

**STUDIES ON FACE DETECTION AND
RECOGNITION FROM STILL AND VIDEO
IMAGES**

THESIS SUBMITTED BY

MANAS GHOSH

DOCTOR OF PHILOSOPHY (ENGINEERING)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

FACULTY COUNCIL OF ENGINEERING AND TECHNOLOGY

JADAVPUR UNIVERSITY

KOLKATA, INDIA

2024

JADAVPUR UNIVERSITY

KOLKATA-700032, INDIA

INDEX NO: 002/18/E

TITLE OF THE THESIS:

STUDIES ON FACE DETECTION AND RECOGNITION FROM
STILL AND VIDEO IMAGES

NAME, DESIGNATION & INSTITUTION OF THE SUPERVISOR:

DR. JAMUNA KANTA SING

PROFESSOR

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

JADAVPUR UNIVERSITY, KOLKATA, INDIA

List of Publications

A. Journals:

1. Manas Ghosh and Jamuna Kanta Sing, "Interval type-2 fuzzy set induced fuzzy rank-level fusion for face recognition," *Appl. Soft Comput.*, vol. 145, no. 110584, p. 110584, 2023. [Impact Factor: 5.1, SCIE Indexed]

B. Conferences:

1. M. Ghosh and J. K. Sing, "Multi-feature-based type-2 fuzzy set induced parallel rank-level fusion in face recognition," in *Proc. Soft Computing: Theories and Applications*, Indian Institute of Information Technology Kota, (Malaviya National Institute of Technology Jaipur), Rajasthan, India, December 17-19, 2021, pp. 181–190.
2. M. Ghosh and J. K. Sing, "Enhanced face recognition using rank-level fusion with multi-feature vectors and interval type-3 fuzzy set," in *Proc. 2023 IEEE 3rd Applied Signal Processing Conference (ASPCON)*, Haldia Institute of Technology, Haldia, West Bengal, India, November 24 – 25, 2023, pp. 118-123.
3. M. Ghosh and J. K. Sing, "Detecting and recognizing faces from video images using multi-deep CNN based rank-level fusion," in *Proc. 2024 2nd International Conference on Advancement in Computation & Computer Technologies (InCACCT)*, Chandigarh University, Chandigarh, Punjab, India, May 2-3, 2024, pp. 283-287.

List of Patents: Nil

List of Presentations in International Conferences

1. M. Ghosh and J. K. Sing, "Multi-feature-based type-2 fuzzy set induced parallel rank-level fusion in face recognition," in *Proc. Soft Computing: Theories and Applications*, Indian Institute of Information Technology Kota, (Malaviya National Institute of Technology Jaipur), Rajasthan, India, December 17-19, 2021, pp. 181–190.
2. M. Ghosh and J. K. Sing, "Enhanced face recognition using rank-level fusion with multi-feature vectors and interval type-3 fuzzy set," in *Proc. 2023 IEEE 3rd Applied Signal Processing Conference (ASPCON)*, Haldia Institute of Technology, Haldia, West Bengal, India, November 24 – 25, 2023, pp. 118-123.
3. M. Ghosh and J. K. Sing, "Detecting and recognizing faces from video images using multi-deep CNN based rank-level fusion," in *Proc. 2024 2nd International Conference on Advancement in Computation & Computer Technologies (InCACCT)*, Chandigarh University, Chandigarh, Punjab, India, May 2-3, 2024, pp. 283-287.

STATEMENT OF ORIGINALITY

I, **Manas Ghosh**, Registration No. **1021804004** of **18-19** registered on **4th June 2018** do hereby declare that this thesis entitled "**Studies on Face Detection and Recognition from Still and Video Images**" contains literature survey and original research work done by the undersigned candidate as part of Doctoral studies.

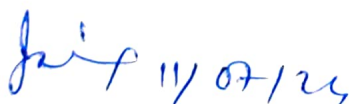
All information in this thesis has been obtained and presented in accordance with existing academic rules and ethical conduct. I declare that, as required by these rules and conduct,

I have fully cited and referred all materials and results that are not original to this work. I also declare that I have checked this thesis as per the "Policy on Anti Plagiarism, Jadavpur University, 2019", and the level of similarity as checked by iThenticate software is **2%**.


Signature of Candidate:

Date: **11.07.2024**

Certified by Supervisor:

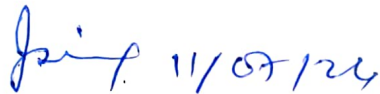


(Signature with date, seal)

Jamuna Kanta Sing, Ph.D.
Professor
Dept. of Computer Science & Engineering
Jadavpur University, Kolkata-700032

CERTIFICATE FROM THE SUPERVISORS

*This is to certify that the thesis entitled “**Studies on Face Detection and Recognition from Still and Video Images**” submitted by **Sri. Manas Ghosh**, who got his name registered on 4th June, 2018 for the award of Ph.D. (Engg.) degree of Jadavpur University is absolutely based upon his own work under the supervision of **Dr. Jamuna Kanta Sing**, Professor, Department of Computer Science & Engineering, Jadavpur University, and that neither his thesis nor any part of the thesis has been submitted for any degree or any academic award anywhere before.*



Signature of Supervisor and date with Official Seal

Jamuna Kanta Sing, Ph.D.

Professor

Dept. of Computer Science & Engineering
Jadavpur University, Kolkata-700032

Dr. Jamuna Kanta Sing

Professor

Department of Computer Science & Engineering

Jadavpur University, Kolkata, India.

Acknowledgments

I would like to express my heartiest gratitude and earnest respect to Prof. Jamuna Kanta Sing, my supervisor, and my guiding light, for meticulously monitoring the course of my work. He has been a mentor extraordinaire.

I would like to thank my dear friend, Mr. Tapas Kumar Tunga, for his unwavering support and belief in my efforts no matter what happened. He has been a speck of hope every time I felt like things were not going as they were supposed to in my life. He has provided me with educational resources ever since I stepped into my academic career.

I would like to express my sincere thanks to Dr. Aniruddha Dey for helping me navigate through the process of my research work and for providing insightful suggestions for a better quality of work.

My sincere regards to Prof. Anirban Mukherjee for his inspirational influence and support in my professional space which helped me devote the best of my time towards my research without having to worry about my assigned academic and administrative responsibilities.

I would like to thank my fellow researchers at Jadavpur University, Dr. Sayan Kahali, and Ms. Madhumita Ray for their constant encouragement and support whenever I felt demotivated. They helped me through each and every phase of my research work both mentally and academically.

My sincere thanks to my friend and colleague, Mr. Biswanath Chakraborty, for having my back and helping me manage my work and academics accordingly. It was for him that I was able to seamlessly balance my professional responsibilities and my research work.

Lastly, I would like to thank my family, the biggest source of motivation and support for me came from them. Their constant encouragement and promising words helped me focus and give my best to the research.

Manas Ghosh

Jadavpur University
Kolkata, India

Manas Ghosh

To my family

Abstract

All tasks for facial analysis, including facial recognition, are initiated with face detection, which embraces precise alignment and augmented modeling of the face. Various factors associated with heterogeneous environmental conditions and imaging settings result in intra-subject diversity and inter-subject resemblance, leading to uncertainties in the recognition process. Different fuzzy sets and systems have proven their potential in managing uncertainty at various levels. Considering its merits, adopting rank-level fusion for face recognition is prudent for improved precision. With the common intention of proposing a more refined approach, the thesis titled “Studies on Face Detection and Recognition from Still and Video Images” encompasses four approaches to facial recognition in which various fuzzy sets are expended in the rank-level fusion involving different feature representations, derived by distinct feature extraction techniques.

The first face recognition methodology entails a general type-2 fuzzy set (GT2 FS) to address intra-class diversity and inter-class resemblance due to environmental conditions and imaging settings. This rank-level fusion framework involves three distinct sets of confidence factors (CFs) from a classifier fed with three different feature representations (vectors) corresponding to the three feature extraction techniques. From each set, the top classes were selected based on the CFs. For each class, a GT2 FS was shaped with feature vectors from the test image and training samples of the respective class. The complement of the type-reduced defuzzified value was regarded as a fuzzy rank. For each class, common to one or more sets, the fuzzy ranks are aggregated by summing them, and then this sum is fused with the complemented mean of the confidence factors to obtain a final score for each class. The test image was recognized based on its final rank.

The second face recognition methodology has taken advantage of an interval type-2 fuzzy set (IT2 FS) that utilizes intra-class face images to generate lower and upper membership functions defining the footprint of uncertainty (FOU) of IT2 FS. Distinct feature vectors derived from different FE techniques for a face image are used to utilize the different underlying discriminant features, which, in turn, address the issues of inter-class similarity. The defuzzified value was used to generate the fuzzy rank for the

class under consideration. The top K classes, which are selected based on the CFs of the classifier, are taken into consideration for the fusion process. For each class, the fuzzy ranks are integrated by summing and fusing with the respective complemented mean of the confidence factors to obtain the final fused rank. If a class does not appear in the top-ordered classes for a feature vector, a penalty is set against it. A facial image is identified based on the final fuzzy rank.

The third methodology employs an interval type-3 fuzzy set (IT3 FS) and multiple feature representations of a facial image to resolve uncertainties arising from intra-class divergence and inter-class similitude due to factors related to imaging and environmental circumstances. Corresponding to the input feature vector, a selected number of classes according to the CFs of the classifier are involved in rank-level fusion. For each class, corresponding training samples were employed to define the secondary membership domain. For each training sample, smaller proximity measures using training samples from all classes were set apart to express the lower and upper membership functions of the scaled Gaussian domain of the FOU of the IT3 FS. Successive direct defuzzifications lead to a crisp value, which is deemed a fuzzy rank for each top-ordered class. We combined these fuzzy ranks judiciously with the respective confidence factors generated by multiple feature vectors to obtain the final rank for that class. The class with the lowest rank was identified as the test-image class.

Another subsequent approach explored a methodology wherein the faces were detected and extracted from an outdoor video clip, and a custom-made face dataset was created. A rank-level fusion mechanism was implemented for facial recognition that uses the probabilities produced by the softmax functions of multiple deep CNN models. The fuzzy ranks are derived for the top-ordered classes from the probabilities produced by the softmax function (termed certainty indices) of the respective deep model by applying a Gaussian distribution function, followed by a complementation operation. Considering three deep CNN models, the complement of the average of the certainty indices is combined with the summation of the fuzzy ranks for each class. A penalty mechanism is implemented for a class not included in the top-ordered classes corresponding to an FV. The class with the lowest final ranking was identified as the test-sample class. All methodologies expounded in this thesis were evaluated on several face databases, such as AT&T, FERET, UMIST, Faces94, and Chokepoint, and demonstrated their pre-eminence over similar related approaches.

Contents

1.	INTRODUCTION	
1.1	Introduction	1
1.2	Generic face recognition framework	2
1.3	Challenges of face recognition	3
1.4	Approaches to face detection and its taxonomy	4
1.5	Approaches to face recognition and its taxonomy	7
1.6	Detecting and recognizing faces using deep learning	
1.6.1	Deep learning	10
1.6.2	Convolutional neural network (CNN)	10
1.6.3	Face recognition using deep convolution neural network	12
1.6.4	Deep face detection	13
1.6.5	Deep face recognition	14
1.7	Multi-biometric fusion	15
1.8	Motivation of the thesis work	21
1.9	Objective of the thesis work	21
1.10	Organization of the thesis	23
1.11	Databases used in the studies	
1.11.1	AT & T face database	25
1.11.2	FERET face database	26
1.11.3	UMIST face database	26
1.11.4	Faces94 face database	27
1.11.5	Chokepoint face database	27
2.	GENERAL TYPE-2 FUZZY SET INDUCED RANK-LEVEL FUSION FOR FACE RECOGNITION	
2.1	Introduction	29
2.2	Related concepts	
2.2.1	Fusion in face recognition revisited	31
2.2.2	Rank-level fusion methods	33
2.2.2.1	Highest-rank method	33
2.2.2.2	Modified highest-rank method	33
2.2.2.3	Borda count method	34
2.2.2.4	Weighted Borda Count or Logistic Regression Method	34

2.2.2.5	Non-linear Weighted Rank Method	34
2.2.3	An overview of Type-2 fuzzy sets	35
2.3	General type-2 fuzzy set induced rank-level fusion for face recognition	
2.3.1	Methodology	38
2.3.2	Mathematical Formulation of the Methodology	39
2.4	Experimental results	43
2.4.1	Experiments with the AT&T Face Database	44
2.4.2	Experiments with the UMIST Face Database	46
2.5	Conclusion	48
3.	INTERVAL TYPE-2 FUZZY SET INDUCED FUZZY RANK-LEVEL FUSION FOR FACE RECOGNITION	
3.1	Introduction	50
3.2	A concise overview of interval type-2 fuzzy sets	54
3.3	Interval type-2 fuzzy set induced fuzzy rank-level fusion for face recognition	
3.3.1	Methodology	56
3.3.2	Mathematical formulation of the methodology	59
3.4	Experimental results	63
3.4.1	Experiments on the AT&T face database	64
3.4.2	Experiments on the UMIST face database	66
3.4.3	Experiments on the FERET face database	68
3.4.4	An illustrative example	70
3.5	Conclusion	75
4.	INTERVAL TYPE-3 FUZZY SET INDUCED FUZZY RANK-LEVEL FUSION FOR FACE RECOGNITION	
4.1	Introduction	76
4.2	A Concise Overview of Interval Type-3 Fuzzy Sets	79
4.3	Interval Type-3 Fuzzy Set Induced FUZZY RANK-Level Fusion for Face Recognition	
4.3.1	Methodology	83
4.3.2	Mathematical Formulation of the Methodology	84
4.4	Experimental Results	87
4.4.1	Experiments on the FERET face database	87
4.4.2	Experiments on the UMIST face database	88

4.4.3	Experiments on the FACES94 face database	89
4.4.4	An Illustrative Example	93
4.5	Conclusion	96
5.	FACE DETECTION AND RECOGNITION FROM VIDEO IMAGES USING DEEP CNN BASED RANK LEVEL FUSION	
5.1	Introduction	97
5.2	Related concepts	
5.2.1	Multitask cascaded convolutional neural network (MTCNN)	101
5.2.2.	VGG16	101
5.2.3	InceptionResNetV2	102
5.2.4	Xception	102
5.3	Face detection and recognition from video images using deep CNN based rank level fusion	
5.3.1	Methodology	103
5.3.2	Mathematical formulation of the methodology	104
5.4	Experimental results	105
5.4.1	Detecting faces from a video clip and preparing a custom face dataset	107
5.4.2	Experiments with the custom face dataset	108
5.4.3	Experiments with the Chokepoint face dataset	110
5.5	Conclusion	111
6.	CONCLUSION AND FUTURE SCOPE	
6.1	Conclusion	113
6.2	Future scope	115

Bibliography

List of Tables

Table 1.1	Summary of a few related studies on sensor-level fusion	16
Table 1.2	Summary of a few related studies on feature-level fusion	17
Table 1.3	Summary of a few related studies on decision-level fusion	17
Table 1.4	Summary of a few related studies on score-level fusion	18
Table 1.5	Summary of studies on rank-level fusion	19
Table 2.1	Performance evaluation of contemporary methods on the AT&T face database.	44
Table 2.2	Statistical significance analysis between the competitive methods and the IT2FSIFRLF method on the AT &T face database.	45
Table 2.3	Performance evaluation of contemporary methods on the UMIST face database.	46
Table 2.4	Performance evaluation of contemporary methods on the UMIST face database.	47
Table 3.1	Summarization of different rank-level fusion techniques.	53
Table 3.2	Recognition rates of the competitive methods on the AT&T face database. The best values are in bold.	64
Table 3.3	Statistical significance analysis between the competitive methods and the proposed IT2FSIFRLF method on the AT &T face database.	65
Table 3.4	Recognition rates of the competitive methods on the UMIST face database.	67
Table 3.5	Statistical significance analysis between the comparative methods and the proposed IT2FSIFRLF method on the UMIST face database.	67
Table 3.6	Recognition rates of the competitive methods on the FERET face database.	68
Table 3.7	Statistical significance analysis between the competitive methods and IT2FSIFRLF method on the FERET face database.	68
Table 3.8	The set of top K classes and the normalized confidence factors corresponding to three feature extraction techniques.	71
Table 3.9	Class-wise normalized degrees of conformity, LMFs, UMFs, and the crisp defuzzified values for top K classes corresponding to the G-2DFLD method.	71
Table 3.10	Class-wise normalized degrees of conformity, LMFs, UMFs, and the crisp defuzzified values for top K classes corresponding to the FG-2DFLD method.	72
Table 3.11	Class-wise normalized degrees of conformity, LMFs, UMFs, and the crisp	72

	defuzzified values for top K classes corresponding to Modular LBP.	
Table 3.12	Calculation of complemented CF sum and complemented fuzzy rank sum	73
Table 3.13	Final ranks of image 70 by the CFGaussFRLF and IT2FSIFRLF methods with $r = 3$.	74
Table 4.1	Comparison of the average rate of recognition success using the FERET database.	88
Table 4.2	Statistical significance analysis of IT3FSIRLF with the other two methods on the FERET database	88
Table 4.3	Comparison of the average rate of recognition success using the UMIST database.	89
Table 4.4	Statistical significance analysis of IT3FSIRLF with the other two methods on the UMIST database	89
Table 4.5	Comparison of the average rate of recognition success using the Faces94 database.	90
Table 4.6	Statistical significance analysis of IT3FSIRLF with the other two methods on the Faces94 database	90
Table 4.7	Outputs generated by RBFNN when applied to test FVs produced by 3 FE techniques	93
Table 4.8	The fuzzy ranks, derived by formulating IT3 FS as described in IT2FSIFRLF	94
Table 4.9	Fused scores for each of the topmost classes for IT2FSIFRLF	94
Table 4.10	Complemented fuzzy ranks derived by formulating IT3 FS as described in IT3FSIRLF	95
Table 4.11	Fused scores for each of the top classes for IT3FSIRLF	95
Table 5.1	the recognition accuracy of the methods on the custom face database	108
Table 5.2	The arithmetic mean of macro averages of the face recognition model that uses the custom face database	109
Table 5.3	statistical significance analysis of the MDCNNRLF with three deep CNN models on the custom database	109
Table 5.4	The recognition accuracy of the methods on the chokepoint face database	110
Table 5.5	The arithmetic mean of macro averages of the face recognition model that uses the custom	111
Table 5.6	Statistical significance analysis of the MDCNNRLF with three deep CNN models on the chokepoint database	111

List of Figures

Fig. 1.1	Face detection from a video frame image	2
Fig. 1.2	A simplified working model of a CNN	12
Fig. 1.3	Representative images from AT & T face database	26
Fig. 1.4	Representative images from FERET face database	26
Fig. 1.5	Representative images from UMIST face database	27
Fig. 1.6	Representative images from the Faces94 facial dataset	27
Fig. 1.7	Representative images from the Chokepoint facial dataset	28
Fig. 2.1	View of a general type-2 Gaussian MF	37
Fig. 2.2	Schematic description of the MFT2FPRLF method	41
Fig. 2.3	Graph illustrating recognition rate versus P (number of top-ordered classes considered in rank-level fusion)	48
Fig. 3.1	Domains of an IT2 FS with Gaussian MF	56
Fig. 3.2	Schematic diagram of the workflow of the IT2FSIFRLF.	58
Fig. 3.3	Flow diagram describing the efficiency of IT2FSIFRLF	62
Fig. 3.4	The plot of average recognition rate versus K (AT & T database with $r=4$)	69
Fig. 3.5	The plot of average recognition rate versus K (FERET database with $r=3$)	70
Fig. 4.1	Three domains of the interval type-3 membership function	82
Fig. 4.2	Schematic description of the present approach.	84
Fig. 4.3	Graph of average recognition rate versus H (UMIST database with $r = 8$)	91
Fig. 4.4	Graph of average recognition rate versus H (FERET database with $r = 3$)	92
Fig. 4.5	Graph of average recognition rate versus K (UMIST database with $r = 8$)	92
Fig. 4.6	Graph of average recognition rate versus K (FERET database with $r = 3$)	93
Fig. 5.1	Diagrammatic outline of the MDCNNRLF method	106
Fig. 5.2 (a)	Detection of faces with various scales from video frames	107
Fig. 5.2 (b)	Detection of faces with various scales from video frames	107
Fig. 5.2 (c)	Extracted faces with partial occlusion, motion blur, and pose	107

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

Identity recognition is normally accomplished using one or more means: possession-based (e.g., passports, IDs), token-based or password-based, and biometric-based authentication [1]. Biometrics are preferred and mostly used for identity recognition because their characteristics or traits are impervious to theft or forfeit [2]. The term ‘biometric’ originated from two Greek words ‘bios’ implying life and ‘metron’ signifying measure. Identifying or recognizing a person using biometric measures relies on their physiological (iris, face, palmprint, ear, etc.) behavioral (gait, signature voice, etc.), or biological (ECG, EEG, DNA, etc.) attributes [3]. These innate characteristics are used to identify a person by leveraging hardware devices with artificial intelligence-driven pattern analysis algorithms. The terms modalities, traits, and cues alternatively denote these characteristics [4]. Compared to other biometric modalities, the face holds a significant edge because it can be captured passively without direct involvement. For biometric identity recognition, face recognition (FR) has garnered significant attention from researchers in diverse fields [5]. Their application has been observed in many secure access control environments, law enforcement departments, and other computer vision domains. This is because the static and dynamic attributes of a face can be utilized to evaluate it [6], and numerous legacy face databases are accessible for

evaluating related methodologies [7]. As technology for sensing and imaging continues to evolve, the use of FR has become increasingly prevalent in various social and commercial applications.

1.2 GENERIC FACE RECOGNITION FRAMEWORK

Typically, a generic FR system passes through different steps that involve (i) acquisition of raw face image data using sensors, accompanied by face detection (FD) from faces captured from still images or video sequences (frames), (ii) extraction of discriminative features, (iii) classification, and (iv) recognition. FD methods result in the coordinates defining the rectangular outlines around each face present in the image. The position, angle, partial coverage, and size of the faces are impediments to FD. Canonical alignment, scaling, and other preprocessing operations may be applied on the clip-out face before FR. Fig. 1.1 depicts an FD from a video frame image. The detected face is enclosed within a coloured rectangular box. The pre-processed facial images are also shown.

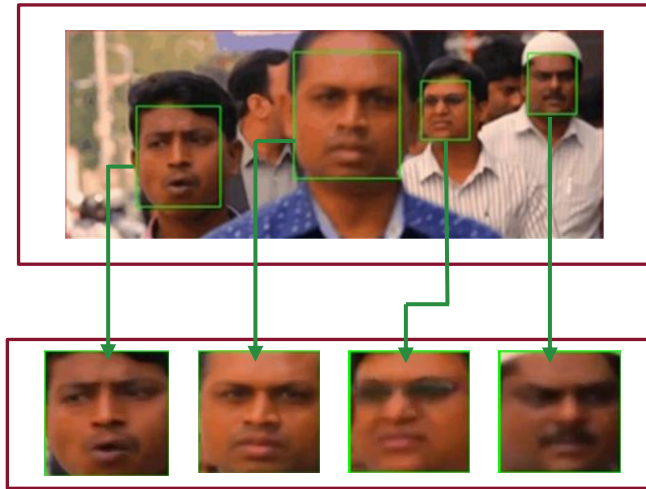


Fig. 1.1. Face detection from a video frame image

In feature extraction (FE), feature sets are drawn from a digital representation of biometric traits. Features acquired from facial images may consider the visual, statistical, transformed, or algebraic aspects of the face [8]. Visual features are characterized by the edges, textures, contours, and selected regions of the image. Statistical features represent the delineation of histograms and other statistical moments. Transformed coefficient-based features refer to features extracted from the transformed version of the image, such as Fourier, direct cosine,

and wavelet transformations. The algebraic facial features are obtained by applying algebraic equations to the representation of spatial facial characteristics such as the distance and angles between landmarks. In [9], various methods for extracting features were debated. The extracted feature vector (FV) from a facial image along with the respective class identifier is saved as a template in the database. To classify a test sample, the FV extracted using the same FE technique was compared with the templates to determine the extent of similarity.

1.3 CHALLENGES OF FACE RECOGNITION

The barriers that impede the effective FR are outlined below [10-15].

- (1) Diverse imaging conditions lead to varying resolution, luminosity, scale, and orientation of the face. These changes are attributed to camera features such as the lens, sensor response, and spectral range. Image quality is largely dependent on the level of illumination and the angle of light. The variability in the scale and orientation was ascribed to the positioning of the camera relative to the face.
- (2) Varied countenance is caused by emotional shifts such as anger, laughing, smiling, and disguise, and facial gestures are also changed by closing the eyes and/or mouth. Expressions can cause significant disfigurement of particular facial areas leading to different facial appearances. When facial appearance varies, the extracted features are also affected.
- (3) Complicated backgrounds, the proliferation of objects, and the assortment of faces in an image create a nodus in detecting faces as well as in FR.
- (4) Morphological variations such as mustache, beards, and spectacles pose difficulties in FR.
- (5) Occlusion is another factor that hinders an effective FR, particularly in unrestricted environments. Such occlusion may ensue when parts of the face are obstructed by objects like masks, hair, or hands.
- (6) Aging and the development of wrinkles can complicate facial recognition.
- (7) The reliability of an FR system is compromised by image counterfeiting.

All of the aforementioned factors lead to noise, inter-class resemblance, and intra-class deviation. Noise can result in sensor defects and user engagement during image acquisition.

Inappropriate FE and aging can give rise to intraclass variations. Untoward feature selection, intrinsic resemblance of subjects (e.g., twins), and substantial enrollment size give rise to interclass resemblance. There is no FR framework available that can address all these barriers and no such contemporary methodology that is completely accurate. The FR precision is still significantly affected by any of the changes mentioned above [16].

1.4 APPROACHES TO FACE DETECTION AND ITS TAXONOMY

Two distinct types of approaches have emerged for FD: feature-based and image-based. The feature-based approach correlates features such as edges, corners, and other localized patterns derived from the image with an established understanding of the facial features. In the image-based approach, pixel-wise image scanning through windowing or sub-framing is the fundamental mechanism for distinguishing a face from a non-face.

The feature-based approach employs either a model-centric or an analysis-centric strategy. The active shape model (ASM) is model-centric and concentrates on the non-rigid characteristics of the face. It involves the automatic mapping of landmarks or key points that describe the shape of a face through statistical modeling, utilizing manual labeling of landmarks or key points of the training images. It falls into the following four groups-

- a) Snakes [17]: Active contours, known as snakes, are used to trace the head boundary and its innate evolution by minimizing the energy function, thereby gradually reshaping the head.
- b) Deformable template matching (DTM) [18] was propounded to resolve the challenges of edge detection, arising from inadequate lighting or contrast. Intensity variation, local valleys, peaks, edges, etc., indicate deformation. The detection process is driven by predefined templates. The size and parameters of the templates were adjusted to delineate the face.
- c) The deformable part model (DPM) [19] relies on the training. In this method, the formation of a face mask involves individual imitation of the discrete rigid elements of the face, like the eyes, mouth, and nose. These parts are constrained by spatial relationships which can be envisioned as springs. Their appearance should evolve linearly in the feature space, maintaining their locations congruous with the

deformation. The model is positioned on the image and warped in a way to find its optimal placement ensuring minimal pressure on springs and revealing the image data beneath it.

- d) Point distribution models (PDM) [20] are concise representations of shapes using statistical modeling. The outline or contour of a PDM is segmented into the arrangements of annotated (labeled) points. The deviations of these points can be delineated from a training dataset made up of faces of different sizes and poses. These feature variations lead to a linearly adaptable model.

Analysis-centric strategies include low-level and feature analysis.

The foundation of low-level analysis lies in evaluating low-level visual traits utilizing pixel characteristics, such as color space, intensity level, edge, and motion information.

- Detecting a face in a color image requires segmentation of the skin region. This segmentation method relies on the skin chromaticity values obtained from various color spaces (e.g., RGB and HIS). This follows thresholding to mask the skin region. The face is isolated by the subsequent demarcation of the bounding box. The authors of [21] and [22] proposed a color-based FD.
- The grey information of a face is regarded as a feature. The gray-scale-based FD makes use of various FE algorithms that rummage around local gray-level minima within the segregated facial area. The low-level grayscale threshold is applied to separate dark patches. It employs a three-tier hierarchical approach to face localization [20].
- An intense variation in brightness is characterized as an edge. In edge-based FD methods, operators and filters such as the Sobel operator [22], steerable filter [23], etc., were employed to detect edges and correlate the edges of an image with a predefined template set to achieve precise detection of a face.
- Motion-based analysis is appropriate for detecting faces from video, and techniques such as frame variance analysis, line clustering algorithm [24], and optimal flow analysis [25] have been employed. McKenna et al. [26] advocated an approach that leverages a spatiotemporal filter to discern the contour of a moving face.

Feature analysis methods rely on comprehensive knowledge of the face model, as well as on the measurements of the normalized distances and angles of the facial elements (nose, eyes,

mouth, etc.). It searches for structural characteristics of the face that are independent of pose and illumination and utilizes them for FD.

- In feature-searching techniques, distinguished structural facial features are sought. It then seeks less notable features by utilizing facial geometry metrics. Commonly used reference features include eye pairs, facial outlines, and face axes [27, 28]. The Viola–Jones method [29] is a prominent feature-searching method. In addition to the Viola-Jones methods, dynamic link architecture (DLA) [30] and EBGM [31] with Gabor features are examples of such methods for FD. The Viola-Jones algorithm can be broadly expounded in three steps: converting the image from colour to grayscale, scouting a face out of the grayscale image, and tracking down the face in the colour image. In the second step, Haar-like features are quickly computed to characterize the integral images, followed by the expeditious separation of the non-face background regions with a cascade framework. The AdaBoost algorithm was subsequently employed to integrate the selected Haar-like features into a composite (ensemble) classifier configured in a cascaded fashion. The performance of the Viola-Jones FD method is compromised when faces are viewed from unusual angles, blurry, or obscure to some extent.
- The constellation analysis method [32] is based on the formation of constellations using face-like features localized in appropriate regions. The optimal constellation is identified using a random graph-matching scheme with vertices denoting features and arcs, indicating the distance between these features. Ranking of the face and non-face constellation is applied based on the probability density function. Other constellations, as proposed by the researchers, include the statistical shape theory and probabilistic shape model.

The image-based approach takes advantage of statistical analysis and machine learning to distinguish the pertinent features of facial and non-facial imagery. The methods following the image-based approach can be grouped into statistical method, linear subspace method, and neural network (NN).

- In statistical methods, discriminative functions such as thresholds, decision surface hyperplanes, etc., or statistical distribution patterns are formulated from the learned features. The analysis of these functions or patterns is eventually applied to discern face from non-face classes [33]. The well-established statistical methods for FD

include principal component analysis (PCA), discrete cosine transform (DCT), and independent component analysis (ICA). Many researchers have successfully exploited support vector machine (SVM) as a classifier for FD [36]. The features are derived using FE methods, such as PCA and histogram of gradients (HOG). Training the SVM with the extracted features enables face and non-face classification.

- Considering the image space as a vector space, the linear subspace is a small section. This subspace is also reckoned as vector space. Linear subspace methods are based on the principle that faces are confined to a subspace within an entire image space. Examples of such linear subspaces include Eigenfaces, Fisher faces, and tensor faces. [34, 35]. Several analysis methods, such as PCA, linear discriminant analysis (LDA), and factor analysis, have been employed.
- NN comprehends face patterns with training images and discerns faces in test images. Various NNs have evolved in recent years. Notable NNs include artificial neural network (ANN), fuzzy NNs (FNN), and decision-based NN (DBNN). Well-established ANNs include feed-forward NN, backpropagation NN, fast NN, radial basis function NN (RBFNN), probabilistic decision-based NN (PDBNN), and Convolutional NN (CNN). Comprehensive descriptions of such NNs, along with related research studies, are reported in [36]. A key benefit of employing NN for FD is their ability to accurately capture intricate density patterns associated with faces. Nevertheless, a potential challenge is that the network should be configured optimally by adjusting its parameters, such as the number of nodes, rate of learning, and layers [37]. FD utilizing CNN is discussed in section 1.6.4.

1.5 APPROACHES TO FACE RECOGNITION AND ITS TAXONOMY

Since researchers from diverse domains have presented various methods of FR in controlled as well as uncontrolled environments, it is onerous to assort these methods in a definitive manner. However, approaches to FR can be categorized as either classical or modern [5]. The classical approach employs global, local, or hybrid methods. The terms global and local represent the facial areas that are regarded as FE. Global implies that the entire facial region or all facial components are taken into consideration for FE. In the case of a local, FE involves considering smaller segments of either the entire face or a subset of facial

components. Global holistic approaches utilize the entire face region and employ linear or nonlinear methods to project facial data into a subspace of smaller dimension. The subspace is characterized by a collection of basis vectors through the learning of the training dataset. Independent component analysis (ICA), linear discriminant analysis (LDA), and frequency domain analysis are some of the common linear methods, along with principal component analysis (PCA) [38, 39]. Kernel-PCA and Kernel LDA, etc. are some of the notable nonlinear methods [40]. Global methods utilizing a transformed version of the facial image evaluate an image in the wavelet or frequency domain. The statistics of the filter responses, such as the Gabor filter or wavelet filter, construe the features [41, 42]. In [40], classical approaches were summarized with criticism with respect to advantages, performance, challenges managed, and limitations. Local feature-based approaches distinguish different distinct regions or key points of a face based on their geometric aspects using statistical patterns or graph-matching algorithms. Local appearance-based methods leverage histograms, pixel orientation, correlation planes, etc. to characterize local features; whereas key-point-based local techniques extract features that are localized on these points [43]. Some noteworthy appearance-based methods include histogram of oriented gradients (HOG), local binary pattern (LBP), speeded-up robust features (SURF), and scale-invariant Feature Transform (SIFT) [43-46]. LBP employs a texture-driven method that delves into the organization of neighboring pixels and the resultant histograms. It is still an often choice for researchers owing to its straightforward calculation, impervious to light variations, and robust separability. Researchers have suggested several refinements to outdo its performance by exploring different dimensions of the classical LBP operator, including the determination of adjacent (neighboring) pixels [47], distinguishing capability [48], and reliability [49]. A comparative evaluation of LBP and its derivatives in FR was presented by Meena et al. [50]. Local key point-based approaches leverage the knowledge of human facial anatomy to develop a graph-driven or statistical model that utilizes the geometric features of facial landmarks or key points. A graph or pattern correlational strategy is applied for the classification. Unlike global methods, these methods are capable of withstanding variations in the scale, alignment, and illumination levels [51]. Such methods include active appearance model (AAM), elastic bunch graph matching (EGBM), and others [52]. Hybrid FR models that integrate global and local approaches have also been documented in the literature to increase recognition accuracy [53,54]. Sing et al. [55] put forward a hybrid method for FR, which embraces local and global features involving PCA and Fisher linear discriminant

(FLD). Many researchers [54, 56, 57, 58] have opted for hybrid models in their studies to improve the reliability of FR.

Classification using an NN entails feeding the network with a diverse set of training images. The NN selects the appropriate distinguishing features through layers that, in turn, are utilized to recognize the class of the face. Multilayer Perceptron (MLP), Self-Organizing map (SOM), and linear and non-linear auto-associative networks are some NNs employed by the researchers [40]. Probabilistic Decision-based NN (PDBNN) was put forward by Lin et al. [59] for FR, in which a multi-tiered network architecture featuring non-linear basis functions and an effective credit distribution method was involved.

FR using modern approaches employs deep learning, fuzzy set theory, or dictionary learning. Deep learning (DL) has become more widespread owing to the advancement of powerful computing systems, and several recent approaches to FR using deep learning have achieved excellent results. DL-based FR has changed the direction of research with respect to the design, implementation, and evaluation framework [60].

Section 1.6.5 discusses FR using DL. The use of fuzzy set theories in FR has showcased improved performance in recent research studies. Li et al. propounded a fuzzy-based 2DPCA method [61] to address the sensitivity of traditional 2DPCA to illumination, pose, and expression in FR. Chowdhury et al. [62] extended the 2DPCA methodology by introducing the generalized 2DFLD method (G-2DFLD), which is proven to outdo in respect of FE and FR than PCA and 2DPCA methods. In the same stride, Dey et al. [63] discussed a methodology for FR that utilized the fuzzy generalized 2DFLD method (FG-2DFLD) method and RBFNN. In recent years, researchers from diverse disciplines have integrated type-2 fuzzy sets into their studies to contend with the inherent uncertainty. Authors of [64] pointed out the influence of uncertainty in FR and proposed an interval type-2 fuzzy LDA framework to mitigate the imprecision due to uncertainty. Murugeswari et al. [65] proposed an image classification methodology in which an interval type-2 fuzzy CNN was exploited. Yadav et al. [66] suggested the utilization of interval type-2 fuzzy logic along with kernel sparse representation in FR.

1.6 DETECTING AND RECOGNIZING FACES USING DEEP LEARNING

1.6.1 DEEP LEARNING

Deep learning (DL) is a sophisticated genre of machine learning composed of several tiers of NNs capable of learning intricate nonlinear (hierarchical) complex feature descriptions from audio, video, and textual data. The orthodox machine-learning algorithm is constrained by its limited feature-learning ability. Consequently, its generalization proficiency for intricate feature matching (classification) is restricted to a certain extent [67]. With the introduction of the deep belief network [68], DL began to evolve, and over the years, it has matured. It has been extensively adopted in research focused on detection (face, vehicle, pedestrian, etc.), recognition (pattern, face, speech, etc.), analysis (medical image, financial, etc.), and control (robot, traffic, vehicle, etc.) [69]. Among several DL models, some momentous models include the deep Boltzmann machine (DBM) [70], CNN [71], and recurrent NN [72]. Recently, more sophisticated DL models, such as attention-based long short-term memory and deep residual shrinkage networks have emerged [67].

1.6.2 CONVOLUTIONAL NEURAL NETWORK (CNN)

Inspired by biological models, deep CNN exploits a multi-strata design to derive the hierarchical characteristics of data. A CNN is characterized by multilevel learning, self-driven FE, and a weight-sharing mechanism through a sliding kernel [73, 74]. LeCun et al. [75, 76] introduced ConvNet, the first multi-layered CNN, in 1989 whose concept was derived from the earlier work of Fukushima's Neocognitron [77]. CNNs provide outstanding learning capabilities that enable the comprehension of visual information and deliver remarkable potential in classification and detection [78]. It leverages the spatiotemporal associations present in the data. The CNN is structured with a group of layers arranged in succession. The fundamental layers comprise convolution, pooling, and fully connected (FC) layers. A specific ordering of the convolution and pooling layers is employed to derive pertinent features from the source image, followed by multiple FC layers that leverage these features for classification or prediction [79]. Convolution operations are carried out using sliding kernels (filters) on the source image leading to a feature map. In the course of

training, the weights of the kernel in each layer are modulated utilizing a backpropagation algorithm to optimize the error between predicted and true labels. Different filter characteristics (type and size), padding strategy, stride (magnitude of displacement of the filter), and convolutional direction result in various convolution operations [80] of distinct types. An activation function enables it to embed nonlinear characteristics within the feature space. Activation functions, such as sigmoid, tanh, exponential linear unit (ELU), rectified linear unit (ReLU), leaky ReLU, etc., contribute to distinguishing the semantic disparities present in the images [73, 81]. The pooling layer diminishes (downsamples) the spatial dimensions of the feature representations (maps) by choosing either the highest or average of the non-intersecting regions. A CNN framework is concluded with FC layers that combine the spatially dispersed features to accomplish classification and regression. FC layers reshape the spatial dimension of the feature map into a unidimensional FV. CNN models are reinforced through loss functions by utilizing optimization approaches (e.g., stochastic gradient descent [82] and Adam [83]) that exalt the discriminative abilities of the features. Aligning with different network architectures, researchers have suggested several loss functions over time that are either classification-based (such as softmax loss) or metric learning-based (contrastive loss, triplet loss, etc.). The softmax function transforms the one-dimensional feature vector provided by the FC layer into probability values corresponding to the candidate classes. To prevent overfitting and lessen the computational burden, additional layers such as batch normalization, dropout, etc., are incorporated into the network structure [84, 85]. Fig. 1.2 shows a typical simplified working model of a CNN. Several researchers came up with various studies to provide improved deep CNN architecture. Such studies have suggested controlling information across different layers and introducing diverse connection mechanisms such as connectivity across the channels of layers and skipping connections. [86]. The contemporary architectural organization leans towards adopting simplified and consistent designs, such as VGG. The inception structure introduced intra-layer branching that can abstract features at various spatial scales. [87, 88] presented ResNet which was brought in the notion of skip connections to train deep networks. Inception-ResNet, Wide ResNet, Xception, Pyramidal Net, PolyNet, etc. were the successors of these designs, which revealed the refinement of multi-tiered revisions on CNNs by varying the cardinality and width [89, 90]. Some other well-known CNN architectures are MobileNet, DenseNet [91,92], etc. A discussion on the underlying architectural principles along with several design parameters was detailed in a paper by HasanPour et al. [93].

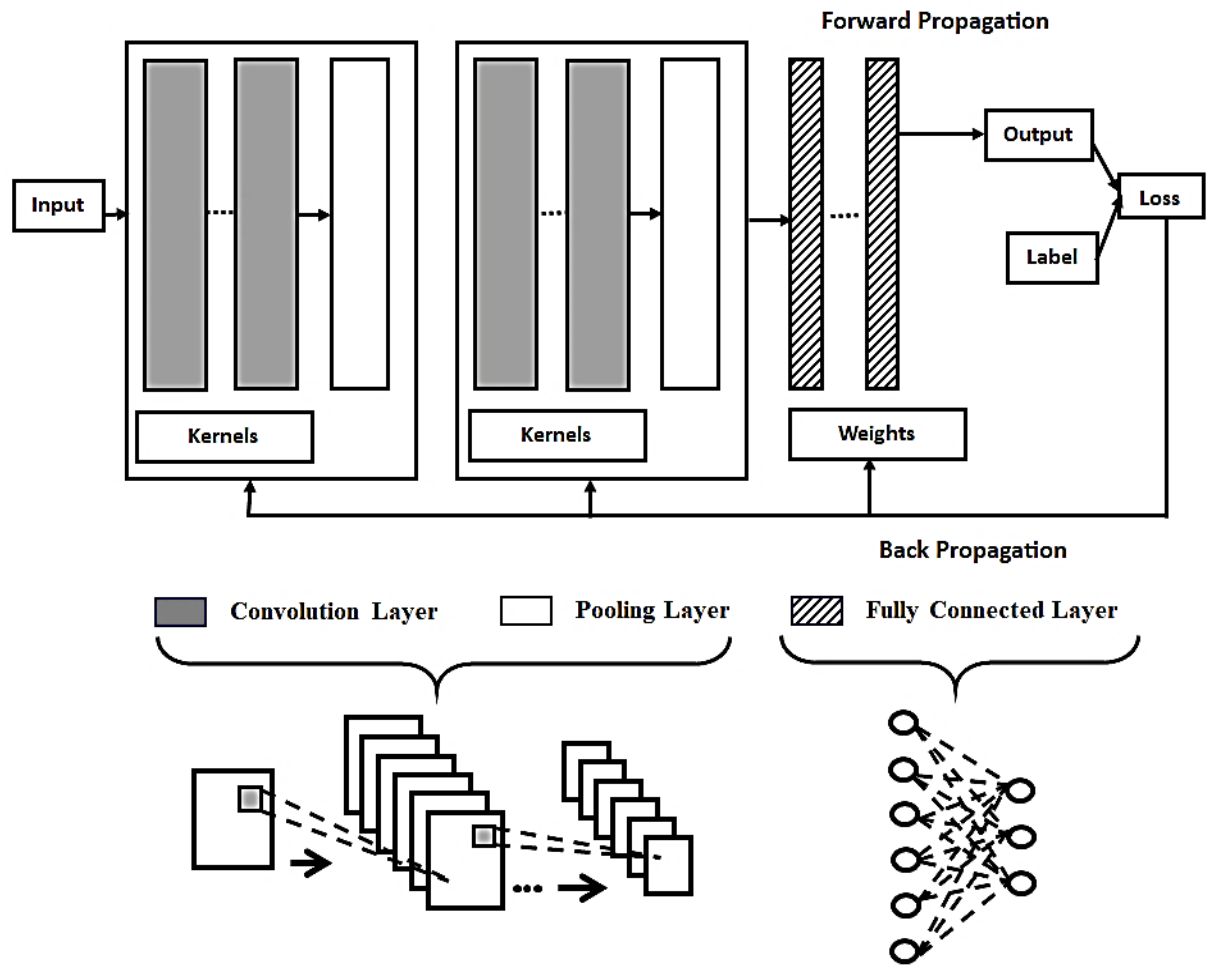


Fig. 1.2. A simplified working model of a CNN

1.6.3 FACE RECOGNITION USING DEEP CONVOLUTION NEURAL NETWORK

Classical FR methods typically rely on handcrafted or learning-oriented local feature representations. These (shallow) feature descriptors impede the recognition process in the presence of complicated nonuniform facial expressions [60]. Deep CNN derives hierarchical features, that acquire data at varied scales in various layers. It finally yields resilient and distinctive features for classification. FR using DL involves two main functional stages: deep FD and deep FR. Deep FD encompasses the localization of circumscribed faces in an image or video frame and the alignment of faces onto standardized reference points by adjusting the scales and orientation. Deep FR often involves preprocessing, deep FE, and face matching.

Augmentation, normalization, or both are employed to preprocess the faces. Augmentation methods include spatial or photometric transformations such as oversampling, rotation, and mirroring [94]. In [95], authors presented a comprehensive exposition of FR using deep networks including domain knowledge, FD challenges in a constrained environment, model dependencies, cost, etc.

1.6.3.1 DEEP FACE DETECTION

Detecting objects in an image can be perceived as a classification across multiple regions [96]. FD frameworks utilizing CNN can be distinguished as either two-stage or one-stage.

A two-stage detection framework is initiated by yielding a limited set of proposals through selective searching or employing a network of proposals (RPN). Each proposal is then endured by trimming and resizing before the FE. In RPN, predetermined anchor boxes with varied shapes and sizes at different loci are employed to engender the region of interest (ROI). It is subsequently inputted into the classifier. Finally, the delineation of the facial regions with bounding boxes occurs through regression. In single-stage detection, each location of the image is deemed as a candidate object and strives to distinguish each ROI as the intended object (face) or background.

The trailblazing two-stage detection model Region-based CNN (R-CNN) [97] was proposed in 2014. To overcome the limitations of extensive training and testing, fast R-CNN [98] was propounded, which involved a multi-task learning strategy. The detection of faces using a faster R-CNN model has been advocated by Jiang et al. [99], which was further revised by Sun et al. [100] and brought forward an improved Faster-RCNN by integrating feature concatenation, tuning of key parameters, Training in multiple scales, hard negative mining, etc. A more flexible model, Mask R-CNN, was promulgated by He et al. [101], that renders the simultaneous bounding box prediction and mask segmentation according to the proposals. Li et al.[102] suggested a framework utilizing a top-down design with lateral links to construct a semantic map with different scales, known as a feature pyramid network. Several studies have been conducted to refine the model [103].

Sermanet et al. [104] introduced OverFeat as an initial contributor to single-stage detection. Another successful detection model was You Only Look Once (YOLO), which has the drawbacks of powerlessness to detect a limited number of faces and the inability to identify small faces. To resolve the constraints of YOLO, a single-shot multi-box detector (SSD) was

propounded by Liu et al. [105]. In SSD, the output space is divided into a group of preset default boxes with varying dimensional proportions for each feature map. During the predictions, it assigns scores to each class of objects present in each box and modifies the box accordingly to embody the contour of the object. In dealing with objects of different extents, the network integrates predictions, derived from various feature maps with varying resolutions. Hu et al. [106] took advantage of SSD and exploited it for FD. A single-stage FD model, RetinaFace, was propounded by Deng et al. [107] precise face localization at the pixel level across faces of different sizes utilizing multi-task self as well as joint extra-supervised learning. Tang et al. [108] came up with a single-shot approach for FD, PyramidBox, to identify tiny, blurry, and partially obscured faces in unconstrained environments. Its extension, PyramidBox++, was proposed by Li et al. [109].

Both detection models employed convolutional and pooling layers to acquire deep feature descriptions. The max-pooling operation transformed the features within the region proposal into feature vectors with definite dimensions. The last FC layer was reinitialized to wrap up the detection. The detection model was subsequently optimized (fine-tuned) using a pre-trained model. In contrast, single-stage detectors use a single feed-forward NN to estimate the probabilities associated with classes and determine the spatial location of the bounding boxes by utilizing the computation of the loss function. Apart from the aforementioned DL-based models, several noteworthy face or object detection models like cascade R-CNN, libra RCNN., etc. have been presented in the literature [110, 111]. A notable FD framework, the multi-task cascaded CNN (MTCNN) [112], was developed using three-stage CNNs, in which facial landmarks are innately correlated with the bounding box of the face.

The literature contains numerous studies that encompass FD using CNNs. Li et al. [113] implemented an FD approach involving six CNNs in a cascading manner. Qin et al. [114] suggested a modified scheme for FD addressing the pitfall of intensive training and applied optimal joint training of the CNN cascade.

1.6.3.2 DEEP FACE RECOGNITION

FR requires training on a labeled facial image dataset of various subjects. Once trained, the CNN is adept at recognizing a new facial image using the learned features by processing the image with the network. Models such as VGGNet with a depth of 16–19 weighted layers, the 22-layer GoogleNet with inception module, (up to) 152-layer ResNet with residual mapping,

etc., were proposed and extensively employed as the foundational models in FR, either unmodified or somewhat revised [115, 116, 117, 118]. In this context, some of the research efforts on FR harnessing DL such as DeepFace, FaceNet, VGGface, sparse ConvNet, UniformFace, etc. have been documented in [119]. Gwyn et al. [120] reviewed the FR performance by employing eight DL architectures viz. AlexNet, two variants of VGGNet (VGG 16 and VGG 19), two models of ResNet (ResNet50 and ResNet101), two adaptations of Inception (Inception V2 and Inception V3) and Xception. Serengil et al. [121] propounded LightFace for FR. Zhou et al. [122] adapted the ResNet model and brought forward a CNN model named ResNet-face18 for recognizing faces.

Researchers have employed a blend of existing architectures that outperform stand-alone deep CNNs [123]. Authors of [124] advocated a Pyramid-Based Scale-Invariant CNN to address the low-resolution and scale disparity of input images. In a recent study by Yang et al. [125], a DL framework, LeNets++, was employed for FR under various lighting conditions.

1.7 MULTI-BIOMETRIC FUSION

Researchers frequently enact fusion strategies to improve FR accuracy. There has been considerable discussion in the literature regarding various fusion levels, including sensor, feature, decision, score, and rank levels [3]. The selection of a fusion level depends on the application-specific requirements, available resources, and individual biometric modalities. Furthermore, the information to be fused (raw data from a sensor, feature data, score, or rank) as well as effective fusion methods are challenges for the implementation of biometric fusion [126]. These biometric fusions can be carried out using unimodal or multimodal data. Multimodal systems exploit biometric traits with dissimilar modalities for human identification. The multi-biometric fusion framework can be accomplished with unimodal biometric traits that involve (i) multiple sensors to acquire biometric data of the same trait, (ii) multiple samples to obtain different recordings of the unimodal trait employing one particular sensor, (iii) various algorithms to treat the same biometric data, or (iv) different instances of a biometric trait [127]. According to recent studies on multi-biometric fusion, fusion approaches fall into three categories: feature-based early fusion, decision-based late fusion, and hybrid fusion. The terms early or late indicate whether fusion is carried out before or after classification.

Early fusion involves the integration of low-level information such as raw image data or feature data. Sensor-level fusion integrates data of a biometric trait acquired by different sensors, or data of a biometric trait acquired by a single sensor by varying the sensor parameters and its surrounding environment. Common methods, utilized in this fusion approach, include mosaicking, weighted aggregation (summation), etc. Such methods often involve alignment and resizing. Some studies that employed sensor-level fusion are presented in Table 1.1. Sensor-level fusion is not always embraced as a widely acknowledged fusion scheme because of the impediment in fusing sensor data from relevant modalities [132].

Table 1.1: Summary of a few related studies on sensor-level fusion

Authors and Year	Fusion approach	Modalities
Yang et al., 2005 [128]	Mosaicking of panoramic faces	Face
Singh et al., 2007 [129]	Mosaicking	Face
Kisku et al., 2009 [130]	Wavelet-based decomposition and fusion of decompositions	Face and palm-print
Zhang et al., 2014 [131]	tensor analysis-based method	Face

The core principle of feature-level fusion (FLF) is to derive features from one or more biometric traits and fuse these features into an integrated FV, thereby accumulating more distinctive information [133]. The resultant FV is input into a classifier. Effective FLF methods include feature concatenation, transformation, weighting, and selection. This type of fusion has serious limitations.

- a) the dimension of the combined FV is substantial leading to inferior performance
- b) incompatibility between FVs from different biometrics causes serious challenges when the feature sets differ considerably in range, distribution, or both [2]. Researchers have employed techniques such as dimension reduction, feature normalization, and feature selection [134]. However, this could result in the loss of crucial feature information.
- c) For the assorted modalities, eliciting this correlation at the feature level is challenging.

Some recent studies on the FLF of biometric traits are summarized in Table 1.2.

Table 1.2: Summary of a few related studies on feature-level fusion

Authors and Year	Fusion strategy	Modalities
Jeng et. al, 2017 [135]	Shuffle coding	Face, iris, and palmprint
Haghighat et al., 2016 [136]	Discriminant correlation analysis	Face, ear, fingerprint, and iris
Ahmad et al., 2016 [137]	Discrete cosine Transform	Face and palmprint
Xing et al., 2015 [138]	Coupled projections-based method	Face and gait
Tiong et al., 2019 [139]	Feature fusion based on deep multimodal learning	Face
Talreja et al., 2021 [140]	Feature fusion using deep hashing	face and iris
Sarangi et al., 2022 [141]	Feature level fusion with kernel discriminative common vector	Face and profile face

Methods for late fusion can be implemented at score, rank, or decision level. Late (post-classification) fusion is preferred because (a) it makes use of the strengths of individual matchers (classifiers), compensating for its shortcomings, and (b) the implementation is uncomplicated from a mathematical perspective [3].

Decision-level fusion methods combine the final outputs (identified class level) of multiple classifiers to obtain a unified decision (single-class level for a test sample). The strategies that have been preferred in this type of fusion are majority voting, weighted majority voting, the Dempster-Shafer method, and Bayesian decision, etc [3]. This fusion level is appropriate for multi-sample and multi-instance biometric systems. Such integration at the decision level is pertinent when an individual matcher exploits distinct feature types; however, the precision of the classifier precision is a pivotal determinant of its robustness [142]. Table 1.3 presents some studies on decision-level fusion methods.

Table 1.3: Summary of a few related studies on decision-level fusion

Authors and Year	Fusion strategy	Modalities
Patil et al., 2016 [143]	AND rule, OR rule	Face and palm-print
Gao et al., 2017 [144].	Dempster-Shafer evidence theory	Face
Rahman et al., 2019 [134]	Decision tree	Face and iris
Devi, 2020 [145]	Weighted majority voting	Palm-print and face

Score-level fusion has been studied extensively. This fusion level identifies a biometric trait by integrating scores from different matchers. Matching or Similarity scores from multiple classifiers with a single biometric or from a single classifier utilizing different feature data of

the single or multiple biometrics are consolidated into a score based on which the recognition decision is finalized. Hence, this approach can be implemented for multi-algorithm, multimodal, and multi-sample biometric systems. Accordingly, the fusion methods based on scores can be derived using either a classifier-based scheme (e.g. using SVM) or transformation-based rules (such as max, min, product, sum, exponential sum, weighted sum, etc.) or density-based measures (e.g. likelihood ratio, etc.) [3]. Some studies on score-level fusion techniques are highlighted in Table 1.4.

Table 1.4: Summary of a few related works on score-level fusion

Authors and Year	Fusion approach	Modalities
Matin et al., 2017 [146]	Weighted score	Face and iris
George et al., 2016 [147]	Weighted score	Eye movement
Sharifi et al., 2016 [148]	Weighted sum rule	Face and iris
Liang et al., 2016 [149]	Order preserving tree	Fingerprint and face
Zhang et al., 2020 [150]	Weighted sum rule-based	Face and voice
Abderrahmane et al., 2020 [151]	Named weighted quasi-arithmetic mean	Fingerprint and face

Owing to the difference in scale or nonlinear relationship of the matching scores, normalization of the outcome of the biometric classification was applied. The min, max, decimal sealing, Z-score, etc. are widely employed score normalization techniques [152]. However, normalization incurs a computational burden, and the selection of a suitable appropriate normalization has an impact on the overall performance.

In rank-level fusion (RLF), the classifier produces an order of matching entities by evaluating their resemblance with the probe. Individual classifier's ranks are unified into the final rank [153]. RLF is regarded as a particular form of decision-level fusion [133]. The RLF in multi-biometrics has been weighed above all other fusion methods because of (i) the reduced impact of sensor-specific variations, (ii) compatibility with a wide range of biometric modalities including face, fingerprint, iris, etc., without any modification or standardization of the individual biometric system, and (iii) capability of handling the relationships between nonlinear or complex scores from different biometric systems by transforming scores into ranks [142, 154]. Despite the adeptness of combining ranks, fewer studies have been conducted on fusion at the rank level [142]. Widely used methods for consolidating ranks are

logistic regression, Borda count, highest rank, and more [157]. Table 1.5 lists some studies on the RLF of biometric traits

Table 1.5: Summary of studies on rank-level fusion

Authors and Year	Fusion approach	Modalities
Talebi and Gavrilova ,2015 [155]	confidence-based RLF	Frontal face, profile face, and ear
Sharma, 2015 [153]	serial and parallel RLF	Face and iris
W. Rahman et al. , 2017 [156]	Borda count and logistic regression	Face and gait
Sing et al., 2019 [157]	Confidence factor weighted RLF	Face

In a previous study, Khan et al. [132] appraised four fusion levels (sensor, feature, score, and decision) using face biometrics.

To maximize its effectiveness, hybrid fusion blends early and late fusion methods into an integrated framework that harnesses the strengths of both methods. Abozaid et al. [158] propounded a bi-modal (face and voice) biometric scheme that exploited both feature- and score-level fusion. Azom et al. [159] suggested a hybrid fusion method with multi-modality (iris and face) that implemented the fusion at the feature level involving five FE techniques and score-level fusion involving two classifiers.

Researchers have also exploited the deep CNN framework to implement the fusion of multi-biometric data. Zangeneh et al. [160] employed nonlinear transformations by leveraging a dual-CNN approach to project facial images with lower and higher pixel density into the image space. A similar image-mapping method was brought forward by Lu et al. [161], which used the ResNet architecture along with coupled mappings. Zhu et al. [162] put forward a fusion strategy where global and local features, derived by 2DPCA and LBP methods respectively, were fused utilizing DL. In another study by AlFawwaz et al. [163], both local and global facial features were reduced by LBP and PCA and fused by applying the multi-resolution DCT method. FLF involving HOG and a local difference binary (LDB) was presented by Wang et al. [164]. Tiong et al. propounded a fusion approach for FR using layers of colour (RGB) information and texture characteristics from the face and periocular in a multimodal framework [165]. Zhang et al. described a fusion methodology that took advantage of supervised feature learning in a hierarchical manner and integrated deep and shallow features by employing VGG and lightweight CNN [166]. To mitigate the challenges of FR in the presence of noise, varied lighting, and different contrast, a scheme was reported

by Juneja et al. [167], in which two different filters (Gaussian and Gabor) were applied to the face images. Then, LBP was used for feature representation, and the combined feature was classified using SVM. A colour 2DPCA-based decision-level fusion employing CNN was advocated by Li et al. [168].

In a previous study, Koo et al. [169] employed ResNet-50 and VGGFace-16 for recognition of the face and body in isolation. They applied the weighted sum rule for score-level fusion to augment the recognition success rate. Cheng et al. [170] reported a fusion strategy that combined similarity scores from different CNN models by applying a heuristic voting scheme. Alay et al. implemented a two-level (score as well as feature-level) fusion involving three cues (iris, face, and finger vein), and the fusion at the score level achieved higher precision relative to that at the feature level [171]. In a study by Soleymani et al., traits from the iris, face, and fingerprint, acquired using modality-specific CNNs, were fused, optimized, and classified for person identification [172].

Multimodality has several advantages that unimodality lacks. Adopting multimodality yields higher recognition precision, more impervious to environmental variability, and more protection against counterfeiting [173]. However, there are certain downsides to biometric systems using multimodality, which include the associated operational expenses and extended registration process [174]. Multiple acquisition steps cause vexation for users. In this context, multi-biometric fusion using unimodal data is favorable. Concerning FR, such fusion can be achieved at all levels by

- capturing faces using optical and infrared cameras (multi-sensor)
- by utilizing facial images of the individual taken with varying environmental conditions (multi-sample or multi-instance)
- acquiring feature data from the same face by employing distinct FE algorithms (multi-algorithm)

The RLF approach is also applicable for unimodal biometrics, while (a) consolidating outputs of the multiple classifiers and (b) integrating outputs of a classifier with varying training sets or with distinct parameters, such as kernels in SVM or K in KNN. This is even germane to/well suited for fusing the output of NNs with varying layer transfer functions [175].

1.8 MOTIVATION OF THE THESIS WORK

Biometric authentication is an essential concern in this digital era. Despite providing several advantages over unimodal recognition systems, multimodal biometric recognition poses certain impediments, such as cost, storage requirements, complicated integration due to lack of standards, and interoperability issues. User acceptance is another barrier due to multiple acquisitions and inconvenience. In contrast, the face, as an unimodal trait, is a boon for biometric identification. It is non-invasive, cost-effective, and well-suited for multi-algorithmic or multi-sample fusion.

In this security-conscious world, FR has made a significant stride in human identification, which is germane to personal device access control, law enforcement, border control, crowd surveillance, attendance tracking, exam monitoring, etc.

Although there has been substantial research on FR, it is still plagued by intrinsic factors (such as countenance, partial occlusion, and aging) and/or extrinsic factors (such as lighting condition, scale, noise, and blurring). These factors contribute to both between-class resemblance and within-class difference. Noise and redundancy in biometric data are other aspects. These determinants induce uncertainty in the decision-making problem domain such as FR. Fuzzy set theory has proven its potential for characterizing uncertainty. The impetus for this thesis work is grounded in improving the accuracy of FR by addressing uncertainty with fuzzy set theory and implementing the fusion strategy.

1.9 OBJECTIVE OF THE THESIS WORK

As discussed in Section 1.2, varied environmental and imaging circumstances, disparate facial countenance, and other factors influence the precision of FR. These parameters result in intra-class diversity and inter-class resemblance. Noise and redundancy in raw biometric data give rise to uncertainty in decision-making. From these studies, it is apparent that fusion contributes significantly to improving the FR success rate. From the discussion on biometric fusion in Section 1.7, it is important to note that there have been fewer studies that employ RLF for human identification compared to other levels of fusion strategies, even though it

can circumvent the constraints inherent to feature and score-level fusion. Considering its merits, it is prudent to implement an RLF for FR to attain a higher precision. Another crucial point to bring up is that the conventional ranking mechanism does not account for the spatial correlation among the outputs of the classifier. Consequently, the resulting rank obtained through fusion may not represent the class appropriately. The different extensions of fuzzy sets and systems have established a coherent way of taking the edge off uncertainty and imprecision across different levels. Instead of orthodox RLF methods, involving a fuzzy ranking mechanism in RLF could prove to be a judicious recourse.

Researchers have conducted assorted studies aimed at designing DL architectures for enhanced FD and FR. Their efforts are oriented toward the exploration of spatial and depth enhancement, managing information across various layers, optimizing cross-layer connectivity, etc. Few studies have explored adopting fusion strategies with CNN-based models. Most efforts have been made to accomplish feature-based fusion. Hence, there is latitude to validate the effect of the fuzzy RLF involving CNN models.

Taking the above points into consideration, the thesis encompasses the subsequent aims.

- (1) To study the effect of employing two variants of type-2 fuzzy sets (general and interval) in deriving fuzzy ranks which are eventually applied in RLF utilizing multiple FVs.
- (2) To explore the effectiveness of a methodology for FR which implements the fusion method with multiple FVs by employing fuzzy ranks, derived by putting interval type-3 fuzzy sets to use.
- (3) To assess the extent of alleviating uncertainty of type-1, interval type-2 fuzzy set (IT2 FS), and interval type-3 fuzzy set (IT3 FS) in deriving fuzzy ranks when implementing fuzzy RLF methods for FR with a similar trial.
- (4) To create a face database with the detected faces from an outdoor video clip and to appraise the FR performance by leveraging fuzzy RLF involving multiple deep CNN models.

1.10 ORGANIZATION OF THE THESIS

This section provides a concise summary detailing the subsequent chapters of the thesis.

Chapter 2: General type-2 fuzzy set induced rank-level fusion for face recognition

This chapter presents a new decision-making model for multi-feature RLF by integrating a type-2 fuzzy set-based approach to mitigate the factors that contribute to FR accuracy. A new scheme is proposed to generate a general type-2 fuzzy set-based rank for each of the top-ordered classes, which are selected based on weights or confidence factors yielded by the classifier. These type-2 fuzzy sets were shaped for each of the top-ordered classes from the feature vectors of the test image and those of the training samples of the respective class. The complement of type-reduced defuzzified value was then conceived as the fuzzy rank of a particular class. Considering the three different feature vectors corresponding to the three FE techniques and the respective outputs of the classifier, the cumulative sum of the fuzzy ranks is fused weighted by the complemented mean of the confidence factors for each of the regarded classes to generate the fused scores. Based on these scores, a class of test images was identified. Well-known databases are used to evaluate the performance of the proposed model. The experimental results demonstrated an improved performance compared to other state-of-the-art methods, despite the varied occlusion, pose, and illumination present within the images of the databases.

Chapter 3: Interval type-2 fuzzy set induced rank-level fusion for face recognition

The third chapter elucidates a framework by introducing interval type-2 fuzzy set-induced fuzzy RLF for FR utilizing multi-feature vectors. It utilizes the outputs of the classifier as confidence factors. We address the wide intra-class variability issue by introducing an interval type-2 fuzzy set that utilizes intra-class face images to generate lower and upper MFs defining the footprint of uncertainty (FOU) of IT2 FS. The defuzzified value was used to generate the fuzzy rank for a class under consideration. It also mitigates the influence of inter-class similarity by excluding inter-class facial images in the computation. Furthermore, multi-feature vectors, derived by different FE techniques for a face image, are used to utilize the different underlying discriminant features, which in turn address the issues of inter-class similarity. For each feature vector, we generated IT2 FS-based fuzzy ranks corresponding to the top K classes that are selected based on CFs. To reduce the complexity, top K classes are taken into consideration for RLF. Likewise, interval type-2 fuzzy set-based multiple fuzzy

ranks were obtained for a class using multiple feature vectors. Finally, the aggregated fuzzy rank is fused with the respective complemented mean of the confidence factors to obtain the final fused rank for each of all top-ordered classes. While aggregating fuzzy ranks, a penalty is set for the class if it is not present in the top-ordered classes corresponding to an FV. A face image is identified based on these final fused ranks. The method was evaluated on several face databases and was found to be superior to many state-of-the-art methods.

Chapter 4: Interval type-3 fuzzy set induced rank-level fusion for face recognition

Very recently, IT3 FS theory has been introduced as a more effective way of addressing uncertainty in decision-making than its fuzzy set counterparts. This chapter describes the design of an approach for FR that utilizes IT3 FSs to yield fuzzy ranks for RLF using multiple feature vectors. The use of multiple feature vectors corresponding to different FE techniques resolves inter-class resemblances, whereas incorporating an interval type-3 fuzzy membership function to generate fuzzy rank reduces the uncertainty due to intra-class divergences. With respect to a feature vector, the classifier generates confidence factors for all classes. The algorithm considers the top-ordered classes, selected according to the confidence factors, to utilize IT3 FS. For each class, the corresponding training samples were employed to constitute the secondary membership domain. Furthermore, with respect to a training sample, a tertiary domain is generated by utilizing its neighboring samples from all classes based on a proximity measure using the Euclidian distance. These distances were employed to express the lower and upper membership functions of the scaled Gaussian domain of uncertainty of the IT3 FS. This mitigates the effects of between-class resemblance. Finally, successive direct defuzzifications result in fuzzy ranks for the top-ordered K classes. We combine these fuzzy ranks judiciously with the respective confidence factors, generated by multiple feature vectors, to obtain the final rank for that class. The class with the lowest rank is identified as the test-image class. Three publicly accessible face databases were used to evaluate the method, and the outcomes of the experiments demonstrated a higher level of effectiveness compared to related methods.

Chapter 5: Face detection and recognition from video images using deep CNN-based rank-level fusion

Chapter 5 explores a methodology wherein a face database is created from an outdoor video clip, and an RLF mechanism is implemented for FR that makes use of the probabilities produced by the softmax functions of multiple deep CNN models. These probabilities are

referred to as certainty indices for potential classes to which a test sample may belong. For each CNN model, fuzzy ranks were obtained for a subset of classes with the top certainty indices by applying a Gaussian distribution function followed by a complementation operation. For each class appearing in one or more subsets corresponding to deep network models, the average of the certainty indices is complemented and fused with the sum of the complemented fuzzy ranks. This results in the final ranking of the respective class. While calculating the summation of the fuzzy ranks, a penalty is put into effect on the class for not being present in any set of classes with respect to the CNN models. This avoids the chance of becoming an unexpected winner. The class with the lowest final ranking was identified as the class of the test sample. In addition to testing with the custom-made outdoor video-based face dataset, the projected method was evaluated with a standard surveillance indoor dataset, and a standard surveillance face database, ChokePoint. For both datasets, the precision and macro-average scores of the proposed method in all cases were higher than those of individual deep CNN models.

Chapter 6: Concluding remarks and future direction of work

In the concluding chapter, the works conducted are summarized, the key contributions are emphasized, and the future directions for future research are discussed.

1.11 DATABASES USED IN THE STUDIES

1.11.1 AT & T FACE DATABASE

This database [176] comprises face images of 40 persons, including both males and females. There were 400 grayscale images, with 10 images per individual. Individuals are with or without mustaches and spectacles as well as with varying hairstyles. These images were captured in a uniform black backdrop with faces in a vertical frontal pose and varied levels of lighting conditions. The images had a resolution of 112×92 pixels, with a tilt and rotation of up to 20 °. The faces exhibited different expressions, including eye-opening, smiling, and frowning. Examples of the facial images are shown in Fig. 1.3.

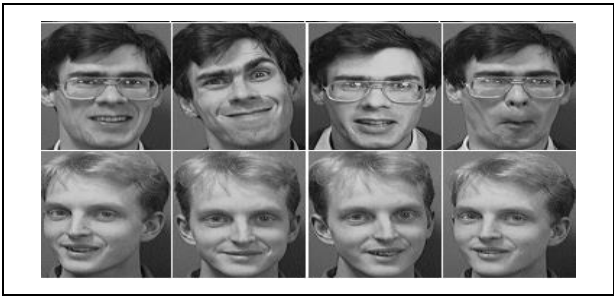


Fig. 1.3. Representative images from AT & T face database

1.11.2 FERET FACE DATABASE

A total of 14126 facial images were included in the FERET database [177, 178] representing 1199 individuals. There were 1564 sets of faces, of which 365 sets were duplicates but were captured on a separate day. The dimensions of each picture were 512×768 in a semi-controlled environment, and 15 different sessions were conducted to capture these images. Fig. 1.4 presents selected images from the FERET FR benchmark database.

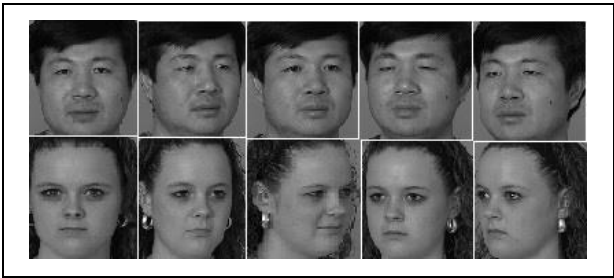


Fig. 1.4. Representative images from FERET face database

1.11.3 UMIST FACE DATABASE

This database [179] consists of 575 grayscale images, featuring 20 distinct males and females with varying numbers of images (19–48) per individual. These images covered a range of races, sexes, and appearances. Facial images were captured from varying angles for both the left and right profiles. Fig. 1.5 displays selected images from the UMIST database.



Fig. 1.5. Representative images from UMIST face database

1.11.4 FACES94 FACE DATABASE

The database comprises 3060 face pictures from 153 people. Of the 153 individuals, 20 were female, 113 were male, and 20 were staff members. There were 20 images per individual, each with a dimension of 180×200 pixels. Each image has a resolution of 180×200 pixels. These images were taken from a position having approximately equal distance from the camera with a homogeneous background. The image-capturing setup used artificial lighting with the melding of tungsten and fluorescent lighting. These images show minor variations in head movement and face position. However, there is considerable variation in expression as they are asked to speak while taking pictures with the camera [180]. The images in Fig. 1.6 are samples sourced from the Faces94 database.



Fig. 1.6. Representative images from the Faces94 facial dataset

1.11.5 CHOKEPOINT FACE DATABASE

The ChokePoint database was curated from surveillance videos. With 48 video sequences, this database contains 64204 facial data points, captured by three cameras above two portals (P1 and P2) with natural chokepoints. Each portal captured the individuals who were coming (E) or going (L) to the portals. With respect to P1 and P2, different sets were compiled with

25 and 29 subjects (individuals), respectively. The collection of images was organized according to the recording context and labelled with a name, for example, P1E_S2_C2, embracing camera (C), sequence (S), and portal (P) numbers. The faces were with varying brightness and different head angles. Two groups, G1 and G2, from case_study_1 were created by merging cropped face images, as detailed and downloaded from [181]. The resolution for all images is set at 800×600 pixels. Fig. 1.7 showcases example images from the ChokePoint database.

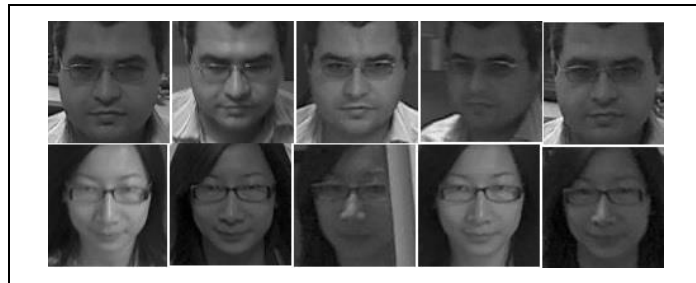


Fig. 1.7. Representative images from the Chokepoint facial dataset

CHAPTER 2

GENERAL TYPE-2 FUZZY SET INDUCED RANK-LEVEL FUSION FOR FACE RECOGNITION

2.1 INTRODUCTION

In image processing systems, when an image is captured by any image acquisition hardware, the quality of the image is disturbed by diverse factors, such as lighting, environment, and distance. As a consequence, noise and redundancy are inevitably injected into the raw image data. In the case of a face image, the recognition accuracy is highly defied by the variation in pose, illumination, and occlusion. This presence of noise and redundancy in raw biometric data could affect the results for precise segmentation, filtering, edge detection, or classification [182]. The aforementioned factors influence intra-class divergence and inter-class resemblance, leading to poor recognition accuracy, thus making FR approaches subpar.

Traditional classical approaches to FR can be grouped into three categories: local, global (holistic), and hybrid [51]. Local methods distinguish the features of specific regions or key points (eyes, nose, and mouth) of the face by utilizing the spatial relationships among these facial regions or landmarks. Several local methods like LBP, SIFT, HOG, SURF, etc. are prevalent in this category [51]. Global techniques project a full-face surface into a low-

dimensional compact feature space (subspace). Linear or non-linear methods entail training images to depict such subspaces. PCA, LDA, ICA, Kernel PCA, Kernel LDA, and others are examples of such local techniques [6]. Chowdhury *et al.* suggested an FR strategy where a G-2DFLD method was used for FE and SVM for classification [62]. Researchers involved hybrid approaches involving global and local features to achieve higher accuracy in FR [54, 55, 57].

Modern FR approaches employ learning-based approaches such as deep learning, dictionary learning, and fuzzy logic theory. [52]. If the size of the training set is small or not suitable enough for training, then the deep learning models would not be an effective solution for FR. High computational resource requirements and overfitting are other factors that limit its implementation [183]. Using fuzzy set theories in FR has showcased improved performance in recent research studies. A fuzzy-based 2DPCA method was presented by Li *et al.* [61] to address the sensitivity of the traditional 2DPCA to illumination, pose, and expression in FR. For the same stride, Dey *et al.* [34] discussed a methodology for FR that utilized the FG-2DFLD method and RBFNN. Another fuzzy model based on linear regression discriminant projection was suggested by Huang *et al.* [184].

In FR, the multi-biometric framework with an information fusion strategy allows for achieving superior FR performance by alleviating the factors stated above. Since the research works, detailed in the subsequent chapters of this thesis, including the current one, are restricted to face as the only modality, multi-algorithm-based multi-instance cues (faces) have been employed to form the multi-biometric framework. From the literature review, it is observed that there are few research efforts in which FVs, derived by multiple FE techniques are utilized in multi-biometric cues, and discriminatory information from these FVs is exploited in information fusion. RLF provides an optimal solution for multi-biometric identification because feature incompatibility and normalization issues are not as impediments as score-level fusion and are not as rigid and sensitive to the classifier's potential as decision-level fusion [133]. An RLF strategy, CFGaussFRLF, was proposed by Sing *et al.*, wherein Gaussian fuzzy ranks, deduced from the classifier's outputs (confidence factors), were multiplied with the respective confidence factors of the classifier to obtain the final fused results [157]. These final fused results were utilized to identify the class of the test sample. The reason behind the use of fuzzy ranking instead of traditional ranking in the fusion process is that fuzzy ranking accounts for the distribution of the classifier's output.

Recently, type-2 fuzzy sets and their logic systems have been widely applied in various research areas and have achieved phenomenal success in handling higher degrees of uncertainty. Fuzzy sets and their logic systems have been successfully applied in image processing [185]. In addition to the application of type-2 fuzzy logic to image classification and quality assessment, various approaches have been expanded to different levels of fusion. However, these methods are seldom used in image fusion.

The rest of the chapter is organized as follows. Section 2.2 introduces related concepts of rank-level fusion and type-2 fuzzy sets. Section 2.3 discusses our methodology. The experimental results and argumentation on the results are explicated in Section 2.4. The experimental findings after the analysis are summarized in the form of a conclusion in Section 2.5.

2.2 RELATED CONCEPTS

2.2.1 FUSION IN FACE RECOGNITION REVISITED

An illustrative discussion on fusion strategies in FR has been debated in Section 1.7 of Chapter 1. For precise FR using the feature-based approach, three aspects are of great impact: (a) the features that characterize the trait and (b) the classifier employed to recognize the class or subject [186]. Researchers have proposed several techniques, including FE, classifier design, and their combination. Studies have revealed that the effectiveness of these methods is influenced by the characteristics of the features and potential of the classifier employed. Hence, exploring approaches for utilizing various feature vectors and classifiers is imperative for attaining higher accuracy. This has brought about the evolution of diverse fusion methods in FR. The levels of fusion prior to classification are fusion at the sensor level (where raw data are melded from different sensors) and fusion at the feature level (in which the features of various biometric traits are combined). A post-classification fusion scheme can be achieved either at the score level (where the match scores from multiple modalities are aggregated to engender new match scores to render a judgement), at the decision level (by consolidating the decision created by a set of classifiers), or at the rank level (in which case, consolidate the rank gain by individual matcher to derive a consensus rank for each identity).

Another crucial note in this context is that early fusion (prior classification) is always more challenging and discriminative than late fusion (post-classification). Kisku et al. [130] proposed a sensor-level combination-based multi-modular individual personality confirmation framework utilizing face and palmprint. FLF is effective when the aggregated features are independent of each other and involve the same type of measurement scale [187]. The impediment to this process is the presence of noisy or redundant data in the raw biometric data, incompatible representation, and the large dimensionality of the FV after fusion. This is why fusion might turn out to be inefficient and less sturdy [186]. Among the post-classification schemes, score-level fusion has been extensively studied. No prior information regarding the features or classifiers is needed for this fusion level [149]. As cited in [188], this type of scheme achieves better performance owing to its conformability and score-combining ability. Heterogeneous scores owing to dissimilar scales, ranges, or probability distributions have been observed to impede this class of fusion. The acceptable performance is contingent on the computationally intensive normalization process as well as the selection of competent techniques. In decision-level fusion, the results of multiple classifiers for each biometric type are aggregated into a final classification by lending decision-fusion methods [127]. The performance of this level of fusion may be limited when adequate information is unavailable. In rank-level fusion, different ranks, rather than match scores, from the distinctive biometric matcher are combined into a single rank, which in turn is used in the decision-making process. Sing et al. [157] addressed various rank-level fusion methods that include the highest rank [189], modified highest rank [189], Borda count [189], logistic regression [190], and non-linear weighted rank fusion [191]. These methods are briefly explained in Section 2.2.2. Rank-level fusion is significantly affected by the quality of the image and is sensitive to inter- and intra-class variations [127]. These variations, which are caused by poor image quality, environmental conditions, and the wide disparity in pose, race, illumination, etc., contribute to the imprecision and uncertainty that degrade FR accuracy. In this context, approaches based on fuzzy set theory provide a practical solution for modeling such uncertainties. The next section illustrates the fundamental concepts of the type-2 fuzzy set.

2.2.2 RANK-LEVEL FUSION METHODS

Similarity scores generated by the classifier for a test sample are sorted from highest to lowest to create the ranking list of the subjects (classes) presented for matching. The highest score is assigned to the lowest rank. A class or subject with a lower score signifies a more suitable match. The ranks from multiple classifiers corresponding to a subject or class are combined to engender a consensus rank. The class for a test sample is determined by the consensus ranks of the subjects (classes).

The following sub-sections delve into some well-established rank-level methods, employed in multi-biometrics. We made use of these methods in the performance analysis of our method, described in this chapter. To explain the different methods for fusing ranks, suppose the database consists of S subjects (or classes). In addition, it is assumed that C matchers (classifiers) were engaged in the recognition process. As determined by the outputs of the c^{th} classifier, let $\mathfrak{R}_{c,s}$ be the rank ascribed to subject s , $c = 1, 2, 3, \dots, C$ and $s = 1, 2, 3, \dots, S$. The rank-level method yields a consensus rank χ_s of for subject s .

2.2.2.1 HIGHEST-RANK METHOD

Using the highest-rank approach, the final fused rank for the subject, s , is determined by Eq. (2.1).

$$\chi_s = \min_c^C \mathfrak{R}_{c,s} \quad (2.1)$$

Multiple classes can always have the same final ranking, which is a limitation of this approach. In this case, the modified highest rank is applied.

2.2.2.2 MODIFIED HIGHEST-RANK METHOD

In this method, Eq. (2.1) is modified as follows to resolve the limitations of the highest-rank approach:

$$\chi_s = \min_c^C \mathfrak{R}_{c,s} + \varepsilon_s \quad (2.2)$$

where the perturbation factor

$$\varepsilon_s = \frac{\sum_c^C \mathfrak{R}_{c,s}}{K}$$

The ε_s is the bias for the consensus rank and takes all the ranks corresponding to subject s into account, and K must be set to a suitably large value to have a very small value.

2.2.2.3 BORDA COUNT METHOD

In this method, the final rank of subject s is obtained by aggregating the individual ranks of all C classifiers. It is articulated in Eq. (2.3)

$$\chi_s = \sum_c^C \mathfrak{R}_{c,s} \quad (2.3)$$

The presumption behind the Borda count method is that there is no correlation between the classifiers from a statistical perspective, and all of these classifiers perform consistently. Practically, the effectiveness of this method relies on the performance of the classifiers.

2.2.2.4 WEIGHTED BORDA COUNT OR LOGISTIC REGRESSION METHOD

To take the efficiency of the classifiers into consideration, a weight, ω_c , is attributed to each classifier, c , reflecting its performance. This method is derived from the Borda count method. The individual ranks for subject s are multiplied by the respective weights of the classifiers and their cumulative sum leads to the fused rank for s . The mathematical expression is given by

$$\chi_s = \sum_c^C \omega_c \times \mathfrak{R}_{c,s} \quad (2.4)$$

Recognition accuracy is highly influenced by the appropriate weight allocation scheme because the classification potentials differ depending on the sample set used.

2.2.2.5 NON-LINEAR WEIGHTED RANK METHOD

For more accurate results, Kumar et al. advocated a nonlinear weighted rank, in which nonlinearity is imposed by applying an exponential rank as follows:

$$\chi_s = \sum_c^C \omega_c \times \exp(\Re_{c,s}) \quad (2.5)$$

This is also called exponential rank-level fusion. This method exploits the varying steepness of nonlinear functions such as \exp and \tanh , allowing certain ranges with slopes exceeding 1 to create a magnifying impact.

2.2.3 AN OVERVIEW OF TYPE-2 FUZZY SETS

Professor Lotfi A. Zadeh is recognized as the proponent of fuzzy set theory for expressing knowledge in a mathematical context [192]. A type-1 fuzzy set (T1 FS) or simply a fuzzy set, S , is formally defined with a membership function (MF), $\varphi_S(x)$, on a fuzzy variable x within the universe X as follows.

$$S = \{(x, \varphi_S(x)) | x \in X, 0 \leq \varphi_S(x) \leq 1\} \quad (2.6)$$

Unlike a crisp set, whose membership function (MF) takes either 0 or 1, the MF of a fuzzy set can range between 0 and 1. The T1 FS was found to be incapable of rationalizing uncertainties in certain practical applications [193]. To address these uncertainties more effectively, the type-2 fuzzy set (T2 FS) was advocated by Zadeh in 1975 as a further development of T1 FS [192]. A general type-2 fuzzy set (GT2 FS) or simply type-2 fuzzy set (T2 FS) is characterized by an MF, $\varphi_{\tilde{S}}(x, a)$, which is intrinsically fuzzy. Such a GT2 FS, \tilde{S} , is formally expressed in terms of primary variable x and secondary variable a from the universes X and A , respectively as follows.

$$\tilde{S} = \{((x, a), \varphi_{\tilde{S}}(x, a)) | x \in X, a \in \beta_x \subseteq [0,1], 0 \leq \varphi_{\tilde{S}}(x, a) \leq 1\} \quad (2.7)$$

The primary membership of x , denoted by β_x , defines the domain for the secondary membership of x . PMF can be formally expressed as

$$\beta_x = \{a \in [0,1] | \varphi_{\tilde{S}}(x, a) > 0\} \quad (2.8)$$

To represent the MF of a GT2 FS three-dimensionally, the primary and secondary variables are to be plotted along the x- and y-axes respectively. The z-axis corresponds to the MF, $\varphi_{\tilde{S}}(x, a)$. The third dimension extends the accessory level of freedom, which allows for

precise modelling of uncertainties. A bounded region signifying the uncertainty of the PMF of a GT2 FS is termed the footprint of uncertainty (FOU), denoted by $FOU(\tilde{S})$. It can be expressed as

$$FOU(\tilde{S}) = \bigcup_{x \in X} \beta_x \quad (2.9)$$

A T1 FS characterizes the secondary membership function (SMF) as shown in (2.10).

$$\varphi_{\tilde{S}(x)}(a) = \{(a, \gamma_x(a)) \mid \gamma_x(a) \in A \quad \forall a \in \beta_x \subseteq [0,1]\} \quad (2.10)$$

where $\gamma_x(a)$ is termed as the secondary membership grade of x and $\gamma_x(b) = \varphi_{\tilde{S}}(x, b) \quad \forall x \in X$ [52]. At each x , the SMF, $\varphi_{\tilde{S}(x)}(a)$, is also referred to as the vertical slice of $\varphi_{\tilde{S}}(x, a)$ and (2.7) may be rewritten as follows.

$$\tilde{S} = \{(x, \varphi_{\tilde{S}(x)}(a)) \mid x \in X\} \quad (2.11)$$

Different forms of MFs are utilized to depict the GT2 FS, such as Gaussian, triangular, trapezoidal, and piecewise linear functions. Gaussian MF is commonly favoured as it ensures the continuity of the subsequent fuzzy set and can be tuned easily with optimization techniques. A three-dimensional portrayal of the MF of a GT2 FS is shown in Fig. 2.1. Apart from expressing a T2 FS in a vertical slice fashion, other representations, such as horizontal slices [194], wavy slices [194], alpha planes [195], and z-slices [196], exist to effectively manage the complexity of the modeling and associated computations [194]. Mendel argued in his paper that utilizing the GT2 FS in vertical slice interpretation should provide an optimal design solution [194]. We utilized the vertical slice design of GT2 FS to implement our methodology. Defuzzification is an imperative task in any fuzzy set-based system. Constructing a type-reduced set (TRS) is the primary step in the type-2 defuzzification. Among the prevalent type reduction methods [197, 198, 199], the vertical slice centroid type reduction (VSCTR) method is a pragmatic approach, as delineated in [200]. The potentials of the VSCTR are as follows: (a) centroids corresponding to the vertical slices represent the membership grades in the TRS and (b) the computation of the centroid of each vertical slice is the same as that of the T1 FS. In VSCTR, the j^{th} vertical slice of \tilde{S} is computed as

$$C_j = \frac{\sum_{r=1}^n a \times \varphi_{\tilde{s}(x)}(a)}{\sum_{r=1}^n \varphi_{\tilde{s}(x)}(a)} \quad (2.12)$$

where n is the count of elements of the j^{th} vertical slice. This value of centroid is the membership of each $x \in X$. These centroid values result in the TRS.

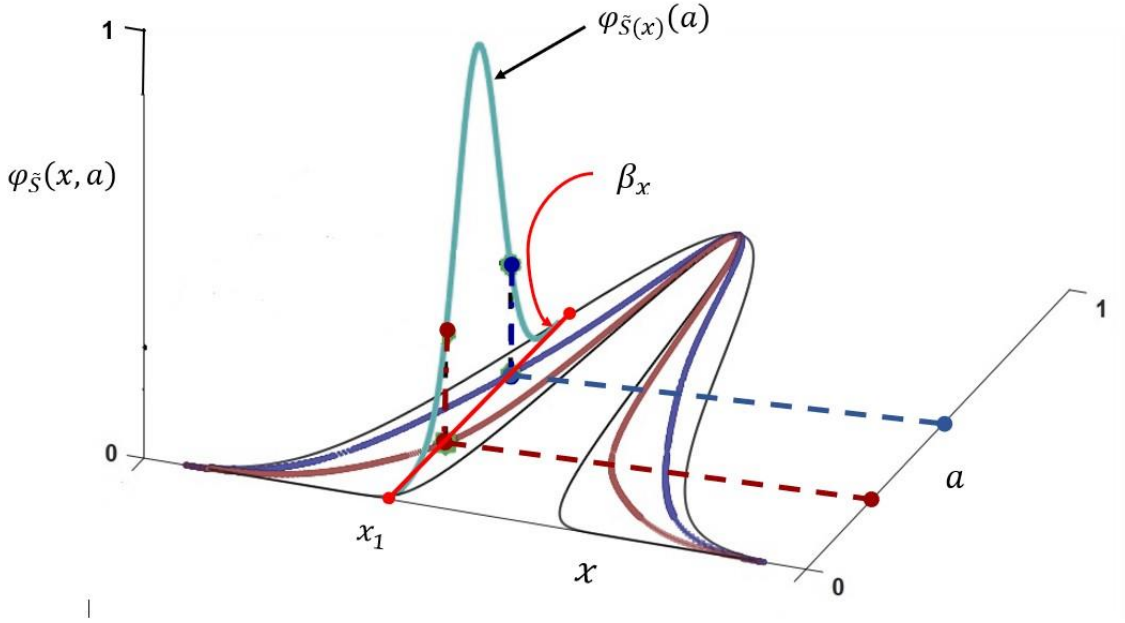


Fig. 2.1. 3D view of a general type-2 Gaussian MF

However, utilizing a GT2 FS-based system framework leads to inherent challenges, owing to the augmented complexity associated with its design and computational constraints. Another aspect impacting the adoption of systems based on GT2 FS is the lack of practical approaches for parameterizing the third dimension. Various developments have been made to alleviate the intricacies of GT2 FS [197, 198, 200, 201]. The progression, development, and applications of T2 FS have been detailed in a recent survey by De et al. [202].

2.3 GENERAL TYPE-2 FUZZY SET INDUCED RANK-LEVEL FUSION FOR FACE RECOGNITION

This section describes the multi-feature type-2 fuzzy set rank level fusion (MFT2FPRLF) method.

2.3.1 METHODOLOGY

In the present study, a multi-feature-based type-2 fuzzy set induced parallel rank-level fusion (MFT2FPRLF) and three FE methods, namely (generalized two-dimensional FLD (G-2DFLD) [62], fuzzy generalized two-dimensional LDA (FG-2DLDA) [63], and modular local binary patterns (LBP) [48] were used. Different FE techniques can yield more useful and relevant features that will help discern the images between classes. Two different neural network-based classifiers, the radial basis function (RBF) neural network [203] and support vector machines (SVM) [204], were used in this research. Each classifier is trained separately using the feature vectors of the training set derived using FE techniques. The feature vectors corresponding to the test image, generated by the same FE technique, are fed into the classifier to output weights or confidence factors corresponding to the different classes under consideration.

Based on the confidence factors, a predetermined number of top-ordered classes were selected. These top classes may differ when feature vectors resulting from different feature-extraction techniques are used. For each of these top-ordered classes, the degree of conformity was computed from the feature vectors of the test image and that of all training samples corresponding to the respective class. To minimize the uncertainty due to the presence of noisy and redundant data in the raw biometric data, a T2 FS is formed using the values of the degree of conformity for a class under consideration. A crisp value results from type reduction, followed by defuzzification for each of the top-ordered classes under deliberation. This value is complemented, leading to the fuzzy rank with respect to a particular FE technique. For a class, by imposing a penalty, in the case of the absence of a class in a set but present in another set, the complemented mean of confidence factors is computed. Similarly, the cumulative sum of complemented fuzzy ranks was calculated for the

same class. These penalties nullify the likelihood of a class becoming an unlikely winner. The fusion of the fuzzy rank sum and complemented mean of the confidence factor is accomplished by multiplication to obtain the fused score. The class with the minimum fused score is assigned as the class of the input test image. The schematic depiction of the methodology is shown in Fig. 2.2.

2.3.2 MATHEMATICAL FORMULATION OF THE METHODOLOGY

Let k different FE techniques (f_1, f_2, \dots, f_k) be applied to generate the feature vectors corresponding to a face image. In the present study, $k = 3$ as three FE techniques namely G2DFLDA, FG2DFLDA, and modular LBP have been employed. There are M and N images in the training set T and test set X , respectively. Let C denote the total number of classes. Let T^{f_k} and X^{f_k} denote the feature vectors for all C classes under consideration resulting from the FE technique f_k corresponding to the training set and test set. T^{f_k} and X^{f_k} can be expressed as follows –

$$T^{f_k} = \{t_{i,j}^{f_k} | i = 1, 2, 3, \dots, M \text{ and } j = 1, 2, 3, \dots, C\} \quad (2.13)$$

$$X^{f_k} = \{x_{i,j}^{f_k} | i = 1, 2, 3, \dots, N \text{ and } j = 1, 2, 3, \dots, C\} \quad (2.14)$$

The mathematical formulation of the steps, that are followed for each test image x_i of X is given below.

1. Using the FE technique f_k , generate feature vectors T^{f_k} and X^{f_k} for all C classes from the training set, T , and test set, X , respectively.
2. When a feature vector of a test image, x_i , $x_i \in X$, is given as input to a classifier to generate confidence factors corresponding to all C classes. Let $\omega_{i,j}^{f_k}$ denote the confidence factor of the j^{th} class corresponding to the test image x_i while applying the FE technique f_k .
3. A set of top-ordered P classes, L^{f_k} , is selected based on the descending order of the confidence factors corresponding to a test image x_i . Normalize these confidence factors by satisfying the following condition-

$$\sum_{p=1}^P w_{i,c}^{f_k} = 1 \quad \forall c, c \in L^{f_k} \quad (2.15)$$

4. Corresponding to a test image x_i , for each of the classes $j \in L^{f_k}$, compute the fuzzy ranks as follows –

4.1. Measure the degree of conformity of the class j using the feature vectors of the test image and that of all training samples corresponding to the class j using the following equation-

$$\theta_{i,j}^{f_k}(u) = \left\| t_{i,j,u}^{f_k} - x_{i,j}^{f_k} \right\| \text{ for } u = 1, 2, \dots, S \quad (2.16)$$

where S is the number of training samples corresponding to the class j . The number of such values depends on the number of training samples per class.

4.2. Normalize the degrees of conformity of the class j satisfying the condition, which is given by

$$\sum_{u=1}^S \theta_{i,j}^{f_k}(u) = 1 \quad \forall u; u = 1, 2, \dots, S \quad (2.17)$$

4.3. Form a type-2 fuzzy set, \tilde{A} , with the normalized degrees of conformity of class j , which is characterized by the Gaussian membership function $\gamma_{\tilde{A}}^{f_k}(j, h)$.

$$\tilde{A} = \{(j, h), \gamma_{\tilde{A}}^{f_k}(j, h) | \forall j \in L^{f_k}, \forall h \in \beta_j^{f_k} \subseteq [0,1], \gamma_{\tilde{A}}^{f_k}(j, h) \subseteq [0,1]\} \quad (2.18)$$

where $\gamma_{\tilde{A}}^{f_k}(j, h)$ is the type-2 membership function and represents the degree of membership of the element (j, h) . It is defined by Eq. (2.19).

$$\gamma_{\tilde{A}}^{f_k}(j, h) = e^{-\frac{(\theta_{i,j}^{f_k}(u) - 1.0)^2}{2 \times 1.0}} \quad (2.19)$$

The Eq. (2.18) represents the point-value representation of the T2FS. The Eq. (2.18) can further be expanded as the union of vertical slices of $\gamma_{\tilde{A}}^{f_k}(j, h)$ as given below-

$$\begin{aligned} \tilde{A} &= \left\{ ((j, h), \gamma_{\tilde{A}}^{f_k}(j, h)) | j \in L^{f_k}, h \in \beta_j^{f_k} \subseteq [0,1], 0 \leq \gamma_{\tilde{A}}^{f_k}(x, a) \leq 1 \right\} \\ &= \left\{ (j, \delta_j^{f_k}(h)) | j \in L^{f_k}, h \in \beta_j^{f_k} \subseteq [0,1] \right\} \end{aligned} \quad (2.20)$$

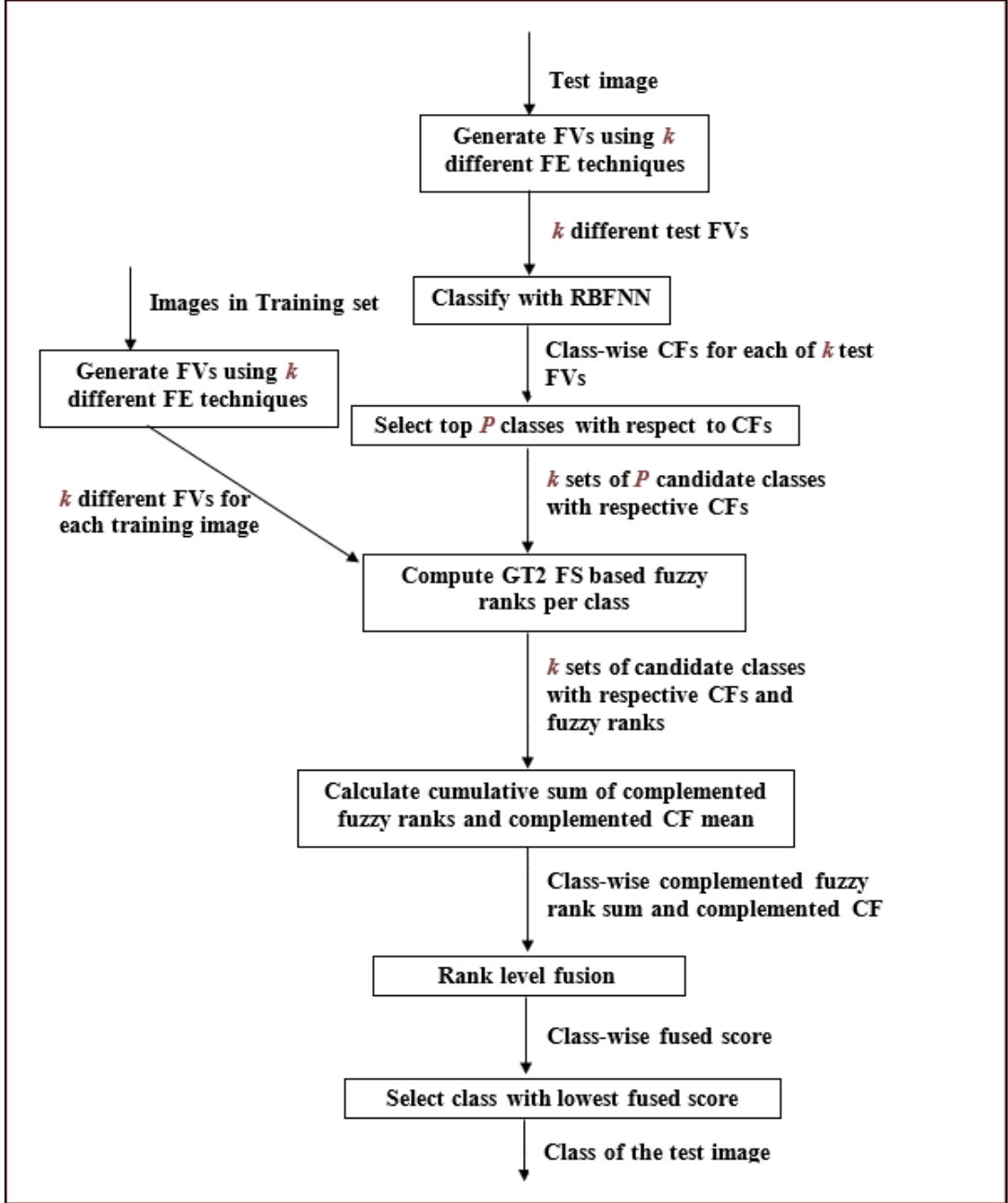


Fig. 2.2. Schematic description of the MFT2FPRLF method

The $\delta_j^{fk}(h)$ is the vertical slices of $\gamma_A^{fk}(j, h)$, and $\delta_j^{fk}(h), 0 \leq \delta_j^{fk}(h) \leq 1$ and called the SMF, and its domain is called the primary membership of j , denoted by $\beta_j^{fk} \subseteq [0, 1]$. $\delta_j^{fk}(h)$ is given by

$$\delta_j^{f_k}(h) = \frac{\gamma_{\tilde{A}}^{f_k}(j, h)}{\varphi_{\tilde{A}}^{f_k}(j)} \quad (2.21)$$

where $\varphi_{\tilde{A}}^{f_k}(j) = \max_h \gamma_{\tilde{A}}^{f_k}(j, h)$.

4.4. Compute the centroid, $C_j^{f_k}$, of the j^{th} vertical slice of \tilde{A} by applying vertical slice centroid type reduction (VSCTR) [200] using Eq. (2.22).

$$C_j^{f_k} = \frac{\sum_{u=1}^S h \times \delta_j^{f_k}(h)}{\sum_{u=1}^S \delta_j^{f_k}(h)} \quad (2.22)$$

4.5. For feature vectors of the test image, $x_i^{f_k}$ yielded by FE technique f_k , deduce the fuzzy rank corresponding to a class j , $R_j^{f_k}$, by taking the complement of $C_j^{f_k}$. That is,

$$R_j^{f_k} = 1 - C_j^{f_k} \quad (2.23)$$

The value of $R_j^{f_k}$ lies $[0, 1]$, and the lowest value is said to be the top-ranked class, similar to the conventional ranking.

5. $k = k + 1$
6. If $k \leq 3$ then repeat steps 2 to 6.
7. Considering the test image x_i , there are at most three sets of confidence factors and fuzzy ranks for each of the top-ordered classes L^{f_k} corresponding to three FE techniques ($k = 1, 2, 3$). The member classes in L^{f_p} and L^{f_q} ($p \neq q$) may vary as they result from different FE techniques. Compute the rank-sum, χ_{R_j} , and complemented mean of confidence factors, η_{w_j} , corresponding to class j , $j \in L^{f_k}$ as follows:

$$\chi_{R_j} = \sum_{k=1}^K \begin{cases} R_j^{f_k} = R_j^{f_k}, & \text{if } R_j^{f_k} \in L^{f_k} \\ R_j^{f_k} = P_j^R, & \text{otherwise} \end{cases} \quad (2.24)$$

and

$$\eta_{w_j} = 1 - \frac{1}{K} \sum_{k=1}^K \begin{cases} w_j^{f_k} = w_j^{f_k}, & \text{if } R_j^{f_k} \in L^{f_k} \\ w_j^{f_k} = P_j^w, & \text{otherwise} \end{cases} \quad (2.25)$$

where P_j^R and P_j^W are penalties that are imposed on class j if it does not come into the top L ranked classes. The value of P_j^R is evaluated to 0.606 by putting $\theta_{i,j}^{f_k}(u) = 0.0$ in the Eq. (2.19), whereas P_j^W is counted as 0.0.

8. Perform the fusion of χ_{R_j} and η_{w_j} with multiplication operation to evaluate the fused score, \mathfrak{J}_j , which is eventually used for the final ranking.

$$\mathfrak{J}_j = \chi_{R_j} \times \eta_{w_j} \quad (2.26)$$

9. Select the class having the least (minimum) fused score, which is designated as the class of input test image, x_i , as defined as follows.

$$Class(x_i) = \arg \min_{j \in L} \mathfrak{J}_j \text{ where } L = \bigcup_{k=1}^K L^{f_k} \quad (2.27)$$

2.4 EXPERIMENTAL RESULTS

The methodology, discussed in the previous section, was implemented in the C programming language under the Linux operating system installed on a computer having an Intel Core i5 (2.4 GHz) and 16 GB DDR3 RAM. We tested the performance of our method using three different holistic FE techniques (G-2DFLD, FG-2DLDA, and modular LBP methods) with two different neural network-based classifiers (RBFNN and SVM). Its performance is also compared with recent state-of-the-art fusion approaches, such as the Borda count method, highest rank method, logistic regression method, nonlinear weighted rank, CFGaussFRLF rank level fusion method [157], and interrelated FE techniques. To endorse the suggested decision-level fusion technique, we used four publicly available face databases: AT&T [176] and UMIST [179]. These databases are expounded in Section 1. 11. The performance of the methods is weighed up based on the average recognition rate. To validate the statistical significance of the performance improvement of our method, a paired t -test (one-tailed method) was applied. In this context, the null hypothesis (H_0) is defined as: ‘the average recognition rate of the MFT2FPRLF method is same as of the compared method’ against the

alternative hypothesis (H_1) as: ‘the average recognition rate of the MFT2FPRLF method is higher as compared to the related method’.

2.4.1 EXPERIMENTS WITH THE AT&T FACE DATABASE

For the experimental setup, a training set was prepared by randomly selecting r images for each subject. Rest $(10 - r)$ images form the test set.

Table 2.1: Performance evaluation of contemporary methods on the AT&T face database.

	Method	Average recognition rate \pm standard deviation (SD)				
		$r=3$	$r=4$	$r=5$	$r=6$	$r=7$
RBENN	G-2DFLD	92.82 ± 2.67	95.94 ± 1.21	97.73 ± 0.87	98.66 ± 0.96	98.42 ± 1.11
	FG-2DLDA	93.54 ± 2.62	96.03 ± 1.59	97.85 ± 1.01	98.75 ± 0.93	98.54 ± 1.04
	Modular LBP	93.43 ± 2.21	95.85 ± 1.71	97.83 ± 0.92	98.56 ± 0.86	98.50 ± 1.07
	Highest rank method	93.50 ± 2.50	96.01 ± 1.50	97.85 ± 1.01	98.65 ± 0.80	98.33 ± 0.81
	Modified highest rank method	93.62 ± 2.32	96.06 ± 1.42	98.08 ± 0.88	98.75 ± 0.93	98.62 ± 0.67
	Borda count	93.58 ± 2.23	96.03 ± 1.64	98.00 ± 1.04	98.65 ± 0.80	98.54 ± 0.66
	Logistic regression	93.66 ± 2.22	96.09 ± 1.63	98.25 ± 1.01	98.91 ± 0.67	98.92 ± 0.48
	Nonlinear weighted rank	94.46 ± 1.75	96.49 ± 1.38	98.50 ± 0.63	99.17 ± 0.56	99.04 ± 0.62
	CFGaussFRLF	94.68 ± 1.59	96.55 ± 1.43	98.58 ± 0.63	99.35 ± 0.59	99.13 ± 0.69
	MFT2FPRLF	94.9 ± 1.36	96.83 ± 1.35	98.70 ± 0.57	99.47 ± 0.65	99.25 ± 0.65
SVM	G-2DFLD	93.04 ± 2.51	95.99 ± 1.23	97.68 ± 0.91	98.72 ± 0.92	98.50 ± 1.07
	FG-2DLDA	93.62 ± 2.61	96.08 ± 1.47	98.00 ± 1.03	98.88 ± 0.63	98.62 ± 0.68
	Modular LBP	93.43 ± 2.21	95.88 ± 1.68	97.85 ± 0.95	98.69 ± 0.79	98.58 ± 0.61
	Highest rank method	93.67 ± 2.30	96.06 ± 1.46	98.00 ± 1.03	98.81 ± 0.90	98.42 ± 0.85
	Modified highest rank method	93.79 ± 2.15	96.09 ± 1.43	98.05 ± 1.05	98.91 ± 0.81	98.71 ± 0.69
	Borda count	93.75 ± 2.13	96.15 ± 1.55	98.08 ± 0.88	98.84 ± 0.71	98.62 ± 0.68
	Logistic regression	93.83 ± 2.06	96.25 ± 1.29	98.28 ± 0.83	99.03 ± 0.48	98.83 ± 0.65
	Nonlinear weighted rank	94.63 ± 1.81	96.66 ± 1.47	98.68 ± 0.59	99.34 ± 0.57	99.04 ± 0.73
	CFGaussFRLF	94.73 ± 1.85	96.73 ± 1.22	98.73 ± 0.62	99.41 ± 0.63	99.25 ± 0.71
	MFT2FPRLF	95.02 ± 1.43	96.95 ± 1.08	98.82 ± 0.57	99.53 ± 0.45	99.37 ± 0.59

For effective evaluation of the presented method, MFT2FPRLF, different values of r , such as 3, 4, 5, 6, and 7, were used to create 20 dissimilar sets of training and test images so that the possibility of any bias was eliminated during the creation of training and test sets and conduct the experiments. The results are reported in Table 2.1.

Table 2.2: Statistical significance analysis between the competitive methods and the IT2FSIFRLF method on the AT &T face database.

	Method	<i>p</i> -value				
		$r=3$	$r=4$	$r=5$	$r=6$	$r=7$
RBFNN	G-2DFLD	0.00021	0.01151	0.00003	0.00059	0.00137
	FG-2DLDA	0.00259	0.01883	0.00010	0.00075	0.00167
	Modular LBP	0.00020	0.00230	0.00015	0.00009	0.00175
	Highest rank method	0.00097	0.00634	0.00010	0.00014	0.00010
	Modified highest rank method	0.00141	0.00023	0.00025	0.00075	0.00364
	Borda count	0.00078	0.00745	0.00039	0.00015	0.00054
	Logistic regression	0.00117	0.00991	0.01363	0.00098	0.00870
	Nonlinear weighted rank	0.03390	0.00220	0.01418	0.00317	0.06919
	CFGaussFRLF	0.07926	0.00797	0.04807	0.02104	0.09321
SVM	G-2DFLD	0.00012	0.00325	0.00000	0.00038	0.00020
	FG-2DLDA	0.00324	0.00231	0.00120	0.00061	0.00015
	Modular LBP	0.00008	0.00060	0.00002	0.00002	0.00000
	Highest rank method	0.00032	0.00027	0.00185	0.00058	0.00002
	Modified highest rank method	0.00037	0.00021	0.00181	0.00103	0.00066
	Borda count	0.00090	0.00089	0.00090	0.00077	0.00015
	Logistic regression	0.00097	0.00120	0.00253	0.00039	0.00056
	Nonlinear weighted rank	0.03984	0.06700	0.06859	0.07383	0.02122
	CFGaussFRLF	0.08355	0.05106	0.05180	0.14721	0.09432

It is apparent from Table 2.1 that the MFT2FPRLF method had the highest average recognition rates in all cases, evidencing the improved performance of MFT2FPRLF over the classical methods in all cases. A performance analysis of the method in terms of *the p*-values is presented in Table 2.2. It is observed that our approach achieves substantially better

significance in most cases than the other methods, at $p < 0.05$. For $r = 3$ and 7, the performance of CFGaussFRLF was significant compared to that of MFT2FPRLF when experimenting with RBFNN and SVM. With the SVM classifier, the performance of the nonlinear weighted rank was significant compared with that of the GT2 FS-based RLF for $r = 4, 5$, and 6. The same is true for the CFGaussFRLF with $r = 6$.

Table 2.3: Performance evaluation of contemporary methods on the UMIST face database.

	Method	Average recognition rate \pm standard deviation (SD)			
		$r=4$	$r=6$	$r=8$	$r=10$
RBFNN	G-2DFLD	86.05 \pm 3.08	92.09 \pm 2.13	95.43 \pm 1.28	96.88 \pm 1.16
	FG-2DLDA	86.30 \pm 3.00	92.43 \pm 1.49	95.59 \pm 1.17	97.05 \pm 1.09
	Modular LBP	86.43 \pm 3.03	92.23 \pm 1.59	95.51 \pm 1.13	97.03 \pm 0.98
	Highest rank method	86.56 \pm 2.89	93.93 \pm 1.14	95.78 \pm 1.11	97.29 \pm 1.21
	Modified highest rank method	86.67 \pm 2.86	93.17 \pm 0.98	96.03 \pm 0.94	97.51 \pm 1.21
	Borda count	86.84 \pm 2.77	93.10 \pm 0.88	96.13 \pm 0.73	97.73 \pm 1.15
	Logistic regression	87.19 \pm 2.78	93.39 \pm 0.93	96.39 \pm 0.90	97.81 \pm 0.88
	Nonlinear weighted rank	87.52 \pm 3.05	93.56 \pm 0.99	96.88 \pm 1.07	98.14 \pm 0.73
	CFGaussFRLF	87.66 \pm 3.02	93.65 \pm 0.91	96.98 \pm 0.98	98.25 \pm 0.86
	MFT2FPRLF	87.92 \pm 3.08	94.15 \pm 1.06	97.21 \pm 0.93	98.60 \pm 0.90
SVM	G-2DFLD	86.22 \pm 3.09	92.28 \pm 2.06	95.54 \pm 1.29	96.92 \pm 1.16
	FG-2DLDA	86.83 \pm 2.83	92.59 \pm 1.51	95.70 \pm 1.11	97.14 \pm 0.95
	Modular LBP	86.77 \pm 2.79	92.50 \pm 1.51	95.62 \pm 1.17	97.01 \pm 1.01
	Highest rank method	86.95 \pm 2.66	92.95 \pm 1.13	95.90 \pm 1.11	97.42 \pm 1.04
	Modified highest rank method	87.06 \pm 2.67	93.08 \pm 1.04	96.14 \pm 0.92	97.55 \pm 1.12
	Borda count	87.14 \pm 2.83	93.24 \pm 0.90	96.22 \pm 0.69	97.65 \pm 1.11
	Logistic regression	87.35 \pm 2.82	93.41 \pm 0.41	96.47 \pm 0.89	97.89 \pm 0.92
	Nonlinear weighted rank	87.61 \pm 3.05	93.60 \pm 0.96	96.91 \pm 1.10	98.28 \pm 0.84
	CFGaussFRLF	87.72 \pm 3.00	93.70 \pm 0.84	97.04 \pm 0.95	98.36 \pm 0.81
	MFT2FPRLF	88.03 \pm 2.86	94.20 \pm 0.86	97.40 \pm 0.69	98.59 \pm 0.72

2.4.2 EXPERIMENTS WITH THE UMIST FACE DATABASE

Similar to the AT&T face database, we tested the performance of our method, MFT2FPRLF, with the UMIST database. The values of r were chosen as 4, 6, 8, and 10 to form different

pairs of training and testing sets. The average recognition rates of the different methods yielded in experimental runs with 20 sets of training and test set pairs for each r are presented in Table 2.3. The results show that the performance of the MFT2FPRLF method is improved compared with that of the classical methods in all cases.

Table 2.4: Performance evaluation of contemporary methods on the UMIST face database.

	Method	<i>p</i> -value			
		$r=4$	$r=6$	$r=8$	$r=10$
RBFNN	G-2DFLD	0.00012	0.00004	0.00001	0.00000
	FG-2DLDA	0.00019	0.00002	0.00002	0.00000
	Modular LBP	0.00032	0.00004	0.00000	0.00000
	Highest rank method	0.00044	0.00001	0.00002	0.00009
	Modified highest rank method	0.00019	0.00013	0.00000	0.00014
	Borda count	0.00135	0.00110	0.00010	0.00000
	Logistic regression	0.00205	0.00214	0.00290	0.00006
	Nonlinear weighted rank	0.03560	0.00734	0.11874	0.00322
	CFGaussFRLF	0.05339	0.01075	0.02248	0.01528
SVM	G-2DFLD	0.00006	0.00018	0.00000	0.00000
	FG-2DLDA	0.00055	0.00008	0.00000	0.00000
	Modular LBP	0.00032	0.00003	0.00000	0.00000
	Highest rank method	0.00026	0.00017	0.00002	0.00000
	Modified highest rank method	0.00039	0.00038	0.00007	0.00000
	Borda count	0.00104	0.00021	0.00001	0.00003
	Logistic regression	0.00380	0.00180	0.00058	0.00011
	Nonlinear weighted rank	0.04358	0.00828	0.06067	0.02288
	CFGaussFRLF	0.09140	0.01130	0.04814	0.06254

The results, reported in Table 2.4, demonstrate that the performance of the MFT2FPRLF method is statistically significant compared to other state-of-the-art methods, except for $r = 4$ and 10. In this case, the CFGaussFRLF method significantly outperformed our method in terms of the performance with the SVM classifier.

The assessment of performance in terms of the average recognition rate as well as statistical significance justifies the improved competence of our fusion approach using a GT2 FS.

It is important to mention here that the top P classes are considered in the RLF based on the confidence factors of the classifiers (step 3 of the algorithm, section 2.3.2). We chose sets with $r = 6$ from the AT&T face database to determine the optimal value of P . Our method was executed repeatedly for each set, with $P = 3, 4, 5, 6$, and 7 . The average recognition rate is plotted against P , as shown in Fig. 2.3. The recognition rate starts increasing from $P = 3$ and reaches a peak at $P = 5$, after which it descends. Therefore, $P = 5$ was selected for the experiments.

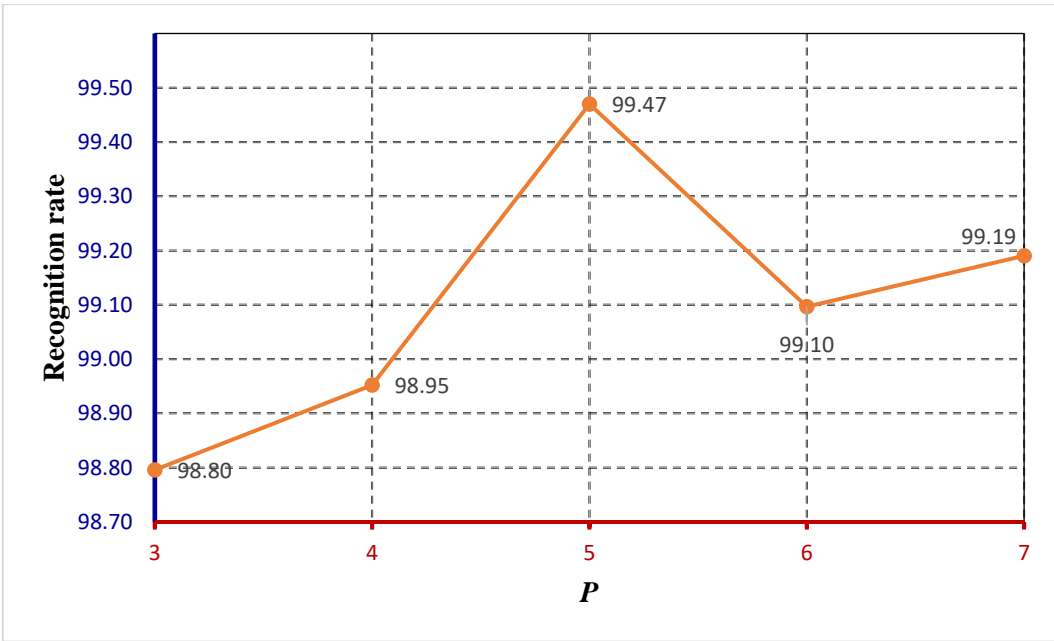


Fig. 2.3. Graph illustrating Recognition rate versus P (number of top-ordered classes considered in rank-level fusion)

2.5 CONCLUSION

In view of the experimental results and comparison of our method, MFT2FPRLF, with not only related individual FE methods but also interrelated fusion methods, it can be argued that our method, illustrated here, has meliorated the recognition precision. It has also demonstrated its potency in resolving highly conflicted classification instances by exercising type-2 fuzzy set-based multi-feature rank-level fusion.

Even though GT2 FS stands out for managing uncertainty and imprecision in the input data, the key constraints entail its intricate structure and intensive computations. Another inherent

problem in this fuzzy set is that the performance of the model relies on parameter selection. These factors have led researchers to be inclined toward interval type-2 fuzzy sets because of their computational efficiency, interpretability, and ease of implementation. Indeed, there is scope for improvement in exploring the predictive effect of multi-feature rank-level fusion with the interval type-2 fuzzy set in dealing with the associated factors that create hindrances in FR.

CHAPTER 3

INTERVAL TYPE-2 FUZZY SET INDUCED RANK-LEVEL FUSION FOR FACE RECOGNITION

3.1 INTRODUCTION

Computer-aided FR has been an active research area for several decades, because its applications have been found in many secure access control environments, law enforcement departments, and other computer vision domains. As FR is used in sensitive areas and deals with human beings, the accuracy of FR systems plays an important role in achieving its acceptability. In order to achieve higher accuracy, even in an uncontrolled environment and in the presence of wide variations in pose, race, illumination, etc., researchers proposed many techniques that include FE, classifier design and their combination. The review suggests that the types of features and classifiers used limit the success of these methods. Therefore, there was a need for finding a way to utilize different feature vectors and classifiers to achieve higher success rates. This has led to the development of different fusion techniques for FR systems. Typically, a generic FR system passes through different steps that involve (i) acquisition of raw face image data using sensors, (ii) extraction of discriminative features, (iii) classification, and (iv) recognition. Fusion techniques can be applied in all steps, as

stated above. Based on the level or step at which fusion is carried out, especially in person recognition, the fusion techniques can be classified into two main different categories [3, 4, 205-209], namely, (a) pre-classification-level, and (b) post-classification-level fusion. The fusions applied at the pre-classification level include (i) sensor-level fusion and (ii) feature-level fusion (FLF).

Image data can be acquired from a subject by using different sensors. In addition, the image data of a biometric trait can be acquired by a single sensor by varying the sensor parameters and surrounding environment. In sensor-level fusion [130, 210-212], suitable image data are fused to accumulate richer information from the image domain. One drawback of this technique is that the dimension of the image data gets increased, which is then reduced by employing dimension reduction algorithms at the cost of possible loss of underlying useful information.

FLF integrates different feature vectors extracted from multiple biometric traits or a single biometric trait using different FE algorithms. The key idea is to extract as many features as possible from a single or multiple biometric trait(s) and fuse them in the form of a feature vector so that richer discriminative information can be accumulated. It aims to collect all possible features before going for the classification process. The main drawback of this fusion method is that the dimensions of the fused feature vectors increase, which creates an extra computational burden in the classification process. Although its dimensions can be reduced using a dimension reduction algorithm, such as PCA, useful information may be lost in the process. Some of the research studies involved FLF with features extracted with various descriptors like PCA, LBP, HOG, and local phase quantization (LPQ), etc. [213-215].

In post-classification-level fusion, the classification results of multiple classifiers, either from a single biometric trait or multiple biometric traits, are aggregated to arrive at a final decision to recognize a person. Based on the manner in which the classifier's outputs are fused, it can be further subdivided into three categories: (i) decision-level fusion, (ii) score-level fusion, and (iii) rank-level fusion.

In decision-level fusion, decisions arising from multiple classifiers are integrated using a combination rule to render a decision [216]. Wang et al. [217] brought forward an approach

in which the LBP and two-dimensional LDA were applied to extract features from faces, and subsequently, matching decisions using SVM were fused to derive the final decision.

In the score-level fusion technique, the classification scores yielded by multiple classifiers are fused to generate final scores, based on which a person is recognized. This method can employ either a density-based scheme, which is based on score distribution estimation, or a transformation-based scheme, which is based on the combination of individual matching scores. Rehmi et al. [218] suggested a recognition framework incorporating the face and palmprint in score-level fusion using the weighted sum rule. Another study involving the face and palmprint was put forward by Rane et al. [219] advocated a score-level fusion scheme using t-norm. Another approach by Zhang et al. [150] propounded the fusion of matching scores from face and voice biometrics by applying the weighted-sum rule.

In the rank-level fusion technique, based on the strengths of the outputs of a classifier, a matcher yields ranks indexed by natural numbers for all possible classes. This is performed for different classifiers, either with unimodal or multimodal biometric traits. Finally, the ranks corresponding to a class are fused to obtain a single final rank, which in turn is used in the decision-making process. A summary of this type of fusion is presented in Table 3.1. Sharma [153] demonstrated a methodology utilizing the face and iris to integrate the results of different rank-level fusion methods in a serial and parallel manner. A methodology was proposed by Talebi [155], wherein the biometric features of frontal and profile faces together with the ear were employed in rank-level fusion using a confidence-based scheme. Devi et al. [220] presented an approach in their study in which palmprint and facial features (both global and local) were deployed in classification, and the Borda count, logistic regression, and highest-rank methods were applied for fusing ranks. A rank-level fusion strategy utilizing a genetic algorithm was suggested by Ahmad et al. [221]. One drawback of the rank-level fusion technique is that it does not consider the spatial correlation among the outputs of the classifier while generating the ranks. As a result, the final rank yielded by the fusion may not be appropriate for the representation of the class. To address this issue, Sing et al. [157] proposed a fusion technique in which fuzzy ranks are formulated using the outputs of a classifier. Then the multiple fuzzy ranks yielded using different types of feature vectors are

suitably fused to generate the final fuzzy rank corresponding to a class. It utilizes the type-1 fuzzy set for generating the fuzzy ranks.

Table 3.1: Summary of different rank-level fusion techniques.

Authors and Year	Fusion method
Ho et al. 1994 [222]	Highest rank
Abaza et al. 2009 [189]	Modified highest rank, Borda count
Monwar et al. 2011 [187]	Logistic regression
Kumar et al. 2010 [191]	Nonlinear weighted
Sing et al. 2019 [157]	Parallel fuzzy rank-level fusion
Ahmad et al. 2022 [221]	Rank level fusion with genetic algorithm

However, owing to the structured shape of the human face and its expressibility with wide variations in poses, emotions, etc., there is always the possibility of a high degree of intraclass variability and interclass similarity, which limits the success of biometric systems. In the work, detailed in this chapter, we address this issue and develop an interval type-2 fuzzy set-based fuzzy rank generation method, which mitigates the effects of intra-class variability, and inter-class similarity. For each feature vector, we generated interval type-2 fuzzy-set-based fuzzy ranks corresponding to the top K classes. To reduce complexity, the top K classes were considered for RLF. The type-2 MFs with respect to a class are generated by excluding the images of the other classes, thereby reducing the effects of inter-class similarity. We used a suitable Gaussian mass function to define the lower and upper bounds of the primary membership grades corresponding to an interval type-2 fuzzy MF. The defuzzified value was used to generate the fuzzy rank for the class being considered. Likewise, for all possible classes, we generated type-1 fuzzy primary membership values, which were then used to generate the fuzzy ranks. Finally, the aggregated fuzzy ranks are fused with the respective complemented mean of the confidence factors to obtain the final fuzzy ranks for the classes being considered, based on which a face image is classified and recognized.

Our method differs from deep learning models such as deep convolutional neural networks (DCNNs) [95] in several ways. First, unlike the presented method, a generic DCNN

model uses a 2D image matrix as input and passes through one or more convolutional and pooling layers to extract features that are more informative and reduce the image matrix simultaneously. It then flattens the feature vector and classifies the image using a fully connected neural network. Second, the DCNNs require a large training set to achieve good accuracy. As a result, its training period is much longer and sometimes that goes up to a few hours. Furthermore, running a DCNN model requires significantly more memory.

The remaining part of this paper is organized as follows. We presented a brief overview of the interval type-2 fuzzy sets in Section 3.2. The detailed framework for our interval type-2 fuzzy set induced fuzzy rank-level fusion (IT2FSIFRLF) method for FR is described in Section 3.3. The simulation results and discussion are presented in Section 3.4. Finally, Section 3.5 draws concluding remarks.

3.2 A CONCISE OVERVIEW OF INTERVAL TYPE-2 FUZZY SETS

An overview of the T2 FS is deliberated in section 2.2.2 of chapter 2. This section revisits the relevant notions of T2 FS for a clear understanding of the interval type-2 fuzzy set (IT2 FS). The fuzzy set, also known as T1 FS, is introduced by Lotfi A. Zadeh in 1965 [223] and T2 FS in 1975 [224]. Since then, T1 FSs and T2 FSs have been extensively studied and used in many computer vision domains, where information uncertainty plays a big challenge to researchers. Unlike a crisp set whose MF takes only 0 or 1, the MF of a fuzzy set can range between 0 and 1. A T1 FS, S , is defined as the ordered pair of member elements and its membership function that describes its belongingness to the fuzzy set S , as defined below:

$$S = \{(x, \varphi_S(x)) | x \in X, 0 \leq \varphi_S(x) \leq 1\} \quad (3.1)$$

where x is the member element, $\varphi_S(x)$ is the membership function of x in the fuzzy set S , and X is the universe of discourse. Each T1 FS is characterized by its membership function. Therefore, T1 FSs mainly differ from each other in the manner in which the MFs are defined. The main difference between a T2 FS, also known as a general T2 FS (GT2 FS),

and a T1 FS is that the membership function of a T2 FS itself is a T1 FS, instead of a crisp membership function. A T2 FS, \tilde{S} , can be defined as follows.

$$\tilde{S} = \{((x, a), \varphi_{\tilde{S}}(x, a)) | x \in X, 0 \leq a, \varphi_{\tilde{S}}(x, a) \leq 1\} \quad (3.2)$$

where values of a and $\varphi_{\tilde{S}}(x, a)$ are characterized as primary and secondary membership grade, respectively. Both grades accept values in $[0, 1]$. The secondary membership corresponding to each primary membership delimits the fuzziness of primary membership. The union of primary memberships forms a bounded region, which is described as the FOU of type-2 MF. However, utilizing a GT2 FS-based system framework leads to inherent challenges, owing to the augmented complexity associated with its design and computational constraints. Another aspect impacting the adoption of systems based on GT2 FS is that there have not been definite practical approaches for parameterizing the third dimension. Mandel et al. [225] conceived the idea of an interval type-2 fuzzy set (IT2 FS) as a derivative of a GT2 FS where all its secondary membership grades remained constant at unity. The computations involved have become less complex, thus making them easier to handle. Studies demonstrated that interval type-2 fuzzy systems are effective in scenarios where the noise level is extreme, the environment is dynamic, or the system holds nonlinearity [226]. Researchers have shown a strong preference for interval type-2 fuzzy systems, which has led to significant advancements in this field. An IT2 FS is constructed based on the mathematical framework of the T1 FS and maintains a secondary grade equal to 1 across all x . An IT2 FS is expressed as an extended version of (3.2) which is given below.

$$\tilde{S} = \{((x, a), \varphi_{\tilde{S}}(x, a) = 1) | x \in X, 0 \leq a \leq 1\} \quad (3.3)$$

The FOU is confined by two T1 FSs known as the lower membership function (LMF) and upper membership function (UMF). An IT2 FS is completely characterized by its FOU because its secondary grades provide no substantial information. Therefore, the IT2 FS can be interpreted as follows:

$$\tilde{S} = 1/FOU(\tilde{A}) \quad (3.4)$$

Since the $FOU(\tilde{S})$ is bounded by the lower membership function (LMF), $\underline{\varphi_{\tilde{S}}}(x)$, and upper membership function (UMF), $\overline{\varphi_{\tilde{S}}}(x)$, (4) can be replaced with the following equation.

$$\tilde{S} = 1 / [\underline{\varphi}_{\tilde{S}}(x), \overline{\varphi}_{\tilde{S}}(x)] \quad (3.5)$$

Based on the aspects of uncertainties, the upper and lower membership functions are selected. In most cases, Gaussian functions are used in IT2 FSs for simplicity and to guarantee continuity. The domains of an IT2 FS with Gaussian MF are depicted in Fig. 3.1. The further discussion of IT2 FSs is beyond the scope of this chapter and our presented methodology as well. Interested readers may refer to [193,227, 228] for further details on IT2 FSs.

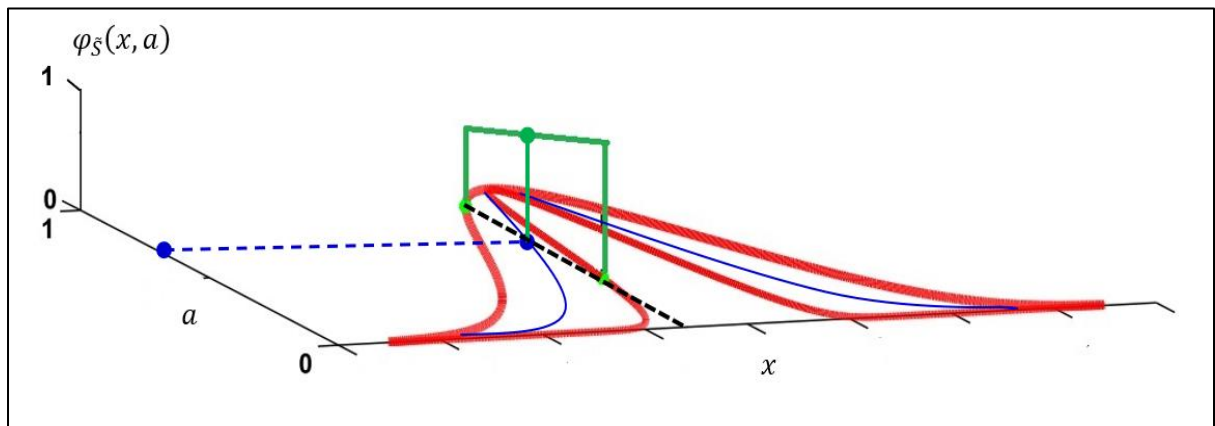


Fig. 3.1. Domains of an IT2 FS with Gaussian MF

3.3 INTERVAL TYPE-2 FUZZY SET INDUCED FUZZY RANK-LEVEL FUSION FOR FACE RECOGNITION

3.3.1 METHODOLOGY

This section presents the interval type-2 fuzzy set induced fuzzy rank-level fusion (IT2FSIFRLF) for FR utilizing the different feature vectors of a single face image. In the previous chapter, we extended our preceding work [157] by introducing the GT2 FS to generate fuzzy ranks among candidate classes while classifying a face image [229]. This is because GT2 FS has proven capability to deal with a higher level of class-specific uncertainty than T1 FS. In a previous study [157], a fuzzy rank was generated from the normalized output

of a classifier using a Gaussian density function. In the work, detailed in this chapter, we employed an IT2 FS because of its simple representation, high computational efficiency, and wide range of use cases across diverse domains. The present effort differs from [229] in two aspects. First, the IT2 FS was formulated with normalized degrees of conformity for the class under consideration using only intra-class face images. Excluding face images of other classes reduces the effects of inter-class similarity while generating fuzzy ranks. A suitable Gaussian mass function is applied to define the lower and upper bounds of the primary membership grades corresponding to an interval type-2 fuzzy MF. Second, direct defuzzification with a simplified version of the Nie-Tan method [228] gives rise to a crisp value circumventing type-reduction. This crisp value was subsequently utilized to obtain the fuzzy rank for the class.

Usually, a feature vector is presented to a classifier to recognize an image based on the strength of the classifier's outputs. These outputs are called the confidence factors (CFs) of the classifier. Owing to the unique structure of the human face and its ability to express different poses, emotions, etc., higher intra-class variability and inter-class similarity often cause more difficulties and challenges in developing a reliable biometric recognition system. To address these issues, we introduced an IT2 FS that utilizes the intraclass variability among the intraclass face images to define the lower and upper bounds of the primary membership grades based on the Gaussian density function with a fixed standard deviation and uncertain means. This mechanism leads us to obtain a primary membership value, which has more influence and is guided by intraclass variability. It also reduces the adverse influence of inter-class similarity by excluding inter-class facial images from the computation. In addition, multi-feature vectors for a face image derived from different FE techniques are used to characterize the underlying discriminant features that address the issues of inter-class similarity. Multi-feature vectors play a role in accumulating complementary yet useful discriminating information from a single face image.

The defuzzified value obtained by applying the Nie-Tan algorithm was complemented to derive the fuzzy rank for a class being considered. For each feature vector, we generated IT2 FS-based fuzzy ranks corresponding to the top-ordered classes that were selected based on

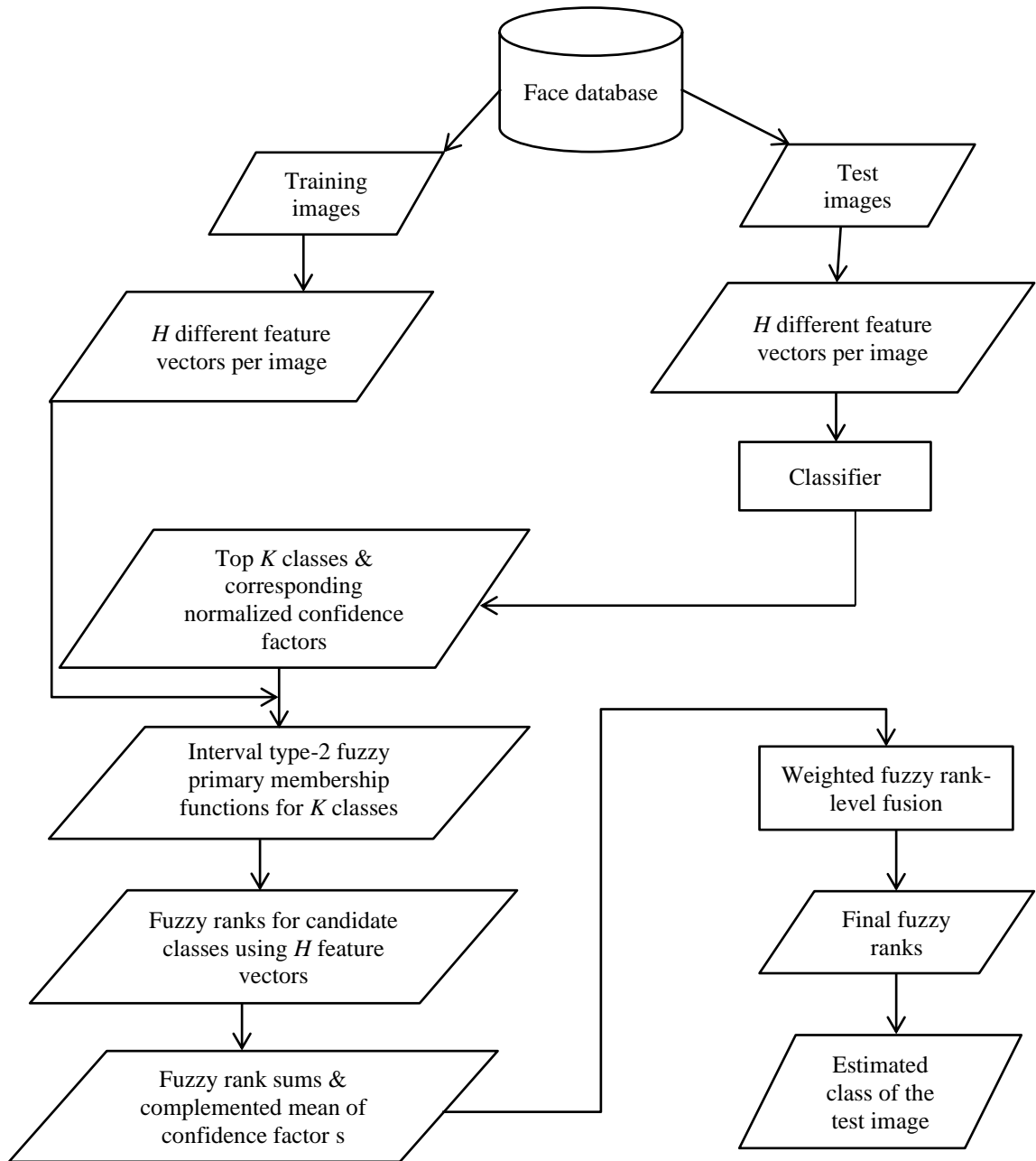


Fig. 3.2. Schematic diagram of the workflow of the IT2FSIFRLF.

the output of the classifier. Likewise, interval type-2 fuzzy set-based multiple fuzzy ranks were obtained using multiple feature vectors. Finally, the sum of these fuzzy ranks was fused with the corresponding complemented mean of the confidence factors to derive the final

fuzzy ranks for the classes based on which a face image is recognized. To better understand our fusion method, the workflow is illustrated in Fig. 3.2.

3.3.2 MATHEMATICAL FORMULATION OF THE METHODOLOGY

The mathematical formulation of the algorithm is as follows.

Let there be M and N images in the training set $T = \{t_1, t_2, \dots, t_M\}$ and test set $X = \{x_1, x_2, \dots, x_N\}$, respectively. Let C be the total number of classes (subjects). Each face image is characterized by a set of different feature vectors $F = \{F_1, F_2, \dots, F_H\}$. In the present study, we took $H=3$ corresponding to three different FE techniques, namely, G-2DFLD [62], FG-2DFLD [63], and modular LBP [48]. Let $T^{F_h} = \{t_i^{F_h}\}, h = 1, 2, \dots, H; i = 1, 2, \dots, M$ denotes the set of all feature vectors in the training set. Likewise, $X^{F_h} = \{x_i^{F_h}\}, h = 1, 2, \dots, H; i = 1, 2, \dots, N$ denotes the same for the test set. The algorithmic description of our method is as follows:

1. Consider any one type of feature vector from the training and test sets and set $h=1$.
2. A test feature vector $x_i^{F_h}$ is given as input to a classifier to generate outputs (confidence factors) corresponding to all C classes. Let $\omega_{i,c}^{F_h}, c = 1, 2, \dots, C$ denotes the confidence factor of the c^{th} class.
3. Let $C_i^{F_h}$ be the set of top K classes corresponding to top K confidence factors and $L_i^{F_h} = \{\omega_{i,c}^{F_h} | c \in C_i^{F_h}\}$, which is normalized satisfying the following condition.

$$\sum \omega_{i,c}^{F_h} = 1 \mid c \in C_i^{F_h} \quad (3.6)$$

4. Generate T2 FS for class c using intraclass face images.

- 4.1 Measure the degrees of conformity of the class $c \in C_i^{F_h}$ using the feature vector of the test image $x_i^{F_h}$ and all that of the training samples $t_j^{F_h} | t_j^{F_h} \in c, j = 1, 2, \dots, M$ of class c using the following equation-

$$\theta_{i,c}^{F_h}(u) = e^{-\frac{\|t_{u,c}^{F_h} - x_i^{F_h}\|^2}{1 \times 2.0}}, u = 1, 2, \dots, S \quad (3.7)$$

where $t_{u,c}^{F_h}$ is the u^{th} training sample out of the total S corresponding to class c .

4.2 Normalize the degrees of conformity of class c , satisfying the following condition:

$$\sum_{u=1}^S \theta_{i,c}^{F_h}(u) = 1 \quad (3.8)$$

4.3 Form an IT2 FS, \tilde{A}_c , with the normalized degrees of conformity of the class c and is defined as follows

$$\tilde{A}_c = \{((y, v), 1) \mid 0 \leq v \leq 1\} \quad (3.9)$$

assuming $x_{i,c}^{F_h}$ as y and $\theta_{i,c}^{F_h}$ as v . There are S training samples in Class c . For each y , we can relate S number of secondary membership grades v_1, v_2, \dots, v_S . However, to every primary membership grade, the secondary membership grade is considered as 1. The Eq. (3.9) can be expressed in terms of $\text{FOU}(\tilde{A}_c)$, which is bounded by $\text{LMF}(\tilde{A}_c)$ and $\text{UMF}(\tilde{A}_c)$ as given below.

$$\tilde{A}_c = \frac{1/\left[\underline{\delta}_{\tilde{A}_c}(y_1), \overline{\delta}_{\tilde{A}_c}(y_1)\right]}{y_1} + \dots + \frac{1/\left[\underline{\delta}_{\tilde{A}_c}(y_S), \overline{\delta}_{\tilde{A}_c}(y_S)\right]}{y_S} \quad (3.10)$$

where $\underline{\delta}_{\tilde{A}_c}(y_u)$ and $\overline{\delta}_{\tilde{A}_c}(y_u)$, $u = 1, 2, \dots, S$, are the consequent $\text{LMF}(\tilde{A}_c)$ and $\text{UMF}(\tilde{A}_c)$, respectively. These are computed as follows.

$$\overline{\delta}_{\tilde{A}_c}(y_u) = \begin{cases} \mathcal{G}(v_u; m_1, \sigma), & \text{if } v_u < m_1 \\ 1, & \text{if } m_1 \leq v_u \leq m_2 \\ \mathcal{G}(v_u; m_1, \sigma), & \text{if } v_u > m_2 \end{cases}$$

$$\underline{\delta}_{\tilde{A}_c}(y_u) = \begin{cases} \mathcal{G}(v_u; m_2, \sigma), & \text{if } v_u \leq \frac{m_1 + m_2}{2} \\ \mathcal{G}(v_u; m_1, \sigma), & \text{if } v_u > \frac{m_1 + m_2}{2} \end{cases}$$

where m_1, m_2 are uncertain means with $m_1 = m - \sigma$, and $m_2 = m + \sigma$, m and σ are the mean and standard deviation of $\{v_u \forall u\}$, respectively.

$$\mathcal{G}(v_u; m, \sigma) = e^{-\frac{(\theta_{i,c}^{Fh}(u) - m)^2}{2 \times \sigma^2}}, m \in [m_1, m_2] \text{ and } v_u = \theta_{i,c}^{Fh}(u).$$

5. Obtain the crisp output for \tilde{A}_c through direct defuzzification using the following equation, which is a simplified version of the Nie-Tan method [228].

$$\mu_c^{Fh} = \frac{\sum_{u=1}^S v_u \times (\delta_{\tilde{A}_c}(y_u) + \overline{\delta_{\tilde{A}_c}(y_u)})}{\sum_{u=1}^S (\delta_{\tilde{A}_c}(y_u) + \overline{\delta_{\tilde{A}_c}(y_u)})} \quad (3.11)$$

6. Repeat steps 4-5 for all classes in C_i^{Fh} .
7. Generate the fuzzy ranks among the K classes, by taking the complement of μ_c^{Fh} as $R_c^{Fh} = 1 - \mu_c^{Fh}$, $c = 1, 2, \dots, K$. The value of R_c^{Fh} lies in $[0, 1]$, and the lowest value is said to be the top-ranked class.
8. Set $h = h + 1$ and repeat steps 2-7 until $h \leq H$.
9. Perform interval type-2 fuzzy rank-level fusion to generate final fuzzy ranks.

9.1 It may be noted that any two sets C_i^{Fh} and $C_i^{Fp} \mid h \neq p$ may be different as they result from two different feature vectors. Compute the fuzzy rank sum, χ_c , for all candidate classes and their complemented confidence factor mean, η_c , as follows:

$$\chi_c = \sum_{h=1}^H \begin{cases} R_c^{Fh} \mid c \in C_i^{Fh} \\ R_c^{Fh} = P_c^{Fh}, \text{Otherwise} \end{cases}, \forall c \quad (3.12)$$

$$\eta_c = 1 - \frac{1}{H} \sum_{h=1}^H \begin{cases} \omega_{i,c}^{Fh} \mid c \in C_i^{Fh} \\ \omega_{i,c}^{Fh} = Q_c^{Fh}, \text{Otherwise} \end{cases}, \forall c \quad (3.13)$$

where P_c^{Fh} and Q_c^{Fh} are the penalties that are imposed on class c if it does not come into the top K ranked classes. The value of P_c^{Fh} is set to 0.606 by

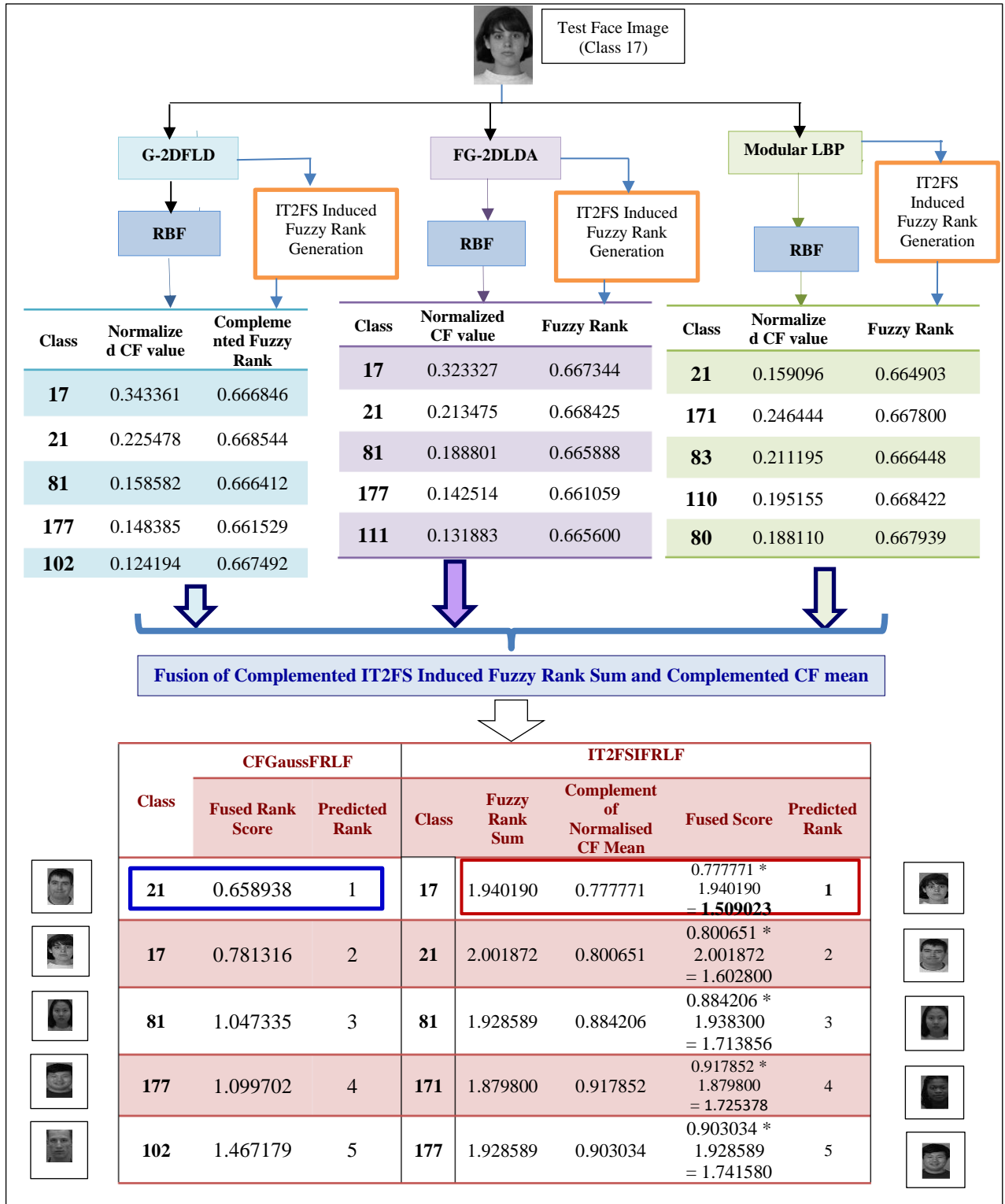


Fig. 3.3. Flow diagram describing the efficiency of IT2FSIFRLF

substituting $x_i^{Fh} = 1.0$ and $\theta_{u,c}^{Fh} = 0.0$ in Eq. (3.7), whereas Q_c^{Fh} value is taken as 0.0.

9.2 Perform the fusion of χ_c and η_c for all candidate classes to generate the final fuzzy rank \mathfrak{J}_c using the following equation. It may be noted that the class, having the least value of \mathfrak{J}_c , is said to be the top-ranked class.

$$\mathfrak{J}_c = \chi_c \times \eta_c, \forall c \quad (3.14)$$

10. Assign the class of the test image x_i as the class with the least value of \mathfrak{J}_c among all the candidate classes and is defined as follows.

$$\text{Class}(x_i) = \arg \min_c \{\mathfrak{J}_c\} \quad (3.15)$$

A more prominent and illustrative flow diagram describing the efficiency of our method is presented in Fig. 3.3 using a test image of class 17 from the FERET database.

3.4 EXPERIMENTAL RESULTS

In this Section, we present the experimental results on five publicly available face databases: AT&T [176], FERET [177, 178], UMIST [179], and Faces94 [180]. Each of these databases addresses different challenges for FR. In particular, the AR database considers these challenges by incorporating variations in facial illumination conditions, expression, and occlusion. The AT&T database includes variations in pose and rotation. The FERET database contains facial images with variations in expression, pose, and illumination. In contrast, the UMIST database provides facial images with large variations in pose. Finally, the Faces94 database not only provides a large number of face images but also provides considerable variations in expressions along with minor variations in head movement and face position. We extracted three sets of feature vectors from each face image using three holistic FE techniques: G-2DFLD [62], FG-2DLD [63], and modular LBP [48]. We employed the RBFNN [203, 230, 231] as the classifier. To demonstrate the efficacy and superiority of the IT2FSIFRLF method, we compared its performance with some of the recent state-of-the-art

fusion methods such as the (i) Highest rank method [189], (ii) Modified highest rank method [189], (iii) Borda count method [189], (iv) logistic regression method [190], (v) nonlinear weighted rank method [191], and (vi) CFGaussFRLF fuzzy rank-level fusion method [157]. Furthermore, we compared its performance with that of the G-2DFLD, FG-2DFLD, and modular LBP methods, which use only feature vectors for FR. The IT2FSIFRLF method was implemented in the C programming language, and the experiments were conducted on a desktop computer with Intel Core i5 (2.4 GHz) and 16 GB DDR3 RAM using Linux operating systems.

3.4.1 EXPERIMENTS ON THE AT&T FACE DATABASE

This database [176] contains facial images of 40 persons, including both males and females, and each one has 10 different images, resulting in 400 images. In this experiment, we randomly took r number of images per person to form the training set and the remaining $(10-r)$ images as the test set. The values of r were taken as 3, 4, 5, 6, and 7.

Table 3.2: Recognition rates of competitive methods on AT&T face database. The optimal values are indicated in bold.

Method	Avg. recognition rate \pm SD				
	$r=3$	$r=4$	$r=5$	$r=6$	$r=7$
G-2DFLD	92.82 \pm 2.67	95.94 \pm 1.21	97.73 \pm 0.87	98.66 \pm 0.96	98.42 \pm 1.11
FG-2DLDA	93.54 \pm 2.62	96.03 \pm 1.59	97.85 \pm 1.01	98.75 \pm 0.93	98.54 \pm 1.04
Modular LBP	93.43 \pm 2.21	95.85 \pm 1.71	97.83 \pm 0.92	98.56 \pm 0.86	98.50 \pm 1.07
Highest rank	93.50 \pm 2.50	96.00 \pm 1.51	97.85 \pm 1.01	98.66 \pm 0.80	98.33 \pm 0.81
Modified highest rank	93.62 \pm 2.33	96.06 \pm 1.42	98.08 \pm 0.88	98.75 \pm 0.93	98.62 \pm 0.68
Borda count	93.58 \pm 2.23	96.03 \pm 1.64	98.00 \pm 1.04	98.65 \pm 0.80	98.54 \pm 0.66
Logistic regression	93.66 \pm 2.22	96.09 \pm 1.63	98.25 \pm 1.01	98.91 \pm 0.67	98.92 \pm 0.48
Nonlinear weighted rank	94.46 \pm 1.75	96.49 \pm 1.38	98.5 \pm 0.63	99.17 \pm 0.56	99.04 \pm 0.62
CFGaussFRLF	94.68 \pm 1.59	96.55 \pm 1.43	98.58 \pm 0.63	99.35 \pm 0.59	99.13 \pm 0.69
IT2FSIFRLF	95.02 \pm 2.11	96.63 \pm 1.19	98.63 \pm 0.84	99.40 \pm 0.61	99.19 \pm 0.75

For each value of r , we constructed 20 different training and test sets to nullify any bias that may be inherited in the training and test sets during the formation and performed the experiments. The results are presented in Table 3.2. The results clearly show that the IT2FSIFRLF method yields better results for each value of r as compared to the other competitive methods.

Additionally, to check whether the performance gained by the IT2FSIFRLF method was statistically significant, we performed a paired t -test (one-tailed method) between the IT2FSIFRLF method and all other comparative methods. The analysis in terms of p -values is presented in a tabular form as shown in Table 3.3. In this context, the null hypothesis (H_0) is defined as: ‘the average recognition rate of the IT2FSIFRLF method is same as of the compared method’ against the alternative hypothesis (H_1) as: ‘the average recognition rate of the IT2FSIFRLF method is higher as compared to the related method’. For $r=3$ and $r=5$, the performance of the IT2FSIFRLF method is statistically significant than all the competitive methods at $p < 0.05$ except the nonlinear weighted rank and CFGaussFRLF methods.

Table 3.3: Statistical significance analysis between the competitive methods and the IT2FSIFRLF method on the AT&T face database.

Method	p -value				
	$r=3$	$r=4$	$r=5$	$r=6$	$r=7$
G-2DFLD	0.00000	0.00973	0.00165	0.00003	0.00101
FG-2DLDA	0.01569	0.06136	0.00038	0.00007	0.00083
Modular LBP	0.00441	0.02833	0.00473	0.00018	0.00056
Highest rank	0.00829	0.03880	0.00038	0.00259	0.00008
Modified highest rank	0.01137	0.06820	0.00181	0.00007	0.00228
Borda count	0.00889	0.05764	0.00179	0.00262	0.00004
Logistic regression	0.01103	0.07813	0.03253	0.00601	0.06627
Nonlinear weighted rank	0.10883	0.35868	0.14326	0.07768	0.24072
CFGaussFRLF	0.23872	0.42539	0.34696	0.34127	0.34318

For $r=4$, the IT2FSIFRLF method is statistically significant compared to all other methods excluding the logistic regression, nonlinear weighted rank, and CFGaussFRLF methods at $p < 0.05$. Similarly, its performance was significant compared to the G-2DFLD, FG-2DLDA,

Modular LBP, and Modified highest rank, logistic regression, and Borda count methods for $r=6$ and $r=7$ at $p \leq 0.05$. The trend of the results suggests that by using a smaller number of training samples and incorporating the IT2 FS, we were able to generate membership values more precisely.

3.4.2 EXPERIMENTS ON THE UMIST FACE DATABASE

This database [179] consists of 575 grayscale images of 20 individuals with varying numbers of images (19 – 48) per individual. These images covered a range of races, sexes, and appearances. Similar to the experiments on the AT & T database, the values of r were chosen as 4, 6, 8, and 10 to form different pairs of training and test sets. We present the average recognition rates of the different methods produced in 20 experimental runs for each value of r in Table 3.4. The results again show that the our IT2FSIFRLF method is superior to all other competitive methods. Similar to the previous databases, we also performed a statistical significance analysis, and the results are presented in Table 3.5. The analysis shows that the performance of the IT2FSIFRLF method is significant over the G-2DFLD, FG-2DLDA, Modular LBP, Highest rank, modified highest rank, and Borda count methods for $r=4, 6, 8$, and 10. For $r=10$, except for the CFGaussFRLF method, its performance was significant over all other methods. Its performance is significant over the logistic regression method for all r values, except $r=6$. Again, for $r=4$ and $r=6$, the performance of our fusion method showed no significant improvement over the nonlinear weighted rank and CFGaussFRLF methods. However, for $r=8$, the IT2FSIFRLF method performs better than all other methods, again demonstrating the efficacy of our fusion technique using an interval type-2 fuzzy set.

In summary, the statistical significance analyses between the IT2FSIFRLF method and other state-of-the-art competitive methods using AT&T and UMIST face databases revealed that the IT2FSIFRLF method is statistically significant over other methods, except for the CFGaussFRLF and nonlinear weighted rank methods. However, for the FERET face database, the IT2FSIFRLF method is statistically significant over all the competitive methods at $p < 0.05$.

Table 3.4: Recognition rates of competitive methods on UMIST face database. The optimal values are indicated in bold.

Method	Avg. recognition rate \pm SD			
	$r=4$	$r=6$	$r=8$	$r=10$
G-2DFLD	86.05 \pm 3.08	92.09 \pm 2.13	95.43 \pm 1.28	96.88 \pm 1.16
FG-2DLDA	86.30 \pm 3.00	92.42 \pm 1.49	95.58 \pm 1.17	97.05 \pm 1.09
Modular LBP	86.43 \pm 3.03	92.23 \pm 1.59	95.51 \pm 1.13	97.03 \pm 0.98
Highest rank	86.56 \pm 2.89	92.93 \pm 1.15	95.78 \pm 1.11	97.29 \pm 1.21
Modified highest rank	86.67 \pm 2.87	93.17 \pm 0.98	96.03 \pm 0.95	97.51 \pm 1.22
Borda count	86.84 \pm 2.77	93.10 \pm 0.88	96.13 \pm 0.73	97.73 \pm 1.15
Logistic regression	87.19 \pm 2.78	93.39 \pm 0.93	96.39 \pm 0.9	97.81 \pm 0.88
Nonlinear weighted rank	87.52 \pm 3.01	93.56 \pm 0.99	96.87 \pm 1.07	98.14 \pm 0.73
CFGaussFRLF	87.66 \pm 3.02	93.65 \pm 0.91	96.98 \pm 0.90	98.25 \pm 0.86
IT2FSIFRLF	87.79 \pm 2.62	93.83 \pm 1.99	97.39 \pm 0.95	98.36 \pm 0.70

Table 3.5: Statistical significance analysis of the comparative methods and IT2FSIFRLF method on the UMIST face database.

Method	p -value			
	$r=4$	$r=6$	$r=8$	$r=10$
G-2DFLD	0.00097	0.00002	0.00000	0.00003
FG-2DLDA	0.00140	0.00011	0.00000	0.00003
Modular LBP	0.00316	0.00011	0.00000	0.00001
Highest rank	0.00343	0.00086	0.00000	0.00053
Modified highest rank	0.01022	0.00993	0.00000	0.00153
Borda count	0.01261	0.04391	0.00000	0.00304
Logistic regression	0.05585	0.09817	0.00005	0.00004
Nonlinear weighted rank	0.27282	0.22235	0.02246	0.04644
CFGaussFRLF	0.38323	0.28785	0.02471	0.21881

Additionally, the average recognition rates of the IT2FSIFRLF method were higher than those of the other competitive methods for all three face databases. The above findings justify the efficiency of our method, presented here.

3.4.3 EXPERIMENTS ON THE FERET FACE DATABASE

This database was sponsored by the US Department of defense through the DARPA program [177, 178]. In this experiment, we used 1400 images of 200 individuals, where seven images of each individual had diverse variations in facial expression, illumination, and pose. Similar to the AT&T database, we experimented by randomly choosing r images per individual to form training and test sets.

Table 3.6: Recognition rates of competitive methods on FERET face database. The optimal values are indicated in bold.

Method	Avg. recognition rate \pm SD		
	$r=2$	$r=3$	$r=4$
G-2DFLD	49.14 \pm 0.92	59.02 \pm 1.22	65.51 \pm 2.69
FG-2DLDA	49.40 \pm 1.47	59.14 \pm 1.27	65.62 \pm 2.63
Modular LBP	49.04 \pm 1.46	59.40 \pm 1.33	65.53 \pm 2.51
Highest rank	49.32 \pm 1.12	60.08 \pm 1.21	65.77 \pm 2.58
Modified highest rank	49.49 \pm 0.92	60.22 \pm 0.65	65.86 \pm 2.04
Borda count	50.12 \pm 1.02	60.44 \pm 0.63	66.18 \pm 1.79
Logistic regression	50.36 \pm 0.80	60.61 \pm 0.64	66.30 \pm 1.68
Nonlinear weighted rank	50.80 \pm 0.62	60.78 \pm 0.50	66.58 \pm 1.48
CFGaussFRLF	52.89 \pm 1.09	64.38 \pm 1.65	69.58 \pm 2.16
IT2FSIFRLF	53.57 \pm 1.05	65.23 \pm 1.88	70.15 \pm 2.21

Table 3.7: Statistical significance analysis of the competitive methods and IT2FSIFRLF method on the FERET face database.

Method	p -value		
	$r=2$	$r=3$	$r=4$
G-2DFLD	0.00000	0.00000	0.00000
FG-2DLDA	0.00007	0.00000	0.00000
Modular LBP	0.00001	0.00000	0.00000
Highest rank	0.00000	0.00001	0.00000
Modified highest rank	0.00000	0.00001	0.00000
Borda count	0.00001	0.00001	0.00000
Logistic regression	0.00007	0.00002	0.00000
Nonlinear weighted rank	0.00009	0.00005	0.00025
CFGaussFRLF	0.00001	0.00381	0.04227

Here, the values of r were chosen as 2, 3, and 4. For each value of r , we performed 10 experiments and presented the average recognition rates in Table 3.6. The results show that the IT2FSIFRLF method is superior to all competitive methods in all experimental strategies. Furthermore, the statistical significance analysis, which is presented in Table 3.7, also suggests that the performance of the IT2FSIFRLF method is significant over all other methods at $p < 0.05$ for $r=2, 3$, and 4.

As mentioned in step (3) of the algorithm, which illustrates the IT2FSIFRLF method, K classes corresponding to the associated FV were used in the RLF. The experimental results, as indicated in the above tables, consider $K = 5$. To determine the best-suited value of K , we studied the recognition rates of our methodology while reiterating the execution of the experiment with varying values of K . For this purpose, 20 sets of training-test pairs of the AT & T database with $r = 4$ were taken, and the experiment was repeated with $K = 2, 3, 5, 4$, and 7. A plot of the recognition rate versus K is shown in Fig. 3.4.

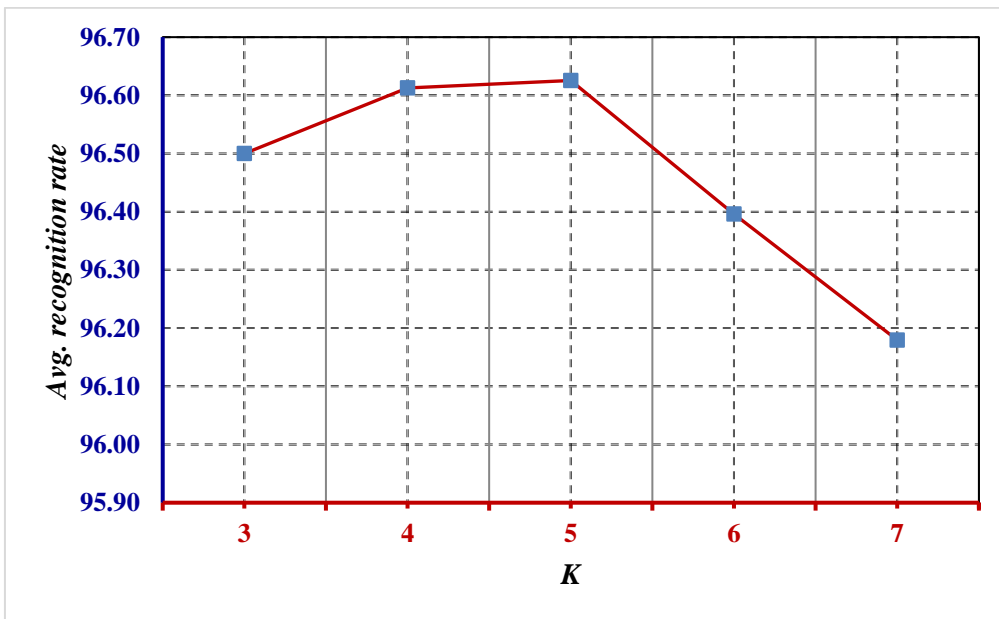


Fig. 3.4. The plot of average recognition rate versus K (AT & T database with $r = 4$)

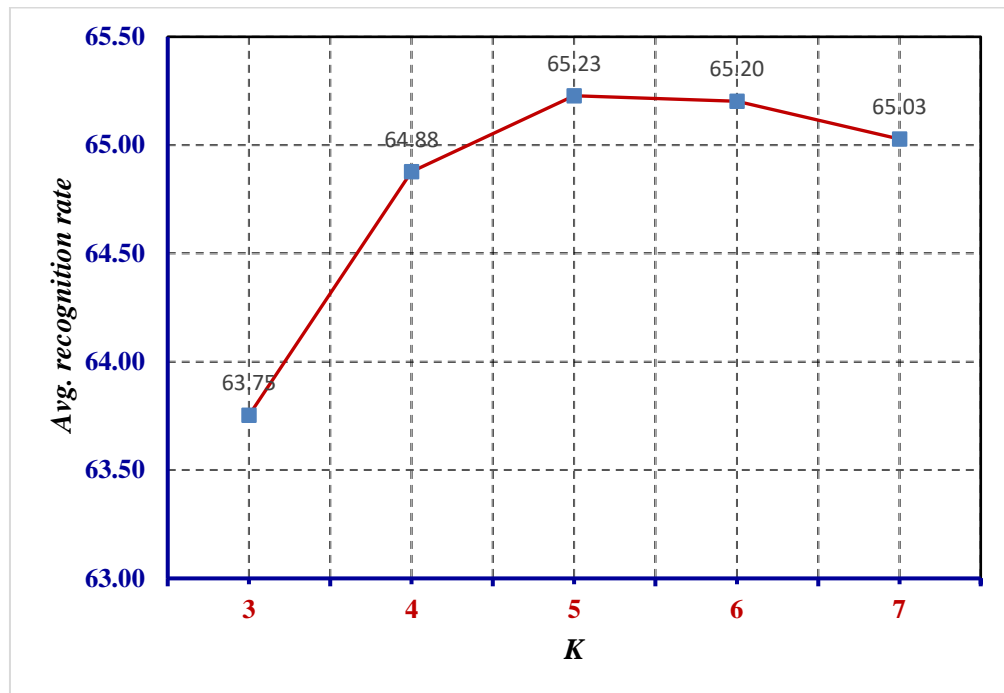


Fig. 3.5. The plot of average recognition rate versus K (FERET database with $r = 3$)

A similar exercise was replicated with the 10 training-test pairs of the FERET database with $r = 3$. A graph comparing the recognition rate with K is shown in Fig. 3.5. The recognition rate improved with higher K values, reaching a maximum at $K = 5$, and then declined as K continued to increase. From both plots, it is apparent that the methodology should achieve optimal recognition performance when the value of K is set to five for the experimentation.

3.4.4 AN ILLUSTRATIVE EXAMPLE

Let us consider an example to illustrate how the IT2 FS is applied to derive the fuzzy rank, and the fusion is performed in a step-by-step manner. To demonstrate the method, we took a test sample (image 70 of class 17) of the experiment with $r=3$ from the FERET face database. The set of top $K=5$ classes and the normalized outputs from the classifier (confidence factors or CF) corresponding to the three FE techniques are presented in Table 3.8.

Table 3.8: Set of top K classes and normalized confidence factors corresponding to the three feature extraction techniques.

G2-DFLD		FG-2DFLD		Modular LBP	
Class	Normalized CF	Class	Normalized CF	Class	Normalized CF
17	0.343361	17	0.323327	171	0.246444
21	0.225478	21	0.213475	83	0.211195
81	0.158582	81	0.188801	110	0.195155
177	0.148385	177	0.142514	80	0.188110
102	0.124194	111	0.131883	21	0.159096

(a) (b) (c)

Table 3.9: Class-wise normalized degrees of conformity, LMFs, UMFs, and crisp defuzzified values for the top K classes corresponding to the G-2DFLD method.

G2DFLD					
Class	Normalized degrees of conformity	LMF	UMF	Defuzzified value	Complemented Fuzzy Rank
17	0.340290	0.455010	0.461385	0.333154	0.666846
	0.326660	0.484357	0.490865		
	0.333050	0.998411	0.998976		
21	0.328646	0.751647	0.758726	0.331456	0.668544
	0.342173	0.365078	0.371590		
	0.329181	0.798831	0.805494		
81	0.323682	0.453496	0.462335	0.333588	0.666412
	0.342584	0.483431	0.492459		
	0.333734	0.998224	0.999025		
177	0.345464	0.766412	0.785233	0.338471	0.661529
	0.345283	0.772323	0.791003		
	0.309253	0.359081	0.376798		
102	0.331175	0.737863	0.741055	0.332508	0.667492
	0.331570	0.816036	0.818919		
	0.337256	0.367673	0.370569		

The degrees of conformity were measured from the feature vector of the test image and all training samples for each of the top K classes using Eq. (7). An interval type-2 fuzzy set is

formed for each of the top K classes with normalized degrees of conformity as explained in step 4.3 of the algorithm. The LMFs, UMFs, and crisp defuzzified values (according to step 5 of the algorithm) are listed in Tables 3.9, 3.10, and 3.11.

Table 3.10: Classwise normalized degrees of conformity, LMFs, UMFs, and crisp defuzzified values for the top K classes corresponding to the FG-2DFLD method.

FG2DFLD					
Class	Euclidian Distance	LMF	UMF	Defuzzified value	Complemented Fuzzy Rank
17	0.341756	0.426029	0.433266	0.332656	0.667344
	0.325996	0.522810	0.530539		
	0.332248	0.984990	0.987130		
21	0.329942	0.864834	0.870720	0.331575	0.668425
	0.342253	0.371521	0.378208		
	0.327804	0.682132	0.689717		
81	0.321229	0.432565	0.443164	0.334112	0.665888
	0.344199	0.508452	0.519622		
	0.334572	0.990115	0.992571		
177	0.346750	0.760446	0.781127	0.338941	0.661059
	0.346214	0.776521	0.796785		
	0.307037	0.358319	0.377669		
111	0.335040	0.985215	0.988585	0.334400	0.665600
	0.345296	0.518477	0.531031		
	0.319664	0.424964	0.436743		

Table 3.11: Classwise normalized degrees of conformity, LMFs, UMFs, and crisp defuzzified values for top K classes corresponding to Modular LBP.

Modular LBP					
Class	Euclidian Distance	LMF	UMF	Defuzzified value	Complemented Fuzzy Rank
21	0.336800	0.854891	0.860838	0.335097	0.664903
	0.338655	0.693026	0.700441		
	0.324545	0.370032	0.376594		

	0.331516	0.983477	0.987058		
171	0.347086	0.422762	0.434551	0.332200	0.667800
	0.321398	0.522012	0.534623		
	0.340883	0.485070	0.492450		
83	0.333677	0.998151	0.998838	0.333552	0.666448
	0.325439	0.453602	0.460820		
	0.330485	0.973911	0.979474		
110	0.350655	0.411951	0.426473	0.331578	0.668422
	0.318859	0.537075	0.552850		
	0.331121	0.926707	0.930815		
80	0.341221	0.387708	0.393872	0.332061	0.667939
	0.327658	0.611278	0.618257		

Taking class 17 from Table 13, the calculation of the defuzzified value is shown below as per step 5 of the algorithm using Eq. (3.11).

$$\frac{0.340290*(0.455010+0.461385)+0.326660*(0.484357+0.490865)+0.333050*(0.998411+0.998976)}{(0.455010+0.461385)+(0.484357+0.490865)+(0.998411+0.998976)} = 0.333154$$

According to steps 9.1 & 9.2 of the algorithm, the calculation of the complemented CF sum, fuzzy rank sum, and fused score using Eq. (3.12) & (3.13) are tabulated in Table 3.12.

Table 3.12: Calculation of complemented CF sum and complemented fuzzy rank sum.

Class	Complemented CF Sum=1.0 - CF Sum/3	Complemented Fuzzy rank sum	Fused Score	Rank
17	$1.0 - (0.343361 + 0.323327 + 0.0)/3 = 0.777771$	$0.666846 + 0.667344 + 0.606 = 1.940190$	$0.777771 * 1.940190 = 1.509023$	1
21	$1.0 - (0.225478 + 0.213475 + 0.159096)/3 = 0.800651$	$0.668544 + 0.668425 + 0.664903 = 2.001872$	$0.800651 * 2.001872 = 1.602800$	2
81	$1.0 - (0.158582 + 0.188801 + 0.0)/3 = 0.884206$	$0.666412 + 0.665888 + 0.606 = 1.938300$	$0.884206 * 1.938300 = 1.713856$	3

177	$1.0 - (0.148385 + 0.142514 + 0.0)/3$ $= 0.903034$	$0.661529 + 0.661059 + 0.606$ $= 1.928589$	$0.903034 * 1.928589$ $= 1.741580$	5
102	$1.0 - (0.124194 + 0.0 + 0.0)/3$ $= 0.958602$	$0.667492 + 0.606 + 0.606$ $= 1.879492$	$0.958602 * 1.879492$ $= 1.801685$	10
171	$1.0 - (0.246444)/3$ $= 0.917852$	$0.667800 + 0.606 + 0.606$ $= 1.879800$	$0.917852 * 1.879800$ $= 1.725378$	4
83	$1.0 - (0.211195 + 0.0 + 0.0)/3$ $= 0.929602$	$0.666448 + 0.606 + 0.606$ $= 1.878448$	$0.929602 * 1.878448$ $= 1.746209$	6
110	$1.0 - (0.195155 + 0.0 + 0.0)/3$ $= 0.934948$	$0.668422 + 0.606 + 0.606$ $= 1.880422$	$0.934948 * 1.880422$ $= 1.758098$	7
80	$1.0 - (0.188110 + 0.0 + 0.0)/3$ $= 0.937297$	$0.667939 + 0.606 + 0.606$ $= 1.879939$	$0.937297 * 1.879939$ $= 1.762060$	8
111	$1.0 - (0.131883 + 0.0 + 0.0)/3$ $= 0.956039$	$0.665600 + 0.606 + 0.606$ $= 1.877600$	$0.956039 * 1.877600$ $= 1.795059$	9

The final ranks for image 70 obtained by the CFGaussFRLF and IT2FSIFRLF methods are presented in Table 3.13. The results show that our method correctly recognized the image; whereas the CFGaussFRLF method misclassified the image as class 21.

Table 3.13: Final ranks of image 70 by CFGaussFRLF and IT2FSIFRLF methods with $r = 3$.

CFGaussFRLF			IT2FSIFRLF		
Class	Fused score	Rank	Class	Fused Score	Rank
21	0.658938	1	17	1.509023	1
17	0.781316	2	21	1.602800	2
81	1.047335	3	81	1.713856	3
177	1.099702	4	171	1.725378	4
102	1.467179	5	177	1.741580	5

3.5 CONCLUSION

In this chapter, an interval type-2 fuzzy set-induced fuzzy rank-level fusion technique based on multi-feature vectors is suggested for FR. It utilizes the outputs of an RBF neural network as confidence factors, which are subsequently used to generate interval type-2 fuzzy membership values. The mitigation of intra-class variabilities and the maximization of inter-class dissimilarities convolve the complemented confidence factors and interval type-2 fuzzy set-induced fuzzy ranks, utilizing weighted fusion to produce the final fuzzy ranks. In particular, by excluding the images of other classes while generating interval type-2 secondary membership functions, the algorithm not only reduces the influence of intra-class variabilities but also maximizes inter-class variabilities. Thus, the algorithm can generate appropriate fuzzy membership values using interval type-2 fuzzy sets. It generates the final fuzzy ranks by fusing the sum of the interval type-2 fuzzy set-induced fuzzy ranks and the complement of the average confidence factor. Based on the final fuzzy ranks, a test image is classified and recognized. The experimental results show that our developed method is superior to all competitive methods. The performance analysis suggests that the improved results are statistically significant compared to most of the comparative methods.

This algorithm utilizes the confidence factors of a single classifier. Utilization of the confidence factors of multiple classifiers and fusing with those of the interval type-2 fuzzy membership values could lead to superior results. Designing a fusion method requires more research. Another important observation from the experiments described in the previous and present chapters is that our rank-level fusion approaches using different forms of type-2 fuzzy sets to derive the fuzzy rank improved recognition precision. As cited in [232], a description with higher levels of fuzziness implies that the system is more adept at logically handling uncertainty. Therefore, employing a type-3 fuzzy set to derive the fuzzy rank is likely to improve recognition success, particularly in challenging or uncertain environments, owing to within-class variation and cross-class resemblances.

CHAPTER 4

INTERVAL TYPE-3 FUZZY SET INDUCED FUZZY RANK-LEVEL FUSION FOR FACE RECOGNITION

4.1 INTRODUCTION

As the technology for sensing and imaging continues to evolve, the use of FR has become increasingly prevalent in various social and commercial applications. Several authors from diverse domains came up with different methods for recognizing faces in unconstrained environments due to differential illumination [233], varied facial posture [234], ageing [235], occlusion [236], etc. It is onerous to assort these methods in a definitive manner. However, approaches to FR can be categorized as classical or modern [5]. The classical approach employs global features, local features, or both (hybrid). Global facial features are projected into a lower-dimensional subspace using linear (e.g. PCA, ICA, LDA, etc.) or non-linear methods (e.g. kernel-PCA, kernel-LDA, etc.) [38-40]. In [12], classical approaches were summarized with criticism with respect to advantages, performance, challenges managed, and limitations. Local feature-based approaches distinguish different distinct regions or keypoints of a face based on their geometric aspects using statistical patterns or graph-matching algorithms. Well-known local methods include HOG, SIFT, EGBM, SURF, and LBP. [52]. Hybrid FR models that integrate global and local approaches have also been documented in

literature [54]. FR, using modern approaches, employs deep learning, fuzzy set theory, and dictionary learning. Deep learning has become more widespread because of the advancement of powerful computing systems, and several recent approaches to FR using deep learning have achieved excellent results [60]. However, FR accuracy degrades with deep neural networks when images are captured at low pixel densities or under poor lighting conditions [237]. The use of fuzzy set theories in FR has showcased improved performance in recent research studies. A fuzzy-based 2DPCA method was presented by Li et al. [61] to address the sensitivity of traditional 2DPCA to illumination, pose, and expression in FR. In recent years, researchers from diverse disciplines have integrated type-2 fuzzy sets into their studies to contend with the inherent uncertainty. Du et al. [64] brought attention to the importance of uncertainty in FR and proposed an interval type-2-based fuzzy LDA model to mitigate the uncertainty.

Research studies have evidenced that the fusion process always leads to elevated recognition performance [238]. Based on the current research trends in multibiometric fusion, fusion methods can be characterized as feature-based early fusion, decision-based late fusion, and hybrid fusion [239]. Early fusion takes raw biometric data (from multiple sensors) or feature data (from different FE techniques or modalities) into account. Then feature matching or classification is carried out on the fused data for the recognition. Extracting and integrating multiple biometric features leads to an improved recognition accuracy [133]. Research based on early fusion, which involves facial features as one of the biometric traits, is cited in [132, 214]. Applying feature-based early fusion approaches is constrained by the dimensionally augmented fused feature space, disparate feature data resulting from different sources, and complex classification processes [240]. Late fusion methods are applied after classification and include decision-level, score-level, or rank-level methods. Decision-based late fusion strategies typically yield a class label and are often exploited for biometric products in the marketplace. The use of this kind of fusion approach in person identification is limited by its rigidity and its performance is influenced by the competence of the classifiers [217]. Researchers have emphasized the importance of score-level fusion in their work [134, 151, 219]. Fusing match scores can mitigate the impact of diverse biometric data, facilitate the retention of every *modus operandi* for matching features, and provide straightforward

calculation. The commonly used strategies for this kind of fusion are mean-score, max-score, and others [241]. It is often necessary to normalize scores at this fusion level, and improper normalization may contribute to subpar accuracy [241]. Rank-level fusion can provide an effective solution in multibiometric identification systems because it can circumvent the challenges associated with normalization and incompatibility, which are more pronounced in score-level fusion [154]. Despite its competence, there has been less research on this level of fusion than on score-level fusion [153]. In this fusion approach, the classifier produces an order of matching entities by evaluating their resemblance with the probe. The ranks of the individual classifiers are unified into a final rank [153]. The widely used methods in rank-level fusion include logistic regression, Borda count, highest rank, and more [142]. Hybrid fusion strives to integrate both early and late fusion methods into a unified, effective framework that leverages their strengths. Such studies involve multiple biometric traits using multiple fusion approaches [242, 158].

Using rank-level fusion, Sing et al. [157] presented an approach in which the top-ordered classes with higher classifier outputs corresponding to a test image were set apart. Multiple sets of such classes were formed for a test image with different feature vectors derived by the three FE methods. Then, the fuzzy ranks of these classes were computed. The fuzzy ranks and classifier outputs were separately aggregated for each class. The class of a face image was determined from the fused score of the candidate classes, calculated with the fuzzy rank and respective classifier's output. A rank-level fusion strategy utilizing a genetic algorithm was suggested by Ahmad et al. [221]. Ghosh et al. [243] delineated a multi-feature rank-level fusion method, referred to as IT2FSIFRLF, in which an IT2 FS was formed with only intra-class training images. This fuzzy set is utilized to derive the fuzzy ranks for each of the classes with the leading classifier's confidence factor (CF) corresponding to the test face image. Distinct IT2 FS-based fuzzy ranks were created per class using multiple feature vectors (FVs). For each of these classes, multiple fuzzy ranks and CFs were combined and a final score was produced by fusing the summated fuzzy ranks with the respective summated CFs.

Rickard et al. [38] put forward the general conceptual framework for type- n fuzzy sets and propounded that the unit interval serves as the domain wherein type- n fuzzy membership

functions assume values from type- $(n-1)$ fuzzy sets. Mohammedzadeh et al. [244] first described the design of an interval type-3 fuzzy set (IT3 FS) where all tertiary grades were set to unity and its secondary membership function was formulated with the IT2 FS. Also, the FOU is extended to be coupled with both the membership functions, primary and secondary, which model uncertainties in a more comprehensive manner [245]. To the best of our knowledge, no effort has been made to utilize IT3 FS for FR. However, very recently, IT3 FS has been used in other areas, including (i) control systems [246–250], prediction [251], [252], and time series [253]. These studies have indicated that the IT3 FS is more competent in addressing uncertainty as well as noise, thereby providing better modeling capabilities in practical real-world systems compared to its type-2 counterparts.

Motivated by recent research on the application of the interval type-3 fuzzy system, the present study exploited the IT3 FS to produce fuzzy ranks that are in turn utilized in rank-level fusion with multiple feature vectors for FR. This study also explores a comparison among T1 FS, IT2 FS, and IT3 FS in effectively mitigating uncertainty in the realm of facial recognition.

The subsequent sections present the remaining content of the chapter. Section 4.2 provides foundational knowledge of interval type-3 fuzzy sets. Section 4.3 describes our rank-level fusion approach for FR utilizing multiple feature vectors and IT3 FS. Section 4.4 confers an analysis of the experimental results to substantiate the effectiveness of the suggested approach. Section 4.5 concludes the chapter.

4.2 A CONCISE OVERVIEW OF INTERVAL TYPE-3 FUZZY SETS

The theoretical notions of the GT2 FS or T2 FS have been expounded in section 2.2.2 of chapter 2, and foundational theories related to IT2 FS have been delineated in section 3.2 of chapter 3. This section elucidates the fundamental portions of these concepts before introducing the IT3 FS for explicit perception.

A (general) type-2 fuzzy set (GT2 FS or T2 FS) is characterized by a MF, $\varphi_{\tilde{S}}(x, a)$, which is intrinsically fuzzy. Such a GT2 FS, \tilde{S} , is formally expressed in terms of primary variable x and secondary variable a from the universes X and A , respectively as follows.

$$\tilde{S} = \{((x, a), \varphi_{\tilde{S}}(x, a)) | x \in X, a \in \beta_x \subseteq [0,1], 0 \leq \varphi_{\tilde{S}}(x, a) \leq 1\} \quad (4.1)$$

The primary membership function (PMF), denoted by β_x , defines the domain for the secondary membership function (SMF). The T1 FS characterizes the SMF, as shown in (4.2).

$$\varphi_{\tilde{S}(x)}(a) = \{(a, \gamma_x(a)) | \gamma_x(a) \in A \quad \forall a \in \beta_x \subseteq [0,1]\} \quad (4.2)$$

where $\gamma_x(a)$ is termed as the secondary membership grade of x and $\gamma_x(b) = \varphi_{\tilde{S}}(x, b) \quad \forall x \in X$ [254]. The SMF, $\varphi_{\tilde{S}(x)}(a)$, is also referred to as the vertical slice of $\varphi_{\tilde{S}}(x, a)$ in IT2 FS which can be formally delineated as given below.

$$\tilde{S} = \{((x, a), \varphi_{\tilde{S}}(x, a) = 1) | x \in X, a \in \beta_x \subseteq [0,1]\} \quad (4.3)$$

where $\varphi_{\tilde{S}}(x, a)$ is the interval type-2 membership function (IT2 MF) and $\varphi_{\tilde{S}}(x, a) = 1$ follows from the fact that $\gamma_x(a) = 1, \forall a \in \beta_x \subseteq [0,1] \quad \forall x \in X$. The expression for the SMF of an IT2 FS is outlined in Eq. (4.4).

$$\varphi_{\tilde{S}(x)}(a) = \{(a, \gamma_x(a) = 1) | a \in [\underline{\gamma}_x(a), \overline{\gamma}_x(a)] \subseteq [0,1]\} \quad (4.4)$$

For both types of T2 FSs, the FOU is regarded as the union of all primary membership functions and embodies the uncertainty intrinsic to a T2 FS [255]. The FOU is confined by two T1 FSs known as the LMF and UMF. An IT2 FS is completely characterized by its FOU because its secondary grades provide no substantial information. Accordingly, \tilde{S} can also be represented using its FOU, as outlined in Eq. (4.5).

$$\tilde{S} = 1/FOU(\tilde{S}) = 1/[\underline{\varphi}_{\tilde{S}}(x), \overline{\varphi}_{\tilde{S}}(x)] \quad (4.5)$$

The design of the IT3 FS follows the GT2 FS framework and leverages the mathematical principles of the IT2 FS. An IT3 FS, \mathbb{S} , is defined formally using a bivariate function $\tilde{\varphi}_{\mathbb{S}}(x, a)$.

$$\mathbb{S} = \{(x, a, \tilde{\varphi}_{\mathbb{S}}(x, a)) | x \in X, a \in A \subseteq [0,1], 0 \leq \tilde{\varphi}_{\mathbb{S}}(x, a) \leq 1\} \quad (4.6)$$

The interval type-3 membership function (IT3 MF), $\tilde{\varphi}_{\mathbb{S}}(x, a)$, is an interval confined by lower and upper T2 FSs with respect to the T2 MFs, $\underline{\varphi}_{\mathbb{S}}(x, a)$ and $\overline{\varphi}_{\mathbb{S}}(x, a)$. That is, $\tilde{\varphi}_{\mathbb{S}}(x, a) \in [\underline{\varphi}_{\mathbb{S}}(x, a), \overline{\varphi}_{\mathbb{S}}(x, a)]$ where $\underline{\varphi}_{\mathbb{S}}(x, a) \leq \overline{\varphi}_{\mathbb{S}}(x, a)$ [246]. Similar to the vertical slice interpretation of T2 FS, IT3 MF can be expressed as a union of vertical slices (cuts) in x .

$$\mathbb{S} = \left\{ \left(x, \tilde{\varphi}_{\mathbb{S}(x)}(a) \right) \mid x \in X, a \in A \subseteq [0, 1] \right\} \quad (4.7)$$

where $\tilde{\varphi}_{\mathbb{S}(x)}(a)$ is the SMF of the IT3 FS, and is an IT2 MF, which is detailed in Eq. (4.8).

$$\begin{aligned} \tilde{\varphi}_{\mathbb{S}(x)}(a) &= \left\{ (a, \tilde{\gamma}_x(a)) \mid x \in X, a \in \tilde{\beta}_x \subseteq [0, 1] \right\} \\ &= \left\{ (a, \tilde{\gamma}_x(a) = 1) \mid x \in X, a \in [\underline{\gamma}_x(a), \overline{\gamma}_x(a)] \subseteq [0, 1] \right\} \end{aligned} \quad (4.8)$$

The PMF of the IT2 MF in x , denoted by $\tilde{\beta}_x$, is derived using the respective PMFs of lower and upper T2 FSs. This serves as a domain for secondary IT2 MF (SMF of $\varphi_{\mathbb{S}(x)}(a)$). At a specific x , say $x=x'$, an embedded secondary IT2 MF, $\tilde{\gamma}_{x'}(a)$, represents an embedded vertical slice. More accurately, $\tilde{\gamma}_{x'}(a) \in [\underline{\gamma}_{x'}(a), \overline{\gamma}_{x'}(a)]$ implies $\tilde{\gamma}_{x'}(a) \in 1/[\underline{\varphi}_{\mathbb{S}}(x', a), \overline{\varphi}_{\mathbb{S}}(x', a)]$. The $\underline{\gamma}_{x'}(a)$ and $\overline{\gamma}_{x'}(a)$ correspond to the embedded secondary T1MFs of $\underline{\varphi}_{\mathbb{S}}(x, a)$ and $\overline{\varphi}_{\mathbb{S}}(x, a)$ respectively at x' .

The IT3 FS consists of FOUs inherent to both its MFs, which can address uncertainties more effectively than the IT2 FS. Vertical slices, characterized by secondary IT2MFs, $\tilde{\varphi}_{\mathbb{S}(x)}(a)$, contribute to the formation of the FOU for the IT3FS. The FOU of an IT3 FS converges to the FOU of a T2 FS when $\underline{\varphi}_{\mathbb{S}}(x, a) = \overline{\varphi}_{\mathbb{S}}(x, a)$. The union of the primary IT2MFs in the (x, a) plane constructs the domain of uncertainty (DOU). The three domains of IT3 MF are depicted in Fig. 4.1.

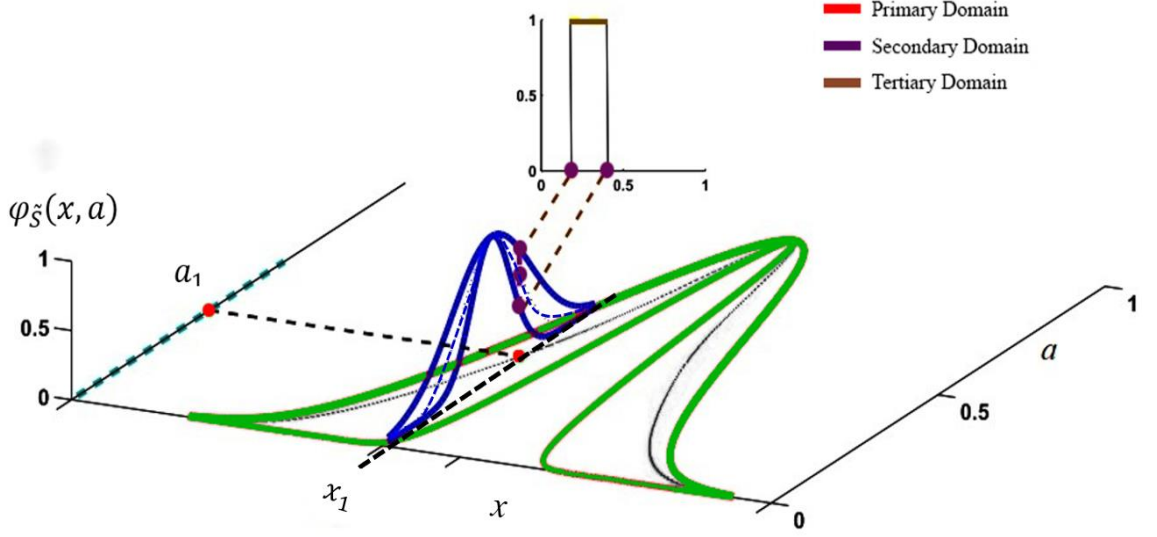


Fig. 4.1. Three domains of the interval type-3 membership function

Castillo et al. [256] suggested various formulations and parameterizations for the design of IT3MFs in their research article. One such recommendation is to characterize the IT3 MF using both primary and secondary variables that are scaled Gaussians. Consequently, its FOU is also Gaussian. To form the DOU, the UMF is parameterized with the standard deviation (σ) and mean (m), and the LMF is delineated with the parameters of lower scale (α) and lower lag (β). The equations below explicate the IT3 MF, $\tilde{\varphi}_{\mathbb{S}}(x, a)$.

$$\bar{a}(x) = e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} \quad (4.9)$$

$$\underline{a}(x) = \alpha \cdot e^{-\frac{1}{2}\left(\frac{x-m}{\sigma^*}\right)^2} \quad (4.10)$$

The upper and lower limits of the DOU are indicated by $\bar{a}(x)$ and $\underline{a}(x)$. σ^* is evaluated as $\sigma^* = \sigma \cdot \sqrt{\frac{\ln(\beta)}{\ln(e)}}$, e being machine epsilon. When β is 0, σ_a and σ_a^* will be the same i.e., $\sigma_a^* = \sigma_a$. As illustrated in (4.10), FOU, circumscribed by LMF, $\underline{\varphi}_{\mathbb{S}(x)}(a)$ and UMF $\overline{\varphi}_{\mathbb{S}(x)}(a)$, is expressed with vertical slices $\tilde{\varphi}_{\mathbb{S}(x)}(a)$. The UMF and LMF are defined in Eq. (4.11) and (4.12), respectively.

$$\overline{\varphi}_{\mathbb{S}(x)}(a) = e^{-\frac{1}{2}\left(\frac{(a-a(x))}{\sigma_a}\right)^2} \quad (4.11)$$

$$\underline{\varphi}_{\mathbb{S}(x)}(a) = \alpha \cdot e^{-\frac{1}{2}\left(\frac{(a-a(x))}{\sigma_a^*}\right)^2} \quad (4.12)$$

The σ_a^* is computed from σ_a as with DOU and also $\sigma_a^* = \sigma_a$ when $\beta=0$.

4.3 INTERVAL TYPE-3 FUZZY SET INDUCED FUZZY RANK-LEVEL FUSION FOR FACE RECOGNITION

4.3.1 METHODOLOGY

This section presents the interval type-3 fuzzy set induced rank-level fusion (IT3FSIRLF) method for FR utilizing multiple feature vectors (FVs) derived from a face image. These feature vectors were derived using G-2DFLD [62], FG-2DFLD [63], and modular LBP (mLBP) [48]. This allows for the extraction of complementary yet meaningful discriminating data from facial images, thereby reducing the challenge of cross-class resemblances. The FV representing the test image was fed into a classifier to produce the outputs, referred to as the confidence factors (CFs), of the potential classes. Utilizing all training FVs, the degrees of proximity of the training samples were computed for each of the top-ordered classes with higher classifier outputs. A subset of smaller degrees of proximity for each training sample of these classes was used to express the UMF and LMF of the scaled Gaussian FOU of the IT3 FS. The defuzzification of this IT3 FS results in a discrete value, which is considered the fuzzy rank of the respective class. For each class, the algorithm yielded multiple fuzzy ranks and CFs corresponding to multiple feature vectors. The final score for each class was generated in three steps. First, the fuzzy ranks were added together to obtain the combined fuzzy rank. Second, we complemented the arithmetic mean of the CFs. Finally, these two parameters were fused to obtain the final score of the class. It may be noted that a class may fall outside of the top-ordered classes with respect to the feature vector. In this case, a penalty is imposed on the class, counteracting the possibility of becoming an unlikely winner. The

test image pertains to the class with the lowest final score. Fig. 4.2 illustrates the suggested methodology in a schematic outline.

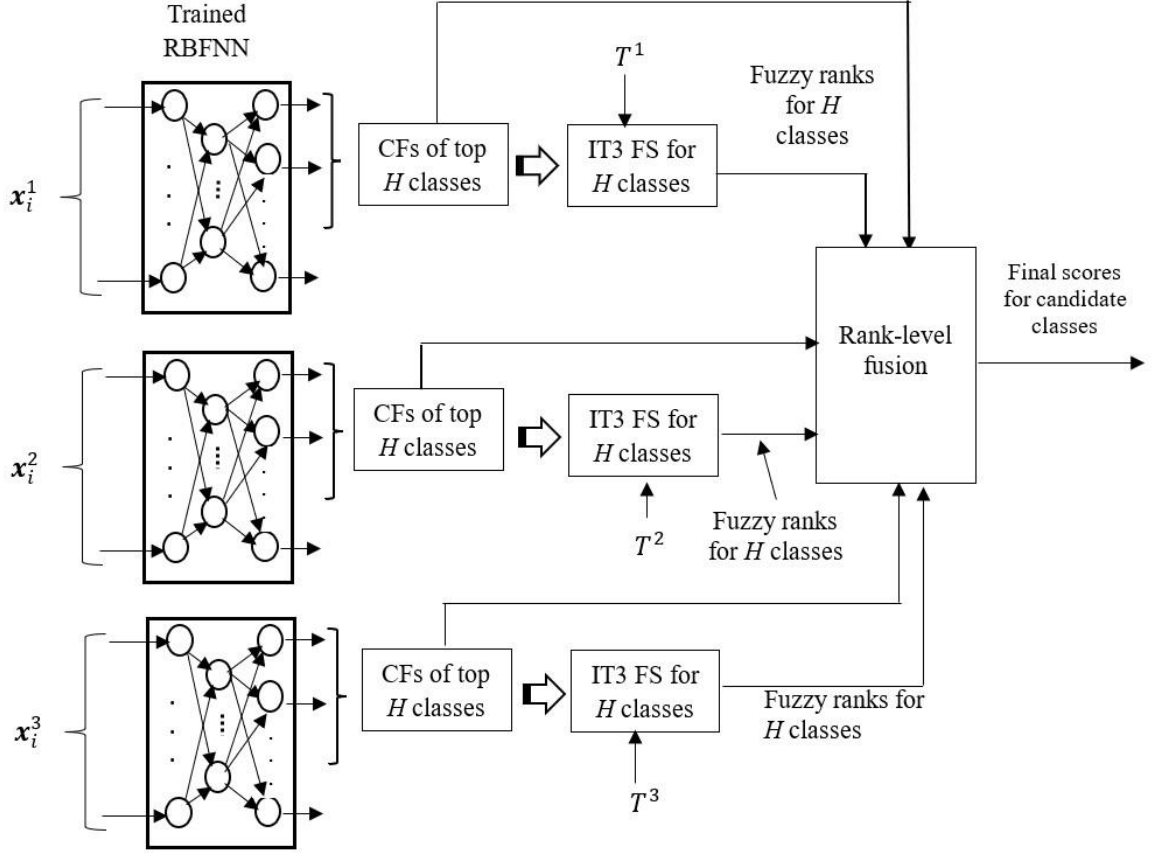


Fig. 4.2. Schematic description of the present approach.

4.3.2 MATHEMATICAL FORMULATION OF THE METHODOLOGY

Considering M and N as the total number of images in the training and corresponding test set and C as the total number of subjects (classes), let $x_{i,c}^p$ and $t_{j,c}^p$ represent the FVs of the i th image of class c in the test set, X^p , and j th image of class c in the training set, T^p , respectively. To create the FV, P different FE techniques are used. Since G2DFLD, FG2DFLD, and modular LBP methods are used in this study, $P = 3$. Let $w_{i,c}^p$ be the output (CFs) of the classifier when the feature vector for the test image, x_i , is fed into it. A_i^p is the set of the topmost H classes that are singled out according to the descending order of CFs. To

ascertain the class of the test image, \mathbf{x}_i , the method, illustrated here, comprises the following steps. In these steps, subscript i is omitted from the notation to simplify the equations.

Step 1: Normalize the confidence factors of all H classes using Eq. (4.13)

$$\sum w_c^p = 1 \mid c \in A^p \quad (4.13)$$

Step 2: Compute the degree of proximity of each training sample for class c with respect to all training FVs using Eq. (4.14).

$$\delta_{u,c}^p(v) = \|t_{u,c}^p - t_{j,c}^p\|, u = 1, 2, \dots, S; j = 1, 2 \dots M \text{ and } j \neq u \quad (4.14)$$

Here S represents the count of training samples for class c .

Step 3: Formulate an IT3 FS, \mathbb{D} , with k smaller degrees of proximities that are associated with each of the S training samples of class c . Assuming $x_{i,c}^p$ and $t_{u,c}^p$ to represent the primary variable y and the secondary variable g , respectively, the IT3 FS, \mathbb{D} is formed as defined in Eq. (4.15).

$$\mathbb{D} = \{(y, g, \tilde{\varphi}_{\mathbb{D}}(y, g)) \mid y \in A^p, g \in [0,1], 0 \leq \tilde{\varphi}_{\mathbb{D}}(y, g) \leq 1\} \quad (4.15)$$

The tertiary grades in an IT3 FS are unity. By definition, IT3 MF $\tilde{\varphi}_{\mathbb{D}}(y, g)$ is constrained by two T2 FSs, $\underline{\varphi}_{\mathbb{D}}(y, g)$ and $\overline{\varphi}_{\mathbb{D}}(y, g)$ and can be represented as $h \in [\underline{\varphi}_{\mathbb{D}}(y, g), \overline{\varphi}_{\mathbb{D}}(y, g)]$ where h is the tertiary variable of \mathbb{D} . Following the formulations and parameterizations to design an IT3 FS discussed in Section 4.2, its DOU is expressed with UMF and LMF as follows: The σ , m are the parameters for the UMF, and λ (scale) and δ (lag) are those for the LMF.

$$\begin{aligned} \overline{g}_{u|v}(y_c) &= \exp \left[-\frac{1}{2} \left(\frac{(\delta_{u,c}^p(v) - m)}{\sigma} \right)^2 \right] \text{ and} \\ \underline{g}_{u|v}(y_c) &= \lambda \cdot \exp \left[-\frac{1}{2} \left(\frac{(\delta_{u,c}^p(v) - m)}{\sigma} \right)^2 \right] \end{aligned} \quad (4.16)$$

where $u = 1, 2, \dots, S$ and $v = 1, 2, 3 \dots, k$. The scale is calculated as $\lambda = \left(\max_v \{\delta_{u,c}^p(v)\} - \min_v \{\delta_{u,c}^p(v)\} \right) / k$ and δ is taken as 0. The vertical slice with Gaussian IT2 MFs, $\varphi_{\mathbb{D}(y_c)}(g) = [\underline{\varphi}_{\mathbb{D}(y_c)}(g_u), \overline{\varphi}_{\mathbb{D}(y_c)}(g_u)]$ are evaluated by Eq. (4.17)

$$\begin{aligned}\bar{\varphi}_{\mathbb{D}(y_c)}(g_u) &= \exp \left[-\frac{1}{2} \left(\frac{(g_u - m_u)}{\sigma_u} \right)^2 \right] \text{ and} \\ \underline{\varphi}_{\mathbb{D}(y_c)}(g_u) &= \lambda \cdot \exp \left[-\frac{1}{2} \left(\frac{(g_u - m_u)}{\sigma_u} \right)^2 \right]\end{aligned}\quad (4.17)$$

where $g_u = \frac{\sum_{v=1}^k \delta_{u,c}^p(v) \times (\bar{g}_{u|v}(y_c) + \underline{g}_{u|v}(y_c))}{(\bar{g}_{u|v}(y_c) + \underline{g}_{u|v}(y_c))}$ and scale is estimated by $\lambda = \left(\max_u \{g_u\} - \min_u \{g_u\} \right) / S$ and δ is taken as 0. Defuzzification using the Nie-Tan algorithm [228] results in the crisp value, θ_c^p , for class c .

$$\theta_c^p = \frac{\sum_{u=1}^S g_u \times (\underline{\varphi}_{\mathbb{D}(y_c)}(g_u) + \bar{\varphi}_{\mathbb{D}(y_c)}(g_u))}{\sum_{u=1}^S (\underline{\varphi}_{\mathbb{D}(y_c)}(g_u) + \bar{\varphi}_{\mathbb{D}(y_c)}(g_u))} \quad (4.18)$$

Step 4: Repeat steps 1-3 for all classes $c \in A^p$.

Step 5: Compute fuzzy ranks for each class $c \in A^p$ by taking the complement of θ_c^p

$$\gamma_c^p = 1 - \theta_c^p, c = 1, 2, \dots, H \quad (4.19)$$

Step 6: Find the sum of fuzzy ranks resulting in a combined fuzzy rank, S_{γ_c} , for each class $c \in A^p$ by applying Eq. (4.20)

$$S_{\gamma_c} = \sum_{p=1}^P \begin{cases} \gamma_c^p & \text{if } c \in A^p \\ \varepsilon, & \text{otherwise} \end{cases} \quad (4.20)$$

where ε is penalty levied on class c for not belonging to any one of $A^p \forall p \in P$. The value of ε is computed by putting $x = 0.0$, $m = 1.0$ and $\sigma = 1.0$ in Gaussian density function.

Step 7: Find the complemented mean of confidence factors, S_{w_c} , for all classes $c \in A^p \forall p \in P$ by applying Eq. (4.21)

$$S_{w_c} = 1 - \frac{1}{P} \sum_{p=1}^P \begin{cases} w_c^p & \text{if } c \in A^p \\ w_c^p = 0, & \text{otherwise} \end{cases} \quad (4.21)$$

Step 8: Fuse S_{γ_c} with S_{w_c} to find the final fuzzy rank, F_c , for each class $c \in A^p$ using Eq. (4.22).

$$F_c = S_{\gamma_c} \times S_{w_c} \quad (4.22)$$

Step 9: The class with the least final fuzzy rank is regarded as the class of x_i .

$$class(x_i) = \arg \min_c F_c \quad (4.23)$$

4.4 EXPERIMENTAL RESULTS

To evaluate our IT3FSIRLF method, three public face databases were employed namely, FERET [177,178], UMIST [179], and Faces94 [180]. The implementation was performed using the Linux-based C language, and the experiment was performed on a desktop computer powered by an Intel Core i5 processor (2.4 GHz) with 16 GB DDR3 RAM. The effectiveness of the IT3FSIRLF approach was evaluated and compared with that of the CFGaussFRLF [157] and IT2FSIFRLF [243] methods. Similar experimental steps were followed for all these methods. These methods were chosen to explore the potential of IT3 FS to address uncertainty while making decisions in FR using rank-level fusion. When generating fuzzy ranks, the CFGaussFRLF used T1 FS whereas the IT2FSIFRLF method utilized IT2 FS. The IT3FSIRLF method employs IT3 FS. To evaluate the method, multiple training and test sets were created with varying numbers of training images per individual for each database.

4.4.1 EXPERIMENTS ON THE FERET FACE DATABASE

The present experiment involved 1400 images from the FERET database belonging to 200 persons, with 7 images for each individual. For this database, r images per person were set aside to create the training set and the remaining $(7 - r)$ images were set aside to form the test set. Ten such pairs for each r were prepared to appraise the method. While implementing these methods on each of the 20 sets, the recognition success rates were noted. A comparison of the IT3FSIRLF method with the other two methods in terms of the average recognition accuracy and standard deviation (SD) is presented in Table 4.1.

Reviewing the results indicates an improved recognition success of the IT3FSIRLF method for all r . In addition to weighing the performance, in terms of recognition accuracy, the

performance of the IT3FSIRLF method is contrasted with CFGaussFRLF and IT2FSIFRLF methods using the one-tailed paired t -test.

Table 4.1: Comparison of the average rate of recognition success using the FERET database.

Avg. recognition success rate \pm SD			
r	CFGaussFRLF	IT2FSIFRLF	IT3FSIRLF
2	52.89 \pm 1.09	53.57 \pm 1.05	53.87 \pm 0.85
3	64.38 \pm 1.65	65.23 \pm 1.88	65.41 \pm 1.90
4	69.58 \pm 2.15	70.15 \pm 2.21	70.74 \pm 2.61

The evaluation of the statistical significance of our method over the CFGaussFRLF and IT2FSIFRLF methods is presented in Table 4.2. The p -values denote that the augmented performance IT3FSIRLF is significant.

Table 4.2: Statistical significance analysis of IT3FSIRLF with the other two methods on the FERET database

Method			
r	p -value	CFGaussFRLF	IT2FSIFRLF
2		0.00070	0.04787
3		0.00125	0.04178
4		0.00302	0.05539

4.4.2 EXPERIMENTS ON THE UMIST FACE DATABASE

The UMIST database consisted of 575 images in total, featuring 20 distinct males and females. Each individual had a varying number of faces. For this database, twenty sets of training-test pairs were created with r and $(L - r)$ images respectively from each individual, where r signifies the count of images per person in the training set, and the value of r was taken as 4, 6, 8, or 10. The value of L differs among individuals and ranges from 19 to 48. The recognition rates were recorded while applying the methods to each of the 20 sets. Table 4.3 shows a comparative analysis of the recognition accuracy between these methods using the UMIST database. Standard deviations (SD) are also mentioned in the tables.

It is apparent from examining the results that the IT3FSIRLF method exhibited an ameliorated average recognition rate over the other two methods.

Table 4.3: Comparison of the average rate of recognition success using the UMIST database.

Avg. recognition success rate \pm SD			
r	CFGaussFRLF	IT2FSIFRLF	IT3FSIRLF
4	87.66 \pm 3.02	87.79 \pm 2.62	88.11 \pm 2.75
6	93.65 \pm 0.91	93.83 \pm 1.99	93.97 \pm 2.07
8	96.98 \pm 0.90	97.39 \pm 0.95	97.52 \pm 0.96
10	98.25 \pm 0.86	98.36 \pm 0.70	98.55 \pm 0.79

The statistical significance evaluation of the experiment using the UMIST database is furnished in Table 4.4. The table indicates that the suggested approach accomplishes substantially better significance compared to both CFGaussFRLF and IT2FSIFRLF methods at $p < 0.05$ for $r = 8$ and 10. For $r = 4$, the IT3FSIRLF method exhibited statistically significant performance when compared to the IT2FSIFRLF method.

Table 4.4: Statistical significance analysis of IT3FSIRLF with the other two methods on the UMIST database

Method			
r	p -value	CFGaussFRLF	IT2FSIFRLF
4		0.16299	0.02906
6		0.17459	0.08242
8		0.00874	0.04635
10		0.03308	0.00914

4.4.3 EXPERIMENTS ON THE FACES94 FACE DATABASE

The database comprises 3060 face pictures from 153 people. Of 153 individuals, 20 were female, 113 were male, and 20 were staff members. There are 20 images per individual. This experiment employed 3000 images from 150 people. For the Faces94 dataset, multiple training and test sets were formed for different values of r , implying a person-wise number of images were randomly selected for training. The respective test set was comprised of $(20 - r)$

images. Twenty sets were created, where r was set to 10, 12, 14, 15, and 17 images. Table 4.5 provides a comparative analysis of our method with the other two related methods in terms of recognition success rate. The results have made it evident that our method (IT3FSIRLF) surpassed the CFGaussFRLF and IT2FSIFRLF methods.

The results of the one-tailed paired t -test for the experiments, implemented with the images of the Faces94 database, are showcased in Table 4.6. The results revealed the significant performance of the IT3FSIRLF method over the other two methods.

Table 4.5: Comparison of the average rate of recognition success using the Faces94 database.

Avg. recognition success rate \pm SD			
r	CFGaussFRLF	IT2FSIFRLF	IT3FSIRLF
10	94.65 \pm 0.57	94.75 \pm 0.47	95.15 \pm 0.60
12	94.30 \pm 0.87	95.08 \pm 0.52	95.27 \pm 0.69
14	93.69 \pm 0.24	93.13 \pm 0.56	93.87 \pm 0.74
15	92.93 \pm 0.50	93.10 \pm 0.79	94.05 \pm 1.14
17	93.78 \pm 0.75	94.13 \pm 0.79	95.20 \pm 0.60

Table 4.6: Statistical significance analysis of IT3FSIRLF WITH the other two methods on the Faces94 database

Method			
r	p -value	CFGaussFRLF	IT2FSIFRLF
10		0.00861	0.01184
12		0.00069	0.18172
14		0.08233	0.01546
15		0.01053	0.01318
17		0.00113	0.01111

The evaluation of the average recognition rate as well as statistical significance justifies the improved competence of the method, as explained in this chapter. According to the presented findings, it has also been observed that generating fuzzy rank utilizing IT3 FS results in improved accuracy while recognizing faces with fusion at the rank level.

According to the algorithm of our methodology, as illustrated in Section 4.3.2, multiple FE techniques were utilized to generate different FVs. For each of the feature vectors, the top H classes with higher CFs were taken into account in the fusion process with IT3 FS-based fuzzy ranks. In the experiments, the results of which are furnished in this section, we considered $H = 5$. To decide the value of H , we analyzed the recognition rates for $H = 3, 4, 5, 6$ and 7 using 20 sets of training-test pairs from the UMIST database with $r = 8$. The same was followed with 10 sets of such pairs using images from the FERET database with $r = 3$. The graphs plotting the average recognition rate against H for the sets from both the UMIST and FERET databases are shown in Fig. 4.3 and Fig. 4.4. From these graphs, it can be seen that the average recognition rate increases with increasing H , reaches its highest at $H = 5$, and then diminishes gradually with increasing values of H .

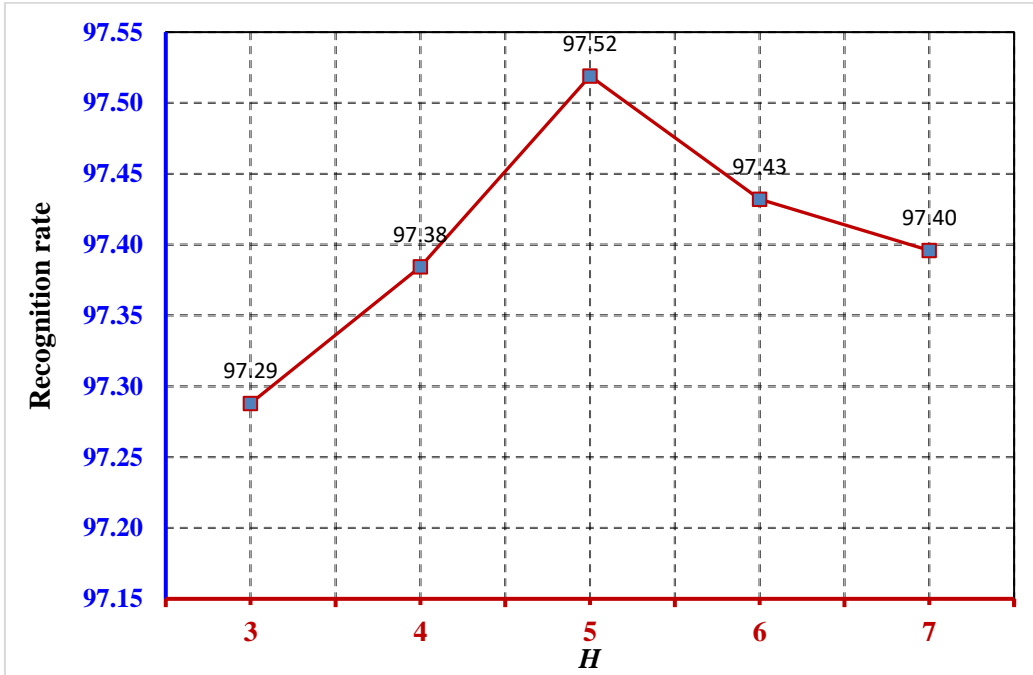


Fig. 4.3. Graph of average recognition rate versus H (UMIST database with $r = 8$)

As specified in step 3 of the algorithm, to express an IT3 FS for each of these top-ordered H classes, K smaller degrees of proximity are regarded for each of the training samples. To determine the value of K , we repeated the experiments using the same sets from the UMIST and FERET databases.

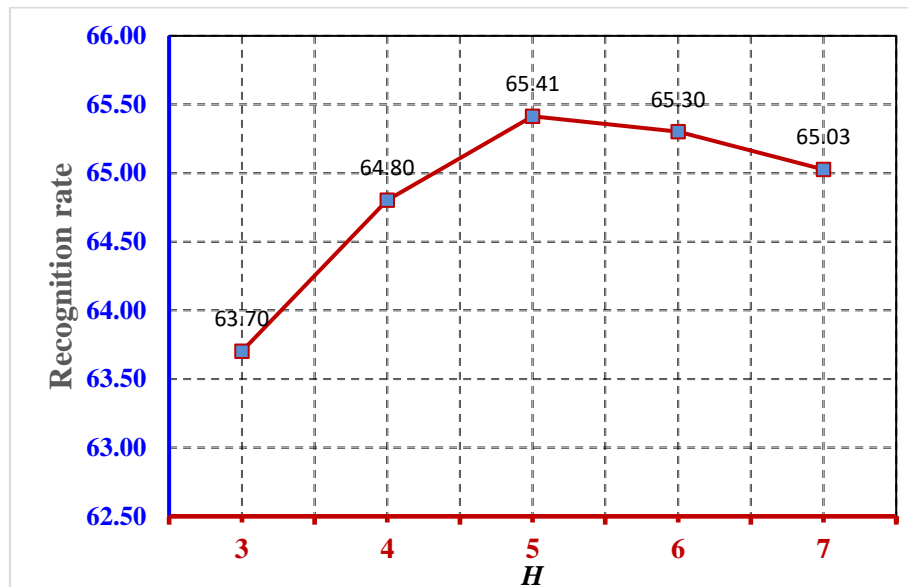


Fig. 4.4. Graph of average recognition rate versus H (FERET database with $r = 3$)

In these experimental runs, we kept H fixed at 5, but executed with $K = 2, 3, 4, \dots$, and 7. Fig. 4.5 and Fig. 4.6 depict the graphs for the average recognition rate versus K . It is apparent that with an increase in K , the average recognition rate rises, followed by a gradual decrease with further increments of K . These observations justify the values of H and K as 5 and 3 respectively for the experiments, carried out to evaluate the efficiency of our methodology.

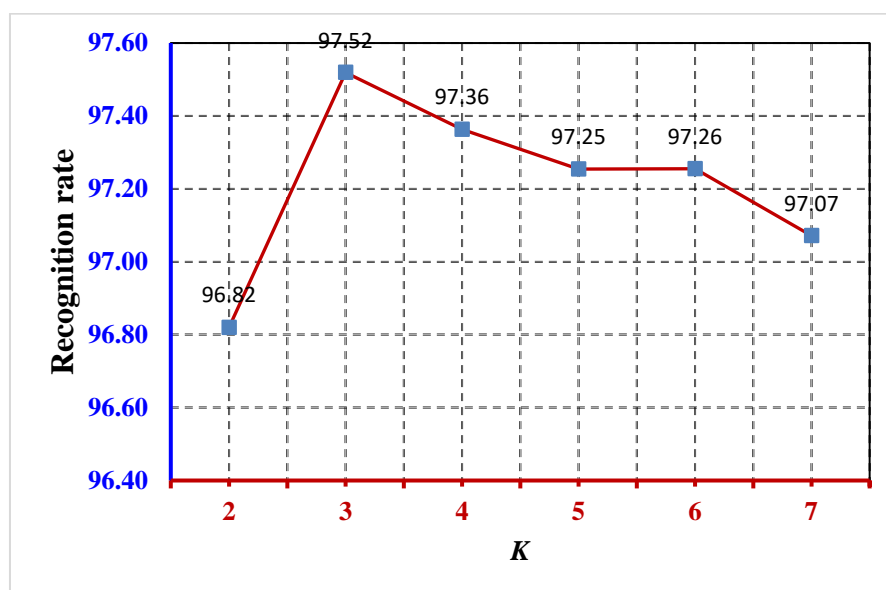


Fig. 4.5. Graph of average recognition rate versus K (UMIST database with $r = 8$)

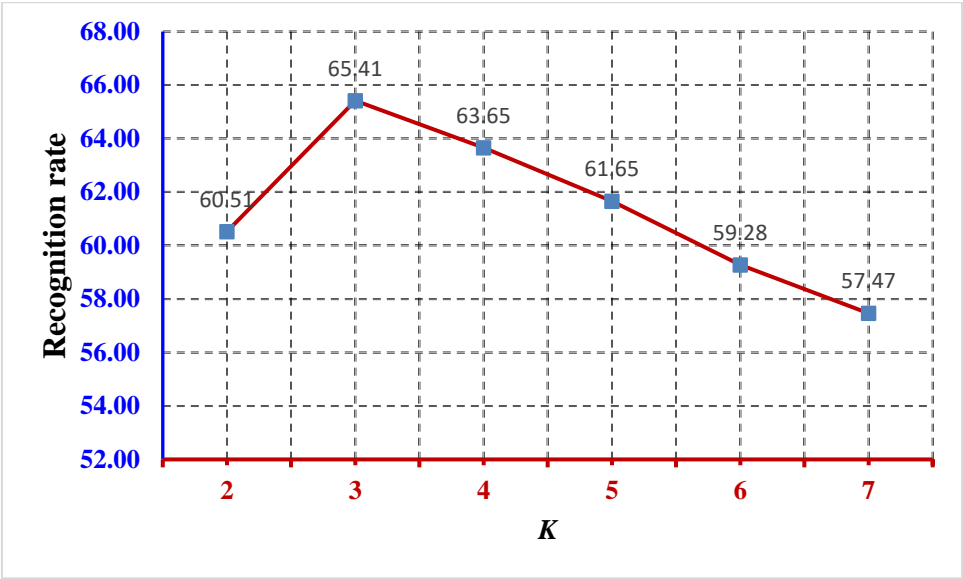


Fig. 4.6. Graph of average recognition rate versus K (FERET database with $r = 3$)

4.4.4 AN ILLUSTRATIVE EXAMPLE

Furthermore, a typical case study is presented below that demonstrates the competence of an IT3 FS to mitigate uncertainty while generating fuzzy ranks in rank-level fusion for FR. The IT3 FS-based approach was compared with a similar trial using IT2 FS (IT2FSIFRLF).

Table 4.7: Outputs generated by RBFNN when applied to test FVs produced by 3 FE techniques

Top 5 CFs from G2DFLD+RBFNN		Top 5 CFs from FG2DFLD+RBFNN		Top 5 CFs from mLBP+RBFNN	
Class	CF	Class	CF	Class	CF
4	0.31048	4	0.29319	19	0.36542
7	0.21753	7	0.22098	7	0.25514
8	0.19225	8	0.20665	13	0.16822
6	0.14155	6	0.14443	4	0.11121
5	0.13818	16	0.13476	6	0.10001

The IT3FSIRLF method succeeded in determining the test image class, whereas the same went wrong in the case of the IT2 FS-based IT2FSIFRLF method. Tables 4.7 – 4.11 show a

typical case of the experiment conducted with the UMIST database. This case study takes into account the test image 178 of set 8 that was formed with $r = 4$. The class of images was 7. The top 5 classifier's outputs (i.e. CF) are shown in Table 4.7 when different test FVs were fed into the RBFNN.

Table 4.8: The fuzzy ranks derived by formulating IT3 FS as described in IT2FSIFRLF

IT2FS-based FRs of the topmost 5 classes with regard to FVs yielded by					
<i>G2DFLD</i>		<i>FG2DFLD</i>		<i>mLBP</i>	
<i>Class</i>	<i>FuzR</i>	<i>Class</i>	<i>FuzR</i>	<i>Class</i>	<i>FuzR</i>
4	0.03090	4	0.03096	19	0.03083
7	0.03089	7	0.03097	7	0.03130
8	0.03153	8	0.03191	13	0.03140
6	0.03085	6	0.03083	4	0.03158
5	0.03081	16	0.03080	6	0.03146

As described in IT2FSIFRLF, Table 4.8 presents the fuzzy ranks (FuzR) derived by formulating the IT2 FS for the classes with the topmost CFs.

Table 4.9: Fused scores for each of the topmost classes for IT2FSIFRLF

Class-wise aCF, cFuzR, and Fused score (only 5 classes are shown)				
<i>Class</i>	<i>aCF</i>	<i>cFuzR</i>	<i>Fused Score = aCF × cFuzR</i>	<i>Rank based on the fused score</i>
4	0.76171	0.09344	0.07117	1
7	0.76878	0.09316	0.07162	2
8	0.86704	0.66944	0.58043	4
6	0.87134	0.09315	0.08116	3
5	0.95394	1.24281	1.18556	5

The combined fuzzy rank (cFuzR), complemented average (mean) of CFs (aCF), and fused score for each of the candidate classes are presented in Table 4.9. The calculation of the aCF for classes 4 and 8 is illustrated below.

$$\text{Class 4: } aCF = 1 - (0.31048 + 0.29319 + 0.11121) / 3 = 0.76171$$

$$\text{Class 8: } aCF = 1 - (0.19225 + 0.20665 + 0.0) / 3 = 0.86704$$

For classes 4 and 8, $cFuzR$ is calculated as follows.

$$\text{Class 4: } cFuzR = 0.03090 + 0.03096 + 0.03158 = 0.09344$$

$$\text{Class 8: } cFuzR = 0.03153 + 0.03191 + 0.606 = 0.66944$$

While computing the sum of fuzzy ranks, a penalty is imposed for classes 5 and 8 for not being present in the top 5 classes of other group(s). Similar tabulations are demonstrated following the IT3FSIRLF method in Tables 4.10 and 4.11.

The aCF and $cFuzR$ for every class under deliberation were determined in a manner similar to that shown above. Table 4.11 confirmed that the IT3FSIRLF method discerned the correct class of the test image whereas Table 4.9 has determined that the class of image 178 is 4, which is incorrect.

Table 4.10: Complemented fuzzy ranks derived by formulating IT3 FS as described in IT3FSIRLF

IT3FS-based FRs of the top 5 classes with regard to 3 FVs					
<i>G2DFLD</i>		<i>FG2DFLD</i>		<i>mLBP</i>	
<i>Class</i>	<i>cFuzR</i>	<i>Class</i>	<i>cFuzR</i>	<i>Class</i>	<i>cFuzR</i>
4	0.66910	4	0.66521	19	0.66494
7	0.65151	7	0.65636	7	0.66300
8	0.65661	8	0.65923	13	0.67052
6	0.66936	6	0.66682	4	0.66833
5	0.64843	16	0.66423	6	0.66653

Table 4.11: Fused scores for each of the top classes for IT3FSIRLF

Class-wise $cCFS$, $cFRS$, and Fused score (only top 5 classes are shown)				
<i>Class</i>	<i>aCF</i>	<i>cFuzR</i>	<i>Fused Score = aCF × cFuzR</i>	<i>Rank based on the fused score</i>
4	0.76171	2.00264	1.52541	2
7	0.76878	1.97087	1.51517	1
8	0.86704	1.92184	1.66630	3
6	0.87134	2.00271	1.74504	4
5	0.95394	1.86043	1.77474	5

4.5 CONCLUSION

This chapter describes a rank-level fusion method for FR wherein interval type-3 fuzzy set-induced fuzzy ranks and classifier outputs are utilized by incorporating multiple feature vectors. Multiple feature vectors were exploited to resolve inter-class resemblances. The leading classes, determined on the basis of higher classifier outputs corresponding to the test feature vector, are taken into account during the fusion process at the rank level. For such a class, smaller degrees of proximity for each training sample, which are calculated using all the training feature vectors, are utilized to create the IT3 FS. The fuzzy rank of the class is then determined by defuzzifying the IT3 FS. This significantly diminishes the uncertainty caused by intra-class divergence. The fuzzy ranks and their corresponding classifier outputs were combined to determine the final fused score, which was used to ascertain the class of the test image. Moreover, the results have established the superior capability of the IT3 FS in mitigating uncertainty as compared to its type-2 equivalents in producing fuzzy ranking during the process of rank-level fusion.

This method utilizes RBFNN as the classifier. Instead of an RBFNN, using a convolution neural network (CNN) or a deep learning model with a suitable fusion method could be a more effective solution to FR.

CHAPTER 5

FACE DETECTION AND RECOGNITION FROM VIDEO IMAGES USING DEEP CNN-BASED RANK LEVEL FUSION

5.2 INTRODUCTION

FR is widely embraced in various fields such as biometrics, surveillance, investigation, human-computer interaction, etc. FR from still and video images posed adversity owing to the unconstrained environment, variability in subjects, subpar sensing quality, etc. [257]. A face image derived from a video can provide not only spatial information but also temporal information, which bolsters recognition performance [258]. Facial analysis, such as FR, detection of expressions, and emotions, relies on FD. FD methods result in the coordinates defining the rectangular outlines around each face present in the image. The position, angle, partial coverage, and size of faces are impediments to face detection. Aligning, scaling, and other pre-processing operations can be applied before FR. To recognize faces, the feature data derived from the test image are compared with the feature data of the database to find a close match. The database was created using the annotated faces of the training images.

Earlier research utilized a window-sliding approach for FD to identify the region of interest [259]. Such techniques are computationally extensive, yet less reliable for detection. Utilizing the AdaBoost technique with a cascade of classifiers, the Viola-Jones algorithm [29] was the first FD algorithm to utilize Haar features. This successful attempt led to several directions for FD thereafter. Of note, most studies have been based on the learning expertise and discernment prowess of convolutional neural networks (CNN) or deep networks for improved precision in detection.

Detection methods involving deep learning are classified as either single-stage or multi-stage methods. Some noteworthy one-stage detectors are YOLO and single-shot detectors (SSD) [260]. These algorithms scan the entire image or video frame once, search for face locations by dense sampling, and obtain a facial map by bounding box regression (BBR) and classification. On the other hand, multi-stage detectors, including region-based CNN (R-CNN) and its variants, such as fast R-CNN, operate in multiple stages. In the multistage FD method, a pool of face regions (region proposals) is generated, and potential cropped region proposals are squared up. In the next stage, the calculated features from the sparse candidate sets were input into the classifier. Multi-stage detectors generally achieve higher accuracy but are computationally costlier and slower [260]. The multitask cascaded convolutional neural network (MTCNN) has been highly valued for its real-time detection capability with face alignment and cascaded lightweight architecture [259]. It used a simple CNN model (P-net) for approximate face region detection and then utilized a more intricate model (R-net) for precise classification and finer adjustment of region box regression to the most probable face candidate regions. The last subnet (O-net) provides a set of bounding boxes together with facial landmarks. According to a study by Nguyen et al. [261] The MTCNN demonstrated superior accuracy over the HOG-SVM and Haar Cascade classifier. It can locate faces even with a partial cover while wearing glasses and at various angles. Guo et al. [262] argued that the model's processing time increases with the number of identified face counts within the input imagery. Several improvements on MTCNN were incorporated to prevail over its limitations and to shore up its performance. To address the limitations of non-frontal FR, Ma et al. [263] suggested a modified version that supports multi-view detection. Khan et al. [264] came up with a modified version of MTCNN which is resilient to facial orientation and capable of detecting faces in low-resolution.

FR leveraging CNN adopts either shallow or deep learning. In shallow learning, handcrafted features such as LBP, Gabor, and SIFT are employed. It often exploits dimension-reduction algorithms such as PCA and LDA. [265]. However, this approach does not suffice in extracting differentiable features that are sufficiently robust against complex real-world variations [60]. The various hidden layers of deep learning models help to learn complex patterns and relationships among data. The loss function plays a pivotal role in deep learning. The model learns to discern the discriminative features between different faces by minimizing the loss function. For classification, a softmax loss is applied to the last fully connected (FC) layer, which produces probabilities for each class. Several studies have been conducted over the last decade to propose different models that employ deep learning for faster and more precise FR. These varied modeling approaches differ in framework, number of parameters, loss function, tuning according to the application domain, etc. Krizhevsky et al. [94] took the lead in introducing AlexNet. The authors of [87, 266] came up with an inception architecture, a deep CNN comprising 22 layers, which was tailored to make the network broaden and widen without compromising computational efficiency. The residual connections in ResNet architecture were suggested by He et al. [267] to sort out the issues that spring up due to vanishing gradient. Huang et al. [92] propounded another deep network model, DenseNet, as a successor to the ResNet. Its dense (feedforward) connections enable efficient feature propagation and representation learning. A brief comparative study on various deep nets, including Alexnet, Xception, certain variants of VGG, Inception, and ResNet, was carried out by Gwyn et al. [120]. Many studies have been steered towards the use of hybrid models to improve FR accuracy. A Neural Aggregation Network (NAN) was put forward in [268] integrating GoogleNet (employed for FE) with Siamese NN architecture (for face verification) and FC layers with softmax (for identification of faces). The hybrid models included in these studies were MDLFace and DeepID3. [60]. In numerous studies, researchers have utilized the fusion mechanism to improve the overall recognition precision [4]. Such fusion can be applied before or after the classification stage of the FR process across different echelons (levels) such as features, scores, and ranks. Considering recent advancements in multi-biometric fusion, Baltrusaitis et al. [239] categorized various fusion techniques into three main types: early, late, and hybridized. Early fusion was primarily feature-based. Dimensional expansion of the feature map resulting from feature-level fusion induces a computing strain while matching features (classification). In late fusion, related methods to accomplish tasks rely on either scores or ranks. Rank-level fusion unraveled the

disparity among the nonlinear scores of different models by reconciling ranks leading to better overall recognition performance [153]. The fuzzy set theory and its logic system have been effectively applied to effectively address problems arising from uncertain and imprecise data across various domains. The rank-level fusion methodology proposed by Ghosh et al. [243] integrates the classification output with fuzzy ranks corresponding to multiple feature vectors. The fuzzy ranking mechanism exploits IT2 FS by utilizing intraclass feature data. Researchers from various domains have also conceived feature-level fusion of deep models. In a review by Zheng et al. [67] explained the recent contributions, scope, and application of fusing fuzzy systems with deep learning. In [269], a fusion at the feature level was brought forward for recognizing faces covered with a mask using DL.

The subsequent sections of the chapter are outlined in the following manner. Section 5.2 covers a comprehensive summary of related concepts. In section 5.3, our methodology is detailed. By examining the results, the performance of the methodology is assessed and showcased in Section 5.4. Conclusions and directions for further research are presented in Section 5.5.

5.3 RELATED CONCEPTS

Deep CNNs are the most widely adopted frameworks in pattern recognition and computer vision domains. The core components of a deep CNN-based architecture are several stacks of layers, convolution layers for extracting and learning features, pooling layers for reducing the spatial dimension of the feature data, activation layers for learning its complex patterns and representations, and fully connected layers for classification or regression tasks. It might also be associated with additional layers for specific tasks, such as batch normalization and skip connections. [60]. Over the years, several deep CNN models have evolved with expansion in size (depth-wise, width-wise, or both) and tuning tailored for applications. The following subsections describe the deep CNN models used in this study:

5.2.1 MULTITASK CASCADED CONVOLUTIONAL NEURAL NETWORK (MTCNN)

The MTCNN, a 3-subnet cascaded DL framework, initiates its processing by resizing the input image at various scales, forming an image pyramid. To detect human faces of different scales from the video frames, they were processed through three sub-networks for training. The image pyramid goes into the first sub-network, the proposal network (or P-net) to produce a group of bounding boxes (candidate frames) that might have faces. P-net is composed of a sequence of convolution and pooling layers, followed by two sibling FC layers. One FC layer predicts the probability of human faces in each candidate frame, and the other obtains and corrects its coordinates using boundary box regression (BBR) analysis. The non-maximum suppression (NMS) algorithm is exercised to exclude the candidate frames of the same target that are imbricate based on the values of intersection over union (IOU). The refine network, alternatively known as R-net, is the second sub-network of MTCNN. It acts towards the fine-tuning of the candidate frames, as produced by the P-net, and obtains more potential facial boundary boxes using BBR analysis and NMS algorithms. R-net consists of layers similar to those of P-net, but the depth and complexity are greater than those of P-net for more precise feature learning. The terminal FC layer contributed to classification (face or non-face) and regression tasks. The final output subnetwork, O-net, decides on the candidate frames to obtain conclusive detections in the form of five key points.

5.2.2 VGG16

Visual Geometry Group (VGG)16 is a 16-layer deep CNN model proposed by Simonyan et al. [270]. A total of 13 convolution layers, each with a filter size of 3×3 and a uniform stride of 1, were spread out into five stacks. The depth of the convolution layers in the consecutive stacks differed in size (64, 128, 256, and 512). A max-pooling layer with a kernel size of 2×2 follows each convolution stack. Subsequent to all convolution stacks, the model culminates with three dense layers (FC layers), with the final layer responsible for classification utilizing softmax activation. The ReLU layer is associated with all hidden layers.

5.2.3 INCEPTIONRESNETV2

InceptionResNetV2 [271] was designed to benefit from multi-scale FE using the inception architecture and gradient flow improvement from the ResNet framework. The components of this deep network include a stem block, three inception blocks with residual connections, two reduction blocks, average pooling, and a fully connected block. The stem block comprises multiple convolution and pooling layers to obtain the basic feature data. Its Inception module substitutes convolution kernels with various series of convolution filters of varied sizes (1×1 , 3×3 , 5×5) and combines its outputs. Residual connections facilitate the circumvention of certain layers during the training phase, thereby ameliorating the speed of the training and alleviating the issues of the vanishing gradient. The reduction block employs convolution and pooling operations to bring down the feature dimension and augment the number of channels. This also contributes to diminishing intensive computations. At the endpoint, there is a global average pooling layer and fully connected block. which subsequently employs softmax activation to aid classification. InceptionResNetV2 offers a noticeably better recognition capability at the expense of costlier hybridization.

5.2.4 XCEPTION

Xception [272], an extreme inception, propounded the notion of inception substantially further by ‘factorizing’ standard convolutions into depthwise separable convolutions (DSC). Every channel is subjected to 1×1 convolutions, followed by the addition of 3×3 convolutions to each of the output tensors. This lessens the computational load while retaining same performance level. An input image was processed through 14 modules that were dispersed into the entry, middle, and exit flows. In addition to the initial and final modules, these modules had residual connections. The entry flow scales down the feature representation using convolutional filtering with strides. It is composed of an array of convolution and separable convolution layers, all of which are subjected to batch normalization and ReLU activation. The middle flow employs DSC layers to preserve the spatial layout of the feature map. The middle flow recurred eight times. The exit flow uses supplementary blocks to spatially subsample the feature data through stridden DSC filters. Finally, a global average pooling layer and FC layers were placed at the end. Activation was applied with a softmax function after the endmost FC layer to facilitate classification.

5.3 FACE DETECTION AND RECOGNITION FROM VIDEO IMAGES USING DEEP CNN-BASED RANK LEVEL FUSION

5.3.1 METHODOLOGY

The work, presented here, is split into two stages. Deep CNN models were employed in both stages. The first stage embraces detecting and extracting faces from video followed by preprocessing. The MTCNN was applied for detection and then extracted and resized to create a facial dataset for FR. Faces were arranged according to each individual. The second phase involved a rank-level fusion framework for FR. The dataset was divided into test, validation, and training sets with varying proportions before applying the deep CNN model for FR. During training, the dataset faces were exposed to random rotations, horizontal and vertical shifts, and flipping. This prevents overfitting and compliments the adaptability of the model to new data. To implement facial recognition, three deep CNN architectures were applied: VGG16, InceptionResNetV2, and Xception. It is worth noting that these models were selected arbitrarily. Any other deep CNN model can substitute these models. The implementation and testing of the models were carried out with a uniform setup and the same dataset.

Inspired by performance enhancement using rank-level fusion and the robust modeling capability of deep CNN models, an approach for detecting and recognizing faces from video images using rank-level fusion with multiple deep CNNs (MDCNNRLF) is outlined here. The methodology utilizes the datasets to recognize faces by exploiting a fusion strategy that has made use of fuzzy ranking. The fuzzy ranks were derived from the probabilities returned by the softmax activation function of a deep CNN-based model. In contrast to the orthodox ranking system, it accounts for the distribution (closeness) of the model's output. In a CNN-based FR model, the softmax activation function yields probabilities that represent the degree of certainty or confidence among the potential classes to which a test sample may belong. This probability is referred to as the certainty index of the class. For each CNN model, a subset consisting of K classes with the top certainty indices was set apart. Each of these classes is assigned a fuzzy rank using the normalized value of the respective certainty index. The fuzzy rank is computed by applying a Gaussian distribution function, followed by a complementation operation. The highest rank was attributed to the value with the smallest magnitude, which was closer to 0. A class that appears in one or more subsets corresponding

to the deep CNN models may have at most three certainty indices, as well as fuzzy ranks. The average of the certainty indices was complemented for each class. Similarly, the fuzzy ranks were summed for each class. A penalty is applied when a class is not included in a specific subset. This prevented the class from becoming an unexpected winner. The final rank of a class is then calculated by multiplying the complemented average of certainty indices with the cumulative sum of fuzzy ranks. Ultimately, the class with the lowest final rank was recognized as the class of the test sample.

5.3.2 MATHEMATICAL FORMULATION OF THE METHODOLOGY

The FR stage involves a fusion approach that is based on fuzzy ranking as delineated in [157]. A Softmax activation function was applied to transform a raw output from a deep CNN model into probabilities within $[0, 1]$ to classify a test sample. These probabilities indicate the confidence or certainty indices of each possible class to which the test sample belongs. These certainty indices are used to calculate the fuzzy ranks of the respective classes. Let M different models be employed separately to predict the class of the test sample, τ . For this study, the value of M was 3, as three deep CNN models, namely, VGG16, InceptionResNetV2, and Xception, were employed. Assuming that the dataset contains N subjects (classes), χ_i^m represents the certainty index (CI) where $i = 1, 2, \dots, N$ and $m = 1, 2, \dots, M$. For each of the M models, a set denoted as S^m , $m = 1, 2, \dots, M$ is created by separating the top K classes, which are determined by their higher values of CI. Therefore, M sets of such classes were obtained for a test image using various CNN models. However, the classes in one set may not match those in another set that corresponds to the same model. The values of CI are then normalized to meet the following conditions:

$$\sum_{k=1}^K \chi_i^m = 1 \mid i \in S^m \quad (5.1)$$

To determine the fuzzy rank (ϑ_i^m) for each class within the set S^m , a Gaussian density function is utilized and then complemented, as demonstrated in Eq. (2) provided below.

$$\vartheta_i^m = 1 - e^{-(\chi_i^m - 1.0)^2 / 2.0} \quad (5.2)$$

The lowest value of the fuzzy rank indicates the highest ranking. A class can have multiple CI corresponding to different CNN models for face recognition. The same is true for the fuzzy ranks. The next steps are carried out to determine the class of the test sample, τ , using a

fusion framework at the rank level for each of the classes, $i \in S^m$. First, by applying Eq. (3), the complement of mean CI, η_{χ_i} , is obtained.

$$\eta_{\chi_i} = 1 - \frac{1}{M} \sum_{m=1}^M \begin{cases} \chi_i^m & \text{if } i \in S^m \\ 0, & \text{otherwise} \end{cases} \quad (5.3)$$

Second, the sum of fuzzy ranks, η_{ϑ_i} , are calculated as per Eq. (4) shown below.

$$\eta_{\vartheta_i} = \sum_{m=1}^M \begin{cases} \vartheta_i^m & \text{if } i \in S^m \\ \rho, & \text{otherwise} \end{cases} \quad (5.4)$$

The class is penalized with ρ , while it appears in a set corresponding to a CNN model but it is not present in the set that corresponds to another model. That is, ρ is applicable only when $i \in S^m$ but $i \notin S^k$, $m \neq k$. The value of ρ is 0.606 which is computed by taking the complement of Equation (2) putting $\chi_i^m = 0$ in it. Finally, the complement of the mean CI (η_{χ_i}) and the sum of the fuzzy rank (η_{ϑ_i}) are fused to obtain the final rank (φ_i) of the class following Eq. (5.5).

$$\varphi_i = \eta_{\chi_i} \times \eta_{\vartheta_i} \quad (5.5)$$

The test image belongs to the class with the lowest final rank.

$$class(\tau) = \arg \min_i \varphi_i \quad (5.6)$$

A schematic of MDCNNRLF is portrayed in Fig. 5.1.

5.4 EXPERIMENTAL RESULTS

For the experimentation of the study, two datasets were used. The first was created from a YouTube video used by G. Barquero et al. [273] as part of the research work. Video was captured in a crowded outdoor environment. The video had a resolution of 1920×1080 and was played at 25 FPS. The original video was trimmed to make an 11-second video clip. The faces from each of the 307 frames were captured using the MTCNN. The extracted faces are prone to blurring owing to motion, and are subjected to partial occlusion and expression.

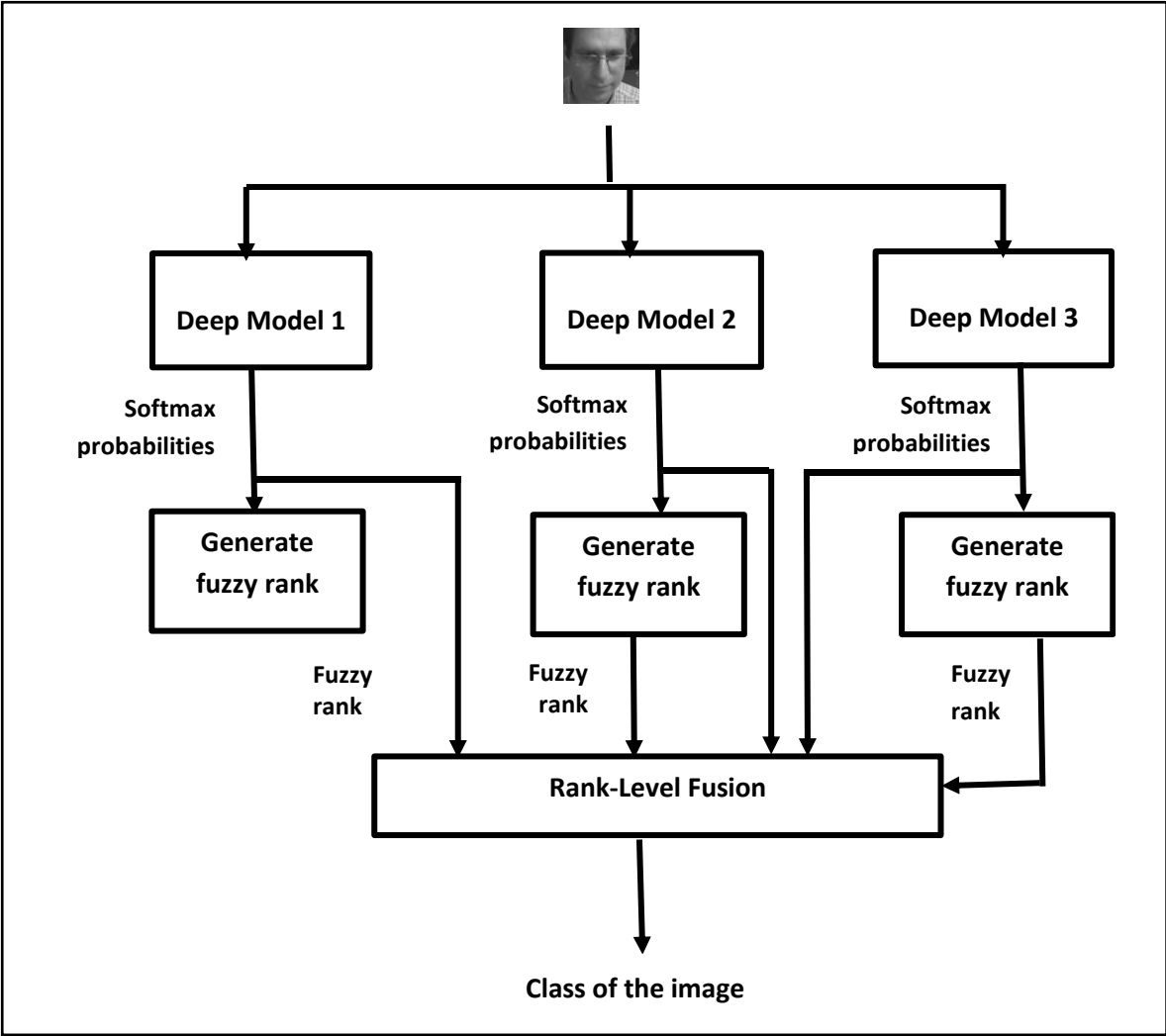


Fig. 5.1. Diagrammatic outline of the MDCNNRLF method

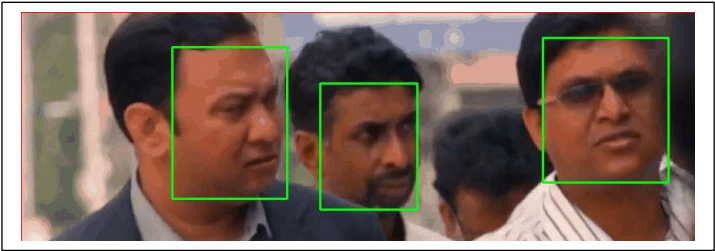
In this study, the ChokePoint dataset [181] was used as the second dataset to evaluate the rank-level fusion approach for facial recognition. This surveillance database may be considered a standard reference for evaluating any FR method. The ChokePoint database was curated from surveillance videos. The faces had varying brightness and head angles. Two groups, G1 and G2, from case_study_1 were created by merging cropped face images, as detailed and downloaded from [25]. In total, 5263 images from 25 people were obtained in G1. These images were used to evaluate the fusion-based face recognition approach.

This study was conducted on a computer with an Intel(R) Core (TM) i5-10400F CPU @ 2.90GHz and 32 GB RAM. For the implementation of the methods, programming was

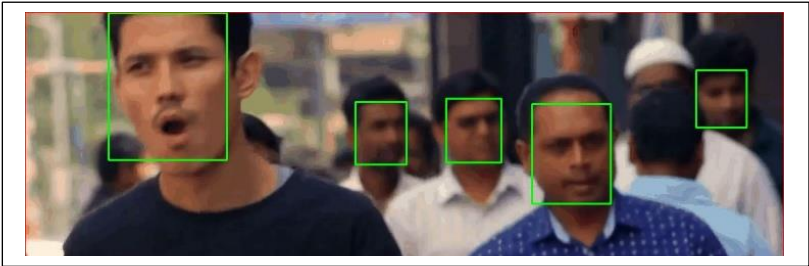
executed utilizing Python and Thonny served as the experimental debugging platform on a Windows 10 operating system. During face detection using the MTCNN, OpenCV was used to draw and show the detected faces.

5.4.1 DETECTING FACES FROM A VIDEO CLIP AND PREPARING CUSTOM FACE DATASET

A subjective manual quality assessment of the extracted faces from the video frames was presented with simple and complex content, as shown in Figs. 5.2(a) – 5.2(c). Complex content includes faces with partial occlusions, motion blur, and varied facial angles.



(a)



(b)

Fig. 5.2. Detection of faces with various scales from video frames



(c)

Fig. 5.2. Detected and Extracted faces with partial occlusion, motion blur, and pose

Three deep neural network models, namely VGG16, InceptionResNetV2, and Xception, were used for face recognition. The network model parameters are trained using TensorFlow and Keras. The learning rate and the number of epochs were fixed at 0.0001 and 10, respectively. During training, random operations such as rotation by 20° , shifting horizontally and vertically by $\pm 20\%$, and horizontal flipping were applied to prevent overfitting and bolster the model's capability with new data.

To assess FR by employing our proposed scheme, a random selection of $m\%$ of the images from each dataset contributed to the training set. The corresponding validation and test sets consisted of $(100 - m)\%$ of the images in the dataset. Five such 'training-validation-test' sets were compiled with $m = 20, 30, 40, 50, 60, 70$, and 80 . The FR accuracies of the CNN models used in this study were evaluated separately using these sets. The experimental setting was kept unchanged so that the potential of the proposed method could be compared with these well-established methods for face recognition

5.4.2 EXPERIMENTS WITH THE CUSTOM FACE DATASET

The experimental outcomes are shown in Tables 5.1 and Table 5.2. The enhanced recognition accuracy in each case establishes the predominance of the proposed rank-level fusion scheme utilizing three deep models over methods that independently utilize deep models.

Table 5.1: The recognition accuracy of the methods on the custom face database

Training-validation-test set split (in %)	VGG16	Inception-ResNetV2	Xception	MDCNNRLF
Average recognition accuracy \pm SD				
20-40-40	0.7191 \pm 0.05	0.5129 \pm 0.06	0.4102 \pm 0.07	0.7517 \pm 0.02
30-35-35	0.8896 \pm 0.07	0.6015 \pm 0.04	0.4008 \pm 0.08	0.9089 \pm 0.06
40-30-30	0.9498 \pm 0.06	0.7014 \pm 0.11	0.6287 \pm 0.11	0.9695 \pm 0.03
50-25-25	0.9648 \pm 0.04	0.7264 \pm 0.17	0.7626 \pm 0.15	0.9824 \pm 0.01
60-20-20	0.9605 \pm 0.03	0.8276 \pm 0.09	0.8987 \pm 0.03	0.9724 \pm 0.02
70-15-15	0.9513 \pm 0.09	0.9183 \pm 0.05	0.9687 \pm 0.01	0.9948 \pm 0.01
80-10-10	0.9925 \pm 0.01	0.9025 \pm 0.09	0.9675 \pm 0.03	0.9975 \pm 0.01

The adeptness of the model in carrying out FR was evaluated in terms of the macro average. The overall macro average of the advocated approach in each case (as shown in Table 5.2) was also higher than that of the methods used to exercise the deep CNN models in isolation.

Table 5.2: The arithmetic mean of macro averages of the face recognition model that uses the custom dataset

Training-validation-test set split (in %)	VGG16	Inception-ResNetV2	Xception	MDCNNRLF
20-40-40	0.5846	0.3644	0.2797	0.6274
30-35-35	0.7448	0.4437	0.2415	0.7649
40-30-30	0.8434	0.5593	0.4698	0.8656
50-25-25	0.8530	0.5921	0.6023	0.8766
60-20-20	0.8677	0.7084	0.7988	0.8794
70-15-15	0.9069	0.8551	0.8679	0.9679
80-10-10	0.9746	0.8449	0.8911	0.9918

To justify the statistical significance of performance improvement, a paired t -test (one-tailed method) is applied. In this context, the null hypothesis (H_0) is defined as: ‘the average recognition rate of the MDCNNRLF method is same as of the compared method’ against the alternative hypothesis (H_1) as: ‘the average recognition rate of the MDCNNRLF method is higher as compared to the related method’. The respective p -values are reported in Table 5.3. The performance of the MDCNNRLF method was found to be significant compared to the implementation utilizing the Xception and InceptionResnetV2 models because the p -values (in boldface) were mostly less than 0.05. Compared with the implementation using the VGG16 model, the MDCNNRLF method's performance was not significant, although the recognition accuracy was improved.

Table 5.3: Statistical significance analysis of the MDCNNRLF with three deep CNN models on the custom database

% of Training-Validation-Test set	VGG16	InceptionResNetV2	Xception
20-40-40	0.11836	0.00009	0.00058
30-35-35	0.04765	0.00071	0.00002
40-30-30	0.10468	0.00288	0.00220
50-25-25	0.11539	0.01404	0.01688
60-20-20	0.05207	0.00663	0.00277
70-15-15	0.15252	0.00861	0.00034
80-10-10	0.08890	0.04454	0.02090

5.4.3 EXPERIMENTS WITH THE CHOKEPOINT FACE DATASET

Tables 5.4 and 5.5 present the average recognition rate and mean of the macro averages of the experiments involving MDCNNRLF and the other three CNN models, VGG16, InceptionResnetV2, and Xception. The tables make it clear that in every case, our presented model for FR, MDCNNRLF, has the highest recognition rates, demonstrating better performance compared to other methods. The elevated rate of recognition success is also justified by the t -test which is explicated in Table 5.6. The respective p -values have manifested that the elevated performance of MDCNNRLF is justified for training-validation-test sets 20-40-40, 30-35-35, and 40-30-30 at $p < 0.05$ compared to InceptionResNetV2 and Xception models. However, in comparison with the VGG16 model, our method has shown statistical significance in performance for the training-validation-test sets 30-35-35, 40-30-30, 60-20-20, and 70-15-15.

Table 5.4: The recognition accuracy of the methods on the Chokepoint face database

Training-validation-test set split (in %)	VGG16	Inception- ResNetV2	Xception	MDCNNRLF
Average recognition accuracy \pm SD				
20-40-40	0.8462 \pm 0.20	0.9211 \pm 0.01	0.8253 \pm 0.07	0.9725 \pm 0.02
30-35-35	0.9755 \pm 0.01	0.9849 \pm 0.01	0.8830 \pm 0.02	0.9968 \pm 0.00
40-30-30	0.9558 \pm 0.04	0.9902 \pm 0.00	0.9763 \pm 0.01	0.9988 \pm 0.00
50-25-25	0.9929 \pm 0.01	0.9928 \pm 0.01	0.9913 \pm 0.01	0.9997 \pm 0.00
60-20-20	0.9961 \pm 0.00	0.9976 \pm 0.00	0.9927 \pm 0.01	1.0000 \pm 0.00
70-15-15	0.9956 \pm 0.00	0.9697 \pm 0.05	0.9998 \pm 0.00	1.0000 \pm 0.00
80-10-10	0.9582 \pm 0.08	0.9989 \pm 0.00	0.9967 \pm 0.00	1.0000 \pm 0.00

The analysis of the experimental outcomes has established a significant performance improvement for our rank-level fusion scheme over the three deep CNN models. It can also be noted from the analysis of the above tables that the MDCNNRLF method manifested its competence when the training set consisted of a smaller number of images. In other words, our rank-level fusion model can exhibit optimized performance despite less training.

Table 5.5: The arithmetic mean of macro averages of the face recognition model that uses the custom dataset

Training-validation- test set split (in %)	VGG16	Inception- ResNetV2	Xception	MDCNNRLF
20-40-40	0.8396	0.9210	0.8192	0.9717
30-35-35	0.9731	0.9849	0.8809	0.9966
40-30-30	0.9543	0.9900	0.9761	0.9987
50-25-25	0.9927	0.9924	0.9911	0.9997
60-20-20	0.9957	0.9976	0.9927	1.0000
70-15-15	0.9953	0.9692	0.9997	1.0000
80-10-10	0.9545	0.9989	0.9967	1.0000

Table 5.6: Statistical significance analysis of the MDCNNRLF with three deep CNN models on the Chokepoint database

% of Training- Validation-Test set	VGG16	InceptionResNetV2	Xception
20-40-40	0.09246	0.00786	0.00339
30-35-35	0.01819	0.01058	0.00018
40-30-30	0.04731	0.00317	0.00215
50-25-25	0.05368	0.11678	0.12778
60-20-20	0.02871	0.06840	0.15032
70-15-15	0.03542	0.11695	0.18695
80-10-10	0.16277	0.10506	0.06095

5.5 CONCLUSION

This chapter put forward an approach that involves developing a face dataset from video clips and utilizing deep learning networks for recognition. It has made use of a fusion strategy with fuzzy ranking. The fuzzy ranks were derived from the probabilities (certainty indices) provided by the softmax activation function of a deep learning model. Fuzzy ranks were calculated for the selected classes with the top certainty indices. The final rank of a class is calculated by combining the complemented average certainty indices and the cumulative sum of the complemented fuzzy ranks. This method identifies the class with the lowest final rank as the recognized class for the test sample. The evaluation of the experimental results confirmed the preeminence of the rank-level fusion scheme compared with three deep CNN

models. This method outperformed the independent deep model, proving its effectiveness across all classes in the dataset. The use of lightweight deep network models could make the proposed methodology more time-efficient. Because the method relies on the predictions of three different deep network frameworks, the execution of the models in a distributed manner makes the modus operandi of the rank-level fusion scheme more optimized.

CHAPTER 6

CONCLUSION AND FUTURE SCOPE

6.1 CONCLUSION

The stimulus for this thesis is to mitigate uncertainty in decision-making systems such as FR. All tasks for facial analysis, including FR, are initiated with FD, which is embraced with precise alignment and augmented modeling of the face. Various factors associated with heterogeneous environmental conditions and imaging settings lead to intra-subject diversity and inter-subject resemblance. As a result, this poses uncertainty in such decision-making. To cope with these issues, different types of fuzzy sets and systems have proven to be a convincing approach for managing uncertainty at various levels. Considering its merits, adopting an RLF for FR is prudent for improved precision. With the common intention of proposing a more refined FR approach, various fuzzy sets are expended in the RLF involving multiple FVs derived from distinct FE techniques. In the first three methodologies presented in this thesis, NN-based classifiers were employed, and deep CNN models were exploited in the fourth methodology.

The first chapter describes the different approaches to FD and FR with related studies and the core concepts of DL with CNN focusing on FD and FR. It continues with multi-biometric fusion, its modality and levels, and related studies where face is one of the traits. Then, the

motivation, objectives, and organization of the thesis are briefly expounded. It ends with a description of the standard databases used to evaluate the proposed methodologies.

The second chapter introduces my first contribution, which is a new decision-making model for multi-feature RLF that integrates the general type-2 fuzzy set (GT2 FS)-based approach to mitigate the factors that contribute to face recognition accuracy. The model selects the top-ordered classes based on the weights or CFs yielded by the classifier. For each of these classes, a GT2 FS was formulated from the feature vectors of the test image and training samples. The complement of type-reduced defuzzified value is conceived as a fuzzy rank. Considering the three different feature vectors corresponding to the three FE techniques, each class might have at most three fuzzy ranks, and CFs. The class-wise cumulative sum of the fuzzy ranks is fused with the respective complemented mean of the confidence factors to yield final rank. While aggregating fuzzy ranks, a penalty is set for the class if it is not present in the top-ordered classes corresponding to an FV. A facial image was identified based on the final ranks.

The third chapter has elucidated an FR framework, which is my second contribution, to address the wide intra-class variability issue by deriving the LMF and UMF describing the FOU of an IT2 FS utilizing the feature vectors of the intra-class face images. The defuzzified value was the fuzzy rank of the class under consideration. Furthermore, multi-feature vectors for a face image are used to utilize the underlying discriminant features, which also address the issues of inter-class similarity. Multiple fuzzy ranks were obtained for each class using multiple feature vectors. To reduce the complexity, top K classes, based on CFs of the classifier, are taken into consideration for RLF. The aggregated fuzzy rank is then fused with the respective complemented mean of the confidence factors to obtain the final fuzzy rank for each of the top-ordered classes. When combining fuzzy ranks, a penalty is set for the class if it is not present in the top-ordered classes corresponding to an FV. A facial image was identified based on the final fuzzy ranks.

In the fourth chapter, an FR framework has put the effectiveness of IT3 FS to the test in dealing with the uncertainty and imprecision caused by various environmental conditions. The framework considers the top-ordered classes, selected according to the confidence factors of the classifier, to ascertain fuzzy ranks utilizing IT3 FSs with multiple feature vectors. For each class, corresponding training samples were employed to generate secondary

membership grades. For each training sample, smaller proximity measures using training samples from all classes are set apart to express the lower and upper membership functions of the scaled Gaussian domain of the FOU of the IT3 FS. Successive defuzzifications yield the fuzzy rank of the class. Corresponding to different FVs, the fuzzy ranks are combined and fused with the arithmetic mean of the confidence factors to determine the final rank. The test images were identified based on these final ranks.

The fifth chapter presents my fourth contribution, which describes a methodology wherein a face database is created from an outdoor video clip using MTCNN and has propounded a fuzzy RLF for FR involving deep CNN models. The fuzzy ranks are derived for the top-ordered classes from the probabilities produced by the softmax function (termed here as certainty indices) of the respective deep model by applying a Gaussian distribution function, followed by a complementation operation. Considering three deep CNN models, the complement of the average of the certainty indices is combined with the summation of fuzzy ranks for each class. A penalty mechanism is implemented for a class not included in the top-ordered classes corresponding to an FV. The class with the lowest final ranking was identified as the test-sample class.

All the methodologies presented in this thesis were evaluated on several face databases, such as AT & T, FERET, UMIST, Faces94, and Chokeypoint, and demonstrated their pre-eminence over similar approaches.

6.2 FUTURE SCOPE

Despite the improved results of the methods detailed in this thesis, several unresolved issues remain. In practice, the perfection of the FR method is elusive, particularly in unconstrained circumstances or outdoor and indoor video surveillance scenarios. This thesis proposes various frameworks aimed at enhancing the performance of FR using still and video images. However, these methods fall short of achieving the desired level of effectiveness and must be optimized or refined. There are still many challenges in video-based FR, such as poor-quality video, small faces, and lighting imbalance for outdoor surveillance. FR and FD using a deep CNN framework have attained a notable level of success in resolving the aforementioned

challenges. Deep CNNs require a large training set for obtaining accurate results, resulting in longer training periods and increased memory usage. Choosing a deep architecture that suits all situations is prudent for effective performance. For effective video-based FR, various subtasks such as detection, prediction of head pose, tracking of face movement, face modeling, and face segmentation play a vital role and necessitate further research.

The methods outlined in this thesis could be extended in future research by taking the aforementioned points into consideration.

BIBLIOGRAPHY

- [1] A. Pradhan, J. He, and N. Jiang, “Score, rank, and decision-level fusion strategies of multicode electromyogram-based verification and identification biometrics,” *IEEE J. Biomed. Health Inform.*, vol. 26, no. 3, pp. 1068–1079, 2022.
- [2] A. A. Ross, K. Nandakumar, and A. K. Jain, *Handbook of multibiometrics*. Springer Science & Business Media, 2006.
- [3] S. K. S. Modak and V. K. Jha, “Multibiometric fusion strategy and its applications: A review,” *Inf. Fusion*, vol. 49, pp. 174–204, 2019.
- [4] M. Singh, R. Singh, and A. Ross, “A comprehensive overview of biometric fusion,” *Inf. Fusion*, vol. 52, pp. 187–205, 2019.
- [5] W. Ali, W. Tian, S. U. Din, D. Iradukunda, and A. A. Khan, “Classical and modern face recognition approaches: a complete review,” *Multimed. Tools Appl.*, vol. 80, no. 3, pp. 4825–4880, 2021.
- [6] M. Taskiran, N. Kahraman, and C. E. Erdem, “Face recognition: Past, present and future (a review),” *Digit. Signal Process.*, vol. 106, no. 102809, p. 102809, 2020.
- [7] A. K. Jain, K. Nandakumar, and A. Ross, “50 years of biometric research: Accomplishments, challenges, and opportunities,” *Pattern Recognit. Lett.*, vol. 79, pp. 80–105, 2016.
- [8] S. Umer, B. C. Dhara, and B. Chanda, “Face recognition using fusion of feature learning techniques,” *Measurement (Lond.)*, vol. 146, pp. 43–54, 2019.
- [9] H. Wang, J. Hu, and W. Deng, “Face feature extraction: A complete review,” *IEEE Access*, vol. 6, pp. 6001–6039, 2018.
- [10] S. Aly, N. Tsuruta, and R.-I. Taniguchi, “Face recognition under varying illumination using Mahalanobis self-organizing map,” *Artif. Life Robot.*, vol. 13, no. 1, pp. 298–301, 2008.
- [11] X. Zhang and Y. Gao, “Face recognition across pose: A review,” *Pattern Recognit.*, vol. 42, no. 11, pp. 2876–2896, 2009.
- [12] U. Park, Y. Tong, and A. K. Jain, “Age-Invariant Face Recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 947–954, 2010.

- [13] M. D. Levine and Y. Yu, "Face recognition subject to variations in facial expression, illumination and pose using correlation filters," *Comput. Vis. Image Underst.*, vol. 104, no. 1, pp. 1–15, 2006.
- [14] X. Tan, S. Chen, Z.-H. Zhou, and J. Liu, "Face recognition under occlusions and variant expressions with partial similarity," *IEEE Trans. Inf. Forensics Secur.*, vol. 4, no. 2, pp. 217–230, 2009.
- [15] U. Jayaraman, P. Gupta, S. Gupta, G. Arora, and K. Tiwari, "Recent development in face recognition," *Neurocomputing*, vol. 408, pp. 231–245, 2020.
- [16] M. Hassaballah and S. Aly, "Face recognition: challenges, achievements and future directions," *IET Comput. Vis.*, vol. 9, no. 4, pp. 614–626, 2015.
- [17] Alvarez, L.; Baumela, L.; Márquez-Neila, P.; Henríquez, P. A real time morphological snakes algorithm. *Image Process. On Line* 2012, 2, 1–7.
- [18] H. Lee, H. Kwon, R. M. Robinson, and W. D. Nothwang, "DTM: Deformable template matching," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 1966–1970.
- [19] R. Ranjan, V. M. Patel, and R. Chellappa, "A deep pyramid Deformable Part Model for face detection," in *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, 2015, pp. 1–8.
- [20] E. Hjelmås and B. K. Low, "Face detection: A survey," *Comput. Vis. Image Underst.*, vol. 83, no. 3, pp. 236–274, 2001.
- [21] Y. Ban, S.-K. Kim, S. Kim, K.-A. Toh, and S. Lee, "Face detection based on skin color likelihood," *Pattern Recognit.*, vol. 47, no. 4, pp. 1573–1585, 2014.
- [22] Zangana, H.M. A New Skin Color Based Face Detection Algorithm by Combining Three Color Model Algorithms. *IOSR J. Comput. Eng.*, 17, 06–125, 2015.
- [23] R. Herpers, M. Michaelis, K.-H. Lichtenauer, and G. Sommer, "Edge and keypoint detection in facial regions," in *Proceedings of the Second International Conference on Automatic Face and Gesture Recognition*, 2002, pp. 212–217.
- [24] B. G. Schunck, "Image flow segmentation and estimation by constraint line clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 10, pp. 1010–1027, 1989.
- [25] Lee, C.H.; Kim, J.S.; Park, K.H. Automatic human face location in a complex background using motion and color information. *Pattern Recognit.*, vol. 29, pp. 1877–1889, 1996.
- [26] S. J. Mc Kenna, S. Gong, and H. Liddell, "Real-time tracking for an integrated face

- recognition system,” In *Second European Workshop on Parallel Modelling of Neural Operators*, vol. 11. 1995.
- [27] Sakai, T. Computer analysis and classification of photographs of human faces. In *Proceedings of the First USA—Japan Computer Conference, Tokyo, Japan, 1972*, pp. 55–62.
 - [28] S.-H. Jeng, H. Y. M. Liao, C. C. Han, M. Y. Chern, and Y. T. Liu, “Facial feature detection using geometrical face model: An efficient approach,” *Pattern Recognit.*, vol. 31, no. 3, pp. 273–282, 1998.
 - [29] P. Viola and M. J. Jones, “Robust real-time face detection,” *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.
 - [30] M. Lades *et al.*, “Distortion invariant object recognition in the dynamic link architecture,” *IEEE Trans. Comput.*, vol. 42, no. 3, pp. 300–311, 1993.
 - [31] L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg, “Face recognition by elastic bunch graph matching,” in *INTELLIGENT BIOMETRIC TECHNIQUES in FINGERPRINT and FACE RECOGNITION*, Boca Raton: Routledge, 2022, pp. 355–396.
 - [32] A. Kumar, A. Kaur, and M. Kumar, “Face detection techniques: a review,” *Artif. Intell. Rev.*, vol. 52, no. 2, pp. 927–948, 2019.
 - [33] M. H. Yang, D. J. Kriegman, and N. Ahuja, “Detecting faces in images: a survey,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 1, pp. 34–58, 2002.
 - [34] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces vs. Fisherfaces: recognition using class specific linear projection,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, 1997.
 - [35] M. A. O. Vasilescu and D. Terzopoulos, “Multilinear Analysis of Image Ensembles: TensorFaces,” in *Computer Vision — ECCV 2002*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2002, pp. 447–460.
 - [36] M. K. Hasan, M. S. Ahsan, Abdullah-Al-Mamun, S. H. S. Newaz, and G. M. Lee, “Human face detection techniques: A comprehensive review and future research directions,” *Electronics (Basel)*, vol. 10, no. 19, p. 2354, 2021.
 - [37] S. Zafeiriou, C. Zhang, and Z. Zhang, “A survey on face detection in the wild: Past, present and future,” *Comput. Vis. Image Underst.*, vol. 138, pp. 1–24, 2015.
 - [38] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, “Face recognition by independent component analysis,” *IEEE Trans. Neural Netw.*, vol. 13, no. 6, pp. 1450–1464, 2002.
 - [39] Sparsh, R. Aggarwal, S. Bhardwaj, and K. Sharma, “Face recognition system using image

enhancement with PCA and LDA,” in *2022 6th International Conference on Computing Methodologies and Communication (ICCMC)*, 2022, pp. 1322–1327.

- [40] Y. Kortli, M. Jridi, A. Al Falou, and M. Atri, “Face recognition systems: A survey,” *Sensors (Basel)*, vol. 20, no. 2, p. 342, 2020.
- [41] C. C. Liu and D.-Q. Dai, “Face recognition using dual-tree complex wavelet features,” *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2593–2599, 2009.
- [42] L. Fang and S. Li, “Face recognition by exploiting local Gabor features with multitask adaptive sparse representation,” *IEEE Trans. Instrum. Meas.*, vol. 64, no. 10, pp. 2605–2615, 2015.
- [43] S. Najafi Khanbebin and V. Mehrdad, “Local improvement approach and linear discriminant analysis-based local binary pattern for face recognition,” *Neural Comput. Appl.*, vol. 33, no. 13, pp. 7691–7707, 2021.
- [44] S. Yallamandaiah and N. Purnachand, “A novel face recognition technique using Convolutional Neural Network, HOG, and histogram of LBP features,” in *2022 2nd International Conference on Artificial Intelligence and Signal Processing (AISP)*, 2022, pp. 1–5.
- [45] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
- [46] G. Du, F. Su, and A. Cai, “Face recognition using SURF features,” in *SPIE Proceedings*, 2009.
- [47] L. Wolf, T. Hassner, and Y. Taigman, “Descriptor based methods in the wild,” in *Workshop on faces in 'real-life' images: Detection, alignment, and recognition*, 2008.
- [48] D. Huang, C. Shan, M. Ardabilian, Y. Wang, and L. Chen, “Local binary patterns and its application to facial image analysis: A survey,” *IEEE Trans. Syst. Man Cybern. C Appl. Rev.*, vol. 41, no. 6, pp. 765–781, 2011.
- [49] X. Tan and B. Triggs, “Enhanced local texture feature sets for face recognition under difficult lighting conditions,” *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1635–1650, 2010.
- [50] K. Meena and A. Suruliandi, “Local binary patterns and its variants for face recognition,” in *2011 International Conference on Recent Trends in Information Technology (ICRTIT)*, 2011, pp. 782–786.
- [51] M. Chihaoui, A. Elkefi, W. Bellil, and C. Ben Amar, “A survey of 2D face recognition techniques,” *Computers*, vol. 5, no. 4, p. 21, 2016.
- [52] I. Adjabi, A. Ouahabi, A. Benzaoui, and A. Taleb-Ahmed, “Past, present, and future of face recognition: A review,” *Electronics (Basel)*, vol. 9, no. 8, p. 1188, 2020.

- [53] S. S. Thenuwara, C. Premachandra, and S. Sumathipala, "Hybrid approach to face recognition system using PCA & LDA in border control," in *2019 National Information Technology Conference (NITC)*, 2019, pp. 9–15.
- [54] H. Cho, R. Roberts, B. Jung, O. Choi, and S. Moon, "An efficient hybrid face recognition algorithm using PCA and GABOR wavelets," *Int. J. Adv. Robot. Syst.*, vol. 11, no. 4, p. 59, 2014.
- [55] J. K. Sing, S. Chowdhury, D. K. Basu, and M. Nasipuri, "An improved hybrid approach to face recognition by fusing local and global discriminant features," *Int. J. Biom.*, vol. 4, no. 2, p. 144, 2012.
- [56] V. P. Vishwakarma and T. Goel, "An efficient hybrid DWT-fuzzy filter in DCT domain based illumination normalization for face recognition," *Multimed. Tools Appl.*, vol. 78, no. 11, pp. 15213–15233, 2019.
- [57] K. Simonyan, O. M. Parkhi, A. Vedaldi, and A. Zisserman, Fisher vector faces in the wild. In *Proceedings of the BMVC 2013—British Machine Vision Conference, Bristol, UK*, 2013.
- [58] M. Moussa, M. Hmila, and A. Douik, "A novel face recognition approach based on genetic algorithm optimization," *Stud. Inform. Contr.*, vol. 27, no. 1, 2018.
- [59] S.-H. Lin, S.-Y. Kung, and L.-J. Lin, "Face recognition/detection by probabilistic decision-based neural network," *IEEE Trans. Neural Netw.*, vol. 8, no. 1, pp. 114–132, 1997.
- [60] M. Wang and W. Deng, "Deep face recognition: A survey," *Neurocomputing*, vol. 429, pp. 215–244, 2021.
- [61] X. Li, "Face recognition method based on fuzzy 2DPCA," *J. Electr. Comput. Eng.*, vol. 2014, pp. 1–7, 2014.
- [62] S. Chowdhury, J. K. Sing, D. K. Basu, and M. Nasipuri, "Face recognition by generalized two-dimensional FLD method and multi-class support vector machines," *Appl. Soft Comput.*, vol. 11, no. 7, pp. 4282–4292, 2011.
- [63] A. Dey and J. K. Sing, "Face recognition by fuzzy generalized 2DFLD method and RBF neural network classifier," in *2015 IEEE Workshop on Computational Intelligence: Theories, Applications and Future Directions (WCI)*, 2015, pp. 1–6.
- [64] Y. Du, X. Lu, W. Zeng, and C. Hu, "A novel fuzzy linear discriminant analysis for face recognition," *Intell. Data Anal.*, vol. 22, no. 3, pp. 675–696, 2018.
- [65] P. Murugeswari and S. Vijayalakshmi, "A new method of interval type-2 fuzzy-based CNN for image classification," in *Computational Vision and Bio-Inspired Computing*, Singapore:

Springer Singapore, 2021, pp. 733–746.

- [66] S. Yadav and V. P. Vishwakarma, “Extended interval type-II and kernel based sparse representation method for face recognition,” *Expert Syst. Appl.*, vol. 116, pp. 265–274, 2019.
- [67] Y. Zheng, Z. Xu, and X. Wang, “The fusion of deep learning and fuzzy systems: A state-of-the-art survey,” *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 8, pp. 2783–2799, 2022.
- [68] G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [69] R. K. Mishra, G. Y. S. Reddy, and H. Pathak, “The understanding of deep learning: A comprehensive review,” *Math. Probl. Eng.*, vol. 2021, pp. 1–15, 2021.
- [70] R. R. Salakhutdinov and G. E. Hinton, “An efficient learning procedure for deep boltzmann machines,” *Neural Comput.*, vol. 24, pp. 1967–2006, 2012.
- [71] Y. LeCun and Y. Bengio, “Convolutional networks for images speech and time series” in *Handbook of Brain Theory & Neural Networks*, MIT Press, Cambridge, MA, USA, 1998, pp. 255–258.
- [72] S. E. Hihi, M. Q. He-J and Y. Bengio, “Hierarchical recurrent neural networks for long-term dependencies”, *Adv. Neural Inf. Process. Syst.*, vol. 8, pp. 493–499, 1995.
- [73] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, “Deep learning for visual understanding: A review,” *Neurocomputing*, vol. 187, pp. 27–48, 2016.
- [74] Q. Abbas, M. E. A. Ibrahim, and M. A. Jaffar, “A comprehensive review of recent advances on deep vision systems,” *Artif. Intell. Rev.*, vol. 52, no. 1, pp. 39–76, 2019.
- [75] Y. LeCun *et al.*, “Backpropagation applied to handwritten zip code recognition,” *Neural Comput.*, vol. 1, no. 4, pp. 541–551, 1989.
- [76] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proc. IEEE Inst. Electr. Electron. Eng.*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [77] Fukushima, Kunihiko. “Neocognitron: A hierarchical neural network capable of visual pattern recognition.” *Neural networks 1*, vol. 2, pp. 119–130, 1988.
- [78] X. Liu, Z. Deng, and Y. Yang, “Recent progress in semantic image segmentation,” *Artif. Intell. Rev.*, vol. 52, no. 2, pp. 1089–1106, 2019.
- [79] M. Krichen, “Convolutional neural networks: A survey,” *Computers*, vol. 12, no. 8, p. 151, 2023.

- [80] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [81] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," *arXiv [cs.LG]*, 2015.
- [82] H. Robbins and S. Monro, "A Stochastic Approximation Method," *Ann. Math. Stat.*, vol. 22, no. 3, pp. 400–407, 1951.
- [83] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv [cs.LG]*, 2014.
- [84] D. Misra, "Mish: A self regularized non-monotonic activation function," *arXiv [cs.LG]*, 2019.
- [85] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on Machine Learning*, 07-09 Jul 2015, vol. 37, pp. 448–456.
- [86] G. Li, M. Zhang, J. Li, F. Lv, and G. Tong, "Efficient densely connected convolutional neural networks," *Pattern Recognit.*, vol. 109, no. 107610, p. 107610, 2021.
- [87] C. Szegedy *et al.*, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1–9.
- [88] M. Shafiq and Z. Gu, "Deep residual learning for image recognition: A survey," *Appl. Sci. (Basel)*, vol. 12, no. 18, p. 8972, 2022.
- [89] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5987–5995.
- [90] X. Zhang, Z. Li, C. C. Loy, and D. Lin, "PolyNet: A pursuit of structural diversity in very deep networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 3900–3908.
- [91] A. G. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *arXiv preprint arXiv:1704.04861*, 2017.
- [92] G. Huang, Z. Liu, G. Pleiss, L. van der Maaten, and K. Q. Weinberger, "Convolutional networks with dense connectivity," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 12, pp. 8704–8716, 2022.
- [93] S. H. Hasanpour, M. Rouhani, M. Fayyaz, M. Sabokrou, and E. Adeli, "Towards principled design of deep convolutional networks: Introducing SimpNet," *Arxiv.org*. [Online]. Available: <http://arxiv.org/abs/1802.06205>. [Accessed: 27-Jun-2024].

- [94] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017.
- [95] R. Ranjan *et al.*, "Deep learning for understanding faces: Machines may be just as good, or better, than humans," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 66–83, 2018.
- [96] X. Wu, D. Sahoo, and S. C. H. Hoi, "Recent advances in deep learning for object detection," *Neurocomputing*, vol. 396, pp. 39–64, 2020.
- [97] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580–587.
- [98] R. Girshick, "Fast R-CNN," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1440–1448.
- [99] H. Jiang and E. Learned-Miller, "Face detection with the faster R-CNN," in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, 2017, pp. 650–657.
- [100] X. Sun, P. Wu, and S. C. H. Hoi, "Face detection using deep learning: An improved faster RCNN approach," *Neurocomputing*, vol. 299, pp. 42–50, 2018.
- [101] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980–2988.
- [102] X. Li *et al.*, "Weighted feature pyramid networks for object detection," in *2019 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom)*, 2019, pp. 1500–1504.
- [103] Q. Zhao *et al.*, "M2Det: A single-shot object detector based on multi-level feature pyramid network," *Aaai.org*. [Online]. Available: <https://cdn.aaai.org/ojs/4962/4962-13-8027-1-10-20190709.pdf>. [Accessed: 27-Jun-2024].
- [104] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, Y. LeCun, Overfeat: Integrated recognition, localization and detection using convolutional networks, in: *arXiv preprint arXiv:1312.6229*, 2013.
- [105] W. Liu *et al.*, "SSD: Single Shot MultiBox Detector," in *Computer Vision – ECCV 2016*, Cham: Springer International Publishing, 2016, pp. 21–37.
- [106] X. Hu and B. Huang, "Face Detection based on SSD and CamShift," in *2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, 2020,

vol. 9, pp. 2324–2328.

- [107] J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou, “RetinaFace: Single-Shot Multi-Level Face Localisation in the Wild,” in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 5202–5211.
- [108] X. Tang, D. K. Du, Z. He, and J. Liu, “PyramidBox: A context-assisted single shot face detector,” in *Computer Vision – ECCV 2018*, Cham: Springer International Publishing, 2018, pp. 812–828.
- [109] Z. Li, X. Tang, J. Han, J. Liu, and R. He, “PyramidBox++: High performance detector for finding tiny face,” *arXiv:1904.00386*, 2019.
- [110] K. Wang, Y. Dong, H. Bai, Y. Zhao, and K. Hu, “Use fast R-CNN and cascade structure for face detection,” in *2016 Visual Communications and Image Processing (VCIP)*, 2016, pp. 1–4.
- [111] J. Pang, K. Chen, J. Shi, H. Feng, W. Ouyang, and D. Lin, “Libra R-CNN: Towards balanced learning for object detection,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 821–830.
- [112] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, “Joint face detection and alignment using multitask cascaded convolutional networks,” *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1499–1503, 2016.
- [113] H. Li, Z. Lin, X. Shen, J. Brandt, and G. Hua, “A convolutional neural network cascade for face detection,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 5325–5334.
- [114] H. Qin, J. Yan, X. Li, and X. Hu, “Joint training of cascaded CNN for face detection,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 3456–3465.
- [115] Y. Zhiqi, “Face recognition based on improved VGGNET convolutional neural network,” in *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, 2021, pp. 2530–2533.
- [116] R. Anand, T. Shanthi, M. S. Nithish, and S. Lakshman, “Face Recognition and Classification Using GoogleNET Architecture,” in *Advances in Intelligent Systems and Computing*, Singapore: Springer Singapore, 2020, pp. 261–269.
- [117] A. Husain and V. P. Vishvakarma, “Face recognition method based on residual convolution neural network,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1228, no. 1, p. 012005, 2022.
- [118] P. Kalaiarasi and P. Esther Rani, “A comparative analysis of AlexNet and GoogLeNet with a simple DCNN for face recognition,” in *Advances in Intelligent Systems and Computing*,

Singapore: Springer Singapore, 2021, pp. 655–668.

- [119] S. Minaee, A. Abdolrashidi, H. Su, M. Bennamoun, and D. Zhang, “Biometrics recognition using deep learning: a survey,” *Artif. Intell. Rev.*, vol. 56, no. 8, pp. 8647–8695, 2023.
- [120] T. Gwyn, K. Roy, and M. Atay, “Face recognition using popular deep net architectures: A brief comparative study,” *Future Internet*, vol. 13, no. 7, p. 164, 2021.
- [121] S. I. Serengil and A. Ozpinar, “LightFace: A hybrid deep face recognition framework,” in *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, 2020, pp. 1–5.
- [122] S. Zhou, C. Chen, G. Han, and X. Hou, “Double additive margin softmax loss for face recognition,” *Appl. Sci. (Basel)*, vol. 10, no. 1, p. 60, 2019.
- [123] E. Lv, X. Wang, Y. Cheng, and Q. Yu, “Deep ensemble network based on multi-path fusion,” *Artif. Intell. Rev.*, vol. 52, no. 1, pp. 151–168, 2019.
- [124] G. P. Nam, H. Choi, J. Cho, and I.-J. Kim, “PSI-CNN: A pyramid-based scale-invariant CNN architecture for face recognition robust to various image resolutions,” *Appl. Sci. (Basel)*, vol. 8, no. 9, p. 1561, 2018.
- [125] Y. Yang and X. Song, “Research on face recognition technology fusion deep learning under different light intensity changes,” in *2021 5th Asian Conference on Artificial Intelligence Technology (ACAIT)*, 2021, pp. 329–332.
- [126] Y. Ma, Z. Huang, X. Wang, and K. Huang, “An overview of multimodal biometrics using the face and ear,” *Math. Probl. Eng.*, vol. 2020, pp. 1–17, 2020.
- [127] A. Lumini and L. Nanni, “Overview of the combination of biometric matchers,” *Inf. Fusion*, vol. 33, pp. 71–85, 2017.
- [128] F. Yang, “Development of a fast panoramic face mosaicking and recognition system,” *Opt. Eng.*, vol. 44, no. 8, p. 087005, 2005.
- [129] R. Singh, M. Vatsa, A. Ross, and A. Noore, “A Mosaicing Scheme for Pose-Invariant Face Recognition,” *IEEE Trans. Syst. Man Cybern. B Cybern.*, vol. 37, no. 5, pp. 1212–1225, 2007.
- [130] D. R. Kisku, J. K. Sing, M. Tistarelli, and P. Gupta, “Multisensor biometric evidence fusion for person authentication using wavelet decomposition and monotonic-decreasing graph,” in *2009 Seventh International Conference on Advances in Pattern Recognition*, 2009, pp. 205–208.
- [131] Z. Zhang, Y. Wang, and Z. Zhang, “Face synthesis from low-resolution near-infrared to high-resolution visual light spectrum based on tensor analysis,” *Neurocomputing*, vol. 140, pp. 146–154, 2014.

- [132] M. H.-M. Khan, V. Makoonlall, and V. S. Solayen, "Implementation and Analysis of Fusion in Multibiometrics," in *2019 13th International Conference on Open Source Systems and Technologies (ICOSST)*, 2019, pp. 1–7.
- [133] Z. Qin, P. Zhao, T. Zhuang, F. Deng, Y. Ding, and D. Chen, "A survey of identity recognition via data fusion and feature learning," *Inf. Fusion*, vol. 91, pp. 694–712, 2023.
- [134] M. Z. Rahman, M. H. Hafizur Rahman, and M. M. Rahman Majumdar, "Distinguishing a person by face and Iris using fusion approach," in *2019 International Conference on Sustainable Technologies for Industry 4.0 (STI)*, 2019, pp. 1–5.
- [135] R.-H. Jeng and W.-S. Chen, "Two feature-level fusion methods with feature scaling and hashing for multimodal biometrics," *IETE Tech. Rev.*, vol. 34, no. 1, pp. 91–101, 2017.
- [136] M. Haghighat, M. Abdel-Mottaleb, and W. Alhalabi, "Discriminant correlation analysis: Real-time feature level fusion for multimodal biometric recognition," *IEEE Trans. Inf. Forensics Secur.*, vol. 11, no. 9, pp. 1984–1996, 2016.
- [137] M. I. Ahmad, W. L. Woo, and S. Dlay, "Non-stationary feature fusion of face and palmprint multimodal biometrics," *Neurocomputing*, vol. 177, pp. 49–61, 2016.
- [138] X. Xing, K. Wang, and Z. Lv, "Fusion of gait and facial features using coupled projections for people identification at a distance," *IEEE Signal Process. Lett.*, vol. 22, no. 12, pp. 2349–2353, 2015.
- [139] L. C. O. Tiong, S. T. Kim, and Y. M. Ro, "Implementation of multimodal biometric recognition via multi-feature deep learning networks and feature fusion," *Multimed. Tools Appl.*, vol. 78, no. 16, pp. 22743–22772, 2019.
- [140] V. Talreja, M. C. Valenti, and N. M. Nasrabadi, "Deep hashing for secure multimodal biometrics," *IEEE Trans. Inf. Forensics Secur.*, vol. 16, pp. 1306–1321, 2021.
- [141] P. P. Sarangi, D. R. Nayak, M. Panda, and B. Majhi, "A feature-level fusion based improved multimodal biometric recognition system using ear and profile face," *J. Ambient Intell. Humaniz. Comput.*, vol. 13, no. 4, pp. 1867–1898, 2022.
- [142] N. Bala, R. Gupta, and A. Kumar, "Multimodal biometric system based on fusion techniques: a review," *Inf. Secur. J. Glob. Perspect.*, vol. 31, no. 3, pp. 289–337, 2022.
- [143] V. H. Patil, M. S. Dhole, "An efficient secure multimodal biometric fusion using palm print and face image," *International Journal of Applied Engineering Research*, vol. 11, no. 10, pp. 7147–50, 2016.
- [144] T. Gao, X.-M. Lei, and W. Hu, "Face recognition based on SIFT and LBP algorithm for

- decision level information fusion,” in *2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, 2017, pp. 2242–2246.
- [145] D. V. R. Devi and K. N. Rao, “Decision level fusion schemes for a Multimodal Biometric System using local and global wavelet features,” in *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, 2020, pp. 1–6.
 - [146] A. Matin, F. Mahmud, T. Ahmed, and M. S. Ejaz, “Weighted score level fusion of iris and face to identify an individual,” in *2017 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 2017, pp. 1–4.
 - [147] A. George and A. Routray, “A score level fusion method for eye movement biometrics,” *Pattern Recognit. Lett.*, vol. 82, pp. 207–215, 2016.
 - [148] O. Sharifi and M. Eskandari, “Optimal face-Iris multimodal fusion scheme,” *Symmetry (Basel)*, vol. 8, no. 6, p. 48, 2016.
 - [149] Y. Liang, X. Ding, C. Liu, and J.-H. Xue, “Combining multiple biometric traits with an order-preserving score fusion algorithm,” *Neurocomputing*, vol. 171, pp. 252–261, 2016.
 - [150] X. Zhang, D. Cheng, P. Jia, Y. Dai, and X. Xu, “An efficient android-based multimodal biometric authentication system with face and voice,” *IEEE Access*, vol. 8, pp. 102757–102772, 2020.
 - [151] H. Abderrahmane, G. Noubel, Z. Lahcene, Z. Akhtar, and D. Dasgupta, “Weighted quasi-arithmetic mean based score level fusion for multi-biometric systems,” *IET Biom.*, vol. 9, no. 3, pp. 91–99, 2020.
 - [152] A. Jain, K. Nandakumar, and A. Ross, “Score normalization in multimodal biometric systems,” *Pattern Recognit.*, vol. 38, no. 12, pp. 2270–2285, 2005.
 - [153] R. Sharma, S. Das, and P. Joshi, “Rank level fusion in multibiometric systems,” in *2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*, 2015, pp. 1–4.
 - [154] E. Marasco, A. Abaza, and B. Cukic, “Why rank-level fusion? And what is the impact of image quality?,” *Int. J. Big Data Intell.*, vol. 2, no. 2, p. 106, 2015.
 - [155] H. Talebi and M. L. Gavrilova, “Confidence based rank level fusion for multimodal biometric systems,” in *Computer Analysis of Images and Patterns*, Cham: Springer International Publishing, 2015, pp. 211–222.
 - [156] M. W. Rahman, F. T. Zohra, and M. L. Gavrilova, “Rank level fusion for Kinect gait and face biometric identification,” in *2017 IEEE Symposium Series on Computational Intelligence*

(SSCI), 2017, pp. 1–7.

- [157] J. K. Sing, A. Dey, and M. Ghosh, “Confidence factor weighted Gaussian function induced parallel fuzzy rank-level fusion for inference and its application to face recognition,” *Inf. Fusion*, vol. 47, pp. 60–71, 2019.
- [158] A. Abozaid, A. Haggag, H. Kasban, and M. Eltokhy, “Multimodal biometric scheme for human authentication technique based on voice and face recognition fusion,” *Multimed. Tools Appl.*, vol. 78, no. 12, pp. 16345–16361, 2019.
- [159] V. Azom, A. Adewumi, and J.-R. Tapamo, “Face and Iris biometrics person identification using hybrid fusion at feature and score-level,” in *2015 Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech)*, 2015, pp. 207–212.
- [160] E. Zangeneh, M. Rahmati, and Y. Mohsenzadeh, “Low resolution face recognition using a two-branch deep convolutional neural network architecture,” *Expert Syst. Appl.*, vol. 139, no. 112854, p. 112854, 2020.
- [161] Z. Lu, X. Jiang, and A. Kot, “Deep coupled ResNet for low-resolution face recognition,” *IEEE Signal Process. Lett.*, vol. 25, no. 4, pp. 526–530, 2018.
- [162] Y. Zhu and Y. Jiang, “Optimization of face recognition algorithm based on deep learning multi feature fusion driven by big data,” *Image Vis. Comput.*, vol. 104, no. 104023, p. 104023, 2020.
- [163] B. M. AlFawwaz, A. AL-Shatnawi, F. Al-Saqqar, and M. Nusir, “Multi-Resolution Discrete Cosine Transform fusion technique face recognition model,” *Data (Basel)*, vol. 7, no. 6, p. 80, 2022.
- [164] H. Wang, D. Zhang, and Z. Miao, “Fusion of LDB and HOG for Face Recognition,” in *2018 37th Chinese Control Conference (CCC)*, 2018, pp. 9192–9196.
- [165] L. C. O. Tiong, S. T. Kim, and Y. M. Ro, “Multimodal facial biometrics recognition: Dual-stream convolutional neural networks with multi-feature fusion layers,” *Image Vis. Comput.*, vol. 102, no. 103977, p. 103977, 2020.
- [166] J. Zhang, X. Yan, Z. Cheng, and X. Shen, “A face recognition algorithm based on feature fusion,” *Concurr. Comput.*, vol. 34, no. 14, 2022.
- [167] K. Juneja and C. Rana, “A feature fusion method for effective face recognition under variant illumination and noisy conditions,” in *Lecture Notes in Electrical Engineering*, Singapore: Springer Singapore, 2019, pp. 945–957.
- [168] J. Li, T. Qiu, C. Wen, K. Xie, and F.-Q. Wen, “Robust face recognition using the deep C2D-

- CNN model based on decision-level fusion,” *Sensors (Basel)*, vol. 18, no. 7, p. 2080, 2018.
- [169] J. H. Koo, S. W. Cho, N. R. Baek, M. C. Kim, and K. R. Park, “CNN-based multimodal human recognition in surveillance environments,” *Sensors (Basel)*, vol. 18, no. 9, p. 3040, 2018.
- [170] Y. Cheng, J. Zhao, Z. Wang, Y. Xu, K. Jayashree, S. Shen, and J. Feng, “Know you at one glance: A compact vector representation for low-shot learning,” in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, 2017, pp. 1924–1932.
- [171] N. Alay and H. H. Al-Baity, “Deep learning approach for multimodal biometric recognition system based on fusion of Iris, face, and finger vein traits,” *Sensors (Basel)*, vol. 20, no. 19, p. 5523, 2020.
- [172] S. Soleymani, A. Dabouei, H. Kazemi, J. Dawson, and N. M. Nasrabadi, “Multi-level feature abstraction from convolutional neural networks for multimodal biometric identification,” in *2018 24th International Conference on Pattern Recognition (ICPR)*, 2018, pp. 3469–3476.
- [173] J. Yang and S. J. Xie, *New trends and developments in biometrics*. BoD – Books on Demand, 2012.
- [174] S. Sahoo, T. Choubisa, and S. R. Mahadeva Prasanna, “Multimodal biometric person authentication : A review,” *IETE Tech. Rev.*, vol. 29, no. 1, p. 54, 2012.
- [175] M. M. Monwar and M. L. Gavrilova, “Multimodal biometric system using rank-level fusion approach,” *IEEE Trans. Syst. Man Cybern. B Cybern.*, vol. 39, no. 4, pp. 867–878, 2009.
- [176] AT&T face database, *AT&T Laboratories*, Cambridge, U. K, [Online] Available: <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>
- [177] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, “The FERET evaluation methodology for face-recognition algorithms,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, 2000.
- [178] PJ Phillips, The Facial Recognition Technology (FERET) database. <http://www.itl.nist.gov/iad/humanid/feret/feret_master.html> (2004).
- [179] D. B. Graham and N. M. Allinson, “Characterising virtual eigensignatures for general purpose face recognition,” in *Face Recognition*, Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, pp. 446–456.
- [180] <https://cmp.felk.cvut.cz/~spacelib/faces/faces94.html>
- [181] Y. Wong, S. Chen, S. Mau, C. Sanderson, and B. C. Lovell, “Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition,” in *CVPR*

- [182] K. Mittal, A. Jain, K. S. Vaisla, O. Castillo, and J. Kacprzyk, “A comprehensive review on type 2 fuzzy logic applications: Past, present and future,” *Eng. Appl. Artif. Intell.*, vol. 95, no. 103916, p. 103916, 2020.
- [183] V. Nguyen, T. Do, V.-T. Nguyen, T. D. Ngo, and D. A. Duong, “How to choose deep face models for surveillance system?,” in *Modern Approaches for Intelligent Information and Database Systems*, Cham: Springer International Publishing, 2018, pp. 367–376.
- [184] P. Huang, G. Gao, C. Qian, G. Yang, and Z. Yang, “Fuzzy linear regression discriminant projection for face recognition,” *IEEE Access*, vol. 5, pp. 4340–4349, 2017.
- [185] O. Castillo, M. Sanchez, C. Gonzalez, and G. Martinez, “Review of recent type-2 fuzzy image processing applications,” *Information (Basel)*, vol. 8, no. 3, p. 97, 2017.
- [186] A. Dey, S. Chowdhury, and J. K. Sing, “Performance evaluation on image fusion techniques for face recognition,” *Int. J. Comput. Vis. Robot.*, vol. 8, no. 5, p. 455, 2018.
- [187] M. M. Monwar and M. Gavrilova, “Markov chain model for multimodal biometric rank fusion,” *Signal Image Video Process.*, vol. 7, no. 1, pp. 137–149, 2013.
- [188] S. N. Tumpa and M. L. Gavrilova, “Score and rank level fusion algorithms for social behavioral biometrics,” *IEEE Access*, vol. 8, pp. 157663–157675, 2020.
- [189] A. Abaza and A. Ross, “Quality based rank-level fusion in multibiometric systems,” in *2009 IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems*, 2009, pp. 1–6.
- [190] A. Kumar, M. Hanmandlu, and S. Vasikarla, “Rank level integration of face based biometrics,” in *2012 Ninth International Conference on Information Technology - New Generations*, 2012, pp. 36–41.
- [191] A. Kumar and S. Shekhar, “Personal identification using multibiometrics rank-level fusion,” *IEEE Trans. Syst. Man Cybern. C Appl. Rev.*, vol. 41, no. 5, pp. 743–752, 2011.
- [192] L.A. Zadeh, “The concept of a linguistic variable and its application to approximate reasoning—1,” *Inform. Sci.* 8, 199–249, 1975.
- [193] J. M. Mendel, “Type-2 fuzzy sets and systems: an overview,” *IEEE Comput. Intell. Mag.*, vol. 2, no. 1, pp. 20–29, 2007.
- [194] J. M. Mendel, “General type-2 fuzzy logic systems made simple: A tutorial,” *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 5, pp. 1162–1182, 2014.

- [195] J. M. Mendel, F. Liu, and D. Zhai, “ α -plane representation for type-2 fuzzy sets: Theory and applications,” *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 5, pp. 1189–1207, 2009.
- [196] C. Wagner and H. Hagrais, “Toward general type-2 fuzzy logic systems based on zSlices,” *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 4, pp. 637–660, 2010.
- [197] C. Gafa and S. Coupland, “A new recursive type-reduction procedure for general type-2 fuzzy sets,” in *2011 IEEE Symposium on Advances in Type-2 Fuzzy Logic Systems (T2FUZZ)*, 2011, pp. 44–49.
- [198] F. Liu, “An efficient centroid type-reduction strategy for general type-2 fuzzy logic system,” *Inf. Sci. (Ny)*, vol. 178, no. 9, pp. 2224–2236, 2008.
- [199] F. Gomide, “Uncertain rule-based fuzzy logic systems: introduction and new directions - Jerry M. Mendel; Prentice-Hall, PTR, Upper Saddle River, NJ, 2001, 555pp., ISBN 0-13-040969-3,” *Fuzzy Sets And Systems*, vol. 1, no. 133, pp. 133–135, 2003.
- [200] L. A. Lucas, T. M. Centeno, and M. R. Delgado, “General type-2 fuzzy inference systems: Analysis, design and computational aspects,” in *2007 IEEE International Fuzzy Systems Conference*, 2007, pp. 1–6.
- [201] S. Greenfield, F. Chiclana, R. John, and S. Coupland, “The sampling method of defuzzification for type-2 fuzzy sets: Experimental evaluation,” *Inf. Sci. (Ny)*, vol. 189, pp. 77–92, 2012.
- [202] A. K. De, D. Chakraborty, and A. Biswas, “Literature review on type-2 fuzzy set theory,” *Soft Comput.*, vol. 26, no. 18, pp. 9049–9068, 2022.
- [203] J. K. Sing, S. Thakur, D. K. Basu, M. Nasipuri, and M. Kundu, “High-speed face recognition using self-adaptive radial basis function neural networks,” *Neural Comput. Appl.*, vol. 18, no. 8, pp. 979–990, 2009.
- [204] V. N. Vapnik, “An overview of statistical learning theory,” *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 988–999, 1999.
- [205] A. Ross, K. Nandakumar, and A. K. Jain, “Introduction to multibiometrics,” in *Handbook of Biometrics*, Boston, MA: Springer US, 2007, pp. 271–292.
- [206] A. K. Jain, P. Flynn, and A. A. Ross, *Handbook of Biometrics*. Springer Science & Business Media, 2007.
- [207] Y. Pan, L. Zhang, X. Wu, and M. J. Skibniewski, “Multi-classifier information fusion in risk analysis,” *Inf. Fusion*, vol. 60, pp. 121–136, 2020.
- [208] V. Bolón-Canedo and A. Alonso-Betanzos, “Ensembles for feature selection: A review and

- future trends,” *Inf. Fusion*, vol. 52, pp. 1–12, 2019.
- [209] T. Meng, X. Jing, Z. Yan, and W. Pedrycz, “A survey on machine learning for data fusion,” *Inf. Fusion*, vol. 57, pp. 115–129, 2020.
 - [210] G. Bebis, A. Gyaourova, S. Singh, and I. Pavlidis, “Face recognition by fusing thermal infrared and visible imagery,” *Image Vis. Comput.*, vol. 24, no. 7, pp. 727–742, 2006.
 - [211] S. G. Kong *et al.*, “Multiscale fusion of visible and thermal IR images for illumination-invariant face recognition,” *Int. J. Comput. Vis.*, vol. 71, no. 2, pp. 215–233, 2007.
 - [212] R. Singh, M. Vatsa, and A. Noore, “Integrated multilevel image fusion and match score fusion of visible and infrared face images for robust face recognition,” *Pattern Recognit.*, vol. 41, no. 3, pp. 880–893, 2008.
 - [213] H. D. Supreetha Gowda, G. Hemantha Kumar, and M. Imran, “Multimodal Biometric Recognition System Based on Nonparametric Classifiers,” in *Data Analytics and Learning*, Singapore: Springer Singapore, 2019, pp. 269–278.
 - [214] Popoola, O. P., & Lasisi, R. A., “A Biometric Fusion System of Face and Fingerprint for Enhanced Human Identification Using HOG-LBP Approach,” *Journal of Engineering Research*, VOL. 25, NO. 2, PP. 197–202, 2020.
 - [215] G. U. Bokade and R. D. Kanphade, “Secure multimodal biometric authentication using face, palmprint and ear: A feature level fusion approach,” in *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2019, pp. 1–5.
 - [216] L. Hong and A. Jain, “Integrating faces and fingerprints for personal identification,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 12, pp. 1295–1307, 1998.
 - [217] Q. Wang, B. Wang, X. Hao, L. Chen, J. Cui, R. Ji, Y. Lei, “Face recognition by decision fusion of two-dimensional linear discriminant analysis and local binary pattern,” *Front. Comput. Sci.*, vol. 10, no. 6, pp. 1118–1129, 2016.
 - [218] Z. Rahmi, M. I. Ahmad, M. N. M. Isa, and Z. I. A. Khalib, “Matching score level fusion for face and palmprint recognition system on spatial domain,” in *2019 IEEE 9th International Conference on System Engineering and Technology (ICSET)*, 2019, pp. 45–49.
 - [219] M. E. Rane and U. S. Bhadade, “Multimodal score level fusion for recognition using face and palmprint,” *Int. J. Electr. Eng. Educ.*, p. 002072092092966, 2020.
 - [220] D. V. R. Devi and K. N. Rao, “A multimodal biometric system using partition based DWT and rank level fusion,” in *2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, 2016, pp. 1–5.

- [221] S. Ahmad, R. Pal, and A. Ganivada, "Rank level fusion of multimodal biometrics using genetic algorithm," *Multimed. Tools Appl.*, vol. 81, no. 28, pp. 40931–40958, 2022.
- [222] T. K. Ho, J. J. Hull, and S. N. Srihari, "Decision combination in multiple classifier systems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 1, pp. 66–75, 1994.
- [223] D. J. Dubois, *Fuzzy Sets and systems: Theory and applications*. San Diego, CA: Academic Press, 1980.
- [224] L. A. Zadeh, "Linguistic variables, approximate reasoning and dispositions," *Med. Inform. (Lond.)*, vol. 8, no. 3, pp. 173–186, 1983.
- [225] Q. Liang and J. M. Mendel, "Interval type-2 fuzzy logic systems: theory and design," *IEEE Trans. Fuzzy Syst.*, vol. 8, no. 5, pp. 535–550, 2000.
- [226] J. E. Moreno *et al.*, "Design of an interval Type-2 fuzzy model with justifiable uncertainty," *Inf. Sci. (Ny)*, vol. 513, pp. 206–221, 2020.
- [227] J. M. Mendel and R. I. B. John, "Type-2 fuzzy sets made simple," *IEEE Trans. Fuzzy Syst.*, vol. 10, no. 2, pp. 117–127, 2002.
- [228] D. Wu and J. M. Mendel, "Recommendations on designing practical interval type-2 fuzzy systems," *Eng. Appl. Artif. Intell.*, vol. 85, pp. 182–193, 2019.
- [229] M. Ghosh and J. K. Sing, "Multi-feature-based type-2 fuzzy set induced parallel rank-level fusion in face recognition," in *Soft Computing: Theories and Applications*, Singapore: Springer Nature Singapore, 2022, pp. 181–190.
- [230] S. S. Haykin, *Neural Networks: A comprehensive foundation*, 2nd ed. Upper Saddle River, N.J. : Prentice Hall, 1999.
- [231] J. K. Sing, D. K. Basu, M. Nasipuri, and M. Kundu, "Face recognition using point symmetry distance-based RBF network," *Appl. Soft Comput.*, vol. 7, no. 1, pp. 58–70, 2007.
- [232] E. Hisdal, "The IF THEN ELSE statement and interval-valued fuzzy sets of higher type," *Int. J. Man. Mach. Stud.*, vol. 15, no. 4, pp. 385–455, 1981.
- [233] G. Y. Chen, "An experimental study for the effects of noise on face recognition algorithms under varying illumination," *Multimed. Tools Appl.*, vol. 78, no. 18, pp. 26615–26631, 2019.
- [234] C. Ding and D. Tao, "A comprehensive survey on pose-Invariant Face Recognition," *ACM Trans. Intell. Syst. Technol.*, vol. 7, no. 3, pp. 1–42, 2016.
- [235] M. M. Sawant and K. M. Bhurchandi, "Age invariant face recognition: a survey on facial aging databases, techniques and effect of aging," *Artif. Intell. Rev.*, vol. 52, no. 2, pp. 981–1008,

2019.

- [236] D. Zeng, R. Veldhuis, and L. Spreeuwers, "A survey of face recognition techniques under occlusion," *IET Biom.*, vol. 10, no. 6, pp. 581–606, 2021.
- [237] K. Grm, V. Štruc, A. Artiges, M. Caron, and H. K. Ekenel, "Strengths and weaknesses of deep learning models for face recognition against image degradations," *IET Biom.*, vol. 7, no. 1, pp. 81–89, 2018.
- [238] P. Punyani and A. Kumar, "Evaluation of fusion at different levels for face recognition," in *2017 International Conference on Computing, Communication and Automation (ICCCA)*, 2017, pp. 1052–1055.
- [239] T. Baltrusaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 2, pp. 423–443, 2019.
- [240] A. S. O. Ali, "Age invariant face recognition using feature level fused morphometry of lip-nose and periocular region features," *LJRCST*, vol. 23, no. 1, pp. 43–68, 2023.
- [241] S. Ahmad, R. Pal, and A. Ganivada, "Score level fusion of multimodal biometrics using genetic algorithm," in *2021 IEEE Congress on Evolutionary Computation (CEC)*, 2021, pp. 2242–2250.
- [242] M. Eskandari and O. Sharifi, "Effect of face and ocular multimodal biometric systems on gender classification," *IET Biom.*, vol. 8, no. 4, pp. 243–248, 2019.
- [243] M. Ghosh and J. K. Sing, "Interval type-2 fuzzy set induced fuzzy rank-level fusion for face recognition," *Appl. Soft Comput.*, vol. 145, no. 110584, p. 110584, 2023.
- [244] A. Mohammadzadeh, M. H. Sabzalian, and W. Zhang, "An interval type-3 fuzzy system and a new online fractional-order learning algorithm: Theory and practice," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 9, pp. 1940–1950, 2020.
- [245] D. Singh, N. Verma, A. Ghosh, and A. Malagaudanavar, "An approach towards the design of interval type-3 T–S fuzzy system," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 9, pp. 3880–3893, 2022.
- [246] O. Castillo, J. R. Castro, and P. Melin, "Interval type-3 fuzzy fractal approach in sound speaker quality control evaluation," *Eng. Appl. Artif. Intell.*, vol. 116, no. 105363, p. 105363, 2022.
- [247] M. Gheisarnejad, A. Mohammadzadeh, and M.-H. Khooban, "Model predictive control based type-3 fuzzy estimator for voltage stabilization of DC power converters," *IEEE Trans. Ind. Electron.*, vol. 69, no. 12, pp. 13849–13858, 2022.

- [248] C. Peraza, P. Ochoa, O. Castillo, and Z. W. Geem, "Interval-type 3 fuzzy differential evolution for designing an interval-type 3 fuzzy controller of a unicycle mobile robot," *Mathematics*, vol. 10, no. 19, p. 3533, 2022.
- [249] A. Taghieh, A. Mohammadzadeh, C. Zhang, N. Kausar, and O. Castillo, "A type-3 fuzzy control for current sharing and voltage balancing in microgrids," *Appl. Soft Comput.*, vol. 129, no. 109636, p. 109636, 2022.
- [250] Z. Liu, A. Mohammadzadeh, H. Turabieh, M. Mafarja, S. S. Band, and A. Mosavi, "A new online learned interval type-3 fuzzy control system for solar energy management systems," *IEEE Access*, vol. 9, pp. 10498–10508, 2021.
- [251] R. H. Vafaie, A. Mohammadzadeh, and M. J. Piran, "A new type-3 fuzzy predictive controller for MEMS gyroscopes," *Nonlinear Dyn.*, vol. 106, no. 1, pp. 381–403, 2021.
- [252] G. M. Mendez, I. Lopez-Juarez, P. N. Montes-Dorantes, and M. A. Garcia, "A new method for the design of interval type-3 fuzzy logic systems with uncertain type-2 non-singleton inputs (IT3 NSFLS-2): A case study in a hot strip mill," *IEEE Access*, vol. 11, pp. 44065–44081, 2023.
- [253] O. Castillo, J. R. Castro, and P. Melin, "Interval type-3 fuzzy aggregation of neural networks for multiple time series prediction: The case of financial forecasting," *Axioms*, vol. 11, no. 6, p. 251, 2022.
- [254] J. Aisbett, J. T. Rickard, and D. G. Morgenthaler, "Type-2 fuzzy sets as functions on spaces," *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 4, pp. 841–844, 2010.
- [255] N. N. Karnik, J. M. Mendel, and Q. Liang, "Type-2 fuzzy logic systems," *IEEE Trans. Fuzzy Syst.*, vol. 7, no. 6, pp. 643–658, 1999.
- [256] O. Castillo, J. R. Castro, and P. Melin, "A methodology for building interval type-3 fuzzy systems based on the principle of justifiable granularity," *Int. J. Intell. Syst.*, vol. 37, no. 10, pp. 7909–7943, 2022.
- [257] S. P. K. Wickrama Arachchilage and E. Izquierdo, "Deep-learned faces: a survey," *EURASIP J. Image Video Process.*, vol. 2020, no. 1, 2020.
- [258] N. A. Jamil, U. U. Sheikh, M. M. Mokji, Z. Omar, and A. A. H. A. Rahman, "Improved face recognition on video surveillance images using pose correction," in *TENCON 2017 - 2017 IEEE Region 10 Conference*, 2017, pp. 1391–1395.
- [259] N. Zhang, J. Luo, and W. Gao, "Research on face detection technology based on MTCNN," in *2020 International Conference on Computer Network, Electronic and Automation*

(ICCNEA), 2020, pp. 154–158.

- [260] T. Edirisooriya and E. Jayatunga, “Comparative study of face detection methods for robust face recognition systems,” in *2021 5th SLAAI International Conference on Artificial Intelligence (SLAAI-ICAI)*, 2021, pp. 1–6.
- [261] O. Nguyen, K. Nguyen, and T. V. Pham, “A comparative study on application of multi-task cascaded convolutional network for robust face recognition,” *Udn.vn*. [Online]. Available: <https://elib.vku.udn.vn/bitstream/123456789/959/1/B5.2019-31-38.pdf>. [Accessed: 02-Jul-2024].
- [262] Q. Guo, Z. Wang, C. Wang, and D. Cui, “Multi-face detection algorithm suitable for video surveillance,” in *2020 International Conference on Computer Vision, Image and Deep Learning (CVIDL)*, 2020, pp. 27–33.
- [263] M. Ma and J. Wang, “Multi-view face detection and landmark localization based on MTCNN,” in *2018 Chinese Automation Congress (CAC)*, 2018, pp. 4200–4205.
- [264] S. S. Khan, D. Sengupta, A. Ghosh, and A. Chaudhuri, “MTCNN++: A CNN-based face detection algorithm inspired by MTCNN,” *Vis. Comput.*, vol. 40, no. 2, pp. 899–917, 2024.
- [265] M. T. H. Fuad *et al.*, “Recent advances in deep learning techniques for face recognition,” *IEEE Access*, vol. 9, pp. 99112–99142, 2021.
- [266] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818–2826.
- [267] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
- [268] J. Yang, P. Ren, D. Zhang, D. Chen, F. Wen, H. Li, and G. Hua, “Neural aggregation network for video face recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4362–4371.
- [269] T. Thaipisutikul, P. Tatiyamaneekul, C.-Y. Lin, and S. Tuarob, “A deep feature-level fusion model for masked face identity recommendation system,” *J. Ambient Intell. Humaniz. Comput.*, vol. 14, no. 12, pp. 16011–16024, 2023.
- [270] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv [cs.CV]*, 2014.
- [271] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, “Inception-v4, Inception-ResNet and the impact of residual connections on learning,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 31, no. 1, 2017.

- [272] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1800–1807.
- [273] G. Barquero, C. Fernandez, and I. Hupont, "Long-term face tracking for crowded video-surveillance scenarios," in *2020 IEEE International Joint Conference on Biometrics (IJCB)*, 2020, pp. 1–8.

Manas Ghosh
11.07.2024

Janip
11/07/24