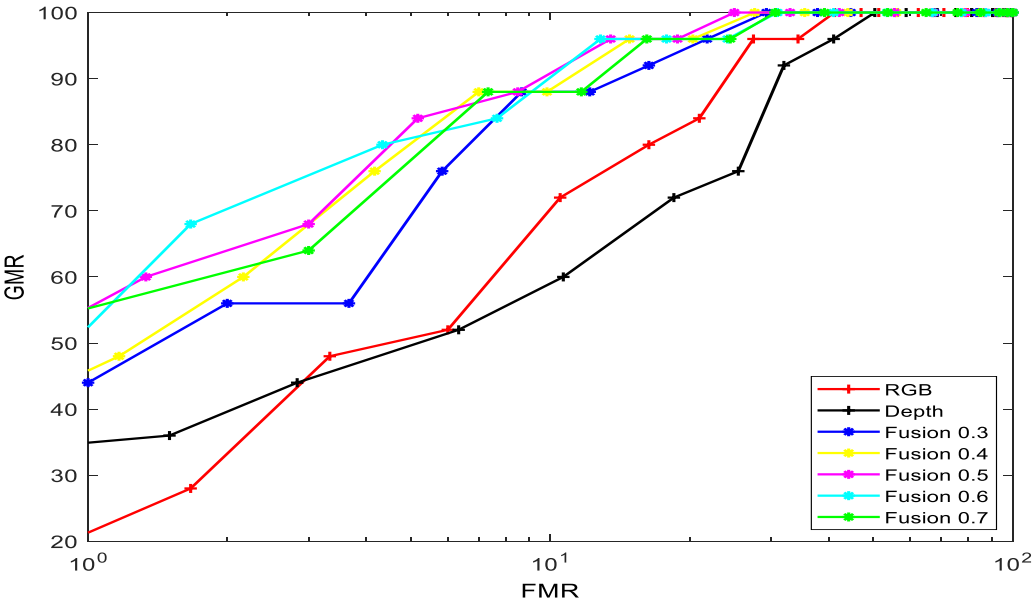


Fig 3. 8: Receiver Operating Curve **(**ROC) plot demonstrating the performance of variation 'mouth open' (session 1) using complementary fusion for α = 0.3 & 0.7

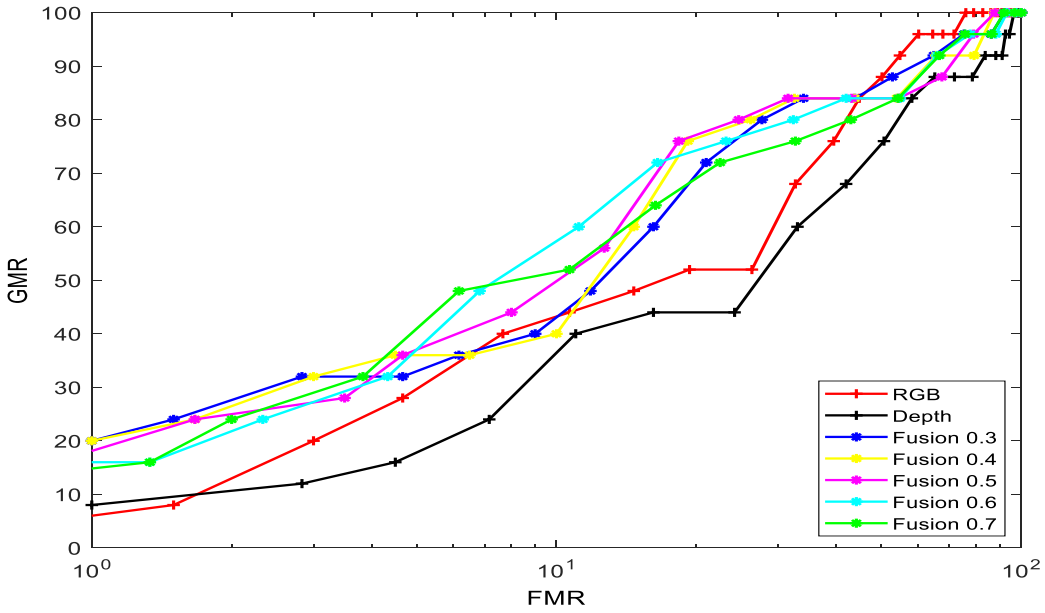


Fig 3. 9: Receiver Operating Curve **(**ROC) plot demonstrating the performance of 'neutral face' (session 2) using complementary fusion for α = 0.3 & 0.7

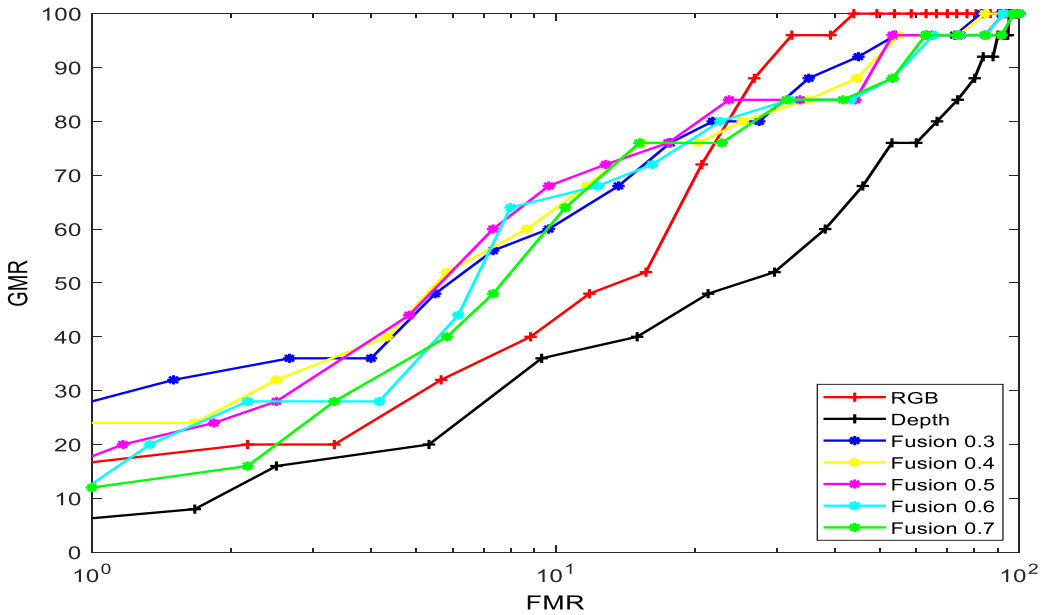


Fig 3. 10: Receiver Operating Curve **(**ROC) plot demonstrating the performance of variation 'illumination' (session 2) using complementary fusion for α = 0.3 & 0.7

The ROC curves of FMR (False Match Rate) v/s GMR (Genuine Match rate) for the different variations and various α values are plotted to represent the verification rates. From these ROC curves, it can be noted that varying the 'α' parameter fusion marks better performance. For variations like a smile from session 1 (Figure 3.7), mouth open (Figure 3.8), neutral face from session 2 (Figure 3.9), illumination variation (Figure 3.10) has better performance as compared to the individual RGB and Depth performance. As in these variations, the full facial triangle is available for computation when compared with the full face triangle of the training dataset.

The fusion method's performance is lower for variations like 'paper occlusion', 'left', and 'right as in these cases, only 50% of the face or even less than that is available in the testing set. But the complementary fusion of the RGB and the depth images have enhanced the performances compared to the individual RGB and Depth verification rates for these variations.

3.3.2.2 Evaluation on GU-RGB-D Database

Similar performance as that of the EURECOM datasets can be seen with the GU-RGB-D dataset with respect to complementary fusion, i.e., performance has been enhanced with the complementary combinations of RGB and Depth scores. The recognition rates computed for the GU-RGB-D database as per methodology 1 are presented in Table 3.6. Some of the observations based on Table 3.6 are as follows:

- The variation like ‘eyes close’ from session I and II and ‘front’ from session II has obtained maximum recognition rates of 100%, 84%, and 96 %, respectively.

- The performance improvement has also been noted in the pose/angle variations like 45º, -45º, 90º, -90º with the implementation of complementary fusion even though the base results of these variations are very low.

Table 3. 6: Recognition Rates computed at Rank 5 for GU-RGB-D database using a complementary fusion approach

Variations	RGB	Depth	Fusion				
			(0.3)	(0.4)	(0.5)	(0.6)	(0.7)
Session I							
Front							
45º	40	52	44	44	44	**52**	**64**
90 º	28	44	36	40	40	**48**	**48**
-45º	40	40	**48**	**48**	**48**	**44**	**48**
-90º	20	24	**24**	**32**	**32**	**40**	**36**
Smile	76	96	84	88	92	**96**	**96**
Eyes Close	84	96	96	96	**100**	**100**	**100**
Paper Occlusion	28	40	32	36	**40**	**52**	**48**
Session II							
Front	76	80	**88**	**92**	**92**	**96**	**92**
45º	44	40	**44**	**44**	**44**	**44**	**48**
90º	32	44	32	36	36	40	**48**
-45º	36	24	**40**	**40**	32	32	32
-90º	24	20	**24**	**24**	**24**	**24**	**24**
Smile	64	80	68	72	72	72	76
Eyes Close	64	64	**72**	**80**	**84**	**88**	**80**
Paper Occlusion	32	44	40	40	**52**	**56**	**56**

The verification rates for different variations of the GU-RGB-D database and for the different values of α are represented through the ROC curves of FMR (False Match Rate) v/s GMR (Genuine Match rate) and are given in Figures 3.11 – 3.14. It is observed from the figures that the performance of full face variations like smile (Figure 3.11), eyes close (Figure 3.12), front face (3.13), etc., are having better performance with the application of the fusion approach. The angular variations and occlusion (Figure 3.14) have low performance. In these cases, the full face is not available for computation; however, the performance increase is also noted with the application of the complementary fusion approach.

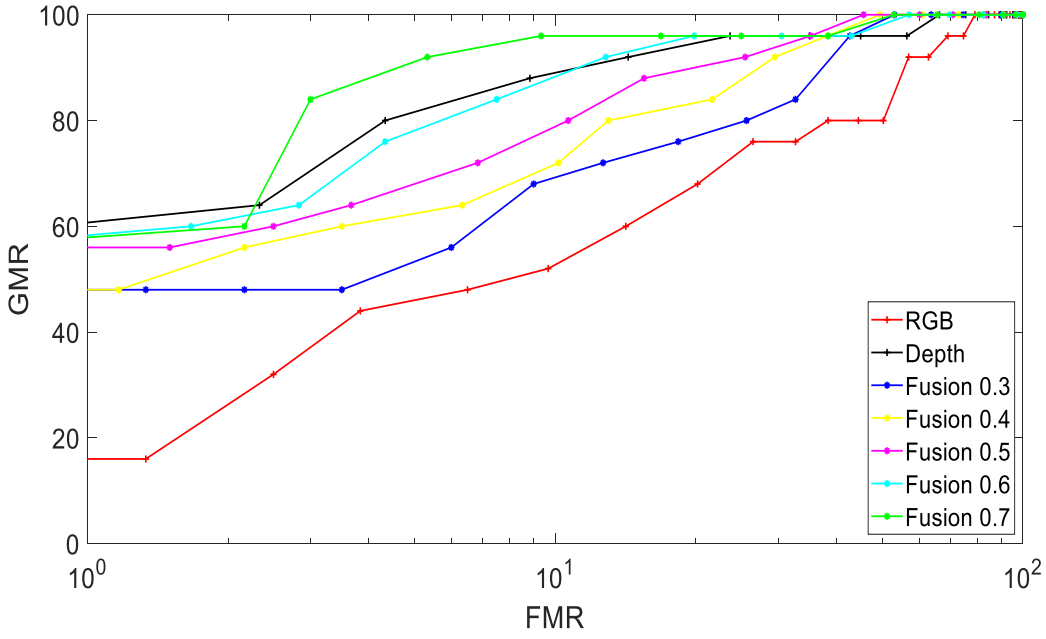


Fig 3. 11: Receiver Operating Curve **(**ROC) plot demonstrating the performance of variation 'smile' (session 1) using complementary fusion for α = 0.3 & 0.7

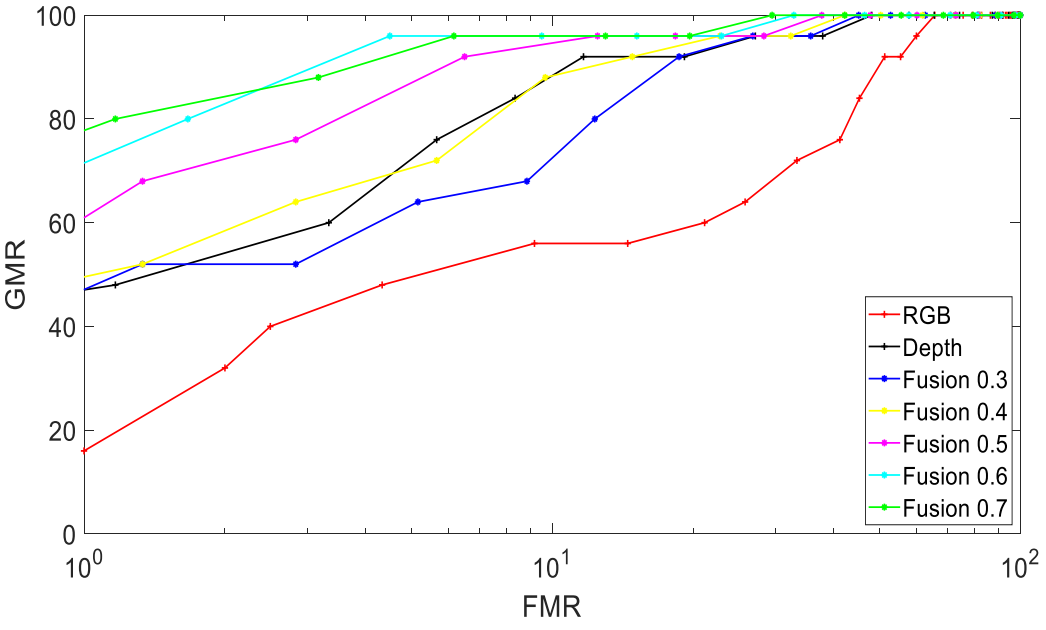


Fig 3. 12: Receiver Operating Curve **(**ROC) plot demonstrating the performance of variation 'eyes close' (session 1) using complementary fusion for α = 0.3 & 0.7

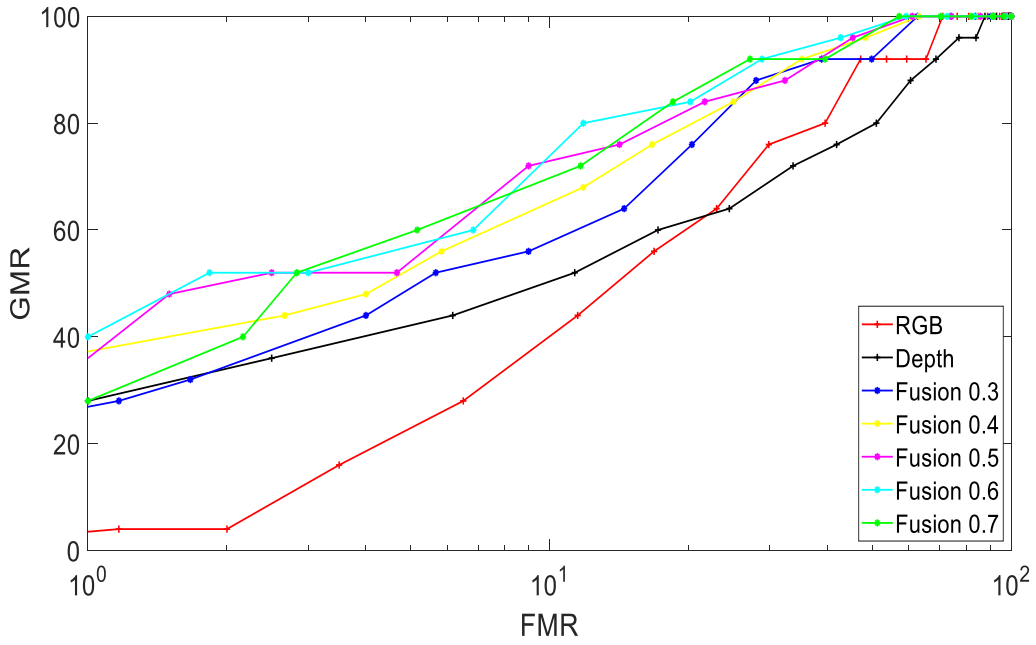


Fig 3. 13: Receiver Operating Curve **(**ROC) plot demonstrating the performance of 'front face' (session 2) using complementary fusion for α = 0.3 & 0.7

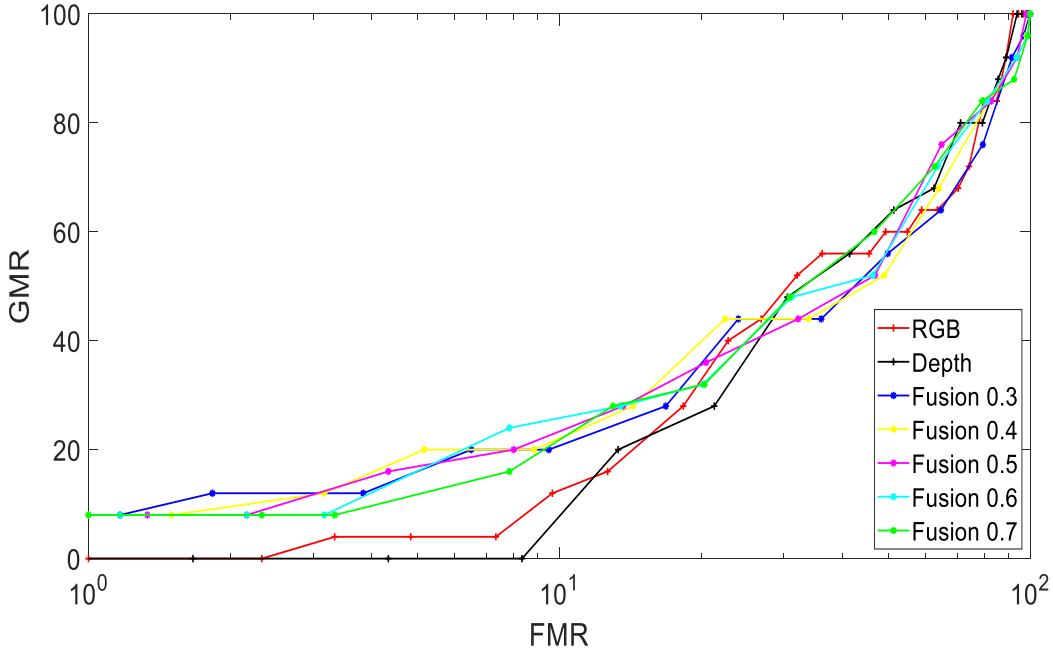


Fig 3. 14: Receiver Operating Curve **(**ROC) plot demonstrating the performance of 45º pose variation (session 2) using complementary fusion for α = 0.3 & 0.7

3.3.3 Methodology 2: Study Based On Image Level Fusion

Here in this study, the cropped images from GU-RGB-D (all 64 subjects) and EURECOM (all 52 subjects) database were taken as input to the gradient filter at the preprocessing stage. The RGB and the depth image outputs obtained from the gradient filter are fused using Pixel level image fusion, where both images' pixel intensities are fused. The gradient filter provides the directional change of image intensity and sharpness of the image [137]. The original image is generally convolved with the pre-defined filter to measure the intensity change of each pixel in the given direction to obtain the gradient image. As mentioned earlier, feature extraction is an essential task in image processing; thus, the features are obtained using PCA in this work. Finally, the recognition rates are computed for different variations in the said databases by calculating the similarity scores from the obtained features. The pictorial description of the workflow can be seen in Figure 3.15.

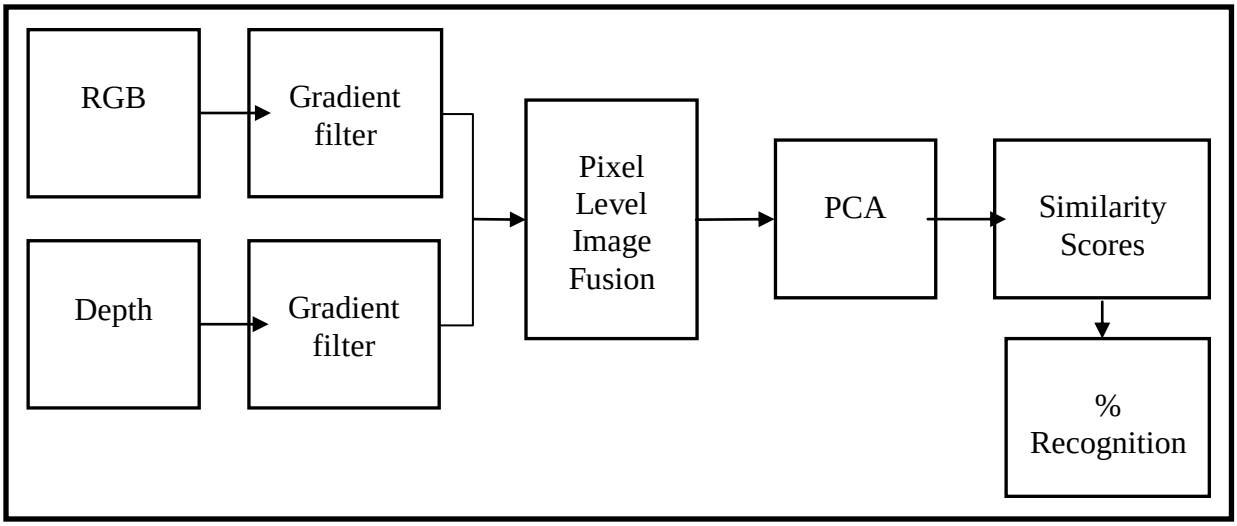


Fig 3. 15: Framework of implemented methodology 2 and fusion system

3.3.4 Results and Discussion

In this approach, we have engaged the entire databases to perform the experiment and to evaluate results, i.e., images of all 52 subjects from the EURECOM database and 64 subjects from the GU-RGB-D database. Similar to the previous approach (section 3.3.1), the cropped images from both databases are downscaled to 96x96 dimensions (to reduce computational

time). The training set is generated using the neutral images (front face) from session 1 of the respective databases. The rest of the database, including all the variations across the sessions (session 1 and session 2), are used as the testing datasets. The same protocol as mentioned in Table 3.4 is used in this approach (section 3.3.2). However, all the images (not only 25) from GU-RGB-D and EURECOM database are engaged to perform this study. Further, the RGB and the depth images obtained after the gradient filter application are fused together to enhance the overall performance, and the same can be seen in Tables 3.7 and 3.8.

3.3.4.1 Evaluation Of The EURECOM Database

Table 3.7 gives the recognition rate computed using the PCA algorithm for the EURECOM database as per the methodology described above. It is seen that the application of Gradient filter to the depth images has improved the performance of the system almost in all the cases precisely where the entire facial triangle is available for computation. This is due to the enhancement of the sharpness of the images with the gradient filter. Further, the fusion of RGB and the depth components have improved the results. The major observations wrt to the Table 3.7 are as follows:

- The base results of the smile variation in session 1 are 33.33 %. With the application of the gradient filter, the performance has increased to 88.24 %, and the maximum enhancement of 98.04% has been obtained. A similar trend is seen in the smile variation in session 2 and in the other places where the entire facial triangle is available.

- The variations with the face occlusions have also shown enhancement with respect to gradient and fusion approach. Occlusion by sunglasses of session 1 has reported the maximum recognition rate of 76.47% among the other occlusions, with the fusion method.

- The left and right face profile performance are marginal as the entire facial triangle is not available for the feature extraction with respect to the ‘front face’.

Table 3. 7: Recognition Rates computed at Rank 5 for EURECOM database using an Image-Level Fusion approach

Variations/Pose	Depth	Depth with application of Gradient filter	RGB-D Fusion
	Session 1		
Neutral	-	-	-
Illumination	52.94	**64.71**	**84.31**
Occlusion by Sunglasses	66.67	**72.55**	**76.47**
Occlusion by Hand	45.10	**47.06**	**60.78**
Occlusion by Paper	15.69	**31.37**	**29.41**
Smile	33.33	**88.24**	**98.04**
Mouth Open	54.90	39.22	49.02
Left	11.76	**13.73**	**17.65**
Right	11.76	**29.41**	**29.41**
	Session 2		
Neutral	25.49	**62.75**	**74.51**
Illumination	27.45	**64.71**	**74.51**
Occlusion by Sunglasses	27.45	**56.86**	**52.94**
Occlusion by Hand	37.25	**37.25**	**43.14**
Occlusion by Paper	13.73	**21.57**	**25.49**
Smile	43.14	**52.94**	**64.71**
Mouth Open	25.49	**25.49**	**27.45**
Left	9.80	**9.80**	**13.73**
Right	19.61	**25.49**	**21.57**

3.3.4.2 Evaluation on GU-RGB-D Database

Table 3.8 has provided the recognition rates computed for the GU-RGB-D database using the methodology as described in section 3.3.3. The recognition rate performance and the trends are similar to that of the trends seen in the case of the EURECOM database. However, the overall performance of the database is less as the database consist of more angular images.

Table 3. 8: Recognition Rates computed at Rank 5 for GU-RGB-D database using an Image-Level Fusion approach

Variations / Pose	Depth	Depth with application of Gradient filter	RGB-D Fusion
		Session 1	
Front	-	-	-
45 º	29.69	18.75	17.19
90 º	20.31	9.38	7.81
-45 º	26.56	23.44	20.31
-90 º	12.50	**17.19**	**17.19**
Smile	87.50	82.81	**93.75**
Eyes Close	81.25	76.56	**89.06**
Paper Occlusion	34.38	**48.44**	**57.81**
		Session 2	
Front	32.81	**42.19**	**78.13**
45 º	15.63	9.38	10.94
90 º	9.38	4.69	7.81
-45 º	9.38	**21.88**	**15.63**
-90 º	14.06	7.81	10.94
Smile	32.81	**39.06**	**79.69**
Eyes Close	39.06	**40.63**	**81.25**
Paper	25.00	17.19	**28.13**

The major observations from Table 3.8 are as follows:

- The application of the gradient filter has shown marginal improvement even in some cases where the entire facial triangle is available for evaluation. The performance degradation is mainly because of the multiple holes which are developed in the images during capturing, and the same has also been reported in the literature. However, with the fusion of the RGB component with the depth, the performance has been enhanced in both the sessions for variations with the entire face triangle.

- For the angular facial images, the performance with the fusion is marginal as the partial face is available for feature extraction and computations.

The issues of the low performance of the system due to the holes (missing pixels) in the depth images need to be resolved to enhance the performance and the reliability of the system. The necessary approach/attempt to resolve this issue has been designed, and the details of the same are described in chapter 4.

CHAPTER 4:

PRE-PROCESSING

&

FEATURE EXTRACTION METHODS

Research in 3D biometric was an expensive task as the expense of system requirements for acquiring 3D images was very high and time-consuming [3] until the development of an efficient, low-cost RGB-D Kinect camera. The captured 3D images have the proficiency in overcoming the limitations due to illumination variation and posing variation, which commonly affects the 2D imaging system [1]. This is mainly because of the inherent property associated with 3D faces, i.e., 3D camera systems can capture more spatial information than 2D systems in the form of depth images (distance from each pixel to the sensor) along with RGB images [2]. The Kinect camera has employed VGA resolution for capturing RGB images, and the depth information is captured with the help of an infrared projector and sensor [8,9]. Both sensors captured images are of low resolution and noisy [6].

The Kinect sensor has low-resolution images, incorporating some noise and inaccuracies in the captured images [15]. Thus, holes (zero pixel values) tend to develop in the captured depth images. Also, the resultant artifacts in the depth image are primarily due to the perturbation in the distance between subject and sensor. The other reason could be due to the low reflectance of the surface to the projected Infra-Red (IR) light. Therefore, the absence of pixel information degrades the image quality and affects the image recognition performance accuracy.

4.1 Contributions

The Literature survey on hole filling in chapter 1 directs that filling holes/missing information is necessary to obtain better performance. Here we present interpolation-based hole filling techniques/filters for computing the missing pixels in the depth images. We have engaged the RGB and Depth images from our own GU-RGB-D database to demonstrate the study. An extensive study has been performed to compute the identification and verification rate of the depth images and the fused RGB-D images by engaging various state-of-the-art feature extraction algorithms such as Histogram of Oriented Gradient (HOG) [11], Principal Component Analysis (PCA) [12], GIST [13], Local Binary Pattern (LBP) [14], LogGabor [15], Local Phase Quantization (LPQ) [16] and Binarized Statistical Image Features (BSIF) [17]. The RGB and Depth images are fused using the Pixel level Average Image fusion technique. Further, to demonstrate the significance of our approach, we present the

performance evaluation results on depth images, RGB-D fused images (Fusion of depth is performed after hole filling), and score level fusion (On two best performing algorithms). Further in due course of this work summarizes the number of contributions as follows:

- A hole filling approach in the depth images acquired from Kinect sensor to enhance the performance of 3*D* face recognition system.
- A simple and effective approach based on variable kernel size for filling the holes with the contribution from neighborhood.
- The study presenting the significance of employing hole filling techniques to improve the performance of the state-of-the-art face recognition methods.
- Experimented on seven different feature extraction methods such as Principal Component Analysis (PCA), Histogram of Oriented Gradient (HOG), Local Binary Pattern (LBP), Local Phase Quantization (LPQ), GIST, Binarized Statistical Image Features (BSIF), and LogGabor to demonstrate the applicability of our approach for improved performance analysis.
- Presents the performance evaluation results in the form of recognition rate and verification rate on depth images alone, a fusion of RGB with Depth images, and Score level fusion extensively.

The rest of the chapter is structured as follows: Section 4.2 presents the mathematical details and description of the different designed hole filling filters (linear interpolation, exponential averaging, and weighted averaging) employed in this work. Experimental protocols are described in section 4.2. Section 4.3 gives the exhaustive experimental evaluation results in verification and recognition rate using seven different state-of-the-art methods for RGB-D face recognition.

4.2 Hole Filling Filter Design

This section presents the method employed to fill the holes in the depth images captured using a Kinect sensor. As said earlier, the Kinect sensor's low resolution incorporates some noise

and inaccuracies in the captured images [21], which results in the missing pixels and affects the overall performance accuracy.

Further, we strongly believe that the neighboring pixel information can be valuable in filling up the missing pixel (Zero pixel) values in place of holes. We, therefore, developed three different hole-filling approaches based on: Linear Interpolation method, Exponential Average method, and Weighted Average method independently in our work. However, the depth images obtained after linear interpolation, exponential average, and the weighted average filter will be represented by the acronyms `LI-Filter', `EA- Filter', and `WA-Filter' respectively, for simplicity. On the other hand, depth images without filtering will be represented by the acronym `WO-Filter'. Table 4.1 provides a detailed description of each of these acronyms used in our work.

Table 4. 1: Summary of acronym illustrating the description of three different filters used for hole filling

Acronym	Description
WO	Depth images without filtering
LI- Filter	Linear interpolation-based filtered depth images
EA- Filter	Exponential averaging based filtered depth images
WA- Filter	Weighted averaging based filtered depth images.

Conceptually, in our work, we first take the depth image of m x m dimension and then append the dummy rows and columns of m/4 pixels to outspread the depth images of the m x m dimension to the higher dimension. This is done so as to circumvent the occurrence of computational errors at the peripheral pixels of the depth images due to the expanding factor of the kernel function used in the algorithm. At the same time, there shall be no contribution

of these dummy pixel elements in the hole filling; thus, these elements are allotted with 'NaN' values to avoid false computations.

Further, to fill holes in the depth image, we traverse through the image to locate the zero-depth pixel value (which is a hole, consists of zero pixel value) that needs to be filled using the interpolation technique employed in our work. To fill the hole, we use the contribution of the neighborhood pixels value by introducing a kernel function. In the context of this work, the kernel is a rectangular window employed to select the region to surround the pixel (zero-valued pixel or hole) in a depth image. Thus, using the kernel function, the filtering operations are introduced, which will take the contribution of neighboring pixel value with a region specified by the kernel to represent the missing pixel value (to fill the hole) in the depth image. Further, the kernel size automatically expands depending upon the size of the hole. The kernel expands until 95% of neighborhood contributing pixels are of non-zero values so as to give the higher importance to the populace of the non zero pixels and to have an overall contribution from different regions of the kernel, thereby giving weighted importance to each non zero pixels to obtain the final missing pixel value in depth image. We also introduced the weighted contribution either with a linear or exponential approach in our work. In a similar manner, we fill the other holes in the depth images. Figure 4.1 present the conceptual illustration of the hole-filling approach employed in this work. Specifically, Figure 4.1 details the pictorial view of the working of hole filling technique's at different locations in the depth image and for different sizes of holes. The mathematical details related to the three designed hole-filling techniques employed in this work are explained in the following sub-sections.

Let $u(x, y) \in R$ be the depth image after pre-processing of dimension m x m, acquired using Kinect sensor with a noisy image having holes, where (x, y) be the spatial coordinates of a depth image. Further, to apply the filtering operation in order to fill the holes, we need to define the kernel function of a rectangular window on depth image $u(x, y)$ such that a specific region surrounds the hole is selected for processing. Let the expression for kernel function for the pixel $u(x,y)$ in a given image can be given using Equation 4.1 as follow:

$$u_k(i,j) = u[(x-k:x+k),(y-k:y+k)] \quad (4.1)$$

Where

$$\text{Kernel}=\begin{cases} \text{valid} & \text{if } 1 \le k \le m/4 \\ \text{not valid} & \text{otherwise} \end{cases}$$

and $i = x - k : x + k$, $j = y - k : y + k$. Once the kernel is defined, we then perform the interpolation operation to fill the hole. In the next sub-sections, we will discuss the filtering methods using the kernel window function defined in Equation 4.1.

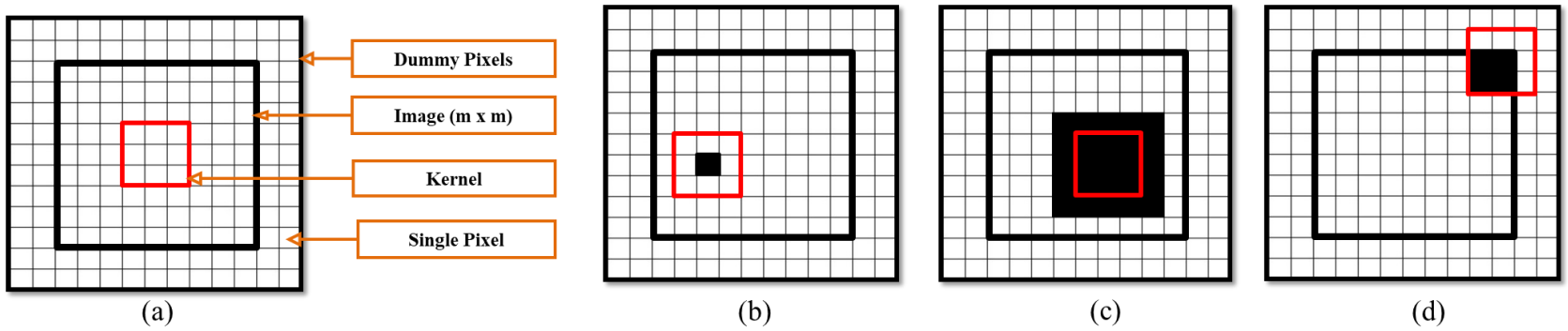


Fig 4. 1: Conceptual illustration of working of the filter using the kernel to fill the holes in a depth image: (a) Presents the labeling of regions such as appending dummy frame, image, kernel, etc., (b) The hole is surrounded by a high population of non-zero depth values (In the figure it shows the window of size 3 x 3 is used for filling the hole), (c) Hole with the kernel is densely surrounded by zeros; thus the kernel function (window) needs to be expanded until the condition (95% and 5% contribution from non-zero and zero pixel values respectively) is reached,(d) Hole is at the corner position, where kernel expand across the dummy rows and columns crossing the image boundaries

4.2.1 LI-Filter: Linear Interpolation

Linear interpolation is one of the simplest filtering method used to fill the missing information. It has been widely used in digital image processing to resize or remap the image from the one pixel grid to another. The method is used in an application where there is a need to increase or decrease the total number of pixels, thereby remapping the image to a new dimension. We use the potential of linear interpolation to fill the hole in a depth image in our work. The linear interpolation method is applied on the kernel function defined in Equation 4.1. In this filtering method, we employ linear interpolation for missing pixels only when the number (nos) of non-zero pixels is 95%, and the zero pixels value is 5% in the expanded kernel. Specifically, we expand the kernel window size until 95% of the contribution is obtained from the non-zero pixel value to compute the missing pixel value.

Let the zero elements within the kernel be denoted by $u_{k_o}(i, j)$ and non-zero elements within the kernel $u_{k_{\bar{o}}}(i, j)$. In our experiment, the kernel window function is expanded till we have 95% of non-zero elements $u_{k_{\bar{o}}}(i, j)$, and 5% of zero elements $u_{k_o}(i, j)$ is 5%, to compute the linear interpolation to fill the missing pixel in the given image. Therefore, Kernel function is expanded and restricted by the condition $n_1 * nos(u_{k_o}(i, j)) \leq n_2 * nos(u_{k_{\bar{o}}}(i, j))$, here in or work we have selected $n_1 = 0.05$ and $n_2 = 0.95$. Once the condition is fulfilled, linear interpolation is employed. The expression for the image after linear interpolation is given by Equation 4.2.

$$\bar{u}_k(i,j) = \sum_i \sum_j u_k(i,j) \qquad (4.2)$$

Where a pixel in the interpolated image is

$$\bar{u}_k(i,j) = \begin{cases} \text{valid} & \text{if } n_1 * nos(u_{k_o}(i,j)) \leq n_2 * nos(u_{k_{\bar{o}}}(i,j)) \\ \text{not valid} & \text{otherwise} \end{cases}$$

After computing, the value of $\bar{u}_k$ is assigned to the new matrix $\bar{u}_{LI}$ *(x, y)* to represent the final filtered depth image. Further, more details related to this approach is given in Algorithm 1.

Algorithm 1: LI-Filter: Linear Interpolation

Result: filtered output depth map image: $\bar{u}_{LI}(x, y)$
require: $n_1 = 0.05, n_2 = 0.95, \bar{u}_{LI} = [\,], u(x, y)$
for $x, y = 1 : 1 : m$ **do**
  while $k = 1 : m/4$ **do**
    compute:
    $nos(u_{k_o}(i, j)), nos(u_{k_{\bar{o}}}(i, j)$
    if $n_1 * nos(u_{k_o}(i, j)) \leq n_2 * nos(u_{k_{\bar{o}}}(i, j)$ **then**
      compute:
      $\bar{u}_{k\_}(i, j) = \sum_i \sum_j u_{k\_}(i, j)$
      $\bar{u}_{LI}(x, y) = \bar{u}_{k\_}(i, j)$
      goto top:
    else
      k=k+1
    end
  end
end

4.2.2 EA-Filter: Exponential Averaging

In the case of exponential average filtering, we start the filter computation for the kernel size where the first non-zero pixel is encountered and continue the computations for all subsequent kernel expansion by counting the number of expansions. We stop the computation once 95% of the elements are non-zero within the kernel. We then combine the filter's contribution exponentially, giving the highest weightage to the nearest neighborhood pixel surrounding the hole to the farthest pixel surrounding the hole.

Algorithm 2: EA-Filter: Exponential Average Filter

Result: filtered output depth map image: $\bar{u}_{EA}(x, y)$
require: $n_1 = 0.05, n_2 = 0.95, \bar{u}_{EA} = [\,], u(x, y)$
for $x, y = 1 : 1 : m$ **do**
  top:
  while $k = 1 : m/4$ **do**
    Kernel:
    compute:
    $nos(u_{k_o}(i, j)), nos(u_{k_{\bar{o}}}(i, j))$
    if $nos(u_{k_{\bar{o}}}(i, j) \geq 1$ **then**
      if $n_1 * nos(u_{k_o}(i, j)) \leq n_2 * nos(u_{k_{\bar{o}}}(i, j)$
      then
        compute:
        $\bar{u}_k\ (i, j) = \sum_i \sum_j \alpha^k * u_k\ (i, j)$
        $\bar{u}_{EA}(x, y) = \bar{u}_k\ (i, j)$
        goto top:
      else
        k=k+1
      end
    else
      k=k+1 **goto** Kernel:
    end
  end
end

Similar to the previous process, we first initialize the kernel function surround the hole and count the zeros elements u_{k_o} (i, j) and non-zeros elements $u_{k_{\bar{o}}}(i, j)$ in the kernel. We start employing the filter soon we get the non-zero element ($nos(u_{k_{\bar{o}}}(i, j) \geq 1$) as we expand the kernel size. However, we continue the computation until the condition $n_1 * nos(u_{k_o}(i, j)) \leq n_2 * nos(u_{k_{\bar{o}}}(i, j)$, where, $n_1 = 0.05$ and $n_2 = 0.95$ is satisfied. Here the weightage is given for every expanded kernel stage. Therefore to meet the requirement, we introduce the α^k variable

(where α is a constant, here we have chosen α= 0.9 and k is the number of kernel expansions) in the Equation 4.2 to define the exponential average filter using the Equation 4.3 as follows:

$$\bar{u}_k(i,j) = \sum_i \sum_j \propto^k * u_k(i,j) \quad (4.3)$$

Where pixel in the interpolated image is

$$\bar{u}_k\ (i,j) = \begin{cases} \text{valid} & \text{if } n_1 * nos(u_{k_o}(i,j)) \leq n_2 * nos(u_{k_{\bar{o}}}(i,j) \\ \text{not valid} & \text{otherwise} \end{cases}$$

The detail represented can be obtained from Algorithm 2. After computing the pixel value of $\bar{u}_k$ is assigned to the new matrix $\bar{u}_{EA}\ (x, y)$ to represent the final filtered depth image with exponential average filtering.

4.2.3 WA-Filter: Weighted Averaging

Similar to the exponential average filtering, we introduce the weighted average filter over the expanding kernel in this section. The process of applying the filter is the same as that of the exponential average filter (Section 4.2.2). Here the weighted average filtering method is a linear approach, and the exponential average filtering is a non-linear approach. Thus instead of gradually giving the non-linear importance to the non-zero elements of the expanding kernel, we introduce linearity, in which the highest weightage is given to the nearest non-zero pixel and the lowest weightage to the farthest non-zero pixel over expanding the kernel function.

Mathematically we present weighted average filtering by equation 4.4, after modifying Equation 4.3 as follows:

$$\bar{u}_k(i,j) = \sum_i \sum_j (1 - (k - \alpha)) * u_k(i,j) \quad (4.4)$$

where pixel in the interpolated image is

$$\bar{u}_k\ (i,j) = \begin{cases} \text{valid} & \text{if } n_1 * nos(u_{k_o}(i,j)) \leq n_2 * nos(u_{k_{\bar{o}}}(i,j) \\ \text{not valid} & \text{otherwise} \end{cases}$$

and α is a constant (here we have chosen α= 0.9), and k is the number of kernel expansions.

Algorithm 3: WA-Filter: Weighted Average Filter

Result: filtered output depth map image: $\bar{u}_{WA}\ (x, y)$
require: $n_1 = 0.05, n_2 = 0.95, \bar{u}_{WA} = [\,], u(x, y)$
for $x, y = 1 : 1 : m$ **do**
  top:
  while $k = 1 : m/4$ **do**
    Kernel:
    compute:
    $nos(u_{k_o}(i, j)), nos(u_{k_{\bar{o}}}(i, j))$
    if $nos(u_{k_{\bar{o}}}(i, j) \geq 1$ **then**
      if $n_1 * nos(u_{k_o}(i, j)) \leq n_2 * nos(u_{k_{\bar{o}}}(i, j)$
      then
        compute:
        $\bar{u}_k\ (i, j) =$
        $\sum_i \sum_j (1 - (k * \alpha)) * u_k\ (i, j)$
        $\bar{u}_{EA}(x, y) = \bar{u}_k\ (i, j)$
        goto top:
      else
        k=k+1
      end
    else
      k=k+1 **goto** Kernel:
    end
  end
end

Finally, the value obtained by computing $\bar{u}_{kc,}$ which is a weighted sum, is then assigned to the new matrix $\bar{u}_{WA}$ *(x, y)* to represent the depth image. The detailed image representation related to this filtering approach can be obtained from Algorithm 3. Further, the graphical demonstration of filling the hole using LI-Filter: Linear Interpolation, EA-Filter: Exponential Averaging, and WA-Filter: Weighted Averaging is illustrated in Figure 4.2. Figure 4.2 clearing showing the successful application of our approach in filling the missing pixel values.

4.3 Experimental Protocol

This section presents in detail the experimental methodology (Figure 4.3), evaluation protocol (Table 4.2), and related results obtained in this work. Basically, we perform the hole filling on the depth image using three different designed filters, i.e., LI-Filter: Linear Interpolation, EA-Filter: Exponential Averaging, and WA-Filter: Weighted Averaging. Further, we make use of the kernel window function to give proper weightage to the neighborhood pixels in the kernel while employing the filtering method. The experimental evaluation results are presented on a GU-RGB-D database consisting 64 subjects. To demonstrate the application of kernel-based filtering approach, we present systematic results on seven different state-of-the-art feature extraction methods, namely, Principal Component Analysis (PCA), Histogram of Oriented Gradient (HOG), Local Binary Pattern (LBP), Local Phase Quantization (LPQ), GIST, Binarized Statistical Image Features (BSIF) and LogGabor, employed on depth images. We present the evaluation results in the form of recognition rate and verification rate in tabular and graphical form.

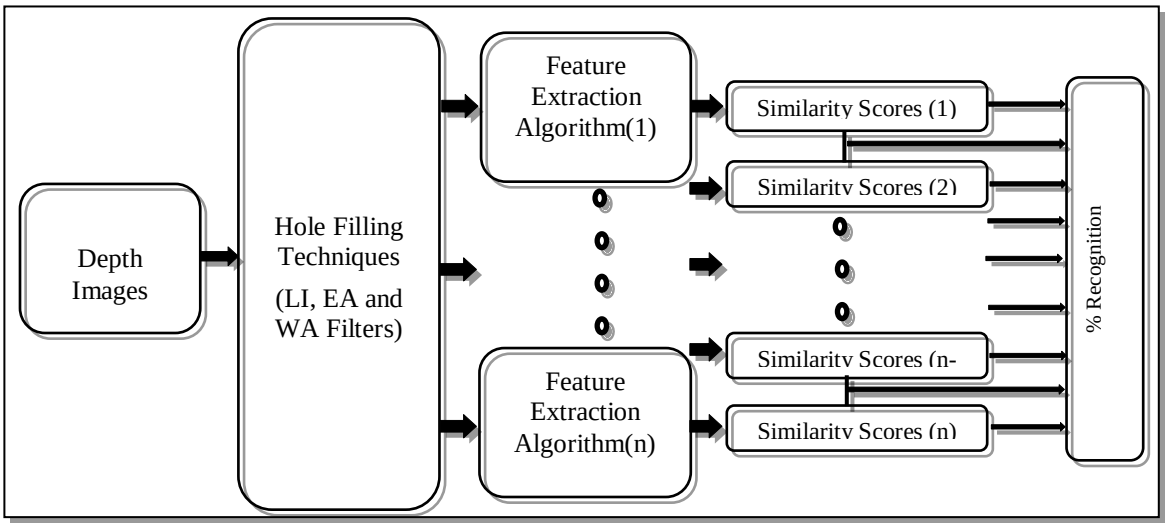


Fig 4. 2: Experimental methodology for evaluating the proposed filters using different feature extraction algorithms

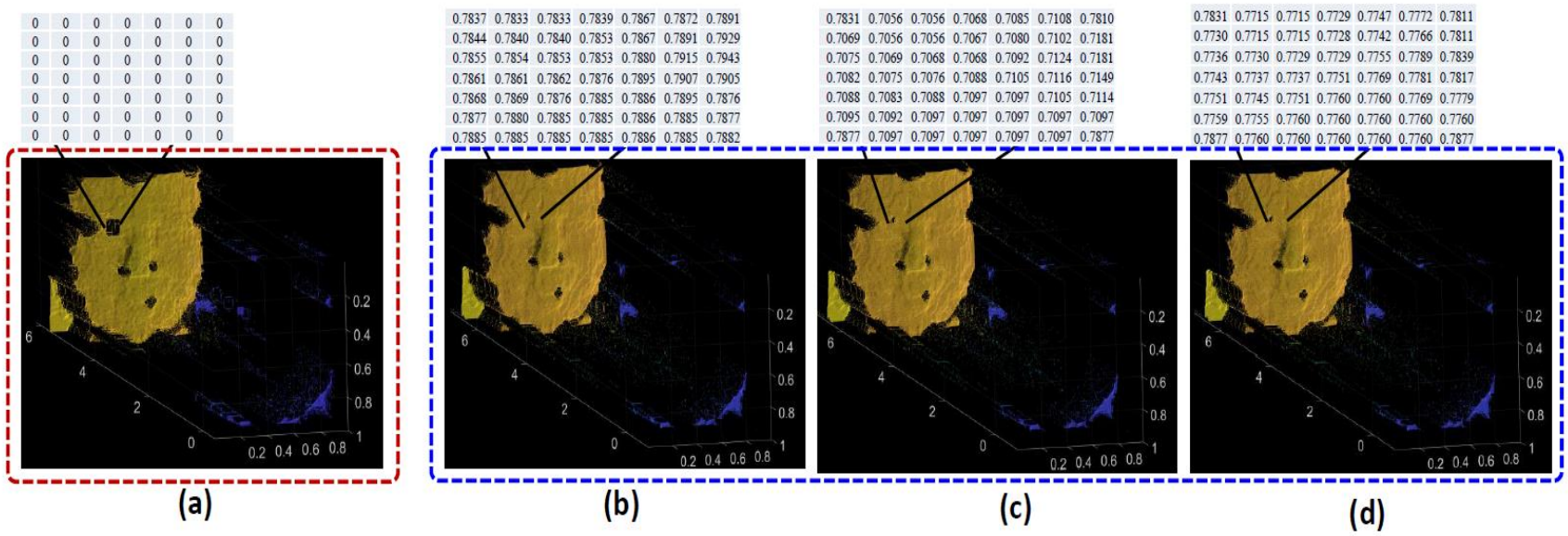


Fig 4. 3: Conceptual illustration of the filter's working using the kernel function to fill the holes in a depth image (point cloud view). (a) Point cloud with the 7 x 7 hole. (b) Hole filing using LI-Filter: Linear Interpolation. (c) Hole filing using EA-Filter: Exponential Averaging. (d) Hole filing using WA-Filter: Weighted Averaging

Table 4. 2: Experimental protocol used to study the effect of the three designed hole-filling filters on the GU-RGB-D database

Pose/Variation	Session	Total Subjects
Training Dataset		
0° pose	Session I	64
Testing Dataset		
0° pose	Session I	64
45° pose	Session I & II	64 each
90° pose	Session I & II	64 each
(-45)° pose	Session I & II	64 each
(-90)° pose	Session I & II	64 each
Paper on face	Session I & II	64 each
Eyes closed	Session I & II	64 each
Smile	Session I & II	64 each

4.4 Evaluation Protocol & Discussion

The experimental evaluation protocol for the database is as described in Figure 4.3. The RGB and depth images from the GU-RGB-D database were cropped to 256 x 256 dimensions using Matlab script. In most of the pose/variations of the RGB-D database, only partial faces are visible, and hence it restricts the use of existing automatic face detection algorithms for cropping. The cropping of RGB and depth images was performed by resizing them to 96 x 96 dimensions to enhance the computational time. Using the GU-RGBD database, we partitioned our data into the training and testing dataset. Training set consists of 64 subjects corresponding to front face (0° pose), including their samples from session 1 of the database, while the testing set consists of corresponding 64 subjects belong to pose/variations over either 45°, -45°, 90°, -90°, smile, eye closed, paper on face occlusion from session 1 of the

database independently when operated on seven different feature extraction methods discussed above. In a similar manner, we also generated the results when the training set belongs to session 1 of the database, and the testing set belongs to session 2 of the database. The details of the training and the testing datasets are given in Table 4.2.

Using the experimental protocol, we present the three sets of evaluations to demonstrate our hole-filling approach. Evaluation 1 presents the results related to depth images. Evaluation 2 presents the results related to the fusion of RGB and Depth images (after filtering). Evaluation 3 illustrates the score level fusion best performing algorithm separately on depth image and on RGB image fused with depth. Also, in our experimental evaluation, the RGB image was first converted to a grayscale image for the sake of processing.

4.4.1. Evaluation 1: Depth Images

In this section, we present the performance accuracy of three different filters using seven feature extraction methods. The idea is to demonstrate the performance accuracy of the face recognition using three different filters and to compare with the raw depth images (without filtering). Thus, in this section, we present the benchmark results on eight different facial variants to present the significance of our approach. Table 4.3, 4.4, and 4.5 presents the recognition rates at Rank-5 for session 1, Figure 4.4 presents the Cumulative Match Curve (CMC) plots, and Figure 4.5 presents the Receiver Operating Curve (ROC) for this set of evaluations. Similarly, Tables 4.6, 4.7 and 4.8, presents the recognition rates computed for session 2 of GU-RGB-D at rank-5. Clearly, a reasonable improvement in performance accuracy of the face recognition system can be seen based on our proposed hole filling approach compared to the depth image without filters. Also, the major improvement with exponential average and weighted average can be observed compared to linear interpolation filtering. This further validates our idea of employing kernel-based filtering to give weighted importance to the neighboring pixel values in filling the hole. Further, based on the evaluation results obtained for facial expression, we present our major observations as follows:

- The variations such as smile, eyes closed (in both the sessions), and ‘0° pose’ variation (in session 2), the obtained recognition rates at rank-5 are higher as compared to

other existing facial variants. The better performance in these cases is due to the presence of full-face geometry compared to the reference depth map image of the frontal face, which is similar in nature.

- Specifically, the recognition rates obtained for variation 'smile, in session 1 using PCA algorithm is 89.06% (without filter) and 90.63% (with LI-filter and WA-filter). Using HOG, the recognition rate obtained is 93.75% (without filter) and 96.88% (with WA-filter), while the lower performance is noted for EA-filter and LI-filter. Further, with LBP, the recognition rate is 48.44% (without filter), and 50% (with LI-filter) compared to the highest recognition rate of 54.69% with EA-filter and WA- filter. On the other hand, the recognition rate of LPQ without filter is poor, but significant improvement can be generated using filters (42.19% (without filter) and 62.50% (with LI-filter and EA-filter), while further maximum enhancement was seen for WA-filter having 64.06% recognition rate). Similarly, the methods such as GIST, BSIF, LogGabor (LG) indicate better results compared to LBP and LPQ, demonstrating the robustness of these methods for face recognition. The maximum recognition rate of 89.06%, 84.38%, and 92.19% has been noted using the GIST algorithm (with LI-filter and EA-filter), BSIF (with EA-filter and WA-filter), and LogGabor (with LI-filter), respectively. Thus, overall it can be seen that the recognition rate for smile using all three designed filters outperforms the baseline results without filter for face recognition algorithms used in this work.

- A similar enhancement trend for 'smile' variation in session 2 (table 4.6, 4.7, and 4.8) can be seen, thus enhancing the liability of the filters across the sessions, i.e., across the environmental and behavioral conditions.

- As mentioned earlier, the 0^{o} pose variation, i.e., the front face from session 2 (Table 4.6, 4.7, and 4.8) has also shown the improvement or has maintained the same performance as that of the base results for almost all the algorithms across all the three designed hole filling techniques with some exceptions.

- On the other hand recognition rate for 45° and -45° pose variation is dominant as compared to 90° and -90° pose variation. As expected, this decrease in the performance is due to the larger angular variation with ±90° as compared to ± 45° for all the three filters. While we note that from the observation table, with the application of filters, the performance is improved reasonably well across most of the face recognition algorithms. Although, due to angular face variability, the performance of 45° and -45° is lower than the full face, the effect of the filters has shown a remarkable noted improvement in the performance accuracy. For the 45° variation from session 1 the performance has been enhanced from 17.19% to 18.75% (by EA filter), from 18.75% to 23.31% (by LI filter), and from 10.94% to 17.91% (by LI filter) for HOG, LBP& LPQ respectively. Similar levels of enhancements are noted in the 90º variation in both sessions, with some exceptions.

- Overall, the application of the filters (at least one of the three) has improved the performance in terms of recognition rate. Even for the angle like 90º pose variation, the filter has shown good performance over base results in both the sessions for the depth images with exceptions at few places.

Table 4. 3: Recognition rate at Rank-5 using depth image after employing LI-Filter (Session 1)

Variation	Filter	PCA	HOG	LBP	LPQ	GIST	BSIF	LG
45º	WO	21.88	17.19	18.75	10.94	20.31	18.75	18.75
	LI	**21.88**	10.94	**20.31**	**17.19**	**26.56**	17.19	12.50
90º	WO	15.63	14.06	14.06	12.50	14.06	15.63	12.50
	LI	**15.63**	**14.06**	**17.19**	**14.06**	**18.75**	**17.19**	10.94
-45º	WO	17.19	18.75	17.19	14.06	18.75	14.06	9.38
	LI	**17.19**	10.94	**25.00**	12.50	14.06	**15.63**	7.81
-90º	WO	12.5	14.06	14.06	14.06	15.63	12.50	17.19
	LI	**15.63**	9.38	**21.88**	6.25	12.50	9.38	14.06
Smile	WO	89.06	93.75	48.44	42.19	89.06	82.81	92.19
	LI	**90.63**	90.63	**50.00**	**62.50**	**89.06**	79.69	**92.19**
Eyes closed	WO	89.06	89.06	34.38	37.50	89.06	78.13	90.63
	LI	85.94	**90.63**	**53.13**	**54.69**	84.38	**78.13**	89.06
Paper on face	WO	32.81	46.88	7.81	7.81	18.75	7.81	31.25
	LI	21.88	23.44	**10.94**	**9.38**	14.06	**10.94**	18.75

Table 4. 4: Recognition rate at Rank-5 using depth image after employing EA-Filter (Session 1)

Variation	Filter	PCA	HOG	LBP	LPQ	GIST	BSIF	LG
45º	WO	21.88	17.19	18.75	10.94	20.31	18.75	18.75
	EA	15.63	**18.75**	10.94	**15.63**	**21.88**	17.19	12.50
90º	WO	15.63	14.06	14.06	12.50	14.06	15.63	12.50
	EA	**18.75**	12.50	**20.31**	**14.06**	**17.19**	**21.88**	**15.63**
-45º	WO	17.19	18.75	17.19	14.06	18.75	14.06	9.38
	EA	**23.44**	14.06	10.94	**20.31**	14.06	**20.31**	**9.38**
-90º	WO	12.5	14.06	14.06	14.06	15.63	12.50	17.19
	EA	**15.63**	7.81	9.38	9.38	10.94	9.38	4.69
Smile	WO	89.06	93.75	48.44	42.19	89.06	82.81	92.19
	EA	**89.06**	90.63	**54.69**	**62.50**	**89.06**	**84.38**	90.63
Eyes closed	WO	89.06	89.06	34.38	37.50	89.06	78.13	90.63
	EA	87.50	**90.63**	**53.13**	**51.56**	85.94	75.00	89.06
Paper on face	WO	32.81	46.88	7.81	7.81	18.75	7.81	31.25
	EA	31.25	28.13	**10.94**	7.81	**21.88**	**20.31**	21.88

Table 4. 5: Recognition rate at Rank-5 using depth image after employing WA-Filter (Session 1)

Variation	Filter	PCA	HOG	LBP	LPQ	GIST	BSIF	LG
45º	WO	21.88	17.19	18.75	10.94	20.31	18.75	18.75
	WA	18.75	**18.75**	**18.75**	9.38	**21.88**	15.63	17.19
90º	WO	15.63	14.06	14.06	12.50	14.06	15.63	12.50
	WA	**18.75**	**15.63**	**20.31**	**15.63**	12.50	**20.31**	**20.31**
-45º	WO	17.19	18.75	17.19	14.06	18.75	14.06	9.38
	WA	**21.88**	15.63	**18.75**	**17.19**	**18.75**	**18.75**	6.25
-90º	WO	12.5	14.06	14.06	14.06	15.63	12.50	17.19
	WA	**17.19**	10.94	7.81	10.94	10.94	10.94	7.81
Smile	WO	89.06	93.75	48.44	42.19	89.06	82.81	92.19
	WA	**90.63**	**96.88**	**54.69**	**64.06**	84.38	**84.38**	87.5
Eyes closed	WO	89.06	89.06	34.38	37.50	89.06	78.13	90.63
	WA	85.94	**90.63**	**48.44**	**53.13**	87.50	**78.13**	85.94
Paper on face	WO	32.81	46.88	7.81	7.81	18.75	7.81	31.25
	WA	26.56	25.00	**9.38**	**7.81**	12.50	**17.19**	14.06

Table 4. 6: Recognition rate at Rank-5 using depth image after employing LI-Filter (Session 2)

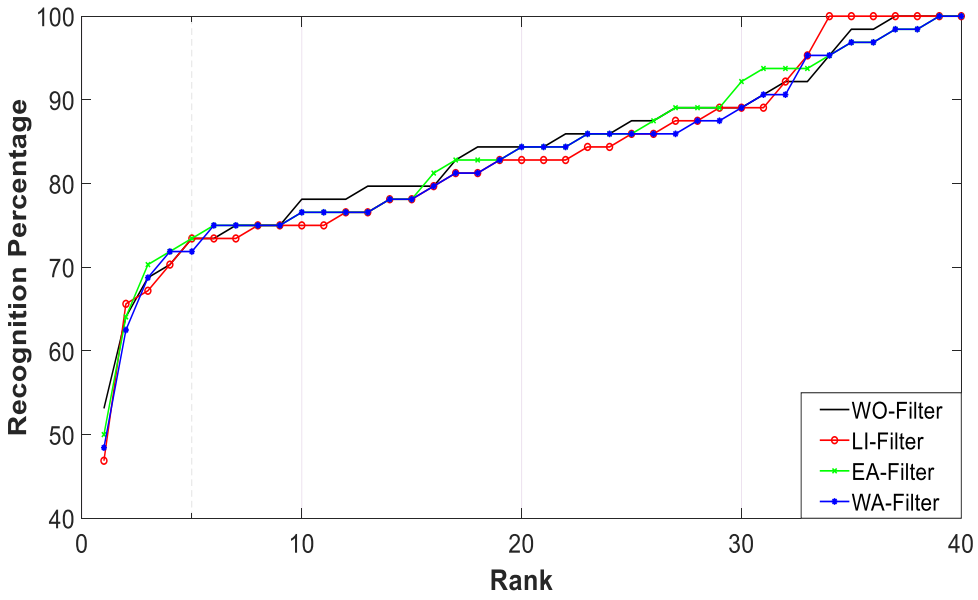
Varation	Filter	PCA	HOG	LBP	LPQ	GIST	BSIF	LG
0º	WO	73.44	65.63	18.75	20.31	25.00	35.94	53.13
	LI	**73.44**	**65.63**	**23.44**	**31.25**	**25.00**	**43.75**	50.00
45º	WO	23.44	17.19	12.50	9.38	10.94	17.19	17.19
	LI	21.88	**17.19**	**15.63**	**21.88**	**10.94**	12.50	14.06
90º	WO	17.19	10.94	10.94	12.50	10.94	15.63	10.94
	LI	**18.75**	**10.94**	**14.06**	7.81	**12.50**	**15.63**	9.38
-45º	WO	14.06	15.63	10.94	17.19	12.50	6.25	15.63
	LI	10.94	**15.63**	**15.63**	10.94	**12.50**	**7.81**	10.94
-90º	WO	10.94	7.81	10.94	12.50	10.94	10.94	9.38
	LI	**10.94**	**7.81**	**14.06**	**12.50**	**12.50**	**12.50**	**9.38**
Smile	WO	73.44	62.50	14.06	14.06	28.13	42.19	54.69
	LI	67.19	**62.50**	**29.69**	**37.50**	**29.69**	**46.88**	45.31
Eyes closed	WO	78.13	60.94	15.63	26.56	26.56	51.56	54.69
	LI	76.56	**60.94**	**35.94**	**35.94**	**26.56**	50.00	51.56
Paper on face	WO	37.50	45.31	17.19	10.94	21.88	9.38	26.56
	LI	34.38	**45.31**	7.81	**18.75**	18.75	**14.06**	20.31

Table 4. 7: Recognition rate at Rank-5 using depth map image after employing EA-Filter (Session 2)

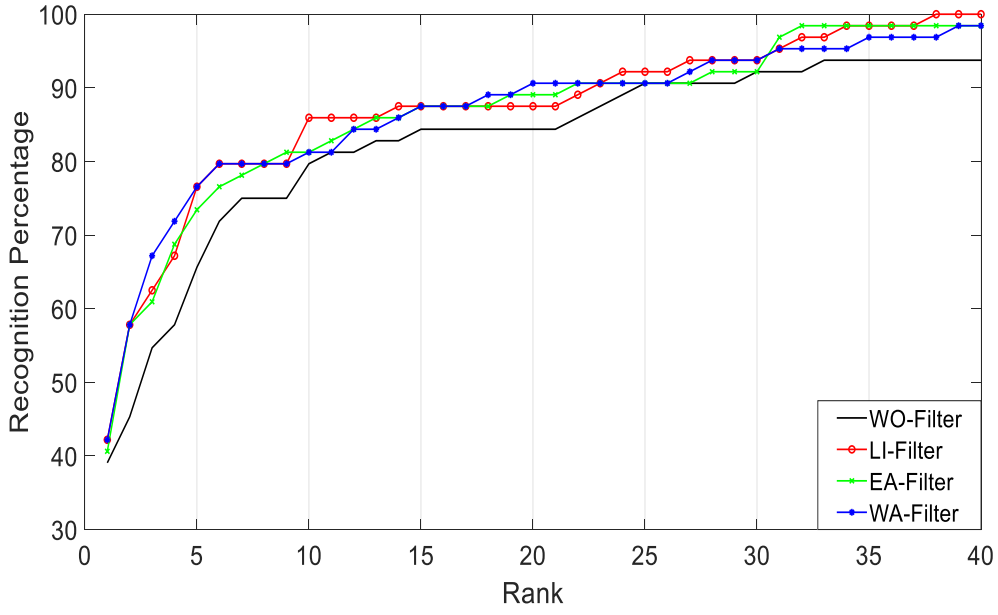
Variation	Filter	PCA	HOG	LBP	LPQ	GIST	BSIF	LG
0º	WO	73.44	65.63	18.75	20.31	25.00	35.94	53.13
	EA	**73.44**	**73.44**	**20.31**	**35.94**	**25.00**	**45.31**	**53.13**
45º	WO	23.44	17.19	12.50	9.38	10.94	17.19	17.19
	EA	18.75	**20.31**	**15.63**	**18.75**	9.38	12.50	14.06
90º	WO	17.19	10.94	10.94	12.50	10.94	15.63	10.94
	EA	**17.19**	**14.06**	**12.50**	9.38	**10.94**	9.38	**10.94**
-45º	WO	14.06	15.63	10.94	17.19	12.50	6.25	15.63
	EA	12.50	10.94	**12.50**	15.63	**14.06**	**6.25**	12.50
-90º	WO	10.94	7.81	10.94	12.50	10.94	10.94	9.38
	EA	9.38	**9.38**	**12.50**	**15.63**	**10.94**	7.81	**9.38**
Smile	WO	73.44	62.50	14.06	14.06	28.13	42.19	54.69
	EA	70.31	**73.44**	**28.13**	**37.50**	**28.13**	**48.44**	46.88
Eyes closed	WO	78.13	60.94	15.63	26.56	26.56	51.56	54.69
	EA	**78.13**	**70.31**	**28.13**	**37.50**	**28.13**	**51.56**	53.13
Paper on face	WO	37.50	45.31	17.19	10.94	21.88	9.38	26.56
	EA	**37.50**	32.81	7.81	**21.88**	17.19	**17.19**	18.75

Table 4. 8: Recognition rate at Rank-5 using depth map image after employing WA-Filter (Session 2)

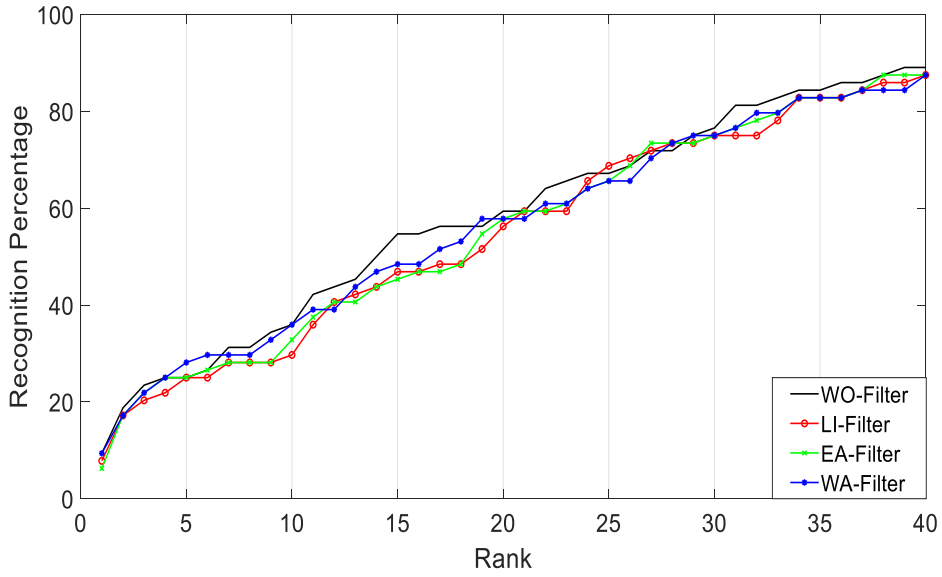
Variation	Filter	PCA	HOG	LBP	LPQ	GIST	BSIF	LG
0º	WO	73.44	65.63	18.75	20.31	25.00	35.94	53.13
	WA	71.88	**76.56**	**21.88**	**34.38**	**28.13**	**39.06**	**57.81**
45º	WO	23.44	17.19	12.50	9.38	10.94	17.19	17.19
	WA	20.31	**23.44**	10.94	**17.19**	7.81	14.06	12.50
90º	WO	17.19	10.94	10.94	12.50	10.94	15.63	10.94
	WA	**17.19**	**12.50**	**17.19**	10.94	9.38	9.38	7.81
-45º	WO	14.06	15.63	10.94	17.19	12.50	6.25	15.63
	WA	**14.06**	9.38	**17.19**	14.06	10.94	**7.81**	9.38
-90º	WO	10.94	7.81	10.94	12.50	10.94	10.94	9.38
	WA	**12.50**	**12.50**	**14.06**	**12.50**	9.38	**10.94**	**12.50**
Smile	WO	73.44	62.50	14.06	14.06	28.13	42.19	54.69
	WA	68.75	**71.88**	**35.94**	**39.06**	26.56	**43.75**	**57.81**
Eyes closed	WO	78.13	60.94	15.63	26.56	26.56	51.56	54.69
	WA	**79.69**	**73.44**	**32.81**	**37.50**	**28.13**	**51.56**	**56.25**
Paper on face	WO	37.50	45.31	17.19	10.94	21.88	9.38	26.56
	WA	35.94	31.25	6.25	**21.88**	14.06	**17.19**	17.19



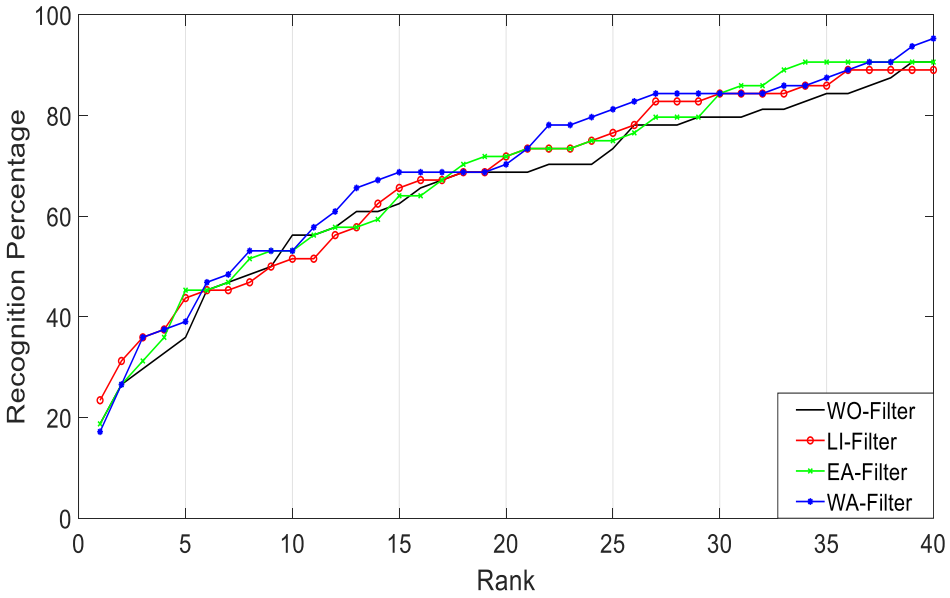
(a) Cumulative Match Curve (CMC) plots generated using PCA as feature extraction algorithm



(b) Cumulative Match Curve (CMC) plots generated using HOG as feature extraction algorithm

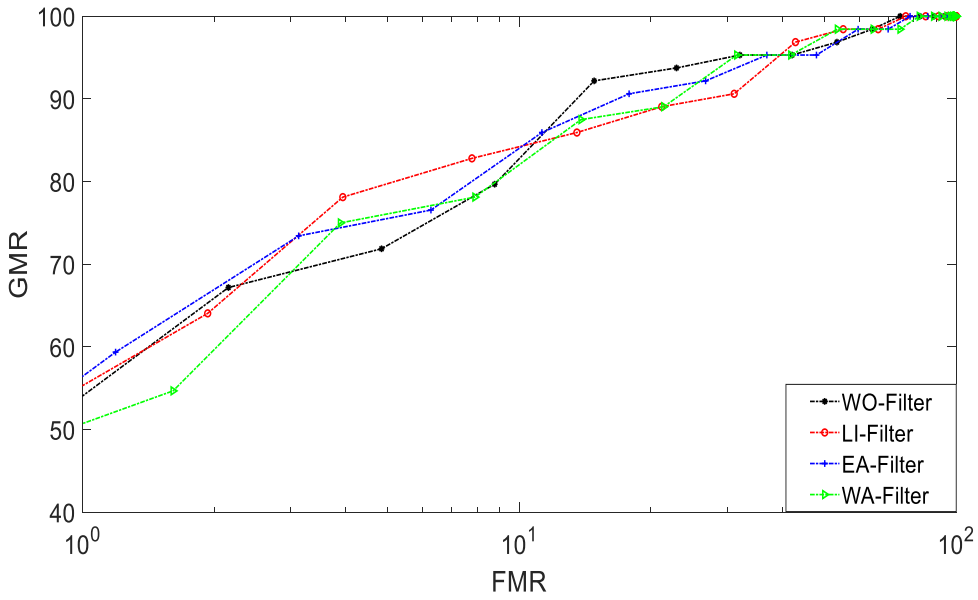


(c) Cumulative Match Curve (CMC) plots generated using GIST as feature extraction algorithm

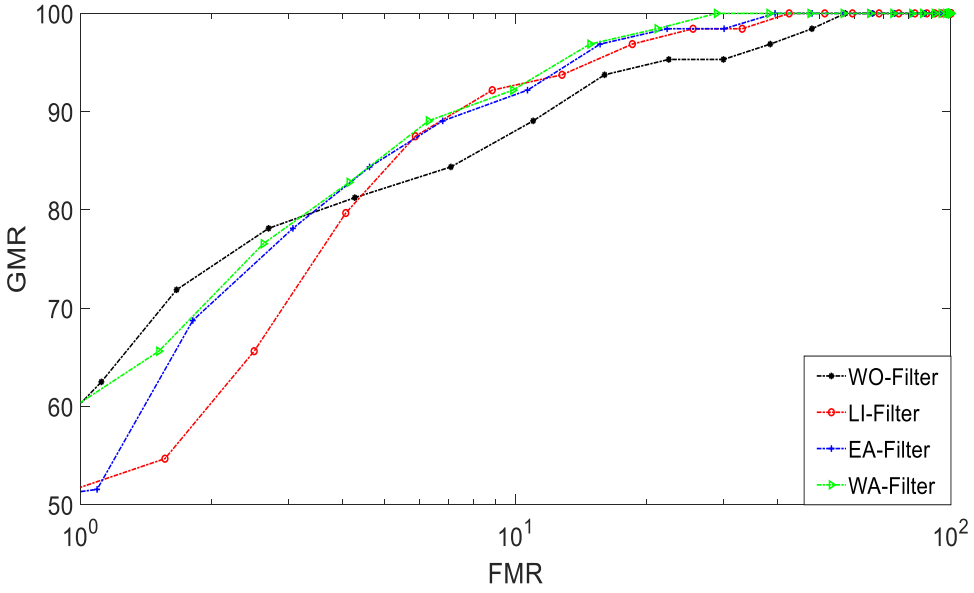


(d) Cumulative Match Curve (CMC) plots generated using BSIF as feature extraction algorithm

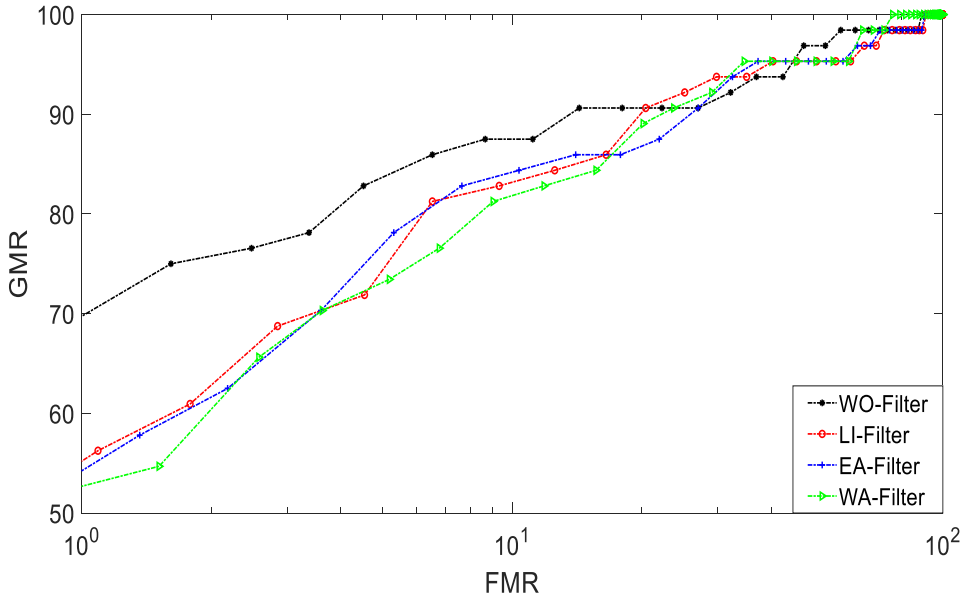
Fig 4. 4: Cumulative Match Curve (CMC) plots demonstrate the face recognition performance on depth images using three different filters and without filter. The best results related to facial variation smile is presented here in (a) – (d)



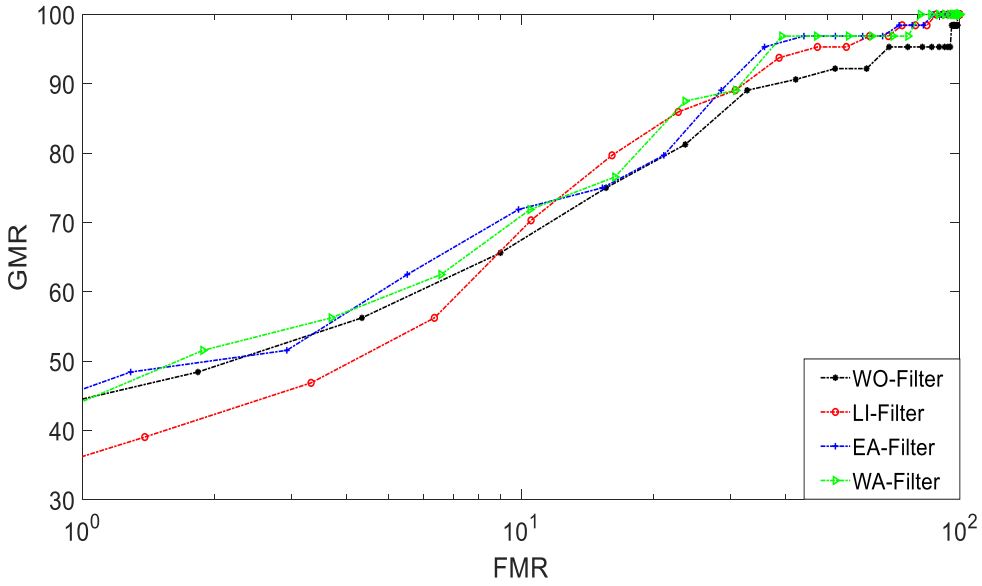
(a) Receiver Operating Curve (ROC) plots generated using PCA as feature extraction algorithm



(b) Receiver Operating Curve (ROC) plots generated using HOG as feature extraction algorithm



(c) Receiver Operating Curve (ROC) plots generated using GIST as feature extraction algorithm



(d) Receiver Operating Curve (ROC) plots generated using BSIF as feature extraction algorithm

Fig 4. 5: Receiver Operating Curve (ROC) plots demonstrate the face recognition performance on depth images using three different filters and without filter. The best results related to facial variation smile is presented here in (a) - (d)

4.4.2 Evaluation 2: Fused (RGB + Depth) Image

This section describes the performance analysis based on fusing RGB image and Depth image; here, the depth images used for fusion are after employing the filters. Although a depth image gives more information than an RGB image, a depth presents less variability among the subjects as compared to the RGB image; thus, combining the RGB and depth map image can improve the performance reasonably. In this evaluation, we used a simple yet effective method to fuse the two images by averaging RGB and depth images. Further, the recognition rates are computed based on the evaluation protocol across seven different face recognition algorithms mentioned earlier. Table 4.9, 4.10, and 4.11 presents the computed recognition rates for session 1 at Rank-5; Figure 4.6 presents the Cumulative Match Curve (CMC) plots, and Figure 4.7 presents the Receiver Operating Curve (ROC) for the said set of evaluations on smile variation. Similarly, Tables 4.12, 4.13 and 4.14, presents the recognition rates computed for session 2 of GU-RGB-D at rank 5 using the said evaluation protocol. The CMC plot demonstrating the face recognition performance on 0º pose variation (session 2) is presented in Figure 4.8. Further to have a fair comparison, we also compute the results on the fusion of RGB and depth images when filters are not employed.

Table 4. 9: Recognition rate at Rank-5 using fused (RBG + Depth) image after employing LI-Filter (Session 1)

Variation	Filter	PCA	HOG	LBP	LPQ	GIST	BSIF	LG
45º	WO	25.00	17.19	15.63	25.00	31.25	26.56	15.63
	LI	**25.00**	**18.75**	12.50	23.44	**32.81**	21.88	**18.75**
90º	WO	14.06	12.50	10.94	7.81	14.06	18.75	7.81
	LI	**14.06**	10.94	7.81	9.38	7.81	9.38	**10.94**
-45º	WO	21.88	18.75	18.75	10.94	14.06	20.31	18.75
	LI	**23.44**	14.06	14.06	7.81	9.38	18.75	**20.31**
-90º	WO	15.63	14.06	10.94	7.81	17.19	10.94	15.63
	LI	**18.75**	6.25	**10.94**	**7.81**	**18.75**	**15.63**	14.06
Smile	WO	95.31	98.44	73.44	81.25	96.88	96.88	98.44
	LI	**95.31**	96.88	59.38	**81.25**	92.19	92.19	**100.00**
Eyes closed	WO	92.19	93.75	79.69	82.81	90.63	93.75	96.88
	LI	**93.75**	**95.31**	67.19	**82.81**	**90.63**	**93.75**	**96.88**
Paper on face	WO	45.31	51.56	10.94	28.13	26.56	15.63	18.75
	LI	28.13	32.81	**10.94**	14.06	21.88	**21.88**	15.63

Table 4. 10: Recognition rate at Rank-5 using fused (RGB + Depth) image after employing EA-Filter (Session 1)

Variation	Filter	PCA	HOG	LBP	LPQ	GIST	BSIF	LG
45º	WO	25.00	17.19	15.63	25.00	31.25	26.56	15.63
	EA	23.44	**20.31**	14.06	18.75	**29.69**	15.63	**17.19**
90º	WO	14.06	12.50	10.94	7.81	14.06	18.75	7.81
	EA	**17.19**	10.94	**10.94**	**7.81**	7.81	14.06	**10.94**
-45º	WO	21.88	18.75	18.75	10.94	14.06	20.31	18.75
	EA	20.31	15.63	17.19	7.81	12.50	17.19	**18.75**
-90º	WO	15.63	14.06	10.94	7.81	17.19	10.94	15.63
	EA	**18.75**	7.81	**10.94**	**7.81**	**21.88**	**12.50**	**15.63**
Smile	WO	95.31	98.44	73.44	81.25	96.88	96.88	98.44
	EA	**95.31**	**98.44**	56.25	**81.25**	93.75	**96.88**	**100.00**
Eyes closed	WO	92.19	93.75	79.69	82.81	90.63	93.75	96.88
	EA	**93.75**	**96.88**	67.19	75.00	**90.63**	**96.88**	**96.88**
Paper on face	WO	45.31	51.56	10.94	28.13	26.56	15.63	18.75
	EA	40.63	**53.13**	**17.19**	18.75	25.00	**17.19**	**18.75**

Table 4. 11: Recognition rate at Rank-5 using fused (RGB + Depth) image after employing WA-Filter (Session 1)

Variation	Filter	PCA	HOG	LBP	LPQ	GIST	BSIF	LG
45º	WO	25.00	17.19	15.63	25.00	31.25	26.56	15.63
	WA	**26.56**	**25.00**	14.06	20.31	20.31	21.88	**17.19**
90º	WO	14.06	12.50	10.94	7.81	14.06	18.75	7.81
	WA	**18.75**	**12.50**	**10.94**	**7.81**	4.69	**18.75**	**12.50**
-45º	WO	21.88	18.75	18.75	10.94	14.06	20.31	18.75
	WA	**21.88**	17.19	**21.88**	7.81	12.50	15.63	**20.31**
-90º	WO	15.63	14.06	10.94	7.81	17.19	10.94	15.63
	WA	**18.75**	10.94	**12.50**	**9.38**	12.50	9.38	9.38
Smile	WO	95.31	98.44	73.44	81.25	96.88	96.88	98.44
	WA	93.75	96.88	68.75	76.56	90.63	**96.88**	96.88
Eyes closed	WO	92.19	93.75	79.69	82.81	90.63	93.75	96.88
	WA	90.63	**95.31**	67.19	78.13	**93.75**	**98.44**	95.31
Paper on face	WO	45.31	51.56	10.94	28.13	26.56	15.63	18.75
	WA	37.50	**56.25**	**20.31**	23.44	15.63	**21.88**	12.50

Table 4. 12: Recognition rate at Rank-5 using fused (RGB + Depth) image after employing LI-Filter (Session 2)

Variation	Filter	PCA	HOG	LBP	LPQ	GIST	BSIF	LG
0º	WO	87.50	93.75	35.94	54.69	45.31	68.75	73.44
	LI	**92.19**	**95.31**	**40.63**	48.44	42.19	**70.31**	**73.44**
45º	WO	23.44	12.50	7.81	15.63	12.50	10.94	18.75
	LI	**25.00**	**15.63**	**14.06**	**18.75**	10.94	**20.31**	17.19
90º	WO	15.63	20.31	7.81	6.25	9.38	17.19	9.38
	LI	12.50	18.75	**9.38**	**7.81**	**14.06**	14.06	**12.50**
-45º	WO	14.06	15.63	6.25	7.81	14.06	10.94	17.19
	LI	**17.19**	14.06	**6.25**	**9.38**	12.50	**17.19**	12.50
-90º	WO	9.38	6.25	6.25	6.25	9.38	12.50	7.81
	LI	**9.38**	**9.38**	**6.25**	**7.81**	**10.94**	**14.06**	**7.81**
Smile	WO	82.81	93.75	39.06	51.56	50.00	62.50	70.31
	LI	**84.38**	92.19	**39.06**	**51.56**	43.75	**70.31**	68.75
Eyes closed	WO	82.81	93.75	34.38	51.56	48.44	56.25	70.31
	LI	**85.94**	**93.75**	**34.38**	46.88	39.06	**67.19**	**73.44**
Paper on face	WO	29.69	48.44	10.94	18.75	25.00	26.56	29.69
	LI	25.00	**54.69**	**15.63**	15.63	18.75	**26.56**	**29.69**

Table 4. 13: Recognition rate at Rank-5 using fused (RGB + Depth) image after employing EA-Filter (Session 2)

Variation	Filter	PCA	HOG	LBP	LPQ	GIST	BSIF	LG
0º	WO	87.50	93.75	35.94	54.69	45.31	68.75	73.44
	EA	**92.19**	**93.75**	**39.06**	51.56	**42.19**	**70.31**	**73.44**
45º	WO	23.44	12.50	7.81	15.63	12.50	10.94	18.75
	EA	21.88	**15.63**	**12.50**	**17.19**	**12.50**	**17.19**	**18.75**
90º	WO	15.63	20.31	7.81	6.25	9.38	17.19	9.38
	EA	14.06	17.19	**7.81**	**6.25**	**14.06**	15.63	**12.50**
-45º	WO	14.06	15.63	6.25	7.81	14.06	10.94	17.19
	EA	**15.63**	14.06	4.69	6.25	10.94	9.38	15.63
-90º	WO	9.38	6.25	6.25	6.25	9.38	12.50	7.81
	EA	**9.38**	**9.38**	**6.25**	**9.38**	**12.50**	**12.50**	**7.81**
Smile	WO	82.81	93.75	39.06	51.56	50.00	62.50	70.31
	EA	81.25	92.19	35.94	**54.69**	46.88	**62.50**	68.75
Eyes closed	WO	82.81	93.75	34.38	51.56	48.44	56.25	70.31
	EA	**84.38**	**93.75**	**34.38**	50.00	43.75	**67.19**	**73.44**
Paper on face	WO	29.69	48.44	10.94	18.75	25.00	26.56	29.69
	EA	28.13	46.88	**15.63**	17.19	18.75	**26.56**	**31.25**

Table 4. 14: Rognition rate at Rank-5 using fused (RGB + Depth) image after employing WA-Filter (Session 2)

Variation	Filter	PCA	HOG	LBP	LPQ	GIST	BSIF	LG
0º	WO	87.50	93.75	35.94	54.69	45.31	68.75	73.44
	WA	**92.19**	**96.88**	**37.50**	**54.69**	**50.00**	**73.44**	**79.69**
45º	WO	23.44	12.50	7.81	15.63	12.50	10.94	18.75
	WA	**23.44**	**15.63**	**15.63**	**12.50**	**12.50**	**18.75**	**21.88**
90º	WO	15.63	20.31	7.81	6.25	9.38	17.19	9.38
	WA	12.50	**20.31**	**9.38**	**6.25**	6.25	12.50	**15.63**
-45º	WO	14.06	15.63	6.25	7.81	14.06	10.94	17.19
	WA	**14.06**	14.06	**7.81**	**12.50**	12.50	**10.94**	**14.06**
-90º	WO	9.38	6.25	6.25	6.25	9.38	12.50	7.81
	WA	**10.94**	**10.94**	**7.81**	**10.94**	**10.94**	**14.06**	**10.94**
Smile	WO	82.81	93.75	39.06	51.56	50.00	62.50	70.31
	WA	**85.94**	**93.75**	35.94	**53.13**	**51.56**	60.94	**75.00**
Eyes closed	WO	82.81	93.75	34.38	51.56	48.44	56.25	70.31
	WA	**84.38**	**93.75**	26.69	**51.56**	**50.00**	**65.63**	**75.00**
Paper on face	WO	29.69	48.44	10.94	18.75	25.00	26.56	29.69
	WA	**28.13**	**50.00**	**15.63**	**18.75**	15.63	25.00	**32.81**

As expected the overall results based on fusing RGB and depth shows the considerable improvements in the performance across all the algorithm, while facial variations such as smile show overall highest recognition accuracy.

Here, we present the specific observation related to different variants in comparison with the previous evaluation result obtained only with depth, as follow:

- For the smile variation from session 1, the implementation of the PCA algorithm has marked the highest recognition rate of 95.31% for fusion approach using LI, EA filter compared to 90.63%, 89.06% recognition rate noted using LI, EA filter for only depth image. For HOG, the highest recognition rate of 98.44% for the fusion approach using EA-filter compared to 90.63%, the recognition rate was noted using EA filter for only depth image. For LBP, the highest recognition rate of 68.75% for the fusion approach using WA-filter compared to the 54.69% recognition rate noted using WA filter for only depth image. For LPQ, the highest recognition rate of

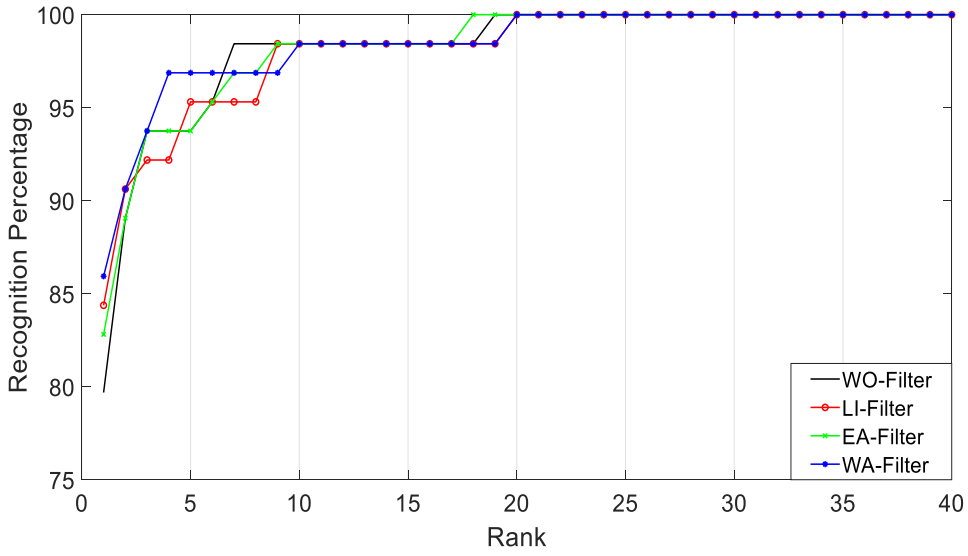
81.25% for fusion approach using LI, EA-filter compared to 62.50% recognition rate noted using LI, EA-filter for only depth image. Compared to the algorithms employed in this work, GIST, BSIF, LG demonstrates the consistently higher performance analysis as compared to the base results for depth images. For GIST, the highest recognition rate of 93.75% for fusion approach using WA-filter compared to 84.38%, recognition rate noted for depth image. For BSIF, the highest recognition rate of 96.88% for fusion approach using EA& WA filter compared to 84.38%, recognition rate noted for only depth image. Similarly, for LG, the highest recognition rate of 100% for fusion approach using LI, EA-filter compared to 92.19%, 90.63%, recognition rate noted using LI, EA filter for only depth image. This observation has been summarized in Table 4.15.

Table 4. 15: Representation of the maximum improvement on smile variation using RGB+Depth fusion

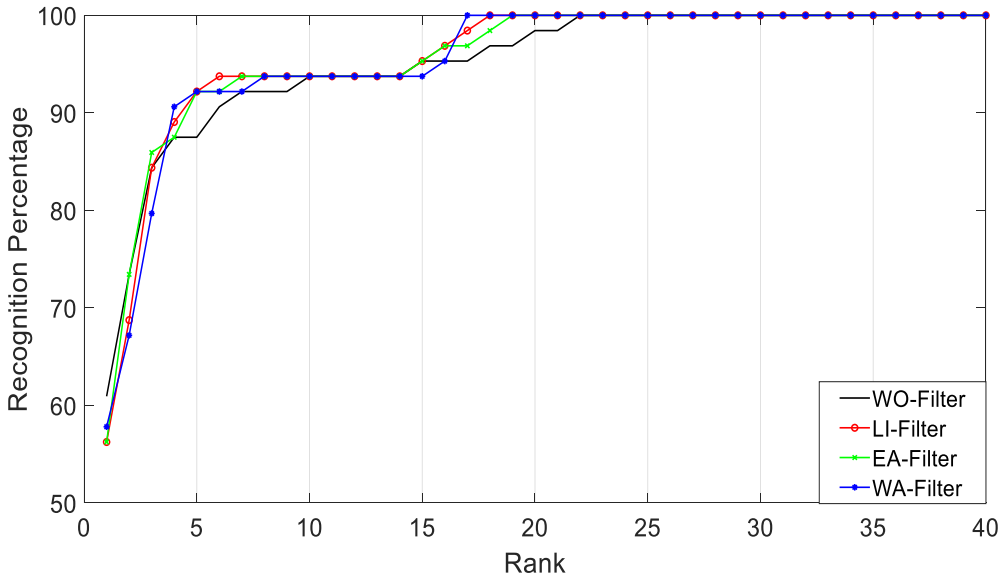
Feature Extraction Algorithm	Designed Filter	Recognition rate computed on Depth Images	Maximum Recognition rate computed on RGB+Depth Fused Images
PCA	LI	90.68	95.31
	WA	89.06	95.31
HOG	EA	90.63	98.44
LBP	WA	54.69	68.75
LPQ	LI	62.50	81.25
	EA	62.50	81.25
GIST	WA	84.38	93.75
BSIF	EA	84.38	96.88
	WA	84.38	96.88
LG	LI	92.19	100.00
	EA	90.63	100.00

- For 45° pose & -45° pose variation the trend is somewhat similar although the performance is marginal but the effect of the fusion methodology can be experienced for most of the algorithms. For 45° the computed recognition rate of 21.88% has been enhanced to 26.56% by application of WA filter using PCA. The maximum performance of 25% (with WA filter), 23.44% (with LI filter), 32.81% (with LI filter), 18.75% (with LI filter) for HOG, LPQ, GIST, & LogGabor, respectively, has been noted for the fused images among the applied filters. Similar marginal enhancement can also be noted for the other angular poses in both sessions, thus justifying the use of filters.

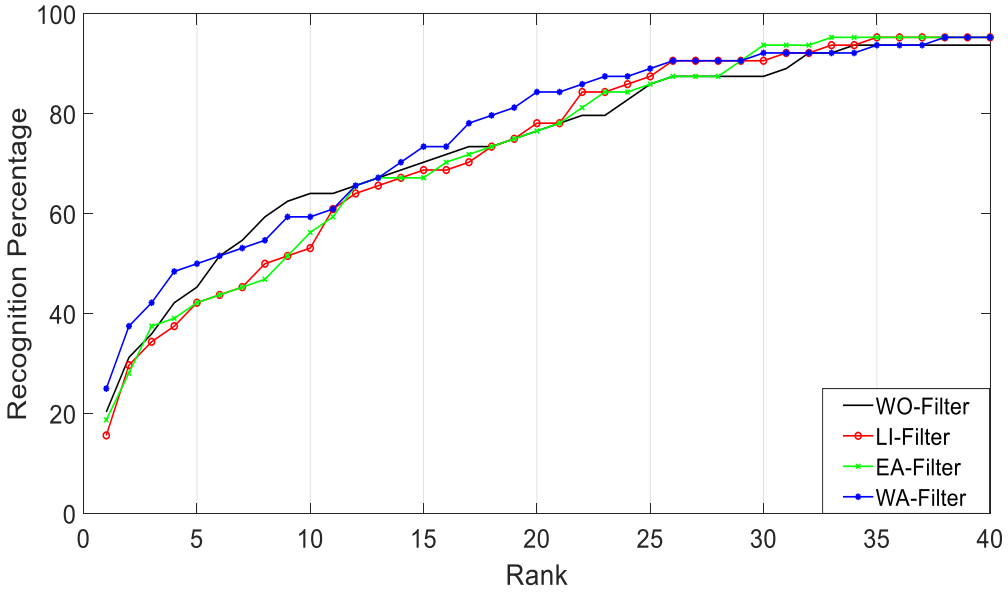
- The enhancement effect in session 2 for 0° pose variation(front face) at rank-5 is demonstrated in Figure 4.8 (a) - (n), representing the CMC curves of the 0° pose variation for various algorithms over depth and fused RGB-D image. Figure 4.8 (a) determines that the recognition rate is maintained as 73.44 by LI & EA filter using PCA. Further on fusion (Figure 4.8 (b)), the performance has been hiked to 92.19%. Using HOG (Figure 4.8 (c) & (d), the recognition rate of 65.63% (WO) has been increased to the maximum of 76.56% with WA filter, and the fusion approach has enhanced the performance in the range of 93% to 96.88%. Using LBP (Figure. 4.8 (e) & (f)) the performance of 18.75% has been an increase of 23.44 % with LI filter, and the maximum increase with the fusion approach is 40.62% with LI filter. Using LPQ (Figure 4.8 (g) & (h)), the lowest performance of 20.31% has been increased to 35.94% with EA filter, and the maximum increase of 54.69% can be seen with the fusion approach. Using GIST, the maximum increase can be seen in Figure 4.8 (i) at WA filter, i.e., 28.13% over 25% and the maximum performance on fusion is 50% with WA filter in Figure 4.8 (j). The maximum performance for BSIF can be seen in Figure 4.8 (k) with EA filter, i.e., 45.31 %, and the maximum enhancement due to fusion is 73.44% in Figure 4.8 (l). Similarly, the maximum performance obtained by using Log Gabor (Figure. 4.8 (m) & (n)) is 57.81% for depth images with WA filter, and it got enhanced to 79.69% on fusion.



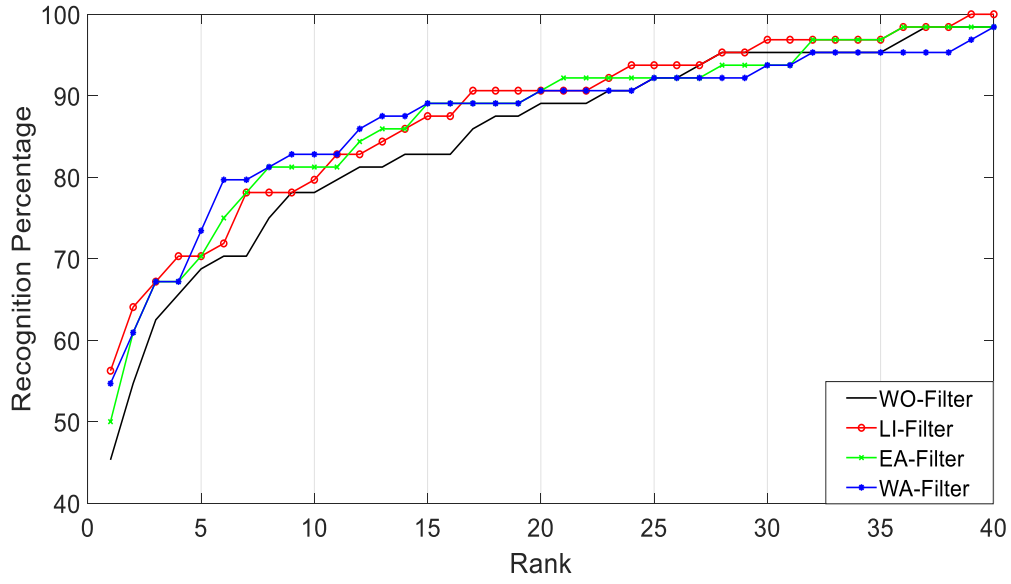
(a) Cumulative Match Curve (CMC) plots generated using PCA as feature extraction algorithm



(b) Cumulative Match Curve (CMC) plots generated using HOG as feature extraction algorithm

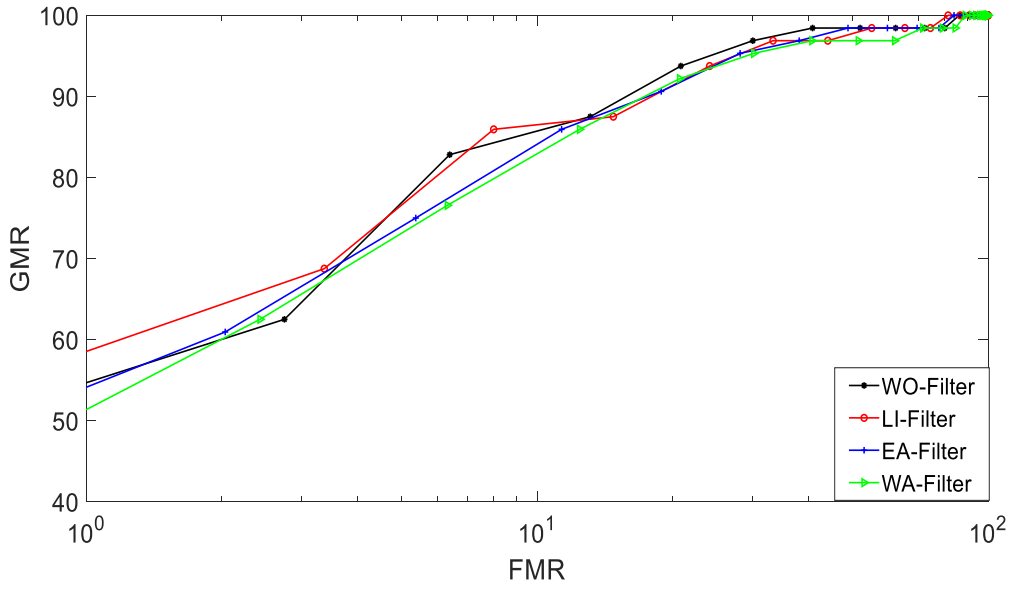


(c) Cumulative Match Curve (CMC) plots generated using GIST as feature extraction algorithm

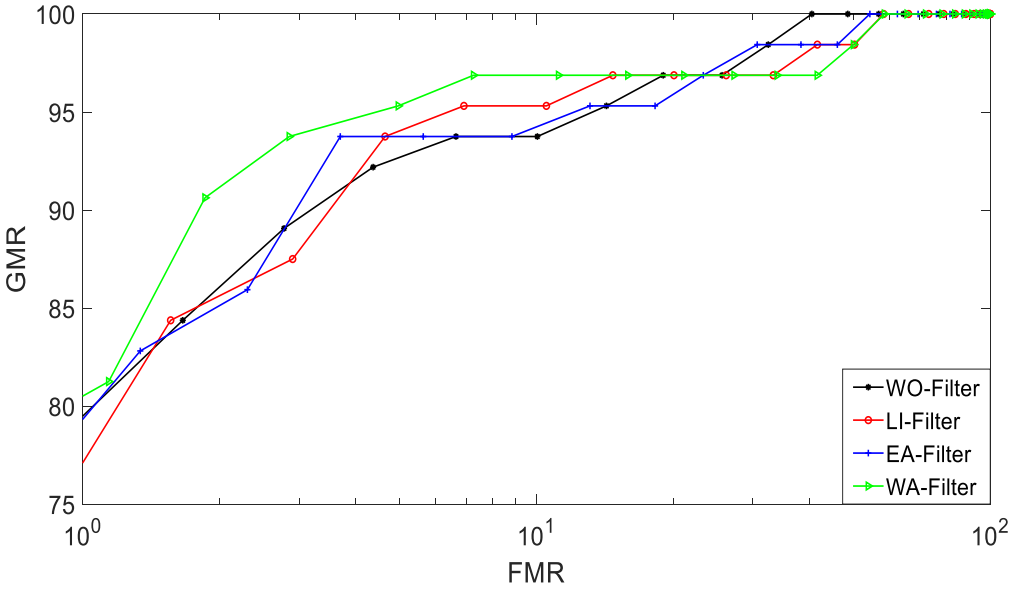


(d) Cumulative Match Curve (CMC) plots generated using BSIF as feature extraction algorithm

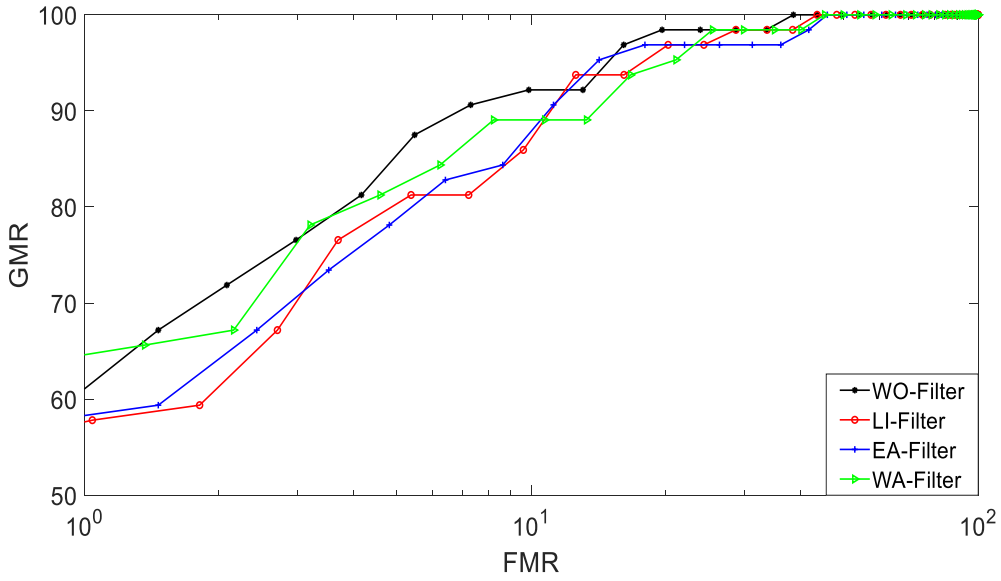
Fig 4. 6: Cumulative Match Curve (CMC) plots demonstrate the face recognition performance on Fused (RGB + Depth) image using three different filters and without filter. The best results related to facial variation smile is presented here in (a)-(d)



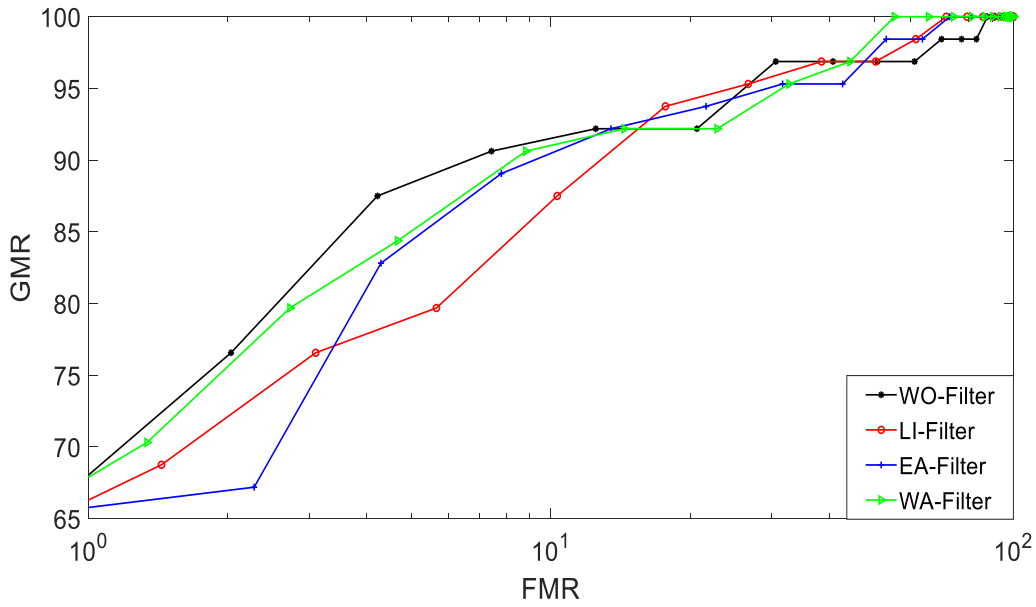
(a) Receiver Operating Curve (ROC) plots generated using PCA as feature extraction algorithm



(b) Receiver Operating Curve (ROC) plots generated using HOG as feature extraction algorithm

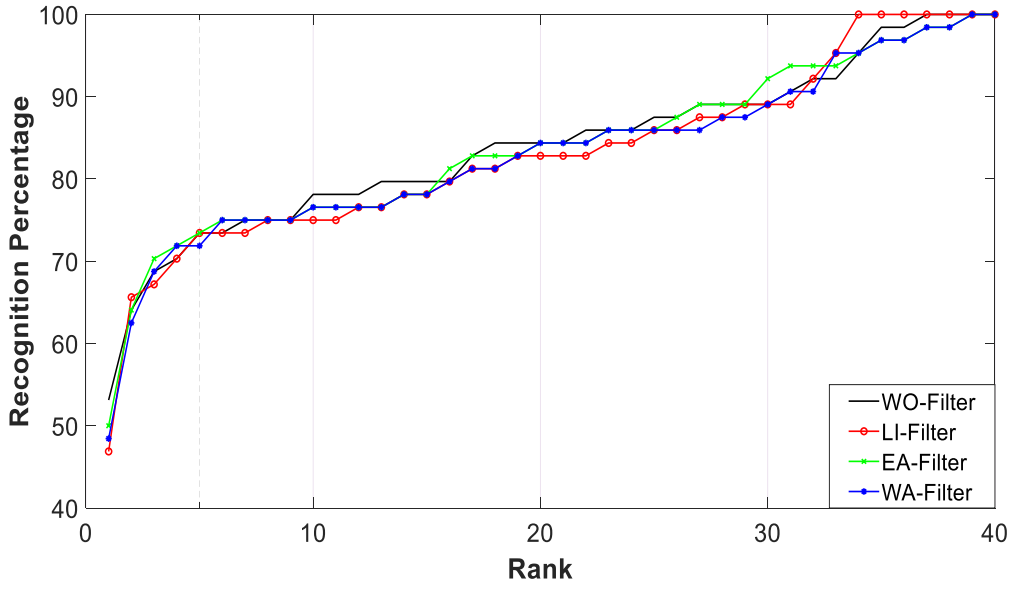


(c) Receiver Operating Curve (ROC) plots generated using GIST as feature extraction algorithm

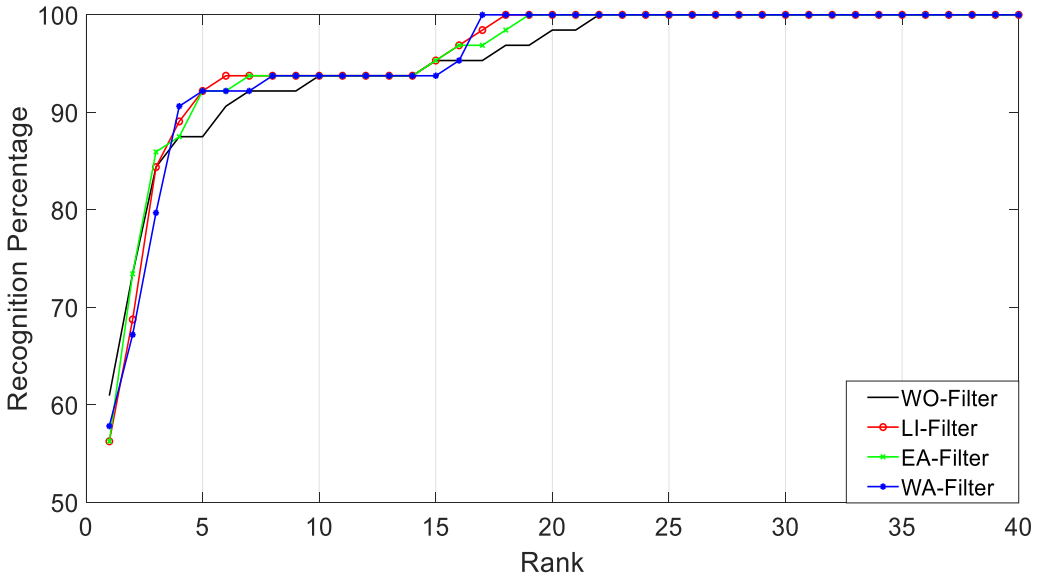


(d) Receiver Operating Curve (ROC) plots generated using BSIF as feature extraction algorithm

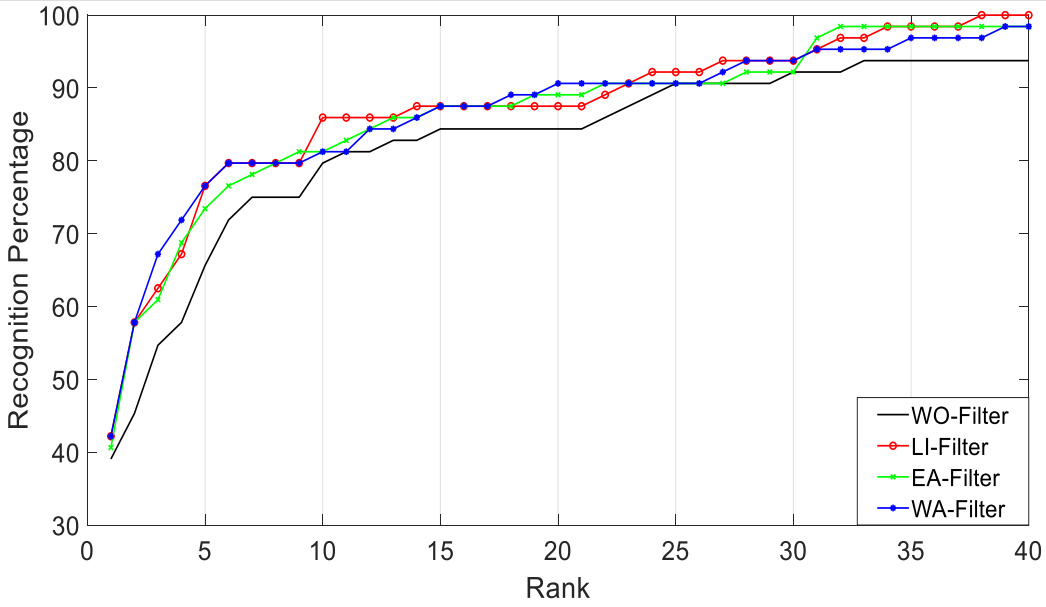
Fig 4. 7: Receiver Operating Curve (ROC) plots demonstrate the face recognition performance on Fused (RGB + Depth) image using three different filters and without filter. The best results related to facial variant smile is presented here in (a) - (d)



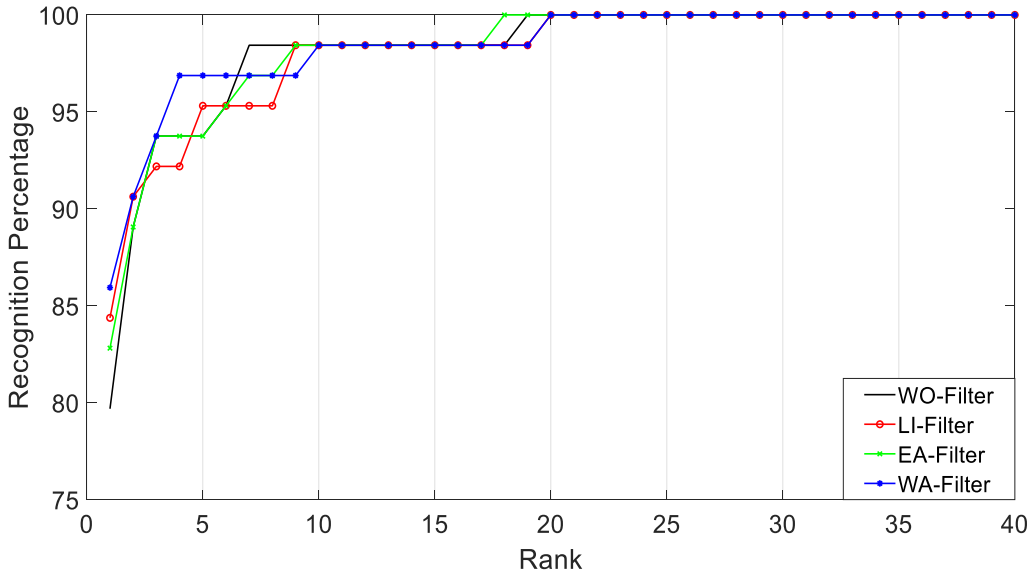
(a) Cumulative Match Curve (CMC) plots generated using PCA as feature extraction algorithm for depth images



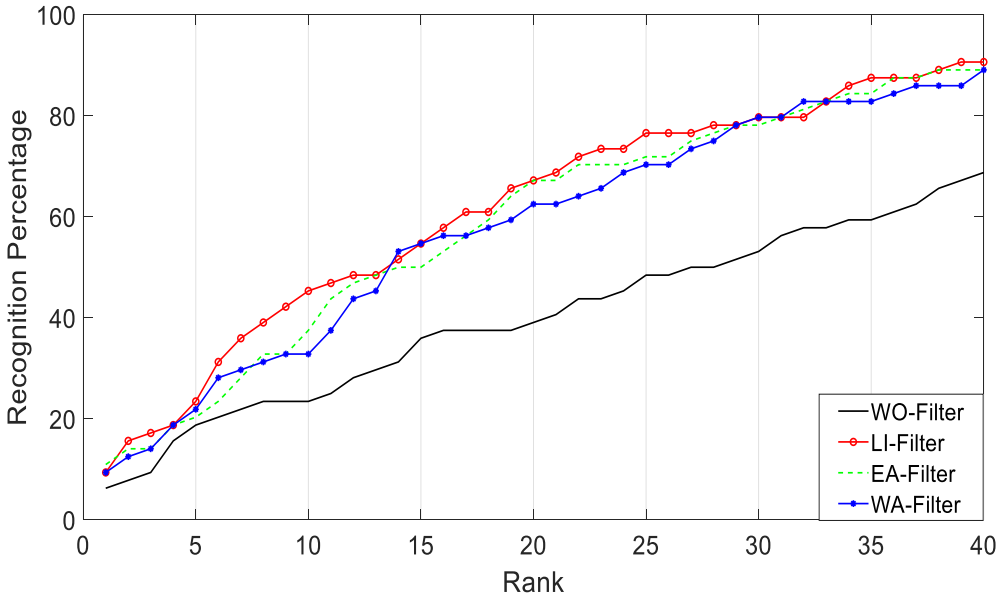
(b) Cumulative Match Curve (CMC) plots generated using PCA as feature extraction algorithm for RGB + depth images



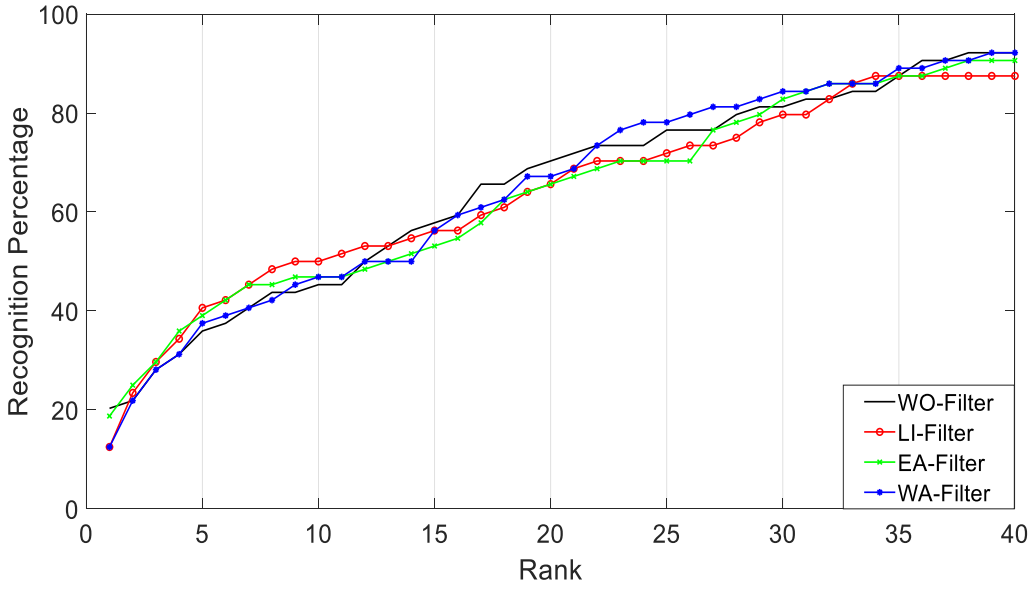
(c) Cumulative Match Curve (CMC) plots generated using HOG as feature extraction algorithm for depth images



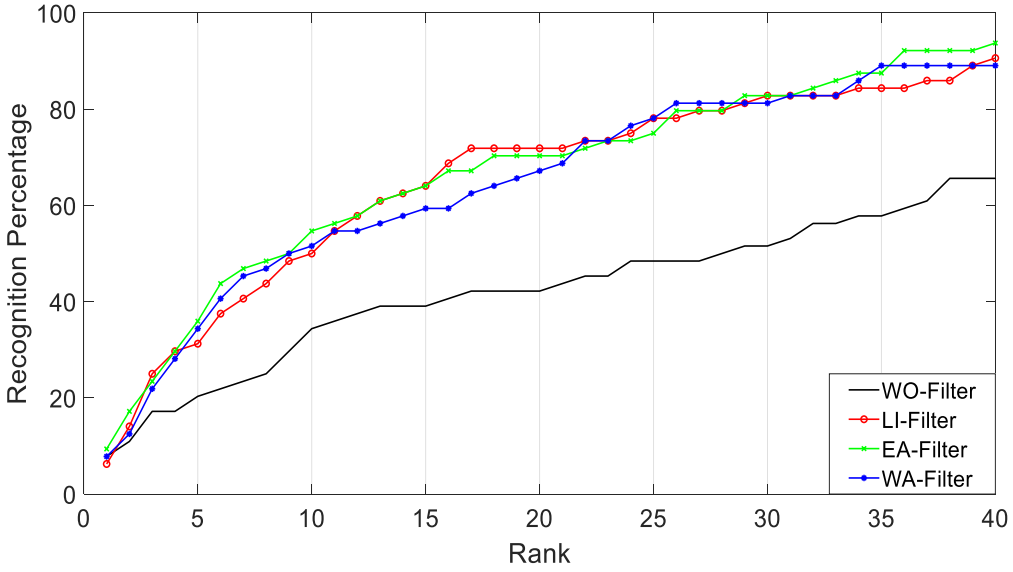
(d) Cumulative Match Curve (CMC) plots generated using HOG as feature extraction algorithm for RGB + depth images



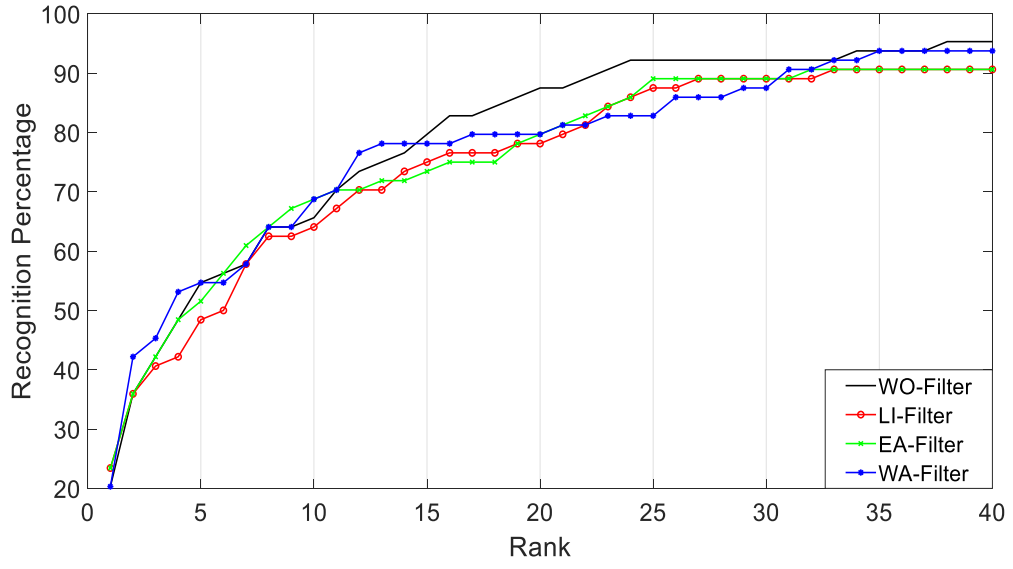
(e) Cumulative Match Curve (CMC) plots generated using LBP as feature extraction algorithm for depth images



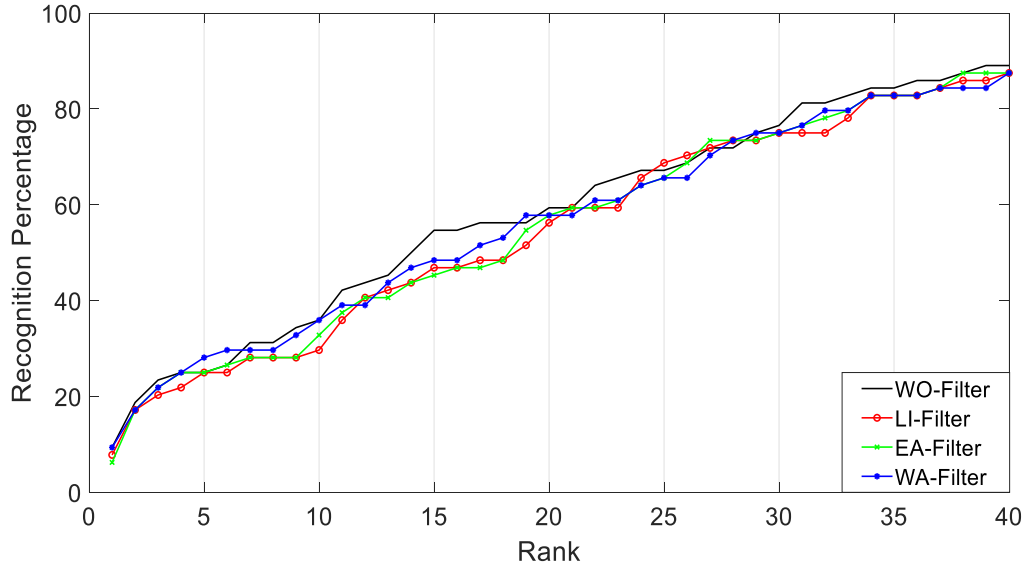
(f) Cumulative Match Curve (CMC) plots generated using LBP as feature extraction algorithm for RGB + depth images



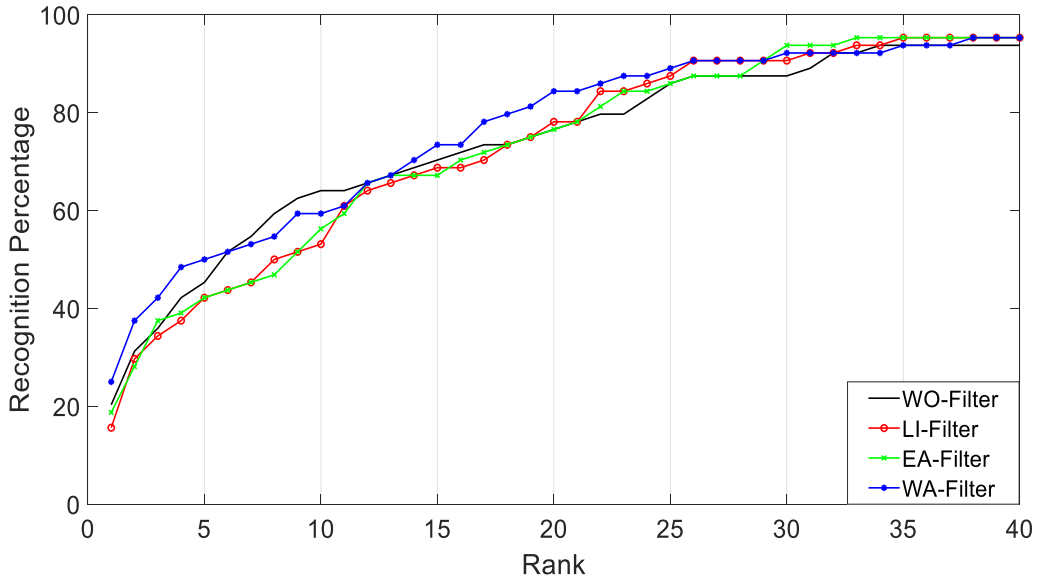
(g) Cumulative Match Curve (CMC) plots generated using LPQ as feature extraction algorithm for depth images



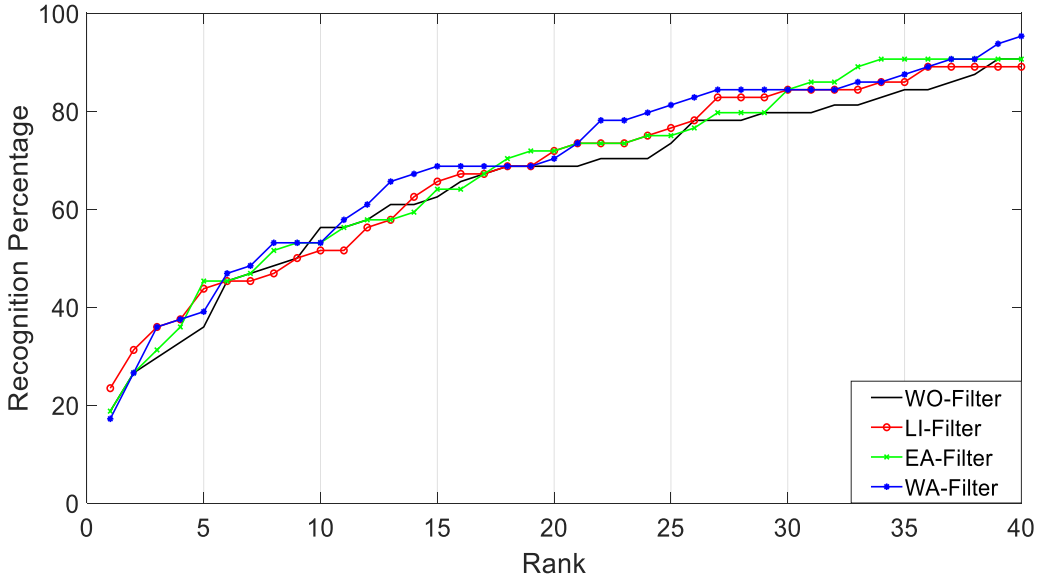
(h) Cumulative Match Curve (CMC) plots generated using LPQ as feature extraction algorithm for RGB + depth images



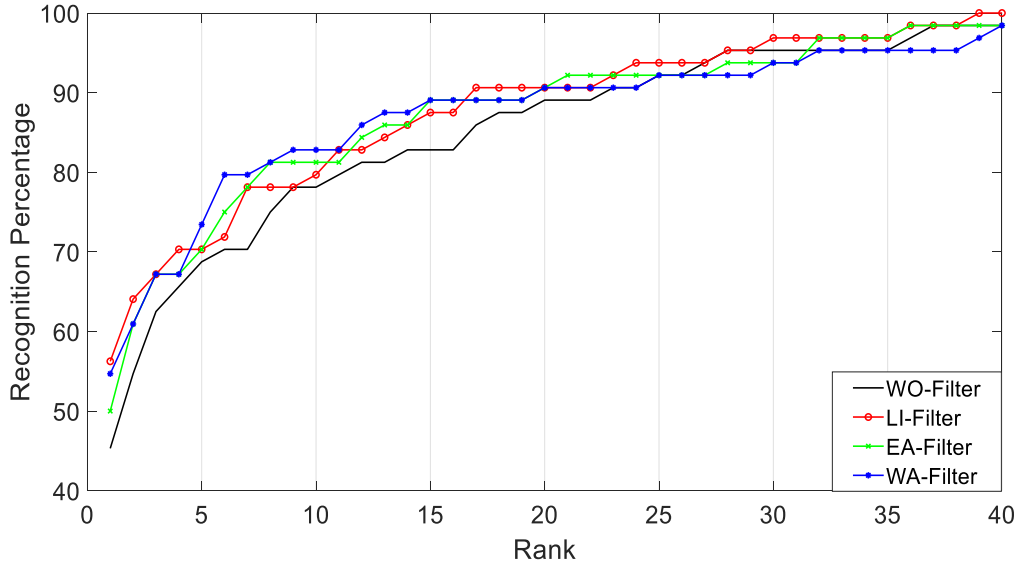
(i) Cumulative Match Curve (CMC) plots generated using GIST as feature extraction algorithm for depth images



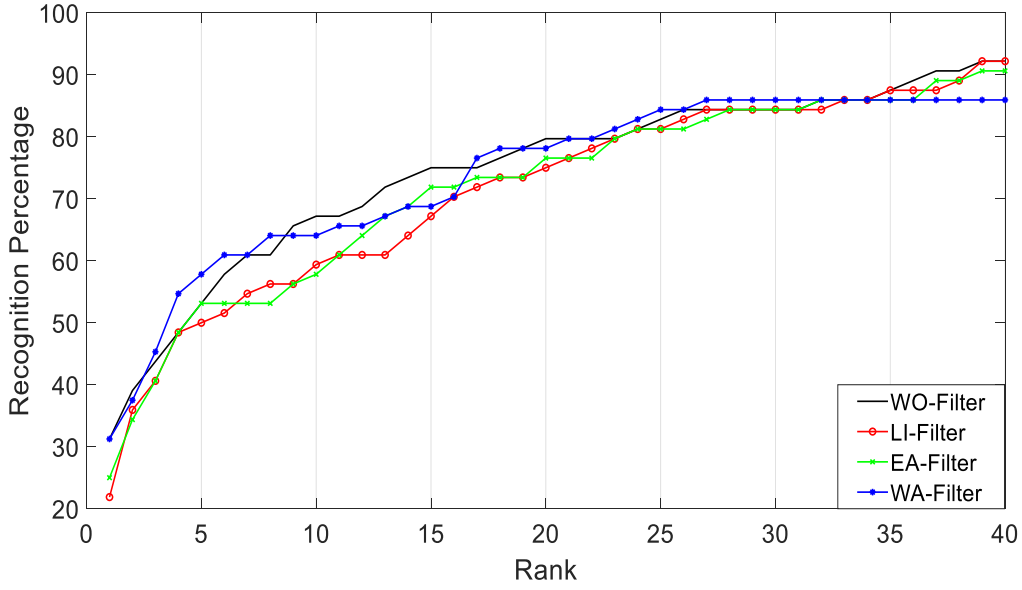
(j) Cumulative Match Curve (CMC) plots generated using GIST as feature extraction algorithm for RGB + depth images



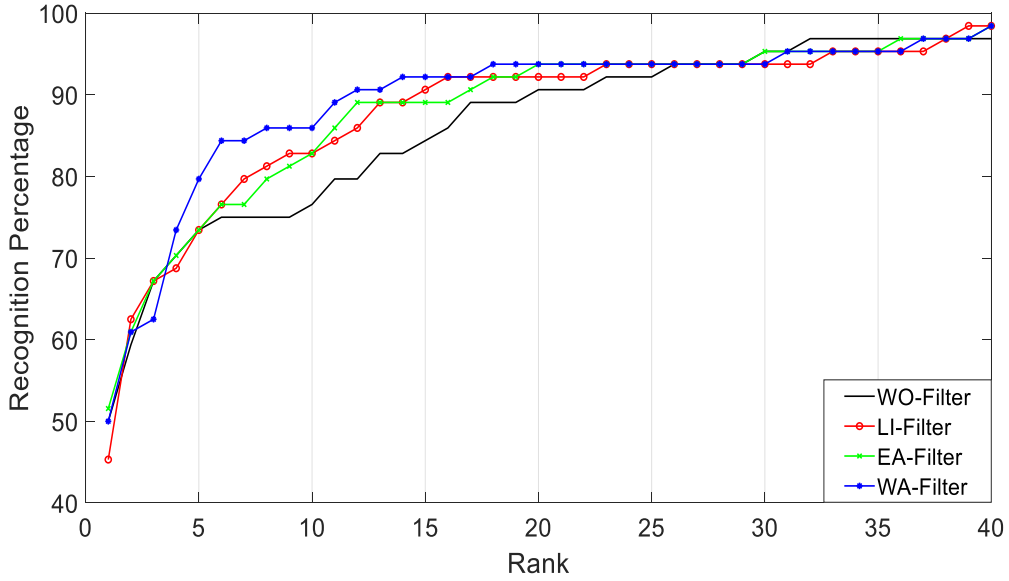
(k) Cumulative Match Curve (CMC) plots generated using BSIF as feature extraction algorithm for depth images



(l) Cumulative Match Curve (CMC) plots generated using BSIF as feature extraction algorithm for RGB + depth images



(m) Cumulative Match Curve (CMC) plots generated using LG as feature extraction algorithm for depth images



(n) Cumulative Match Curve (CMC) plots generated using LG as feature extraction algorithm for RGB + depth images

Fig. 4.8: Cumulative Match Curve (CMC) plots demonstrate the face recognition performance on Depth and Fused (RGB + Depth) image using three different filters and without filter for various state-of-the-art algorithms. The best results related to facial variant 0^{o} pose (session 2) are presented here in (a) – (n)

4.4.3 Evaluation 3: Fusion Based On Scores

Similar to the previous section, we present the evaluation results based on score level fusion. Precisely, we fused the scores of the best performing algorithm based on the previous two evaluation results. Based on the previous results, we employed PCA and HOG score level fusion approach to demonstrate the set of results. Further, the results were repeated for depth images alone and fusion of RGB with depth images. The scores are fused using a simple sum rule to demonstrate the results for session 1 and session 2. Table 4.16, and Table 4.17, presents the recognition rate at Rank-5 for depth images and the fusion of RGB with depth images.

The results obtained with the score level fusion are found to be better in almost all the evaluation results performed in this experiment with some exceptional cases. Hence we present the observation related to some varients in this section:

- In comparison with the smile of Session 1 (PCA and HOG Column), it is observed that the base result for PCA without filter is 89.06% and for HOG is 93.75%, and it has increased maximum to 90.63% for PCA and 96.88% for HOG using filters whereas by fusing the scores of PCA and HOG for the depth images the base results obtained are 93.75% (without filter), while but with the application of WA-filter the performance has enhanced to 98.44%. Similarly, the maximum value of 98.44% has been noted for the RGB & depth image fusion.

- For 0° pose variation of session 2, an enhancement in results can be seen. Here the base results (refer Table 4.6) for depth images using PCA and HOG without filters are 73.44% and 65.63%, respectively, which has been enhanced maximum to 76.56% for HOG using WA filter (refer Table 4.8), and the further enhancement up to 78.13% by fusing the scores of the two algorithms. For the fused RGB-D (refer Table 4.12) images, the base result of PCA is 87.50% while HOG is 93.75%, which is increased to 92.19% for EA filter (refer Table 4.13) and 96.88% for WA filter (refer Table 4.14), for PCA, and HOG, respectively, using filters. These results are further enhanced to 98.44% (refer Table 4.17) by WA filter, with score

level fusion. A similar trend is obtained for most of the variations in the table, with some exceptions.

Table 4. 16: Recognition rates of depth images computed at Rank-5 after score level fusion of PCA+HOG, with its implicit designed filters

Variation	Session 1				Session 2			
	WO	**LI**	**EA**	**WA**	**WO**	**LI**	**EA**	**WA**
0°	-	-	-	-	71.88	**78.13**	**76.56**	**78.13**
45°	26.56	**26.56**	25.00	**26.56**	25.00	23.44	23.44	23.44
90°	17.19	15.63	**20.31**	**21.88**	15.63	14.06	**17.19**	14.06
-45°	15.63	**18.75**	**18.75**	**23.44**	9.38	**12.50**	7.81	**12.50**
-90°	12.50	**20.31**	**18.75**	**17.19**	9.38	**12.50**	**9.38**	**9.38**
Smile	93.75	**95.31**	**96.88**	**98.44**	78.13	**82.81**	**82.81**	76.56
Eyes Closed	92.19	90.63	**92.19**	90.63	76.56	**85.94**	**85.94**	**81.25**
Paper on face	35.94	25.00	34.38	32.81	48.44	45.31	**48.44**	46.88

Table 4. 17: Recognition rates of RGB-D (Fused) images computed at Rank 5 after score level fusion of PCA+HOG, with its implicit designed filters designed filters

Variation	Session 1				Session 2			
	WO	LI	EA	WA	WO	LI	EA	WA
0°	-	-	-	-	98.44	96.88	96.88	**98.44**
45°	32.81	**32.81**	**37.50**	**35.94**	28.13	25.00	25.00	26.56
90°	18.75	15.63	**18.75**	17.19	17.19	15.63	15.63	**20.31**
-45°	28.13	**28.13**	**32.81**	**29.69**	17.19	**17.19**	**20.31**	**18.75**
-90°	14.06	**14.06**	12.50	**17.19**	4.69	**6.25**	**7.81**	**7.81**
Smile	98.44	**98.44**	**98.44**	96.88	95.31	**95.31**	**95.31**	**95.31**
Eyes Closed	96.88	**96.88**	**96.88**	**98.44**	95.31	93.75	93.75	93.75
Paper on face	64.06	43.75	60.94	59.38	54.69	**56.25**	**56.25**	**56.25**

To summarize, employing filters based on kernel function indicates the improvement in the performance accuracy, demonstrating the applicability of our approach for 3D depth images; the presence of holes in the image significantly degrades the quality and overall performance of the biometric face recognition system. The improvement has been noted for almost all poses and angles, with some exceptions where the full-face triangle is not available. It may be noted that the other published results also indicate the poor performance for the angular variation and occlusions. Hence the results obtained are in unison with the published results in the literature [11]. The implementation of the above discussed fusion strategies has further shown the improvement in the computed results.

CHAPTER 5:

ADVANCED CLASSIFICATION APPROACH FOR PERFORMANCE ANALYSIS

CHAPTER 5

5.1: Classification With Collaborative Representation

Image/pattern recognition has attracted attention due to various practical applications in face recognition, medical diagnoses, etc., and accordingly, multiple methods are developed for attending the classification task. The conventional methods applied for image classification problems are nearest neighbor (NN) [138] and Nearest subspace (NS) [139], where the testing sample is represented with the training samples and then assigned to the nearest class. Furthermore, the sparse representation based classification (SRC) method has been developed by Wright et al. [140], which gives the sparse representation of the testing sample with the training samples and assigns the sample to the class with the least residual error. Further, Collaborative Representation Classifier (CRC) [85] has emerged as a robust feature classification method in the face recognition domain. It is an extended version of the Sparse Representation Classifier (SRC), where the l_1-norm in SRC is replaced by l_2-norm. This approach computes the maximum likelihood ratio between the test sample image and the other classes in a joint manner and classifies the test sample to a class with the least reconstruction error. In order to perform the final feature classification, the maximum likelihood of the test sample is computed against the other classes from the training set.

The study has been performed on the GU-RGB-D [141] database and IIIT-D [59] database. The proposed hole-filling techniques are used in the pre-processing stage for filling the holes in the depth images in both databases. The resultant depth images are then fused with the RGB image (we used the grayscale image) using 2D-Discrete Wavelet transform. The fused composite images corresponding to the training set and the testing set is processed to extract features using feature extraction algorithms mentioned in the previous chapter. The study has been extended for feature extraction using Convolution Neural Network.

Further, the comparison based on collaborative subspace is implemented, and the set of obtained scores are treated as comparison scores to either accept or reject the subject. Employing the proposed scheme based on fusing the depth and RGB image, followed by a collaborative representation classifier, presents the recognition system's improvement and applicability of our RGBD face recognition approach.

5.1.1 Contributions

This chapter presents the proposed scheme where the RGB images and filtered depth (filtered using the designed hole filling filters described in chapter 4) images are fused using 2D-Discrete wavelet transform, followed by a collaborative representation classifier for RGBD face recognition. To the best of our knowledge, CRC is introduced for the first time for RGBD face recognition. Further, to demonstrate our study and the significance of using the hole filling method of kernel size, we present an extensive experimental evaluation based on our GU-RGB-D database and publicly available IIIT-D database. The GU-RGB-D database comprises a series of challenges such a pose variation, occlusion, and expression, while the IIIT-D database is frontal, having slight variations in the pose. Also, to present the significance of our approach, the proposed scheme is demonstrated on eight different state-of-the-art feature extraction methods, including Histogram of Oriented Gradient (HOG) [64], Local Phase Quantization (LPQ) [70], GIST [72], Local Binary Pattern (LBP) [142], LogGabor [77], Principal Component Analysis (PCA) [14], Binarized Statistical Image Features (BSIF) [75], and deep convolutional neural network features extracted at `conv5' layer, so as to have a fair comparison, with the algorithms and the three hole-filling filters employed in this work. All the evaluation results are presented in the form of verification rate and recognition rate at Rank5 using the GU-RGB-D and IIIT-D databases. The contributions can be summarised as follows:

- Present a proposed scheme that combines the RGB and depth image (after hole filling) using 2D-Discrete wavelet transform, which is followed by a robust collaborative representation classifier (CRC) for RGBD based face recognition.

- Present extensive experimental evaluation based on designed hole filling filters on our GU-RGB-D and IIIt-D databases.

- An experiment in the form of verification and recognition rate is performed on eight different feature extraction methods such as Local Phase Quantization (LPQ), Local Binary Pattern (LBP), Histogram of Oriented Gradient (HOG), GIST, LogGabor, Binarized Statistical Image Features (BSIF), Principal

Component Analysis (PCA), and deep convolutional neural network features extracted at `conv5' layer to demonstrate the applicability of our approach for improved performance analysis.

The rest of section 5.1 is structured as followed; sub-section 5.1.2 describes the proposed RGBD face recognition scheme using wavelet transform and Collaborative representation classifier. Sub-section 5.1.3 introduces the experimental evaluation results in the form of verification and recognition rate using eight different state-of-the-art methods to present the potential of our scheme for RGB-D face recognition.

5.1.2 Scheme of Evaluation

This section presents the detailed proposed scheme for RGB-D face recognition employed in this work. Figure 5.1 illustrates the conceptual representation of our proposed scheme. In general, this section presents the proposed scheme for comparing train RGB-D face image against the test RGB-D face image. The samples corresponding to the training set and testing set are disjoint. The training set and testing set consist of RGB-D images processed using wavelet transform to form a composite image. Specifically, the depth image is processed independently through the hole filling filters, which is then fused with RGB image (we used grayscale image) using 2D-Discrete wavelet transform to form a training set. In a similar line, the test composite RGB-D image is formed using wavelet transform to form a testing set. The composite images corresponding to the training set and testing set are processed to extract features using feature extraction methods (we used eight different feature extraction methods) and perform the comparison based on collaborative subspace. The set of obtained scores are treated as comparison scores, to either accept or reject the subject. A detailed discussion of the mathematics of the proposed scheme employed in this work is presented as follows:

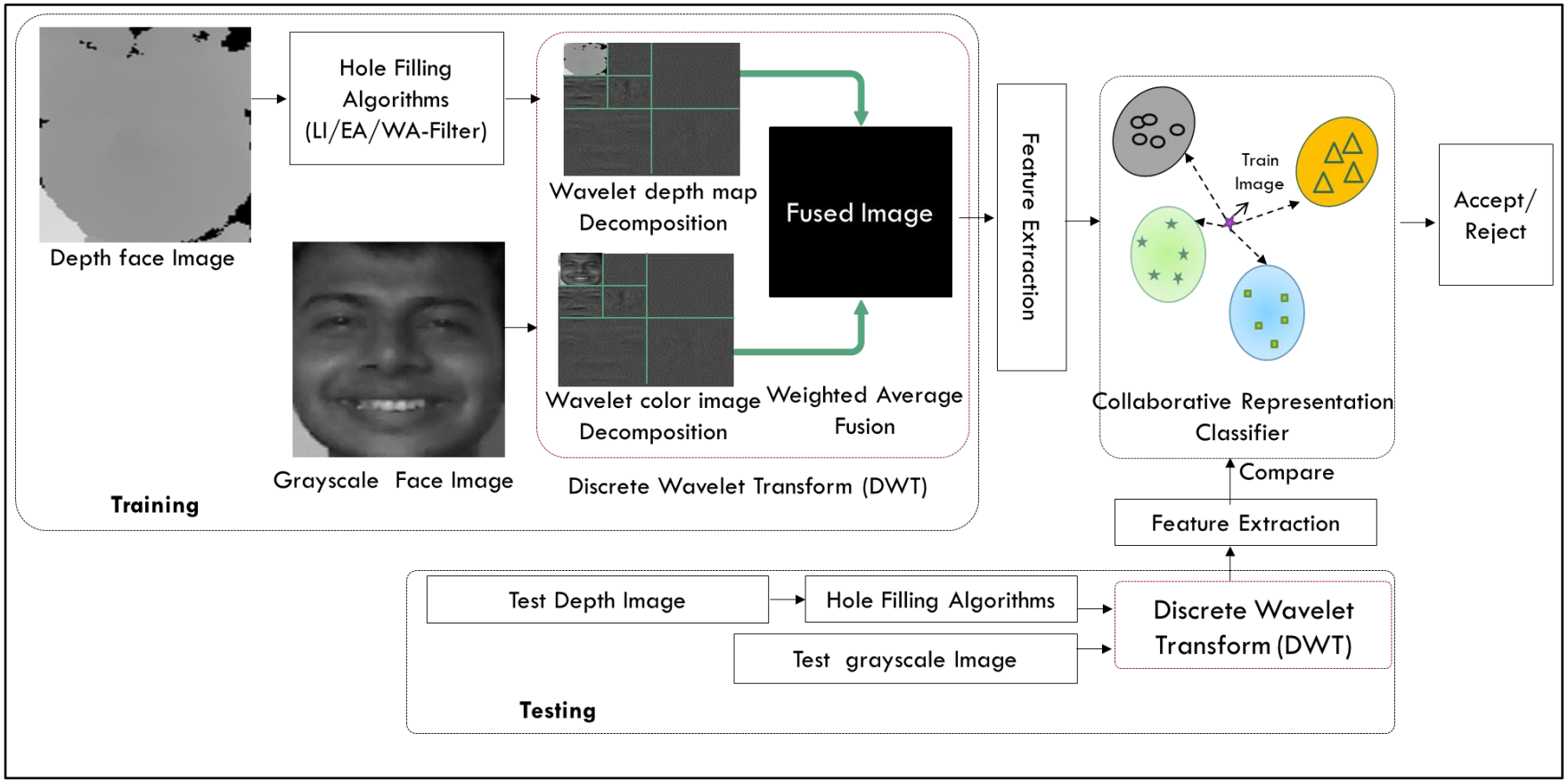


Fig 5.1: Schematic block diagram illustrating the proposed framework based on DWT and CRC

Let the $\bar{u}(x,y)$ represents the depth image obtained after applying the filter, and *R(x, y)* represents the color image (we have converted the RGB image to a grayscale image in this work). Using the depth and color image, we employ 2-Level Discrete Wavelet Transform (DCT) to process the complementary image information into a single composite image [143][144]. The idea is to extract another set of features in a composite manner. In this work, we have used 2-Level Discrete Wavelet Transform carried out using Haar mother wavelet function. The output of Wavelet decomposition provides the wavelet coefficients in seven sub-band images that correspond to one approximation, two horizontal, two vertical, and two diagonal coefficient details. Mathematically, 2-Level DCT can be represented by using Equation 5.1.

$$Q = \{a, h, h', v, v', d, d'\} \tag{5.1}$$

where *a* - wavelet coefficient represents the approximation detail; *h, h'* - wavelet coefficient represents the horizontal details; *v, v'* - wavelet coefficient represents the vertical details; and *d, d'* - wavelet coefficient represents the horizontal details. The final composite image comprises the depth and color image information is obtained by performing the weighted average of wavelet coefficients corresponding to depth and color image. It can be mathematically expressed as follows using Equation 5.2.

$$k = 0.5 * \begin{Bmatrix} w_1 * (a_D + a_C), w_2 * (h_D + d_C), w_3 * (h'_D + h'_C), w_4 * (v_D + v_C), \\ w_5 * (v'_D + v'_C),\ w_6 * (d_D + d_C), w_7 * (d'_D + d'_C) \end{Bmatrix} \tag{5.2}$$

where { $a_D, h_D, h'_D, v_D, v'_D, d_D, d'_D$ } represents the wavelet coefficients belongs to depth images and $\{a_C, h_C, h'_C, v_C, v'_C, d_C, d'_C\}$ represents the wavelet coefficients belongs to the color image. The final composite image is then obtained after performing an inverse wavelet transform on the output of Equation 5.2, and we represent this output by variable *K(x, y)*.

Further, to extract dominant local and global features from the composite image, texture descriptor methods are employed. In general, we have employed eight feature extraction methods such as Local Binary Pattern (LBP), Histogram of Oriented Gradient (HOG) , Local Phase Quantization (LPQ), Binarized Statistical Image Features (BSIF), LogGabor, Principal Component Analysis (PCA), GIST, and deep convolutional neural network features extracted at `conv5' layer. In this work, we independently employed the feature extraction method to perform the evaluation results.

Further, we make use of the Collaborative Representation Classifier in our work to compare the training sample against the testing sample for RGB-D face recognition. In CRC, a test image is generated as a collaborative representation that belongs to training samples. The algorithm then classifies a test image to a class that has a minimum distance between the collaborative represented test image and its projection within the class. In this work, the features of composite RGB-D images are obtained using the above texture descriptor methods, and the extracted feature vectors can be represented by Equation 5.3 as

$$Z = \{Z_1, Z_2, \ldots, Z_b\} \in \mathbb{R}^{m \times N} \tag{5.3}$$

where: b - total no of classes, m - dimension of the feature vector of each image, N - total number of images across b classes.

The expression for CRC can be represented by a general model given by Equation 5.4.

$$\alpha' = \arg\min_{\alpha}(\| x - Z\alpha \|_2^2 + \mu \| \alpha \|_2^2) \tag{5.4}$$

where $\alpha = \alpha_1 \ldots. \alpha_b$ is the coefficient vector, μ is the regularization parameter, and x is the input test image is given by $x \in \mathbb{R}^m$.

The solution of Equation 5.4 by Least Square Method is given by Equation 5.5

$$\alpha' = (Z^T Z + \mu . I)^{-1} Z^T x \quad (5.5)$$

The residual between the input test image and each class is given by Equation 5.6

$$r_s = \frac{\|x - Z_s \propto_s' \|_2^2}{\|\propto_s' \|_2^2} \quad (5.6)$$

where: r_s - residual corresponding to each class, s – 1, 2, 3, . . . , b

The class is attributed to each input test image by computing the minimum residual between the input test image and each class given by Equation 5.7.

$$Class\ (g) = arg \min_s r_s \quad (5.7)$$

All the expressions in this chapter are simplified for the sake of convenience. The detailed information regarding the method is available in the paper [32].

5.1.3 Experiment Protocol And Results

This section presents the experimental evaluation protocol and related experiments performed in this work for RGBD face recognition. Basically, the entire results are based on the proposed scheme for RGB-D based face recognition. The approach combines filtered depth image with the color image using 2-Discrete Wavelet Transform to form a composite image, followed by a collaborative representation classifier for RGBD face recognition. The experimental results are based on the GU-RGB-D database and IIIT-D publicly available database consisting of depth and color images collected from the Kinect sensor. To demonstrate the applicability of using kernel-based filtering approach along with the proposed scheme, the systematic results on eight different state-of-the-art feature extraction

methods like Local Phase Quantization (LPQ), Local Binary Pattern (LBP), Histogram of Oriented Gradient (HOG), GIST, LogGabor (LG), Binarized Statistical Image Features (BSIF), Principal Component Analysis (PCA), and deep convolutional neural network features obtained at layer `CONV5' are presented. The said evaluation results are presented in the form of recognition rate in tabular and graphical form. Since the results are obtained using two different databases, the evaluation protocol and results, along with the discussions, are presented in the following two subsections.

Table 5. 1: Recognition rate at Rank-5 on GU-RGB-D and IIIT-D face database using WO-Filter across eight different feature descriptor methods.

METHODS	GU-RGB-D							IIIT-D
	45	**90**	**-45**	**-90**	**Smile**	**Close Eyes**	**Occlusion**	**Frontal Random Images**
LBP-CRC	14.84	8.59	17.19	8.59	65.63	57.81	13.28	7.17
LPQ-CRC	12.50	8.59	16.41	9.38	75.78	75.00	12.5	11.32
HOG-CRC	12.50	13.28	7.03	11.72	**94.53**	95.31	32.81	11.89
GIST-CRC	10.94	8.59	11.72	7.81	41.41	35.94	18.75	9.43
LG-CRC	**24.22**	14.06	19.53	**15.63**	95.31	**97.66**	33.59	64.91
BSIF-CRC	21.09	**17.19**	14.06	14.84	92.97	94.53	25.78	20.19
PCA-CRC	18.05	8.59	12.50	9.38	89.84	91.41	18.75	54.53
CONV5-CRC	21.88	**17.97**	**20.31**	**15.63**	92.19	96.09	**48.44**	**96.42**

Table 5. 2: Recognition rate at Rank-5 on GU-RGB-D and IIIT-D face database using LI-Filter across eight different feature descriptor methods.

METHODS	GU-RGB-D							IIIT-D
	45	90	-45	-90	Smile	Close Eyes	Occlusion	Frontal Random Images
LBP-CRC	25.78	10.16	21.88	14.84	76.56	82.03	19.53	4.91
LPQ-CRC	33.59	13.28	25.78	15.63	**100.00**	98.44	38.28	11.89
HOG-CRC	14.06	10.16	14.84	11.72	98.44	99.22	**73.44**	16.056
GIST-CRC	19.53	13.28	13.28	14.84	92.97	89.06	35.16	8.87
LG-CRC	**38.28**	15.63	18.75	**17.19**	99.22	**100.00**	46.09	63.21
BSIF-CRC	31.25	9.38	**28.91**	15.63	**100.00**	**100.00**	68.75	18.30
PCA-CRC	25.00	13.28	14.06	10.16	95.31	98.44	10.94	54.34
CONV5-CRC	33.59	**17.97**	26.56	14.06	**100.00**	**100.00**	71.88	**96.98**

Table 5. 3: Recognition rate at Rank-5 on GU-RGB-D and IIIT-D face database using EA-Filter across eight different feature descriptor methods.

METHODS	GU-RGB-D							IIIT-D
	45	90	-45	-90	Smile	Close Eyes	Occlusion	Frontal Random Images
LBP-CRC	28.13	10.94	25.78	12.5	77.34	84.38	26.56	5.47
LPQ-CRC	32.81	14.06	25.78	15.63	**100.00**	98.44	36.72	11.89
HOG-CRC	14.06	10.94	14.06	10.16	98.44	**100.00**	70.31	11.32
GIST-CRC	11.72	10.16	11.72	8.59	88.28	87.5	29.69	8.68
LG-CRC	**38.28**	15.63	18.75	**17.19**	99.22	**100.00**	45.31	63.02
BSIF-CRC	31.25	9.38	**29.69**	14.84	**100.00**	99.22	**71.09**	19.62
PCA-CRC	25.00	13.28	14.06	10.16	95.31	98.00	10.94	53.96
CONV5-CRC	32.81	**17.97**	27.34	14.06	99.22	**100.00**	**71.88**	**96.79**

Table 5. 4: Recognition rate at Rank-5 on GU-RGB-D and IIIT-D face database using WA-Filter across eight different feature descriptor methods.

METHODS	GU-RGB-D							IIIT-D
	45	90	-45	-90	Smile	Close Eyes	Occlusion	Frontal Random Images
LBP-CRC	24.22	13.28	24.22	14.06	77.34	80.47	24.22	5.09
LPQ-CRC	32.03	14.84	26.56	15.63	**100.00**	98.44	35.16	13.4
HOG-CRC	13.28	9.38	13.28	10.16	98.44	99.22	**72.66**	17.92
GIST-CRC	13.28	10.94	10.94	9.38	89.06	92.97	38.28	8.87
LG-CRC	**38.28**	15.63	18.75	**17.19**	99.22	**100.00**	38.28	63.58
BSIF-CRC	32.03	9.38	**28.13**	14.06	**100.00**	**100.00**	68.75	19.43
PCA-CRC	25	13.28	14.06	10.16	95.31	98.44	10.94	53.77
CONV5-CRC	32.03	**17.97**	26.56	14.06	**100.00**	**100.00**	**72.66**	**96.42**

5.1.3.1 Evaluation Based on GU-RGB-D Database

The evaluation results based on the GU-RGB-D face database are presented in this section. The said face database is partitioned into a training and testing set. The training set consists of 64 subject correspondings to front face (0º pose), including the samples from session 1 and session 2, while the test set consists of 64 subjects belongs to 45º, -45 º, 90 º, -90 º, smile, eye closed, paper occlusion variations from session 1 and session 2 operated independently using eight different feature extraction methods mentioned above. Based on this evaluation protocol (Table 5.5), this section presents the experimental result independently with three different designed hole filling filters discussed in chapter 4 and the benchmark results performed without employing filter to have a fair comparison with our proposed approach. Table 5.2, 5.3, 5.4, presents the recognition rate at Rank-5, and Figure 5.2 presents graphical representation results in the form of Cumulative Match Rate (CMC). On the other hand,

Table 5.1 presents the recognition rate at Rank-5 and corresponding CMC plots in Figure 5.2, representing the benchmark results without employing filter operation.

Table 5. 5: Evaluation protocol

	Training	**Testing**						
Variation	Front face (0^o pose)	45^o	90^o	-45^o	-90^o	Smile	Eyes Closed	Paper on face occlusion
Number of subjects	64	64	64	64	64	64	64	64
Session	Session 1 & 2	Session 1 & 2						

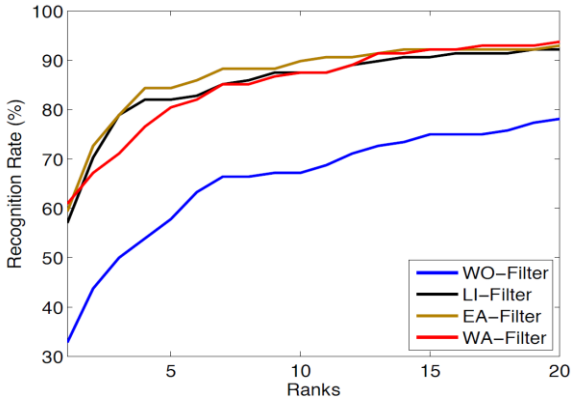
Based on the obtained results using the GU-RGB-D database, we present our major observations as follow:

- The overall results obtained based on three different hole-filling algorithms and our proposed RGBD face recognition scheme demonstrate the improvement in recognition accuracy compared to output results obtained without employing any filter, thus presenting the significance of our proposed approach in this work. This improvement in the performance analysis can be observed clearly in Figure 5.2.

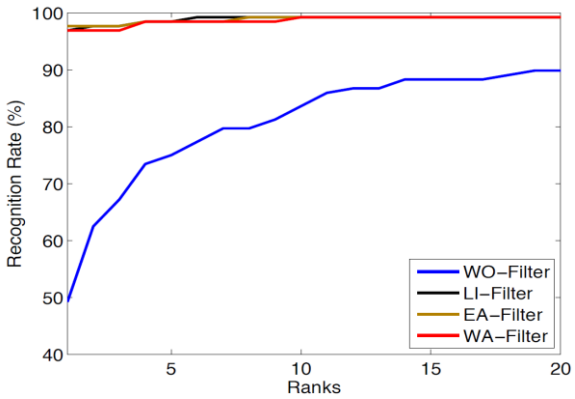
- The highest recognition rate at Rank-5 obtained for WO - Filter is 97.66% for `Close Eye' variation using the LG-CRC algorithm. In comparison, 100% accuracy is obtained with all the three filters, i.e., LI-Filter, EA-Filter, and WA-Filter, for the same variation and algorithm. Further, one can also note that the 100% recognition rate is also obtained for other algorithms such as LPQ-CRC, BSIF-CRC, CONV5-CRC for `Smile' variation using three proposed hole-filling filters.

- Among the algorithms, except for the LBP-CRC algorithm, the rest of the algorithm performs reasonably better for `Smile' and `Close Eyes' variation. At the same time, the performance of all the employed feature extraction methods degrades for variations such as angle and occlusion, where the entire face triangle is not available for computation. Even though the performance is low in these variations, it is reasonably better with three filters than without applying filter results.

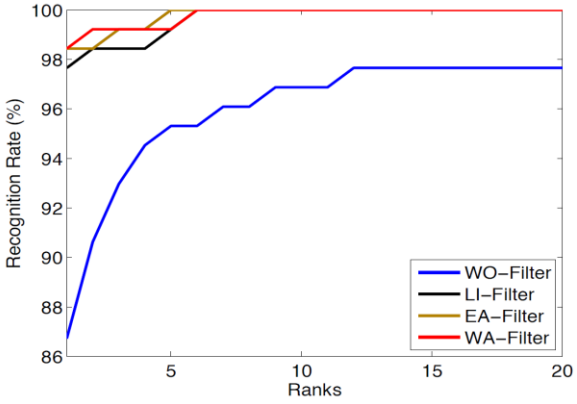
- Although all the filtering technique proposed in this work outperforms the without filtering, non of the technique outperforms each other, showing the similar performance behavior.



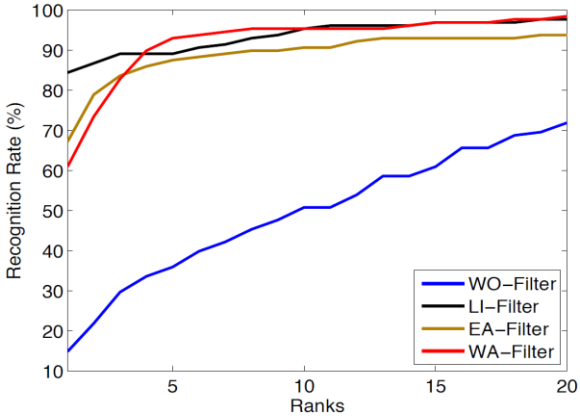
(a) LBP – CRC



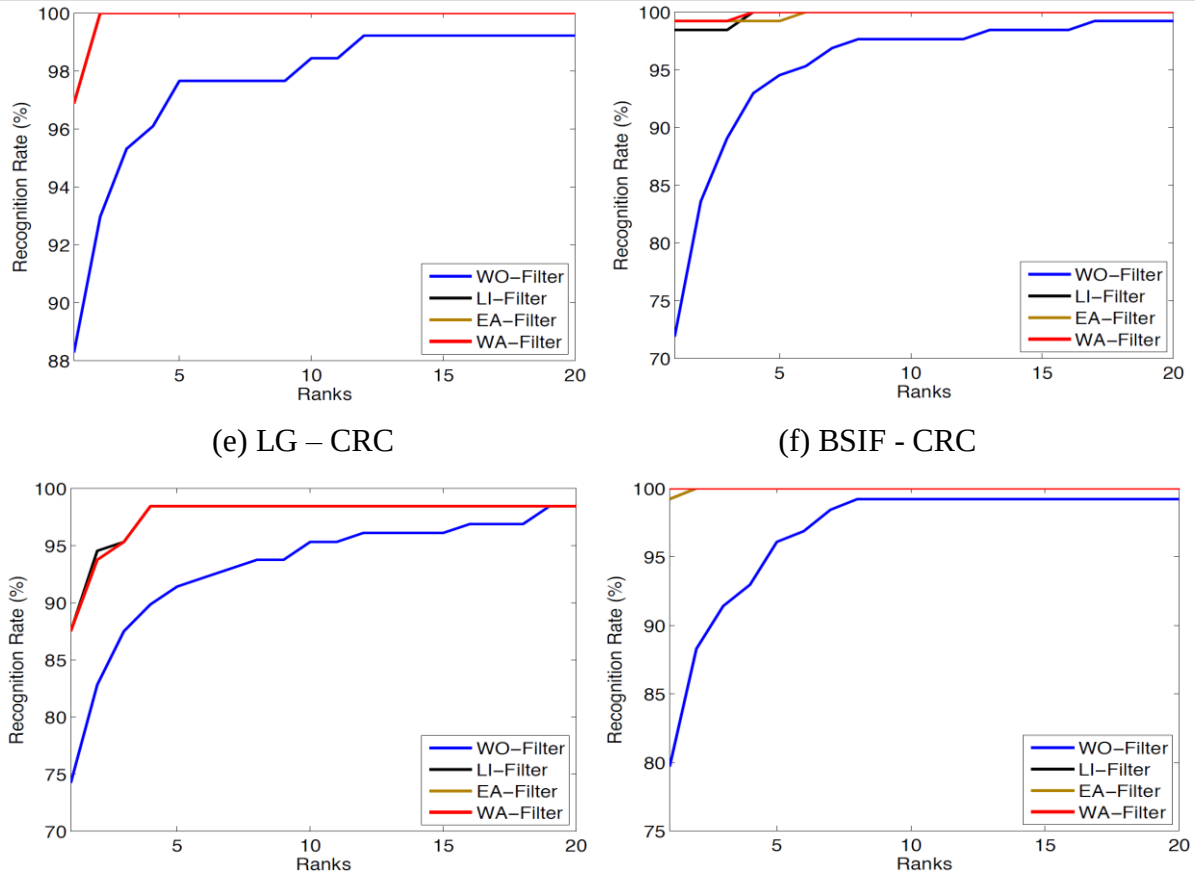
(b) LPQ - CRC



(c) HOG – CRC



(d) GIST - CRC



(e) LG – CRC (f) BSIF - CRC

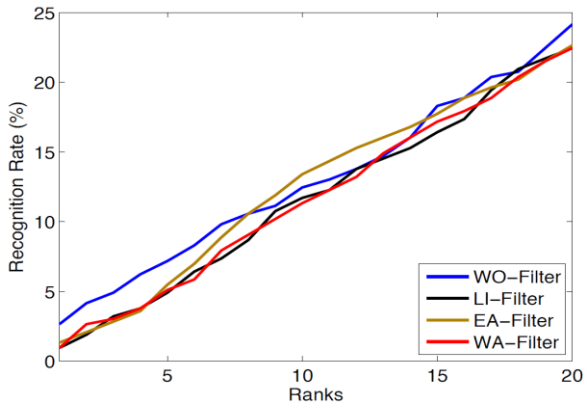
(g) PCA – CRC (h) CONV5 – CRC

Fig 5. 2: Cumulative Match Curve (CMC) plots demonstrating RGB-D face recognition on GU-RGB-D face database using three different filters and without filter. For simplicity, the best results corresponding to the `Close Eye' variation are presented.

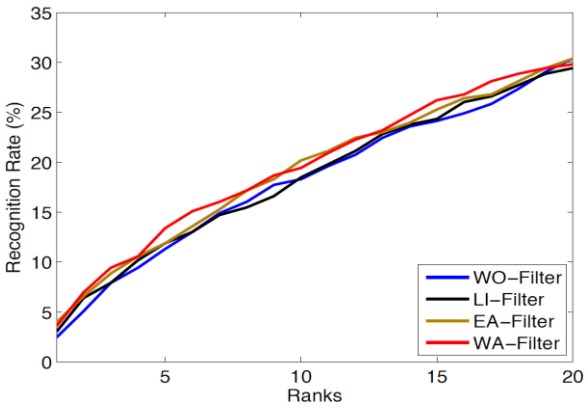
5.1.3.2 Evaluation Based on IIIT-D Database

This section presents the experimental results based on the publicly available IIIT-D face database. Compared to other publicly available RGB-D databases such as the EURECOM database, the IIIT-D database has been selected in this work for evaluation. The IIIT-D database has a reasonable amount of holes in the depth images, while the EURECOM database is available in the prepossessed form without having holes in the depth images,

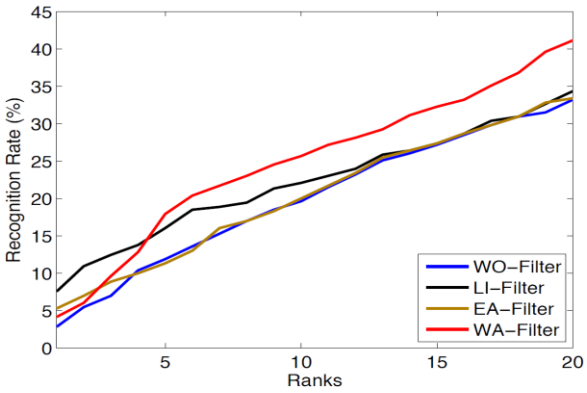
which is the reason why we intended to use the IIIT-D database in this work to demonstrate the potential of our proposed filters. Considering the IIIT-D database, having 106 subjects having a minimum of 11 to a maximum of 254 images per subject over a single sessions, here presented an evaluation protocol which consists of training set having randomly selected four images per subject, while the remaining samples of the database form the testing set.



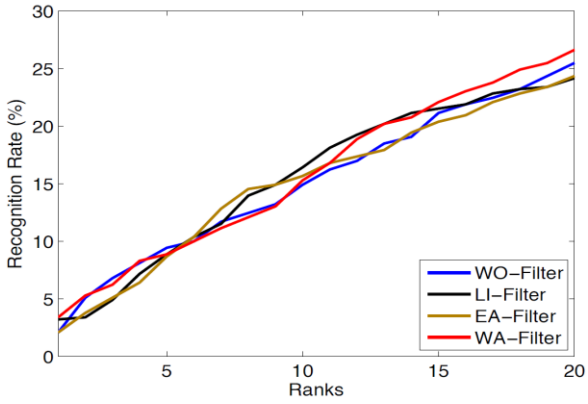
(a) LBP – CRC



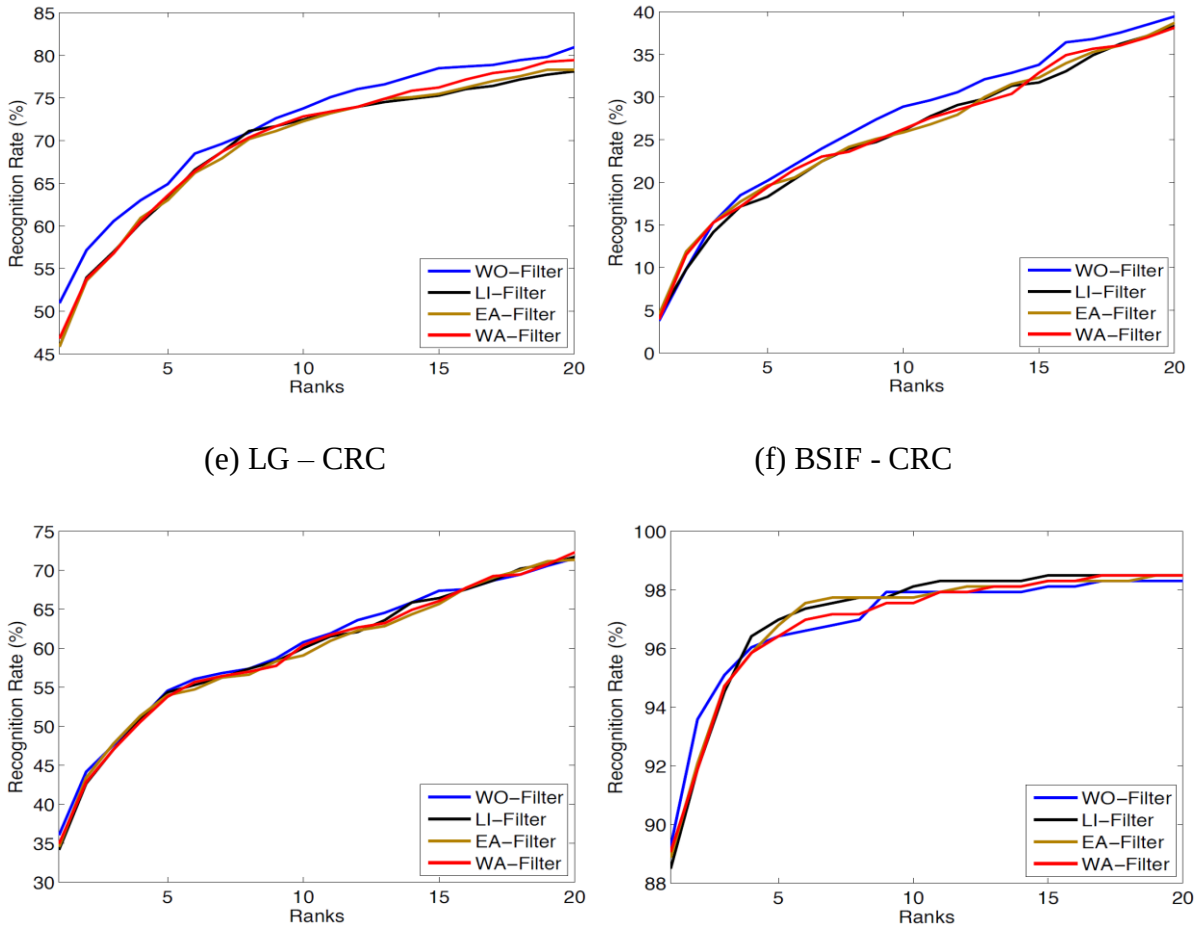
(b) LPQ - CRC



(c) HOG – CRC



(d) GIST - CRC



(e) LG – CRC (f) BSIF - CRC

(g) PCA – CRC (h) CONV5 - CRC

Fig 5. 3: Cumulative Match Curve (CMC) plots demonstrating RGB-D face recognition on IIIT-D face database using three different filters and without filter.

Table 5.2, 5.3, 5.4 presents the recognition rate at Rank-5, and Figure 7 presents a graphical representation of results in the form of Cumulative Match Rate (CMC). On the other hand, Table 5.1 presents the recognition rate at Rank-5 and corresponding CMC plots in Figure 5.3, represents the benchmark results without employing filter operation. Based on the obtained results using the IIIT-D face database, the significant observations are as follow:

- As expected, the evaluation results obtained indicate the significance of the filter and proposed scheme for RGB-D face recognition. However, we observe a marginal improvement in the recognition accuracy at Rank-5 with all three filters compared to without filters. This may be due to the fact that depth images in the IIIT-D database have bigger holes, as a result of which these filtering techniques present the lower performance. Thus restricting the designed hole-filling approach to the holes of smaller dimensions in the depth images.

The highest recognition rate at Rank-5 is obtained for WO-Filter is 96.42% using CONV5-CRC algorithm, while 96.98%, 96.79%, 96.42% recognition accuracy with LI-Filter, EA-Filter, and WA-Filter respectively for same CONV5-CRC algorithm, demonstrating the potential of employing deep learning feature obtained using convolutional neural network features at CONV5 layer. While we also note that except for CONV5 features, LG-CRC and PCA-CRC, and other methods show poor performance.

5.2: Classification With Image Set Algorithms

Face recognition based on multiple images can be formulated as an Image Set Classification problem. Each set contains images belonging to the same subject but consisting of a wide range of variations, for example, the images captured using a video-based surveillance system, multiple modalities, images acquired from multiple image cameras, etc. Generally, the image set classification has two significant steps: i) Finding the way to represent the set images. ii) applying necessary distance matrics to these representations. The image set methods are classified into parametric and non-parametric model methods considering the type of image representations.

In parametric model methods [145], the image set is represented in terms of certain parametric distribution (statistical distribution) followed by similarity measurement between the two image sets. Kull-back-Leibler (KL) divergence approach is used in these techniques to measure the similarity between the distributions. However, there is a need for a strong statistical relationship between training and testing image sets of the parametric model methods for obtaining good performance, and this acts as the limitation of this approach. To overcome these limitations, the non-parametric methods for image set classification have been developed, independent of any statistical assumptions of the data.

The non-parametric model methods approximate an image set in several different ways, including the set mean, a linear subspace, adaptively learnt set samples, a mixture of subspaces, and complex non-linear manifolds. Depending upon the type of representations, different matrices determine the distance between the sets distance. For example, the distance between the sets can be defined by the Euclidian distance between the set representatives, such as adaptively [88] or the set mean [92]. Cevikalp et al. [88] has termed the set to set distance as Affine Hull Image Set Distance (AHISD) or Convex Hull Image Set Distance (CHISD), which learn the set samples from the affine hull or convex hull models of the set images. A principle angle approach is used to determine the distance between the image sets represented by a linear subspace. The principle angle is the smallest angle between any

vectors in the two given subspaces. Further, the sum of cosines of the principle angles defines the similarities between the sets. Distance metrics such as the geodesic distance [146], the projection kernel metric [147] are adopted for image set representations on the Grassmann manifold and the log-map distance metric [148] on the Lie group of Riemannian manifold.

Non-parametric methods also have various classification strategies to decide the class of an image set and are divided into two categories:

i) The method which makes the decision based on Nearest Neighbor (NN) classification by computing the one-to-one set distance. Here the one to one set distance is computed between the set representatives such as set mean (Manifold to Manifold Distance (MMD) [92]), subspace or mixture of subspaces (Mutual Subspace Method *(*MSM) [149], Orthogonal Subspace Method (OSM) [150]), adaptively learnt set samples (*e. g.* AHISD and CHISD [88]), etc. However, these methods could be computationally slow as one to one match of the testing set with all the training sets is needed.

ii) The method where the discriminant function is learnt first, and then this function is used to classify the image set. The examples of these methods include Covariance Discriminative Learning (CDL) [89], Discriminative Canonical Correlations (DCC) [151], Manifold Discriminant Analysis (MDA) [94], and Graph Embedding Discriminant Analysis (GEDA) [152].

5.2.1 Contributions:

Section 5.2 of chapter 5 presents the study performed on the three sets of images, i.e., depth image, RGB-D image fused using pixel-level image fusion (averaging), and RGBD image fused using CNN fusion. We have not used the set of RGB images separately as our focus is on depth images. The said fusion methodologies are used to fuse the RGB and the depth images after employing the hole-filling filters on the depth images (mentioned in chapter 4). Further, the features are extracted by the state-of-the-art algorithms, and then these features

are given as the inputs to the Image set classification algorithms for classification. The results show that for the RGBD images, fused using pixel-level average image fusion and CNN fusion improves the performance of most of the Image Set Classification algorithms for various feature extraction methods with hole-filling filters.

The major contributions of section 5.2 of chapter 5 are as follows:

- Present Image Set Classification study based on various Image Set Classification algorithms, i.e., MMD: Manifold-Manifold Distance [92], MDA (Manifold Discriminant Analysis) [94], CDL (Covariance Discriminative Learning), AHISD (Affine Hull Based Image Set Distance) [88]; CHISD (Convex Hull Based Image Set Distance) [88], SANP (Sparse Approximated Nearest Point) [86].

- Present image set classification study on depth images, fused RGB and depth images using image-level pixel fusion and CNN based image fusion.

- Presents experimental results performed using seven different feature extraction algorithms, i.e., Histogram of Oriente Gradient (HOG), Principal Component Analysis (PCA), Local Phase Quantization (LPQ), GIST, Binarized Statistical Image Features (BSIF), Local Binary Pattern (LBP), and Convolution Neural Networks (CNN).

- The study is presenting the significance of employing hole filling techniques to improve the performance of the system.

The rest of the flow is as follows: section 5.2.2 presents the experimental protocol used to perform the image set classification study. The pixel-level image fusion-based evaluation protocol has been explained along with the computed results in section 5.2.3. Similarly, section 5.2.4 presents the evaluation protocol based on CNN-based fusion and the computed results.

5.2.2 Experimental Protocol And Results

This section presents the experimental evaluation protocol and related experimental results obtained in this work. The experimental study and evaluation of Image Set Classification algorithms have been performed on the GU-RGB-D database. The database is divided into training and testing sets. In the training set, the eight variations (i.e., front face, 45, -45º, 90º, -90º, smile, close eyes, and occlusion) per subject (total 64 subjects) from session 1 are grouped together to form the sets of images. While the similar sets generated protocol is used for session 2 for the generation of testing sets.

Initially, at the pre-processing stage, the designed hole-filling operations, i.e., LI-Filter: Linear Interpolation, EA-Filter: Exponential Averaging, and WA-Filter: Weighted Averaging are performed over the depth image. Here the kernel function gives proper weightage to the neighboring pixels in order to fill the hole by computing the appropriate missing value/s. Further in the experiment, the fusion methodologies are employed to enhance the performance of the system. Here the RGB component of the Image is fused with the depth image using Pixel level image fusion (averaging) and Convolution Neural Network (CNN) based fusion independently.

In the face recognition system, feature extraction is one of the essential stages, and the same have been extracted from the depth and fused images independently using seven different state-of-the-art feature extraction methods, namely, Principal Component Analysis (PCA), Histogram of Oriented Gradient (HOG), Local Binary Pattern (LBP), Local Phase Quantization (LPQ), GIST, Binarized Statistical Image Features (BSIF), LogGabor and CNN (Conv-5). The extracted features are further given as inputs to the different image set classification algorithms, namely, MMD: Manifold-Manifold Distance MDA (Manifold Discriminant Analysis), CDL (Covariance Discriminative Learning), AHISD (Affine Hull Based Image Set Distance); CHISD (Convex Hull Based Image Set Distance) and SRC (Sparse Approximated Nearest Point) to compute the final performance of the system, i.e., the recognition rate. The schematic view of the experimental protocol is presented in Figure 5.4.

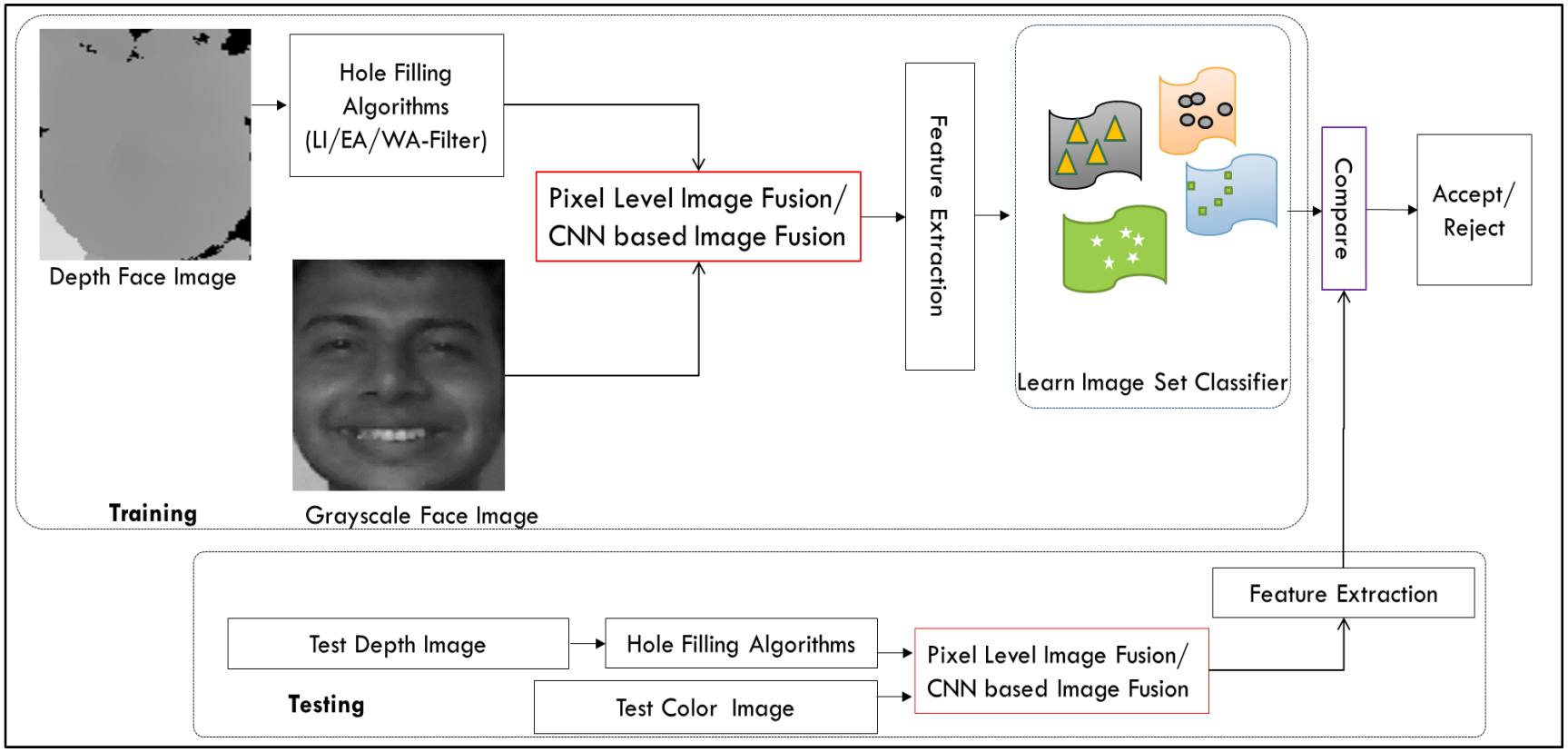


Fig 5. 4: Schematic block diagram illustrating the Framework of Image Set Classification approach

Table 5. 6: Evaluation protocol

Data	Number of Image Sets	Number of Images per set
Training (Session 1)	64	8
Testing (Session 2)	64	8

Using this experimental protocol, the following sections present the two sets of experimental evaluations to demonstrate the effect of the hole-filling and fusion approach on image set classification. Evaluation 1 presents the results at rank-5 related to image set classification algorithms for depth and fused images obtained from the implementation of pixel-level image fusion. Evaluation 2 presents the results at rank-5 related to the image set classification algorithms for depth and fused images obtained from Convolution Neural Network (CNN) based image fusion

5.2.2.1 Evaluation 1: Pixel Level Image Fusion (Averaging)

This section discusses the results obtained from the pixel-level fusion methodology employed for generating data for Image Set Classification. After hole filling with the designed filters at the pre-processing level, the depth images are fused with the RGB images. The pixel-level image fusion strategy using averaging approach has been employed to obtain a fused image. This is followed by feature extraction algorithms and the Image Set Classification algorithms as discussed in section 5.2.2. The evaluation results are presented in the form of recognition rate in tabular form for the various state-of-the-are image set classification algorithms, i.e., AHSID, CHSID, CDL, MMD, MDA, and SANP employed on the features extracted by seven feature extraction algorithms independently. Finally, the results demonstrate the effect of the designed filters on the depth and the fused images. Tables 5.7, 5.10, 5.13, 5.16, 5.19, 5.22 presents the recognition rates computed for depth images at Rank-5 using the different image set classification algorithms after employing hole-filling filters and seven feature extraction algorithms. Tables 5.8, 5.11, 5.14, 5.17, 5.20,

5.23 presents the recognition rates computed for pixel-level image fusion at Rank-5 using the different image set classification algorithms after employing hole-filling filters and seven feature extraction algorithms. The major observations deduced from the tables are as followed:

- Overall it has been noted that the base results of depth images for different Image set classification algorithms without application of the hole filling filter have been improved with the LI, EA, and WA hole-filling filters for all the feature extraction algorithms with very few exceptions. This improvement effect has also been noted with respect to the filters for the fused images, thus justifying the applicability of the proposed filters. For example, consider the AHISD algorithm where the without filter results for the BSIF feature extraction algorithm is 62.50% which is enhanced to 65.63% with filter (WA) for the depth images. Further, for the RGB and depth fused images, the filter results are 100% as compared with 96.88%.

- The fusion results obtained for different image set classification algorithms are much higher than the base results computed on depth images almost for all the feature extraction algorithms. For example, using PCA for CHISD algorithm, the recorded performance for WO-filter, LI-filter, EA-filter, and WA-filter is 62.50%, 68.75%, 65.63% and 75.00%, respectively. On the other hand, the enhanced results on fusion are 92.19%, 98.44%, 96.88%, and 98.44% for WO-filter, LI-filter, EA-filter, and WA-filter, respectively. This trend can be seen for all the image set classification algorithms, i.e., AHISD, CHISD, CDL, MDA, MMD, and SNAP.

- There are very few cases in the tables where the effect of filters is not seen on the fused images, i.e., there is no improvement in the performance. However, these results are still higher than the depth image results. For example, the performance of CDL for LBP feature extraction is 40.63% for LI-filter, 50% for EA- filter, and 48.44% for WA-filter, which is lower than 54% of without filter (Table5.13).

However, these results are much higher than the 20.31% of without filter, 17.19% of LI-filter, 15.63% of EA-filter, and 25% of WA-filter results obtained from depth images (Table 5.14).

- BSIF has recoded the highest performance of 100% for the fused images for AHISD (with all three filters), CHISD (with all three filters), MMD (with LI and WA filter), and SNAP (with LI filter).

- The results obtained using CNN as a feature extraction method are also quite promising. It can be seen that CNN shows improvement in the results of depth and fused images (with the application of filter). Further, the trend of improvement due to fusion over the depth images is also noted here. CNN has shown result enhancement for all the image set classification algorithms for all three hole filling filters accept for MDA fused images; however, these fused images results are much higher than the depth image results.

- Overall this evaluation protocol has generated well distinguishable and promising results for almost all the image set classification algorithms employed over the features extracted from different state-of-the-art feature extraction algorithms.

Table 5. 7: Recognition rate computed for depth images at Rank-5 using AHISD algorithm after employing hole-filling filters and seven feature extraction algorithms

	PCA	**HOG**	**LBP**	**BSIF**	**LPQ**	**GIST**	**CNN**
WO- Filter	76.56	18.75	20.31	62.50	35.94	29.69	37.50
LI Filter	71.88	**20.31**	18.75	**62.50**	31.25	20.31	**43.75**
EA Filter	**84.38**	**18.75**	18.75	**62.50**	**40.63**	23.44	**39.06**
WA Filter	**82.81**	**20.31**	14.06	**65.63**	**40.63**	26.56	**42.19**

Table 5. 8: Recognition rate computed for pixel level fused (RGB+D) images at Rank-5 using AHISD algorithm after employing hole-filling filters and seven feature extraction algorithms

	PCA	HOG	LBP	BSIF	LPQ	GIST	CNN
WO- Filter	93.75	46.88	39.06	96.88	67.19	45.31	75.00
LI Filter	**95.31**	**46.88**	29.69	**100.00**	65.63	**53.13**	**84.38**
EA Filter	**95.31**	43.75	31.25	**100.00**	**68.75**	**51.56**	**82.81**
WA Filter	**95.31**	**60.94**	32.81	**100.00**	**79.69**	**57.81**	**89.06**

Table 5. 9: Recognition rate computed for CNN based fused (RGB+D) images at Rank-5 using AHISD algorithm after employing hole-filling filters and seven feature extraction algorithms

AHISD	PCA	HOG	LBP	BSIF	LPQ	GIST	CNN
WO- Filter	89.25	46.88	35.94	96.88	67.19	40.63	67.19
LI Filter	**92.23**	45.31	26.56	**100.00**	**67.19**	**46.88**	**89.06**
EA Filter	**90.25**	43.75	34.38	**100.00**	**70.31**	**48.44**	**82.81**
WA Filter	**95.23**	**53.13**	**39.06**	**100.00**	**76.56**	**51.56**	**96.88**

Table 5. 10: Recognition rate computed for depth images at Rank-5 using CHISD algorithm after employing hole-filling filters and seven feature extraction algorithms

	PCA	HOG	LBP	BSIF	LPQ	GIST	CNN
WO- Filter	62.50	20.31	14.06	62.50	35.94	26.56	39.06
LI Filter	**68.75**	**20.31**	**17.19**	**62.50**	29.69	20.31	**43.75**
EA Filter	**65.63**	**20.31**	**20.31**	**62.50**	**40.63**	23.44	**39.06**
WA Filter	**75.00**	**20.31**	**15.63**	**65.63**	**40.63**	**26.56**	**42.19**

Table 5. 11: Recognition rate computed for pixel level fused (RGB+D) images at Rank-5 using CHISD algorithm after employing hole-filling filters and seven feature extraction algorithms

	PCA	HOG	LBP	BSIF	LPQ	GIST	CNN
WO- Filter	92.19	43.75	39.06	96.88	67.19	43.75	75.00
LI Filter	**98.44**	**45.31**	32.81	**100.00**	65.63	**53.13**	**84.38**
EA Filter	**96.88**	42.19	29.69	**100.00**	**68.75**	**51.56**	**84.38**
WA Filter	**98.44**	**59.38**	**39.06**	**100.00**	**79.69**	**59.38**	**89.06**

Table 5. 12: Recognition rate computed for CNN based fused (RGB+D) images at Rank-5 using CHISD algorithm after employing hole-filling filters and seven feature extraction algorithms

CHISD	PCA	HOG	LBP	BSIF	LPQ	GIST	CNN
WO- Filter	90.63	45.31	40.63	96.88	65.63	43.75	65.63
LI Filter	**98.44**	**48.44**	28.13	**100.00**	**67.19**	**46.88**	**89.06**
EA Filter	**96.88**	42.19	37.50	**100.00**	**68.75**	**48.44**	**82.81**
WA Filter	**96.88**	**56.25**	39.06	**100.00**	**76.56**	**54.69**	**95.31**

Table 5. 13: Recognition rate computed for depth images at Rank-5 using CDL algorithm after employing hole-filling filters and seven feature extraction algorithms

	PCA	HOG	LBP	BSIF	LPQ	GIST	CNN
WO- Filter	26.56	37.50	20.31	39.06	20.31	46.88	34.38
LI Filter	20.31	**37.50**	17.19	**46.88**	**35.94**	**46.88**	**35.94**
EA Filter	20.31	**37.50**	15.63	**48.44**	**37.50**	**60.94**	**34.38**
WA Filter	25.00	**40.63**	**25.00**	**45.31**	**37.50**	**51.56**	**37.50**

Table 5. 14: Recognition rate computed for pixel level fused (RGB+D) images at Rank-5 using CDL algorithm after employing hole-filling filters and seven feature extraction algorithms

	PCA	HOG	LBP	BSIF	LPQ	GIST	CNN
WO- Filter	67.19	81.25	54.69	89.06	79.69	82.81	59.38
LI Filter	64.06	78.13	40.63	**96.88**	68.75	81.25	**62.50**
EA Filter	**70.31**	**81.25**	50.00	**95.31**	**81.25**	76.56	**65.63**
WA Filter	**67.19**	**92.19**	48.44	**100**	**84.38**	**92.19**	**84.38**

Table 5. 15: Recognition rate computed for CNN based fused (RGB+D) images at Rank-5 using CDL algorithm after employing hole-filling filters and seven feature extraction algorithms

CDL	PCA	HOG	LBP	BSIF	LPQ	GIST	CNN
WO- Filter	70.31	82.81	56.25	89.06	76.56	84.38	67.19
LI Filter	68.75	79.69	43.75	**96.88**	70.31	81.25	**96.88**
EA Filter	**71.88**	**82.81**	50.00	**95.31**	**85.94**	76.56	**96.88**
WA Filter	67.19	**92.19**	**60.94**	**100.00**	**84.38**	**92.19**	**98.44**

Table 5. 16: Recognition rate computed for depth images at Rank-5 using MDA algorithm after employing hole-filling filters and seven feature extraction algorithms

	PCA	HOG	LBP	BSIF	LPQ	GIST	CNN
WO- Filter	17.19	29.69	23.44	12.50	34.38	20.31	10.94
LI Filter	10.94	28.13	20.31	6.25	29.69	10.94	**14.06**
EA Filter	12.50	21.88	17.19	**14.06**	**40.63**	17.19	**15.63**
WA Filter	12.50	**29.69**	21.88	9.38	**42.19**	**20.31**	**15.63**

Table 5. 17: Recognition rate computed for pixel level fused (RGB+D) images at Rank-5 using MDA algorithm after employing hole-filling filters and seven feature extraction algorithms

	PCA	HOG	LBP	BSIF	LPQ	GIST	CNN
WO- Filter	6.25	53.13	29.69	17.19	56.25	17.19	34.38
LI Filter	**12.50**	**53.13**	25.00	10.94	**62.50**	**25.00**	28.13
EA Filter	**7.81**	**64.06**	25.00	10.94	45.31	**34.38**	15.63
WA Filter	**10.94**	48.44	25.00	**21.88**	50.00	**25.00**	23.44

Table 5. 18: Recognition rate computed for CNN based fused (RGB+D) images at Rank-5 using MDA algorithm after employing hole-filling filters and seven feature extraction algorithms

MDA	PCA	HOG	LBP	BSIF	LPQ	GIST	CNN
WO- Filter	10.94	56.25	34.38	23.44	59.38	14.06	17.19
LI Filter	**12.50**	**67.19**	31.25	18.75	53.13	**29.69**	**42.19**
EA Filter	6.25	**65.63**	31.25	15.63	51.56	**32.81**	**40.63**
WA Filter	9.38	**62.50**	29.69	17.19	**59.38**	**21.88**	**51.56**

Table 5. 19: Recognition rate computed for depth images at Rank-5 using MMD algorithm after employing hole-filling filters and seven feature extraction algorithms

	PCA	HOG	LBP	BSIF	LPQ	GIST	CNN
WO- Filter	23.44	23.44	9.38	78.13	4.69	39.06	31.25
LI Filter	21.88	21.88	**18.75**	76.56	**20.31**	**46.88**	**42.19**
EA Filter	18.75	18.75	**20.31**	76.56	**26.56**	35.94	**40.63**
WA Filter	18.75	18.75	**17.19**	**78.13**	**26.56**	**45.31**	**43.75**

Table 5. 20: Recognition rate computed for pixel level fused (RGB+D) images at Rank-5 using MMD algorithm after employing hole-filling filters and seven feature extraction algorithms

	PCA	HOG	LBP	BSIF	LPQ	GIST	CNN
WO- Filter	68.75	68.75	42.19	98.44	21.88	75.00	71.88
LI Filter	65.63	65.63	34.38	**100.00**	20.31	65.63	**79.69**
EA Filter	67.19	67.19	37.50	**98.44**	**21.88**	71.88	**78.13**
WA Filter	**76.56**	**76.56**	**43.75**	**100.00**	**28.13**	**81.25**	**85.94**

Table 5. 21: Recognition rate computed for CNN based fused (RGB+D) images at Rank-5 using MMD algorithm after employing hole-filling filters and seven feature extraction algorithms

MMD	PCA	HOG	LBP	BSIF	LPQ	GIST	CNN
WO- Filter	56.25	70.31	50.00	98.44	21.88	68.75	12.50
LI Filter	**67.19**	60.94	40.63	**100.00**	**23.44**	60.94	**18.75**
EA Filter	**62.50**	65.63	48.44	**98.44**	20.31	**68.75**	**21.43**
WA Filter	**71.88**	**75.00**	42.19	**100.00**	**31.25**	**76.56**	**18.75**

Table 5. 22: Recognition rate computed for depth images at Rank-5 using SANP algorithm after employing hole-filling filters and seven feature extraction algorithms

	PCA	HOG	LBP	BSIF	LPQ	GIST	CNN
WO- Filter	34.38	18.75	10.94	60.94	32.81	25.00	32.81
LI Filter	29.69	**18.75**	**18.75**	**64.06**	28.13	20.31	**51.56**
EA Filter	**43.75**	**18.75**	**18.75**	**62.50**	**39.06**	23.44	**45.31**
WA Filter	**37.50**	**18.75**	**15.63**	**68.75**	**32.81**	23.44	**45.31**

Table 5. 23: Recognition rate computed for pixel level fused (RGB+D) images at Rank-5 using SANP algorithm after employing hole-filling filters and seven feature extraction algorithms

	PCA	HOG	LBP	BSIF	LPQ	GIST	CNN
WO- Filter	67.19	43.75	40.63	95.31	59.38	48.44	76.56
LI Filter	**70.31**	**43.75**	25.00	**100.00**	45.31	**48.44**	**79.69**
EA Filter	**68.75**	**45.31**	25.00	**96.88**	**60.94**	**48.44**	**76.56**
WA Filter	**76.56**	**56.25**	37.50	**98.44**	**76.56**	**54.69**	**84.38**

Table 5. 24: Recognition rate computed for CNN based fused (RGB+D) images at Rank-5 using SANP algorithm after employing hole-filling filters and seven feature extraction algorithms

SANP	PCA	HOG	LBP	BSIF	LPQ	GIST	CNN
WO- Filter	62.50	40.63	34.38	95.31	54.69	40.63	64.06
LI Filter	**76.56**	**42.19**	23.44	**100.00**	45.31	**43.75**	**70.25**
EA Filter	**70.31**	**45.31**	32.81	**96.88**	**62.50**	**43.75**	**67.33**
WA Filter	**76.56**	**51.56**	**35.94**	**98.44**	**75.00**	**50.00**	**74.56**

5.2.2.2 Evaluation 2: Convolution Neural Network (CNN) Based Image Fusion

This section presents the Image Set Classification results obtained from the set of data after implementing Convolution Neural Network (CNN) based image fusion. First, as discussed earlier, the designed hole-filling filters are applied to the depth images at the pre-processing level. These filtered depth images are then fused with the RGB component of the image using the CNN based image fusion. In the next level, the features are extracted using the seven feature extraction algorithms employed in this study, i.e., PCA, HOG, LBP, LPQ,

GIST, BSIF, and CNN. These extracted features are then independently engaged in the different state-of-the-art Image Set Classification algorithms (AHISD, CHISD, CDL, MMD, MDA, and SANP) to compute the system's performance in the form of recognition rate and verification rate. The computed results are presented in the tabular form, and they demonstrate the effect of hole-filling filters and the CNN based fusion approach. Tables 5.9, 5.12, 5.15, 5.18, 5.21, 5.24 presents the recognition rates computed for CNN based image fusion at Rank-5 using the different image set classification algorithms after employing hole-filling filters and seven feature extraction algorithms. The major observations deduced from the tables are as followed:

- At first glance on comparing with the Pixel Level Image Fusion (averaging), it is observed that a similar trend of increase in the result with the application of designed hole-filling filters has been maintained. Further, these CNN based image fusion results are also higher than the results obtained for the depth images. For example, for CIHSD using PCA, the results obtained for WO, LI, EA, and WA filters are 62.50%, 68.75%, 65.63%, and 75%, and these are enhanced to 90.63%, 98.44%, 96.88%, and 98.88% respectively using the CNN based image fusion approach.
- Overall, the system has shown a great enhancement in performance regarding HOG, BSIF, LPQ, GIST, and CNN feature extraction algorithms for almost all the image set classification, with BSIF having a maximum of 100% performance on many occasions in the tables.
- As discussed earlier, the cases where the results for depth images were not that promising with respect to the designed filters have also shown good performance with the CNN fusion, same as that of other pixel-level image fusion approach.
- CNN based image fusion followed by feature extraction using CNN has shown higher performance than the pixel-based image fusion followed by CNN feature extraction. This justifies the ability of CNN based approaches to enhance the system performances.

CHAPTER 6:

SUMMARY & CONCLUSION

CHAPTER 6

The technological advancement and the low-cost acquisition system have made the biometric community explore 3D biometrics research much deeper. Authentication based on facial biometric traits has been widely used in various security applications due to its non-intrusive nature of image capture in a covert manner. Although face recognition shows great potential, face recognition performance is challenged by multiple covariates such as pose, expression, illumination, etc. Considering these issues, 3D biometric finds an alternative approach over traditional face recognition methods operated in 2D.

We have generated and added one more RGB-D database to the 3D biometrics researcher community through this thesis, i.e., the GU-RGB-D database. The said database has a wide variation in pose (-90^0, -45^0, 0^0, $+45^0$, $+90^0$), expressions (smile, eyes closed), and occlusion (paper covering half part of the face) and was collected in two sessions. Also, presented kernel-based hole filling filters for the depth images at the pre-processing level that enhanced the recognition performance of the system. The experimental evaluation was performed on the GU-RGB-D database and also on the publicly available EURECOM & IIITD RGB-D databases. The state-of-the-art local and global features extractor algorithms such as PCA, HOG, LBP, LPQ, GIST, BSIF Log Gabor, and CNN were engaged to extract the features. These features are further used to compute the recognition rate and the verification rates for the depth images.

A similar protocol has been enforced on the fused RGB-D images obtained by fusing the RGB and the depth images by the pixel-level average image fusion, Wavelet fusion, and CNN fusion. Further, the Euclidian distance and collaborative representation classifier-based approach has been implemented for the computation of scores. The classification study has also been performed on the depth and RGB-D databases using image set classification algorithms. It is observed that the recognition rates were enhanced by the implementation of various methodologies discussed, and then designed hole filling filters seem to be performed quite promising and have fulfilled the purpose.

CHAPTER 6

6.1 Conclusion of Chapter 3

For the development of any real-time solution, background research is essential, and it is only possible when the databases of all the practical scenarios are available. With the literature survey, it is understood that the Kinect-based 3D face databases are limited as compared to the 3D face databases acquired from other expensive 3D scanning devices. In this research work, we have presented a Kinect-based GU-RGB-D database. Here in this database, we have tried to cover most practical scenarios like pose/angle variations, expression variations, occlusion, and the effect of controlled and uncontrolled environmental conditions so as to improve the robustness of the performance algorithm for a good piece of research work.

Two preliminary studies are performed on the GU-RGB-D database and also on the publicly available EURECOM database. In the first study, the features are extracted from the depth and RGB images using the PCA algorithm, and the computed scores are fused using complementary fusion. Thereafter, the recognition rates are computed. It is observed that the complementary fusion strategy has enhanced the recognition rates almost for all the variations of the EURECOM database depending upon the weightages of the RGB and depth images. The performance enhancement has also been noted for the GU-RGB-D database but has a lower performance as compared to the EURECOM database.

In the second study, the RGB and the depth images are pre-processed with the gradient filter, and the images are fused using pixel-level image fusion. The fused images are further used for feature extraction using PCA, and subsequently, the recognition rates are computed. It is observed that with the application of gradient image, the performance has been enhanced, and further improvement is observed with the fusion strategy. In the case of the GU-RGB-D database, the performance enhancement due to the application of gradient filter is not uniform across the variations. This is mainly because the database has more angular variations and missing information in the form of holes. Thus, making it more challenging for research problems and more suitable for performing realistic, practical scenarios base research.

6.2 Conclusion of Chapter 4

The depth captured using the Kinect sensor, having holes in it, significantly degrades the overall biometric system's performance. This thesis presents three different filtering techniques: linear interpolation, exponential averaging, and weighted averaging filter for depth images. Further, we used the kernel function to demonstrate the filtering technique efficiently. Specifically, we employ filters on variable kernel size to give appropriate weightage to the nearest neighbors that surround the hole such that the contribution of the neighborhood pixels is considered to fill the missing pixel values.

We experimented on our GU-RGB-D database consisting of 64 subjects collected across seven facial variants such as 45^0, -45^0, 90^0, -90^0 , smile, close eyes, and paper on face occlusion. To present our results, using three different filtering approaches, we presented the results with seven different face recognition algorithms such as PCA, HOG, LBP, LPQ, GIST, BSIF, and Log Gabor used in 3D biometrics. The extensive experimental analysis is obtained independently with depth image, fusion of RGB and depth, and score level fusion of two best-performing algorithms. All the results related to the each of the evaluation is presented using recognition rate and verification rate. The results obtained after employing kernel-based filtering outperform the without filtered depth image performance.

6.3 Conclusion of Chapter 5

The study based on a collaborative representation classifier and image set classification approach has been presented in this thesis. The collaborative representation approach has been performed on GU-RGB-D and IIIT-D databases. Here the filtered training and testing depth images (filtered using LI, EA, & WA filters) are fused with their corresponding RGB images using Discrete Wavelet Transform fusion independently. Further, these fused images are employed with the eight feature extraction algorithms, i.e., PCA, HOG, LBP, LPQ, GIST, BSIF, Log Gabor, and CNN (conv5). These obtained features are classified using a collaborative representation classifier. The extensive experimental analysis is obtained independently with LI, WA & EA filters and without filter data. It is observed that the results

obtained after employing filters outperform the without filtered depth image performance. Mostly in the cases where the full face triangle is available for computation in the GU-RGB-D database. Further, the feature extraction using CNN has also shown enhancement on filtered images as compared to the WO (without filter). Improvement in performance over the IIIT-D database is marginal; however, the effect of the filter can be seen on it.

The image set classification approach has been performed on the GU-RGB-D database. Here the filtered depth images are fused with the RGB images using pixel-level image fusion and CNN based image fusion. The features are extracted from the fused images using PCA, HOG, LBP, LPQ, GIST, BSIF, and CNN, and these features are used to learn the different image set classifiers (AHSID, CHSID, CDL, MMD, MDA, SANP). Further, the test features are compared with these learnt classifiers to quantify the performance. It is observed that the system outperforms most of the cases with respect to filters. Further, the performance obtained using both the fusion approaches is almost similar in most of the cases.

Overall, it can be concluded that the proposed filter LI, EA & WA can enhance the quality of the RGB-D databases and the performance of the system. Different studies and schemes based on feature extractions, fusions, classifications have outperformed and have justified the objective of the research.

Future Works

The 3D facial biometric has the ability to be in line with the ever-increasing needs of high security from an individual level to the highly confidential areas (defense system) to the best of its capacity. The day-to-day research development in 3D face biometrics has provided the public domains with highly secure security systems. The major challenge to these systems could be spoofing, disguise, and presentation attacks from the 3D models or 3D masks. This area needs to be more explored and experimented on to develop robust algorithms and systems free from presentation attacks. Future plans will be to initiate research in the line of 3D disguise and spoofing as an extension of this research work.

REFERENCES

1 Jain, A.K., Flynn, P., Ross, A.A.: ‘Handbook of Biometrics’ (Springer n, 2007).

2 Prabhakar, S., Pankanti, S., Jain, A.K.: ‘Biometric Recognition: Security and Privacy Concerns Biometrics’IEEE Secur. Priv., 2003, pp. 33–42.

3 O’Gorman, L.: ‘Comparing passwords, tokens, and biometrics for user authentication’, in ‘Proceedings of the IEEE’ (2003), pp. 2021–2040.

4 Jain, A.K., Ross, A., Prabhakar, S.: ‘An Introduction to Biometric Recognition’IEEE Trans. Circuits Syst. Video Technol., 2004, 14, (1), pp. 4–20.

5 Wayman, J., Jain, A., Maltoni, D., Maio, D.: ‘Biometric Systems’ (2005).

6 P Tripathi, K.: ‘A Comparative Study of Biometric Technologies with Reference to Human Interface’Int. J. Comput. Appl., 2011, 14, (5), pp. 10–15.

7 Alsaadi, I.M.: ‘Study On Most Popular Behavioral Biometrics , Advantages , Disadvantages And Recent Applications : A Review’Int. J. Sci. Technol. Res., 2021, 10, (January 2021).

8 Barbosa, I.B.: ‘Unconventional biometrics’. 2020.

9 Li, B.Y.L., Mian, A., Liu, W., Krishna, A.: ‘Using Kinect for Face Recognition Under Varying Poses , Expressions , Illumination and Disguise’, in ‘IEEE Workshop on Applications of Computer Vision (WACV)’ (2013), pp. 186–192.

10 Jain, A., Hong, L., Bolle, R.: ‘Online fingerprint verification’IEEE Trans. Pattern Anal. Mach. Intell., 1997, 19, (4), pp. 302–314.

11 Spreeuwers, L.: ‘Fast and accurate 3D face recognition’ Int. J. Comput. Vis., 2011, 93, (3), pp. 389–414.

12 Kumar, A., Passi, A.: ‘Comparison and combination of iris matchers for reliable personal authentication’Pattern Recognit., 2010, 43, (3), pp. 1016–1026.

13 Bledsoe, W.W.: ‘The Model Method in Facial Recognition’ (1964).

14 Turk, M.A., Pentland, A.P.: ‘Face recognition using eigenfaces’IEEE, 1991, 810, pp. 586–591.

15 B. Moghaddam, T. Jebara, A.P.: ‘Bayesian Face Recognition’Pattern Recognit., 2002, **33**, (11), pp. 1771–1782.

16 Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: ‘Eigenfaces vs. fisherfaces: Recognition using class specific linear projection’IEEE Trans. Pattern Anal. Mach. Intell., 1997, 19, (7), pp. 711–720.

17 Frey, B.J., Colmenarez, A., Huang, T.S.: ‘Mixtures of local linear subspaces for face recognition’Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 1998, (June), pp. 32–37.

18 Wiskott, L., Fellous, J.M., Krüger, N., Von der Malsburg, C.: ‘Face recognition by elastic bunch graph matching’IEEE Trans. Pattern Anal. Mach. Intell., 1997, 19, (7), pp. 775–779.

19 Mpiperis, I., Malassiotis, S., Strintzis, M.G.: ‘3-D face recognition with the geodesic polar representation’IEEE Trans. Inf. Forensics Secur., 2007, 2, (3), pp. 537–547.

20 Bowyer, K.W., Chang, K., Flynn, P.: ‘A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition’Comput. Vis. Image Underst., 2006, 101, pp. 1–15.

21 Medioni, G., Waupotitsch, R.: ‘Face Modeling and Recognition in 3-D’, in ‘IEEE International Workshop on Analysis and Modeling of Faces and Gestures’ (2003).

22 Zhu, X., Lei, Z., Yan, J., Yi, D., Li, S.Z.: ‘High-Fidelity Pose and Expression Normalization for Face Recognition in the Wild’, in ‘Computer vision and pattern recognition (CVPR)’ (2015), pp. 787–796.

23 Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: ‘Face Recognition: A Literature Survey’ACM Comput. Surv., 2003, 35, (December), pp. 399–458.

24 Abate, A.F., Nappi, M., Riccio, D., Sabatino, G.: ‘2D and 3D face recognition: A survey’Pattern Recognit. Lett., 2007, 28, (14), pp. 1885–1906.

25 Zhou, X., Kuijper, A., Busch, C.: ‘Template Protection For 3D Face Recognition’ (2010).

26 Cartoux, J.Y., Lapreste, J.T., Richetin, M.: ‘Face authentification or recognition by profile extraction from range images’, in ‘Workshop on Interpretation of 3D Scenes’ (1989), pp. 194–199.

27 Gordon, G.G., Drive, W.B.: ‘Face Recognition from Frontal and Profile Views’Int. Work. Autom. face gesture Recognit., 1995, pp. 26–28.

28 Goswami, G., Vatsa, M., Singh, R.: ‘RGB-D Face Recognition with Texture and Attribute Features’IEEE Trans. Inf. Forensics Secur., 2014, 9, (10), pp. 1692–1640.

29 Min, R., Choi, J., Medioni, G.G., Dugelay, J.L.: ‘Real-time 3D face identification from a depth camera’, in ‘Int. Conf. on Pattern Recognition (ICPR)’ (2012).

30 Granger, S., Pennec, X.: ‘Multi-scale EM-ICP : A Fast and Robust Approach for Surface Registration’, in ‘European Conference on Computer Vision’ (2002), pp. 418–432.

31 Min, R., Kose, N., Dugelay, J.L.: ‘KinectfaceDB: A kinect database for face recognition’IEEE Trans. Syst. Man, Cybern. Syst., 2014, 44, (11), pp. 1534–1548.

32 Mantecon, T., Del-Blanco, C.R., Jaureguizar, F., Garc´ıa, N.: ‘Depth-Based Face Recognition USing Local Quantized Patterns Adapted For Range Data’, in ‘IEEE Int. Conf. on Image Processing (ICIP)’ (2014), pp. 293–297.

33 Mantecón, T., Carlos, R., Jaureguizar, F., García, N.: ‘Visual Face Recognition Using

Bag of Dense Derivative Depth Patterns'IEEE Signal Process. Lett., 2016, 23, (6), pp. 771–775.

34 Neto, J.B.C., Marana, A.N.: 'Face Recognition Using 3DLBP Method Applied to Depth Maps Obtained from Kinect Sensors', in 'Workshop de Vis˜ao Computacional WVC 2014' (2014), pp. 168–172.

35 Huynh, T., Min, R., Dugelay, J.: 'An Efficient LBP-Based Descriptor for Facial Depth Images Applied to Gender Recognition Using RGB-D Face Data', in 'ACCV 2012 Workshops' (2012), pp. 133–145.

36 Kim, D., Hernandez, M., Choi, J., Medioni, G.: 'Deep 3D face identification'IEEE Int. Jt. Conf. Biometrics, IJCB 2017, 2017, pp. 133–142.

37 Lee, Y., Chen, J., Tseng, C.W., Lai, S.H.: 'Accurate and robust face recognition from rgb-d images with a deep learning approach'Br. Mach. Vis. Conf. 2016, BMVC 2016, 2016, pp. 123.1-123.14.

38 Zhang, L., Xia, H., Qiao, Y.: 'Texture Synthesis Repair of RealSense D435i Depth Images with Object-Oriented RGB Image Segmentation'Sensors, 2020, pp. 1–15.

39 Xu, K., Zhou, J., Wang, Z.: 'A method of hole-filling for the depth map generated by Kinect with moving objects detection', in 'IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, BMSB' (2012).

40 Suarez, J., Murphy, R.R.: 'Using the Kinect for search and rescue robotics', in 'IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)' (IEEE, 2012), pp. 1–2.

41 Mao, Y., Cheung, G., Ortega, A., Ji, Y.: 'Expansion Hole Filling In Depth-Image-Based Rendering Using Graph-Based Interpolation', in 'IEEE International Conference on Acoustics, Speech and Signal Processing' (2013), pp. 1859–1863.

42 Solh, M., Alregib, G.: 'Hierarchical Hole-Filling For Depth-Based View Synthesis in

FTV and 3D Video'IEEE J. Sel. Top. Signal Process., 2012, 6, (5), pp. 495–504.

43 Wang, D., Zhao, Y., Wang, Z., Chen, H.: 'Hole-Filling for DIBR Based on Depth and Gradient Information Regular Paper'Int. J. Advanved Robot. Syst., 2015.

44 Wang, D., Zhao, Y., Wang, J., Wang, Z.: 'A Hole Filling Algorithm for Depth Image Based Rendering Based on Gradient Information', in 'Ninth International Conference on Natural Computation (ICNC)' (2013), pp. 1209–1213.

45 Feng, L., Po, L., Xu, X., Ng, K., Cheung, C., Cheung, K.: 'An Adaptive Background Biased Depth Map Hole-Filling Method For Kinect', in '39th Annual Conference of the IEEE Industrial Electronics Society' (2013), pp. 2366–2371.

46 Atapour-abarghouei, A., Garanderie, G.P. de La, Breckon, T.P.: 'Back to Butterworth - a Fourier Basis for 3D Surface Relief Hole Filling within RGB-D Imagery', in '23rd International Conference on Pattern Recognition (ICPR)' (2016), pp. 2813–2818.

47 Wang, L., Jin, H., Yang, R., Gong, M.: 'Stereoscopic Inpainting : Joint Color and Depth Completion from Stereo Images', in 'IEEE Computer Society Conference on Computer Vision and Pattern Recognition' (2008).

48 Breckon, T.P., Fisher, R.B.: 'Three-Dimensional Surface Relief Completion via Nonparametric Techniques'IEEE Trans. Pattern Anal. Mach. Intell., 2008, 30, (12), pp. 2249–2255.

49 Abebe, H.B., Hwang, C.: 'RGB-D face recognition using LBP with suitable feature dimension of depth image'IET Journa;, 2019, 4, pp. 189–197.

50 Sang, G., Li, J., Zhao, Q.: 'Pose-invariant face recognition via RGB-D images'Comput. Intell. Neurosci., 2016.

51 Zhang, H., Han, H., Cui, J., Shan, S., Chen, X.: 'RGB-D Face Recognition via Deep Complementary and Common Feature Learning', in 'IEEE International Conference on Automatic Face & Gesture Recognition RGB-D' (2018), pp. 8–15.

52 Krizhevsky, A., Hinton, G.E.: 'ImageNet Classification with Deep Convolutional Neural Networks', in 'Advances in Neural Information Processing Systems' (2012).

53 Ioffe, S., Szegedy, C.: 'Batch Normalization : Accelerating Deep Network Training by Reducing Internal Covariate Shift', in 'International Conference on Machine Learning' (2015).

54 Simonyan, K., Zisserman, A.: 'Very Deep Convolutional Networks For Large-Scale Image Recognition Karen', in 'ICLR 2015' (2015).

55 Uppal, H., Sepas-moghaddam, A., Greenspan, M., Etemad, A.: 'Two-Level Attention-based Fusion Learning for RGB-D Face Recognition', in 'International Conference on Pattern Recognition (ICPR) (2020).

56 Ajmera, R., Nigam, A., Gupta, P.: '3D Face Recognition using Kinect', in 'ICVGIP (2014).

57 Aly, S., Trubanova, A., Abbott, L., White, S., Youssef, A.: 'VT-KFER: A Kinect-based RGBD+time dataset for spontaneous and non-spontaneous facial expression recognition'Proc. 2015 Int. Conf. Biometrics, ICB (2015) (June), pp. 90–97.

58 Muhammad, S., Mousavi, H., Mirinezhad, S.Y.: 'Iranian kinect face database (IKFDB): a color - depth based face database collected by kinect v . 2 sensor'SN Appl. Sci., 2021, (December 2019).

59 Goswami, G., Bharadwaj, S., Vatsa, M., Singh, R.: 'On RGB-D face recognition using Kinect', in 'IEEE 6th International Conference on Biometrics: Theory, Applications and Systems, BTAS (2013).

60 Kumar, G., Bhatia, P.K.: 'A detailed review of feature extraction in image processing systems'Int. Conf. Adv. Comput. Commun. Technol. ACCT, 2014, pp. 5–12.

61 Paul, L.C., Sumam, A. Al: 'Face Recognition Using Principal Component Analysis Method', in 'International Journal of Advanced Research in Computer Engineering &

Technology (IJARCET)' (IEEE, 2012), pp. 135–139.

62 Turk, M.A., Pentland, A.P.: 'Eigenfaces for Recognition'J. Cogn. Neurosci., 1991, 3, (1), pp. 71–86.

63 Vijaya Lata, Y., Kiran Bharadwaj Tungathurthi, C., Ram Mohan Rao, H., Govardhan, A., Reddy, L.P.: 'Facial recognition using eigenfaces by PCA'Int. J. Recent Trends Eng., 2009, **1**, (1), pp. 587–590.

64 Dalal, N., Triggs, B.: 'Histograms of Oriented Gradients for Human Detection', in 'IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)' (2005).

65 Kobayashi, T., Hidaka, A., Kurita, T.: 'Selection of Histograms of Oriented Gradients', in 'ICONIP 2007' (2008), pp. 598–607.

66 Prakasa, E.: 'Texture Feature Extraction by Using Local Binary Pattern'J. INKOM, 2016, 9, (2), pp. 45–48.

67 Malhotra, A., Sankaran, A., Mittal, A., Vatsa, M., Singh, R.: 'Fingerphoto authentication using smartphone camera captured under varying environmental conditions', in 'Human Recognition in Unconstrained Environments: Using Computer Vision, Pattern Recognition and Machine Learning Methods for Biometrics' (Elsevier Ltd, 2017, 1st edn.), pp. 119–144.

68 Gupta, D., Agrawal, U., Arora, J., Khanna, A.: 'Bat-inspired algorithm for feature selection and white blood cell classification', in 'Nature-Inspired Computation and Swarm Intelligence' (Elsevier, 2020), pp. 179–197.

69 Brahnam, S., Nanni, L., Shi, J., Lumini, A.: 'Local phase quantization texture descriptor for protein classification', in 'International Conference on Bioinformatics and Computational Biology' (2010), pp. 159–165.

70 Ojansivu, V., Heikkilä, J.: 'Blur Insensitive Texture Classification Using Local Phase

Quantization', in 'Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)' (2008), pp. 236–243.

71 Heikkilä, J., Ojansivu, V.: 'Methods for local phase quantization in blur-insensitive image analysis'2009 Int. Work. Local Non-Local Approx. Image Process. LNLA (2009) (2), pp. 104–111.

72 Oliva, A., Torralba, A.: 'Modeling the shape of the scene: A holistic representation of the spatial envelope'Int. J. Comput. Vis., 2001, 42, (3), pp. 145–175.

73 Al-Akam, R., Paulus, D.: 'Local and global feature descriptors combination from RGB-Depth videos for human action recognition'ICPRAM 2018 - Proc. 7th Int. Conf. Pattern Recognit. Appl. Methods, (2018), (Icpram), pp. 265–272.

74 Gabor, D.: 'Theory of communication. Part 1: The analysis of information'J. Inst. Electr. Eng. - Part III Radio Commun. Eng., 1946, 93, (26), pp. 429–441.

75 Kannala, J., Rahtu, E.: 'BSIF: Binarized statistical image features', in 'Proceedings - International Conference on Pattern Recognition (ICPR)' (2012), pp. 1363–1366.

76 Field, D.J., Hubel, S.: 'Relations between the statistics of natural images and the'America (NY)., 1987, 4, (12).

77 Cook, J., Mccool, C., Chandran, V., Sridharan, S., Box, G.P.O., Qld, B.: 'Combined 2D / 3D Face Recognition using Log-Gabor Templates Queensland University of Technology 1 Introduction 2 Log-Gabor Filters', in 'Proceedings of the International Conference on video and signal based surveillance' (2006).

78 Teuwen, J., Moriakov, N.: 'Convolutional neural networks', in 'Handbook of Medical Image Computing and Computer Assisted Intervention' (Elsevier Inc., 2020), pp. 481–501.

79 Bodapati, J.D., Veeranjaneyulu, N.: 'Feature extraction and classification using Deep

convolutional Neural Networks'J. Cyber Secur. Mobil., 2019, 8, (2), pp. 261–276.

80 Matsugu, M., Mori, K., Mitari, Y., Kaneda, Y.: 'Subject independent facial expression recognition with robust face detection using a convolutional neural network'Neural Networks, 2003, 16, (5–6), pp. 555–559.

81 Girshick, R., Donahue, J., Darrell, T., Malik, J.: 'Rich feature hierarchies for accurate object detection and semantic segmentation'Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2014, pp. 580–587.

82 Rani, K., Sharma, R.: 'Study of Different Image fusion Algorithm'Int. J. Emerg. Technol. Adv. Eng., 2013, 3, (5), pp. 288–291.

83 Sharma, M.: 'A review: image fusion techniques and applications'Int J Comput Sci Inf Technol, 2016, **7**, (3), pp. 1082–1085.

84 Mallat Stephane G: 'A Theory for Multiresolution Signal Decomposition : The Wavelet Representation'IEEE Trans. PATTERN Anal. Mach. Intell., (1989), **11**, (july), pp. 674–693.

85 Zhang, L., Yang, M., Feng, X.: 'Sparse Representation or Collaborative Representation : Which Helps Face Recognition ? Sparse Representation or Collaborative Representation : Which Helps Face Recognition ?', in 'IEEE International Conference on Computer Vision. IEEE International Conference on Computer Vision' (2011).

86 Hu, Y., Mian, A.S., Owens, R.: 'Face recognition using sparse approximated nearest points between image sets'IEEE Trans. Pattern Anal. Mach. Intell., 2012, 34, (10), pp. 1992–2004.

87 Zhao, Z.Q., Xu, S.T., Liu, D., Tian, W.D., Jiang, Z. Da: 'A review of image set classification'Neurocomputing, 2018, 335, pp. 251–260.

88 Cevikalp, H., Triggs, B.: 'Face recognition based on image sets'Proc. IEEE Comput.

Soc. Conf. Comput. Vis. Pattern Recognit., 2010, pp. 2567–2573.

89 Wang, R., Guo, H., Davis, L.S., Dai, Q.: 'Covariance discriminative learning: A natural and efficient approach to image set classification'Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2012, pp. 2496–2503.

90 Baudat, G., Anouar, F.: 'Generalized discriminant analysis using a kernel approach'Neural Comput., 2000, 12, (10), pp. 2385–2404.

91 Rosipal, R., Krämer, N.: 'Overview and recent advances in partial least squares'Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 2006, 3940 LNCS, pp. 34–51.

92 Wang, R., Shan, S., Chen, X., Gao, W.: 'Manifold-manifold distance with application to face recognition based on image set', in 'IEEE Conference on Computer Vision and Pattern Recognition, CVPR' (2008)

93 Wang, R., Shan, S., Chen, X., Dai, Q., Gao, W.: 'Manifold-manifold distance and its application to face recognition with image sets'IEEE Trans. Image Process., 2012, 21, (10), pp. 4466–4479.

94 Wang, R., Chen, X.: 'Manifold Discriminant Analysis', in '2009 IEEE Conference on Computer Vision and Pattern Recognition' (IEEE, 2009), pp. 429–436.

95 Sharma, P., Goyani, M.: '3D Face Recognition Techniques: A Review'Int. J. Eng. Res. Appl., 2012, 2, (1), pp. 787–793.

96 Gross, R.: 'Chapter 13. Face Databases', in 'Handbook of Face Recognition' (2005), pp. 301–327.

97 Georghiades, A.S., Belhumeur, P.N., Kriegman, D.J.: 'From few to many: Illumination cone models for face recognition under variable lighting and pose'IEEE Trans. Pattern Anal. Mach. Intell., 2001, 23, (6), pp. 643–660.

98 Lee, K.C., Ho, J., Yang, M.H., Kriegman, D.: 'Video-based face recognition using probabilistic appearance manifolds'Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2003, 1, (May 2014).

99 Lee, K.C., Ho, J., Yang, M.H., Kriegman, D.: 'Visual tracking and recognition using probabilistic appearance manifolds'Comput. Vis. Image Underst., 2005, 99, (3), pp. 303–331.

100 Phillips, P.J., Wechsler, H., Huang, J., Rauss, P.J.: 'The FERET database and evaluation procedure for face-recognition algorithms'Image Vis. Comput., 1998, 16, (5), pp. 295–306.

101 Phillips, P.J., Flynn, P.J., Scruggs, T., et al.: 'Overview of the face recognition grand challenge', in 'Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005' (2005), pp. 947–954

102 Martnez, A., Benavente, R.: 'The AR face database'Tech. Rep. 24 CVC Tech. Rep., 1998, (January).

103 Samaria, F.S., Harter, A.C.: 'Parameterisation of a stochastic model for human face identification', in 'IEEE Workshop on Applications of Computer Vision - Proceedings' (1994), pp. 138–142.

104 Gross, R., Matthews, I., Cohn, J., Kanade, T., Baker, S.: 'Multi-PIE', in '2008 8th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2008' (2008).

105 Gary, B.H., Ramesh, M., Berg, T., Learned-Miller, E.: 'Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments', in 'Workshop on Faces in "Real-Life" Images: Detection, Alignment, and Recognition' (IEEE, 2008).

106 Grgic, M., Delac, K., Grgic, S.: 'SCface - Surveillance cameras face

database'Multimed. Tools Appl., 2011, 51, (3), pp. 863–879.

107 Mccool, C., Matˇ, P., Ahonen, T., Cernock, J., Larcher, A., Christophe, L.: 'On the Results of the First Mobile Biometry (MOBIO)'ICPR 2010, LNCS, 2010, (6388), pp. 210–225.

108 Grgic, M., Delac, K.: 'Face Recognition Homepage: Databases', http://www.face-rec.org/databases/.

109 Faltemier, T.C., Bowyer, K.W., Flynn, P.J.: 'Using a multi-instance enrollment representation to improve 3D face recognition'IEEE Conf. Biometrics Theory, Appl. Syst. BTAS'07, 2007.

110 Yin, B., Sun, Y., Wang, C., Ge, Y.: 'BJUT-3D large scale 3D face database and information processing'2009, (15).

111 Moreno, A.., Sanchez, A.: 'GavabDB: a 3D face database', in 'Workshop on biometrics on the internet: fundamentals, advances and applications' (2004), pp. 77–85.

112 Panchal, K.K., Shah, H., Makwana, R.M.: '3D Face Recognition on GAVAB Dataset'Int. J. Eng. Res. Technol., 2013, **2**, (6), pp. 1353–1357.

113 Colombo, A., Cusano, C., Schettini, R.: 'UMB-DB: A database of partially occluded 3D faces', in 'Proceedings of the IEEE International Conference on Computer Vision' (2011), pp. 2113–2119.

114 Yin, L., Wei, X., Sun, Y., Wang, J., Rosato, M.J.: 'A 3D facial expression database for facial behavior research', in 'FGR 2006: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition' (2006), pp. 211–216.

115 Gupta, S., Castleman, K.R., Markey, M.K., Bovik, A.C.: 'Texas 3D Face Recognition Database'Proc. IEEE Southwest Symp. Image Anal. Interpret., 2010, pp. 97–100.

116 Hg, R.I., Jasek, P., Rofidal, C., Nasrollahi, K., Moeslund, T.B., Tranchet, G.: 'An

RGB-D Database Using Microsoft ' s Kinect for Windows for Face Detection in 'IEEE 8th International Conference on Signal Image Technology & Internet Based Systems' (2012), pp. 42–46.

117 Merget, D., Eckl, T., Schwoerer, M., Tiefenbacher, P., Rigoll, G.: 'Capturing facial videos with Kinect 2.0: A multithreaded open source tool and database'2016 IEEE Winter Conf. Appl. Comput. Vision, WACV (2016).

118 Piemontez, R.A., Comunello, E.: 'RAP3DF - One shoot 3D face dataset'Data Br., 2020, 32.

119 Cortes, C., Vapnik, V.: 'Support-Vector Networks'Mach. Leaming, 1995, 20, pp. 273–297.

120 Nielsen, M.: 'Neural Networks and Deep Learning' (2015).

121 Zhou, S., Xiao, S.: '3D face recognition: a survey'Human-centric Comput. Inf. Sci., 2018, 8, (1).

122 Patil, H., Kothari, A., Bhurchandi, K.: '3-D face recognition: features, databases, algorithms and challenges'Artif. Intell. Rev., 2015, 44, (3).

123 Savran, A., Alyüz, N., Dibeklioğlu, H., et al.: 'Bosphorus database for 3D face analysis'Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 2008, 5372 LNCS, pp. 47–56.

124 Beumier, C., Acheroy, M.: 'Automatic 3D face authentication'Image Vis. Comput., 2000, 18, (4), pp. 315–321.

125 Hesher, C., Srivastava, A., Erlebacher, G.: 'A novel technique for face recognition using range imaging'Proc. - 7th Int. Symp. Signal Process. Its Appl. ISSPA 2003, 2003, 2, pp. 201–204.

126 Xu, C., Tan, T., Li, S., Wang, Y., Zhong, C.: 'Learning effective intrinsic features to

boost 3D-based face recognition'Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 2006, 3952 LNCS, pp. 416–427.

127 Conde, C., Serrano, Á., Cabello, E.: 'Multimodal 2D, 2.5D & 3D face verification', in 'Proceedings - International Conference on Image Processing, ICIP' (2006), pp. 2061–2064.

128 Lu, X., Jain, A.: 'Deformation modeling for robust 3D face matching', in 'IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)' (2006).

129 Wang, Y., Pan, G., Wu, Z., Wang, Y.: 'Exploring facial expression effects in 3D face recognition using partial ICP'Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 2006, 3851 LNCS, pp. 581–590.

130 Heseltine, T., Pears, N., Austin, J.: 'Three-dimensional face recognition using combinations of surface feature map subspace components'Image Vis. Comput., 2008, 26, (3), pp. 382–396.

131 Vijayan, V., Bowyer, K.W., Flynn, P.J., et al.: 'Twins 3D face recognition challenge'2011 Int. Jt. Conf. Biometrics, IJCB (2011).

132 Chhokra, P., Chowdhury, A., Goswami, G., Vatsa, M., Singh, R.: 'Unconstrained Kinect video face database'Inf. Fusion, 2018, 44, pp. 113–125.

133 Zennaro, S.., Munaro, M.., Milani, S.., et al.: 'Performance Evaluation Of The 1st And 2nd Generation Kinect For Multimedia Applications', in 'Proceedings of the 2015 IEEE International Conference on Multimedia and Expo' (2015).

134 Mutto, C., Zanuttigh, P., Cortelazzo, G.: 'Time-of-Flight Cameras and Microsoft Kinect' (2012).

135 Sarbolandi, H., Lefloch, D., Kolb, A.: 'Kinect range sensing: Structured-light versus Time-of-Flight Kinect'Comput. Vis. Image Underst., 2015, 139, pp. 1–20.

136 Sireesha, V., Sandhyarani, K.: ‘Overview of Fusion Techniques in Multimodal’Int. J. Eng. Res. Technol., 2014, pp. 3–8.

137 Ryan, R., Baldridge, B., Schowengerdt, R.A., Choi, T., Helder, D.L., Blonski, S.: ‘IKONOS spatial resolution and image interpretability characterization’Remote Sens. Environ., 2003, 88, (1–2), pp. 37–52.

138 Duda, R.O., Hart, P.E., Stork, D.G., Wiley, J.: ‘Pattern Classification’ (2000).

139 Tao, D., Li, X., Wu, X., Maybank, S.J.: ‘Geometric mean for subspace selection’IEEE Trans. Pattern Anal. Mach. Intell., 2009, 31, (2), pp. 260–274.

140 Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: ‘Robust face recognition via sparse boosting representation’IEEE Trans. Pattern Anal. Mach. Intell., 2009, 31, (2), pp. 210–227.

141 Gaonkar, A.A., Gad, M.D., Vetrekar, N.T., Tilve, V.S., Gad, R.S.: ‘Experimental evaluation of 3D kinect face database’Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 2016, 10481 LNCS, pp. 15–26.

142 Ojala, T., Pietikäinen, M., Mäenpää, T.: ‘Multiresolution gray-scale and rotation invariant texture classification with local binary patterns’IEEE Trans. Pattern Anal. Mach. Intell., 2002, 24, (7), pp. 971–987.

143 Pajares, G., de la Cruz, J.M.: ‘A wavelet-based image fusion tutorial’ Pattern Recognit., 2004, 37, (9), pp. 1855–1872.

144 Amolins, K., Zhang, Y., Dare, P.: ‘Wavelet based image fusion techniques - An introduction, review and comparison’ ISPRS J. Photogramm. Remote Sens., 2007, 62, (4), pp. 249–263.

145 Arandjelović, O., Shakhnarovich, G., Fisher, J., Cipolla, R., Darrell, T.: ‘Face recognition with image sets using manifold density divergence’, in ‘IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005’ (2005),

pp. 581–588

146 Sra, S.: ‘Positive definite matrices and the Symmetric Stein Divergence’(2011).

147 Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: ‘DeepFace: Closing the gap to human-level performance in face verification’, in ‘IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2014).

148 Toshev, A., Szegedy, C.: ‘DeepPose: Human pose estimation via deep neural networks’ IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., (2014), pp. 1653–1660.

149 Yamaguchi, O., Fukui, K., Maeda, K.I.: ‘Face recognition using temporal image sequence’, in ‘International Conference on Automatic Face and Gesture Recognition (1998), pp. 318–323.

150 Oja, E.: ‘Subspace methods of pattern recognition’ (1983).

151 Kim, T.K., Kittler, J., Cipolla, R.: ‘Discriminative learning and recognition of image set classes using canonical correlations' (2007).

152 Harandi, M.T., Sanderson, C., Shirazi, S., Ovell, B.C.: ‘Graph embedding discriminant analysis on Grassmannian manifolds for improved image set matching’, in ‘IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2011), pp. 2705–2712

153 Liu, Y., Chen, X., Peng, H.,Wang Z.: ‘ Multi-focus image fusion with a deep convolutional neural network’, in ‘Information Fusion’, 2017, 36, pp. 191-207