Emerging Covariates of Face Recognition

By

Himanshu Sharad Bhatt

Advisors

Richa Singh, PhD
Mayank Vatsa, PhD
Afzel Noore, PhD
Nalini K. Ratha, PhD

Dissertation submitted to the
Indraprastha Institute of Information Technology-Delhi (IIIT-Delhi)
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Computer Science & Engineering

February, 2014

Keywords: Face Recognition, Covariates, Co-transfer Learning, Plastic Surgery, Forensic Sketches, Cross-resolution Images, Video Based Face Recognition.

Abstract

A covariate in face recognition can be defined as an effect that independently increases the intra-class variability or decreases the inter-class variability or both. Covariates such as pose, illumination, expression, aging, and disguise are established and extensively studied in literature and are categorized as existing covariates of face recognition. However, ever increasing applications of face recognition have instigated many new and exciting scenarios such as matching forensic sketches to mug-shot photos, faces altered due to plastic surgery, low resolution surveillance images, and individual from videos. These covariates are categorized as emerging covariates of face recognition, which is the primary emphasis of this dissertation. One of the important cues in solving crimes and apprehending criminals is matching forensic sketches with digital face images. The first contribution of this dissertation is a memetically optimized multi-scale circular Weber's local descriptor (MCWLD) for matching forensic sketches with digital face images. This dissertation presents an automated algorithm to extract discriminative information from local regions of both sketches and digital images using MCWLD. An evolutionary memetic optimization is proposed to assign optimal weights to every local facial region to boost the identification performance. Since, forensic sketches and digital images can be of poor quality, a pre-processing technique is also used to enhance the quality of images. Results on different sketch databases, including forensic sketch database, illustrate the efficacy of the proposed algorithm. Widespread acceptability and use of biometrics for person authentication has instigated several techniques for evading identification such as altering facial appearance using surgical procedures. These procedures modify both the shape and texture of facial features to varying degrees and thus degrade the performance of face recognition when matching pre- and post-surgery images. The second contribution of this dissertation is a multi-objective evolutionary granular algorithm for matching face images altered due to plastic surgery procedures.

The algorithm first generates non-disjoint face granules at multiple levels of granularity. The granular information is assimilated using a multi-objective genetic algorithm that simultaneously optimizes the selection of feature extractor for each face granule along with the weights of individual granules. On IIIT-D plastic surgery database, the proposed algorithm yields the state-of-the-art performance. Face recognition performance degrades when a low resolution face image captured in unconstrained settings, such as surveillance, is matched with high resolution gallery images. The primary challenge is to extract discriminative features from the limited biometric content in low resolution images and match it with information-rich high resolution face images. The problem of cross-resolution face matching is further alleviated when there is limited labeled low resolution training data. The third contribution of this dissertation is co-transfer learning framework, a cross pollination of transfer learning and co-training paradigms, for enhancing the performance of cross-resolution face recognition. The transfer learning component transfers the knowledge that is learned while matching high resolution face images during training for matching low resolution probe images with high resolution gallery during testing. On the other hand, co-training component facilitates this knowledge transfer by assigning pseudo labels to unlabeled probe instances in the target domain. Experiments on a synthetic, three low resolution surveillance quality face databases, and real world examples show the efficacy of the proposed co-transfer learning algorithm as compared to other approaches. Due to prevalent applications and availability of large intra-personal variations, videos have gained significant attention for face recognition. Unlike still face images, videos provide abundant information that can be leveraged to compensate for variations in intra-personal variations and enhance face recognition performance. The fourth contribution of this dissertation is a video based face recognition algorithm which computes a discriminative video signature as an ordered (ranked) list of still face images from a large dictionary. A three stage approach is developed for optimizing ranked lists across multiple video frames and fusing them into a single composite ordered list to compute the video signature. The signature embeds diverse intra-personal variations and facilitates in matching two videos across large variations. Results obtained on Youtube and MBGC v2 video databases show the effectiveness of the proposed algorithm.

Dedicated to my family and friends.

Acknowledgements

As I complete my PhD degree, I would like to thank people around me who have been thoughtful, caring, and helpful. The best and worst moments of this journey are shared with many people. I would rather say that this dissertation is a product of the most cherished phase of my life i.e. my PhD life.

Above all, I would like to express my deepest gratitude to my PhD advisors, Dr. Richa Singh and Dr. Mayank Vatsa. They are my first mentor in this journey and taught me many things. Their vision, encouragement, and constant motivation always kept me enthused to take up challenges and make the best out of the opportunities. I am grateful to my advisors for taking extra efforts in providing the best environment, facilities, and support during the course of my doctoral program. I always look up to them for their knowledge, insightful discussions, and suggestions in my PhD and beyond.

I would like to thank Prof. Afzel Noore for providing me a unique opportunity to work under his guidance. His perspective on conducting research, vast knowledge, critical reasoning, and constructive feedbacks during our discussions helped me to grow as a better researcher. I will always be grateful to him for inviting me to West Virginia University and expanding my prospects and thoughts about research.

I would like to express my gratitude to Dr. Nalini K. Ratha for his constant support and feedback. I always appreciated the discussions with him for his understanding of problems, their implications, and his viewpoint, being from the industry. The perspective he brings with his understanding of biometrics and immense experience has always been an indispensable source of learning.

I would like to acknowledge Dr. A. Lanitis, Dr. A. Martinez, Dr. X. Wang, Mislav Grgic, Ralph Gross, CVRL University of Notre Dame, and NICTA for granting us access to the face databases used in this research.

My grandmother, parents, sister, and brother-in-law bestowed me with unconditional love and support for which an expression of thanks does not suffice. I

would also like to thank all my friends from IIIT-Delhi and “Chestnut” family, especially, Sam, KD, Tejas, Anush, Praful, Shruti, Megha, Paridhi, Anjali, and Raj. I would like to give special mention to Sam for being there with me in the worst as well as the best times of my life. I am thankful to the shooting star which now and then fulfils all my wishes. I hope that this work makes my family and friends proud.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

One of the most common visual patterns that one comes across every day is a human face. Humans have a remarkable property of recognizing faces and identifying a face appears to be one of the most effortless human activities. For the last four decades, face recognition has been an active research problem and researchers have been motivated to develop algorithms to emulate the recognition capability of human mind [12]. Imparting this intelligence to machines has led to the development of several automated face recognition algorithms. However, human face is not a rigid object and can have a lot of differences due to inter-personal or intra-personal variations. Inter-personal variations can be attributed to changes in race or genetics, while intra-personal variations can be attributed to changes in pose, illumination, expression, aging, hair, cosmetics, and facial accessories.

As a part of *identity science*, face biometrics has the benefit of being non-intrusive and passive as compared to other biometric modalities such as fingerprint and iris. Face images can be easily captured from a distance without much co-operation from the user. It has received a lot of attention from both academicians as well as industry because of its ever increasing applications in surveillance, access control, law enforcement, cross border security, multimedia, forensics, and many more. With advancements in technology and reduction in sensor cost (camera), new applications of face recognition have become prevalent. Verification based on face images captured through built-in cameras is used to allow access to personal devices such as laptops and mobile phones. With development in face recognition technology, it is now used for cross border security. Hong Kong-SAR border has the worlds first drive-thru face recognition system. Smart-Gate at Australia, US Visit, and Japan Visit programs also collect face for all visitors. Face recognition is also used in kiosk applications to allow access to ATM machines, server rooms, and e-commerce applications (online banking). Face recognition has found applications in large

social welfare programs where a new user is matched against all existing users to check for duplicates. Currently, two states in United States (Massachusetts and Connecticut) use face identification for large scale de-duplication. In India, UIDAI is also collecting face biometric (along with other biometric modalities) to issue a unique identification number to all the citizens. Face, being a non-invasive biometric, is widely used for surveillance. In surveillance applications, face images are captured without active co-operation from the user and are matched to a watch list database of individuals. Surveillance cameras now have a profound presence at public places like airports, railway stations, shopping malls, and banks. A face recognition system has different stages. As shown in Figure 1.1, face recognition [13] starts by detecting the facial region in an image, i.e. face detection. Once the face is detected, features are extracted to generate a template that captures the discriminative information from the face image. The template of the probe (query) image is then matched with the templates stored in database. The match scores thus obtained are used to establish the identity of an individual. Depending on the context, a face recognition system can operate either in a verification (1:1 matching) or an identification (1:N matching) mode. Verification involves confirming or denying the identity claimed by an individual, whereas identification involves determining the identity of an individual from a list of N individuals enrolled in the database.

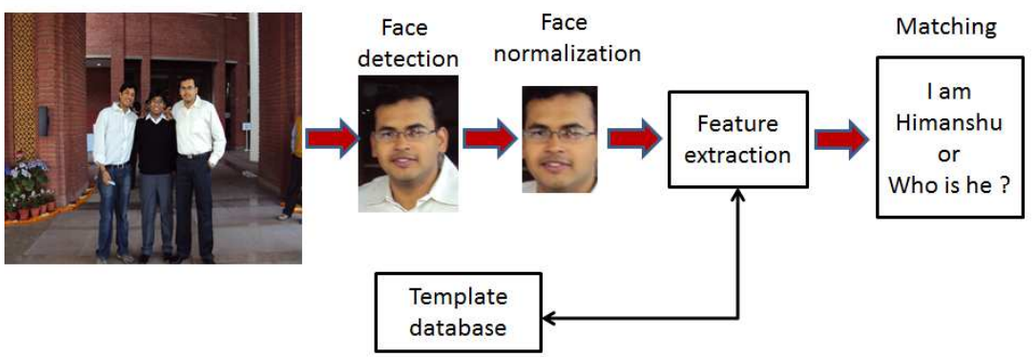

Figure 1.1: Illustrating different stages in a face recognition system i.e. image acquisition, face detection, face normalization, feature extraction, and matching.

1.1 Face Recognition Literature

As mentioned previously, face detection is the first stage in an automated face recognition system. Given an input image, the goal of face detection is to detect all the faces present in the image irrespective of its position, orientation, and lighting conditions. Face detection is a challenging task because of the variability in scale, location, orientation, facial expression, occlusion, and lighting conditions. Yang *et al.* [14] categorized the techniques for face detection into four classes: knowledge-based, feature invariant, template matching, and appearance-based methods. Knowledge-based methods comprise a set of rules that encode human knowledge of what constitutes a face and generally consist of relationships between facial features. Feature invariant methods aim to find structural features that exist even when the pose and lighting conditions vary, and then use these features to locate faces. In template matching methods, standard patterns of a face are stored and the correlations between an input image and the stored patterns are used for detection. In appearance-based methods, models (or templates) learned from a set of training images to capture the representative variability of facial appearance are used for face detection. Further in literature, there has been two widely used face detectors: 1) proposed by Rowley *et al.* [15] and 2) proposed by Viola and Jones [16]. Face detector proposed by Rowley *et al.* [15] is a neural network based technique which is fast and efficient. On the other hand, the Adaboost face detector proposed by Viola and Jones [16] uses Haar-like features along with a cascade of boosted decision tree classifiers as a statistical model which is fast, reliable and computationally less expensive. Zhang and Zhang [17] presented a survey on recent advances in face detection where several techniques are categorized based on the feature extraction and learning algorithms utilized for robust face detection.

From the detected face images, facial features are extracted which are matches with the stored templates (database). In an attempt to categorize different face recognition algorithms, Klare and Jain [18] proposed the taxonomy of facial features by grouping the salient information available in 2D face images into feature categories: level 1, level 2, and level 3. Level 1 facial features capture the holistic nature of a face such as skin color, gender, and general appearance of a face such as principal component analysis (PCA) and linear discriminant analysis (LDA) approaches. Level 2 features are locally derived and describe facial structures that are relevant for recognition such as Gabor wavelets, local binary patterns (LBP), and scale invariant feature transform (SIFT). Level 2 features are the most discriminative face features and are predominantly used for face recognition. Level 3 features comprise unstructured micro level features on the face such as scars, moles,

and facial marks. These features are especially efficient for matching look-alike faces [19], biologically identical twins [20, 21, 22] and faces across different age variations. Most of the research effort has gone into level 1 and 2 features and it is quite recent that level 3 information is used in applications where level 1 and 2 features cannot perform efficient face recognition.

Zhao *et al.* [23] presented a survey of face recognition algorithms and existing challenges. They categorized the face recognition techniques (using still images) into holistic, feature based, and hybrid approaches. Holistic approaches use the global appearance of a face image and extract features from the full face, whereas in feature based approaches, local features such as eyes, nose, and mouth are extracted and their characteristics such as local geometry and appearance are utilized. Hybrid approaches, based on human perception, use both local features and the full facial region for recognition. They have identified pose and illumination variations as the two major issues in face recognition. Kong *et al.* [24] divided techniques for face recognition into visible and infrared domain. They presented a review of 2D face recognition techniques in visible spectrum and showed that these algorithms can achieve significant performance in controlled settings with cooperative users. However, the performance of these algorithms degrade when face images are captured in uncontrolled environment with large variations in pose, illumination, and expression. Their survey also presents a comprehensive review of algorithms proposed for robust face recognition in infrared imagery. Several approaches such as detecting disguise variations using thermal imagery and multi-spectral fusion for illumination normalization are presented. Face recognition techniques in infrared imagery have shown to improve the overall performance in uncontrolled environments; however, one limitation of infrared sensing methods is their high dependency on the environmental illumination.

Belhumeur [25] presented some ongoing challenges in face recognition such as pose, illumination, and expression and described several techniques proposed to address these challenges. Techniques proposed for matching face images across these variations are categorized as feature based, appearance based, and 3D face recognition techniques. Feature based methods using geometric relations (e.g. distances and angles) between facial features such as eyes, mouth, nose, and chin are used for efficient face recognition because of their economical representation. However, feature-based methods are dependent on the reliability of the feature extraction algorithm. The subspace based methods differ from feature-based techniques in their low-dimensional representation. Subspace based methods recognize a face only if the face has been previously seen under similar circumstances. In 3D face recognition, the images acquired during enrollment are used to estimate the

models of the 3D shape of a face. These 3D models can then be used to synthetically render the images of each face under arbitrary pose and lighting conditions, effectively increasing the gallery set for each subject. Out of several approaches proposed for face recognition, 3D face recognition has gone a long way towards addressing challenges due to pose, lighting, and expression. This observation is also discussed by Abate *et al.* [26] in their comprehensive review of techniques for 2D and 3D face recognition.

The progression in face recognition literature has been analyzed from different points of view. As discussed above, one view categorizes face recognition algorithms into holistic, feature-based, and hybrid techniques, another categorizes them as visible and infrared domain techniques; one divides them as 2D and 3D face recognition techniques whereas, the other proposes a grouping based on the salient information into hierarchical feature category. There are also a few papers that review face recognition across pose variations [27], illumination variations [25, 28], aging [29] and forensic applications [30]. The progression in face recognition is dependent on the availability of large publicly available databases. Table 1.1 lists the widely used databases in face recognition literature. With the availability of the CMU PIE [31] and CMU Multi-PIE [9] databases, there has been a significant development in the algorithms for addressing pose, illumination, and expression variations. Similarly, with the availability of databases for age variations such as FG-NET [32] and MORPH [33], researchers are trying to model the biological process of aging for simulating age and developing age invariant face recognition algorithms. The real application of face recognition involves matching face images in unconstrained setting such as arbitrary pose, uncontrolled illumination and expression or in the presence of one or more covariate simultaneously. Recent advances in this direction led to development of few large unconstrained databases such as LFW [34], Pubfig [35], YouTube faces [6]. These large scale unconstrained databases facilitate development and evaluation of algorithms for the more general application of face recognition.

Table 1.1: List of widely used publicly available face databases for different covariates.

Covariate	Database	Description
Pose	CMU-PIE [31]	13 poses within ±66° in yaw and ±15° in tilt.
	AT&T ([36])	10 random poses within ± 20° in yaw and tilt.
	XM2VTS [37]	5 poses:0° ±30° in yaw and tilt.
	Multi-PIE [9]	15 poses within ±66° in yaw and ±15° in tilt.
	FERET [38]	18 poses, $0° to \pm 90°$.
	Yale-B [39]	9 different poses $0°, 12°$ and $24°$.
	CAS-PEAL [40]	21 different poses.
	CMU-PIE [31]	Different illumination from 13 light sources.
Illumination	AR [41]	Left, right, and all side lights on.
	CASIA-FaceV5 [42]	2,500 color facial images of 500 subjects.
	Yale-B [39]	64 lighting conditions and 1 ambient illumination.
	CASIA NIR [43]	3,940 images of 197 subjects.
	PolyU NIR [44]	34,000 images of 335 subjects.
Expression	JAFFE [45]	7 different facial expressions.
	Cohn-Kanade [46]	Neutral to a peak expression.
	BU-3DFE [47]	7 expression variations for 100 subjects.
	CMU-AMP ([48])	75 images showing different expressions.
Aging	FG-NET ([32])	6-18 images per subject from 0-69 years of age.
	Morph [33]	46 days to 29 years.
Sketch	CUHK face sketch [49]	606 viewed sketches.
	CUHK face sketch FERET [50]	1194 viewed sketches.
	IIIT-D sketch database [51]	238 viewed sketches, 140 semi-forensic sketches, and 6 forensic sketches.
Cosmetics	IIIT-D plastic surgery [8]	900 subjects with different plastic surgery cases.
	YouTube makeup [52]	2 images before makeup and 2 images after makeup for 151 subjects.
	Virtual Makeup [52]	204 images for 51 subjects.
	Makeup in the "wild" [53]	154 images with and without makeup.
Look-alikes & Twins	IIIT-D look-alike [19]	50 subjects with 5 genuine and 5 look-alikes.
	3D twins expression challenge [54]	428 images of 107 twin pairs.
	ND-Twins [21]	24050 color photographs of the faces of 435 attendees at the Twins Days Festivals.
Low resolution/video	SCface [10]	4160 surveillance images of 130 subjects.
	MBGC video challenge problem ([7])	399 walking sequence and 202 standard sequence (720×480).
	Honda UCSD [55, 56]	Dataset 1: 75 video of 20 subjects.
		Dataset 2: 30 video of 15 subjects.
	ChokePoint [11]	54 video sequences and 64,204 labeled face images.
	COX-S2V [57]	1000, subjects with 1 high quality photo and 4 surveillance video.
Unconstrained	Labeled faces in the wild [34]	13,233 images of 5749 subjects.
	PubFig [35]	58,797 images of 200 subjects.
	YouTube faces [6]	3,425 videos of 1,595.

1.2 Covariates of Face Recognition

The generality in the applications of face recognition introduces several challenges (covariates) such as pose, illumination, expression, aging, and disguise. These covariates significantly alter the inter-class and intra-class dynamics thus resulting in reduced face recognition performance. Figure 1.2 illustrates the concepts of intra-class and inter-class variability. Based on it, *a covariate in face recognition is defined as an effect that independently increases the intra-class variability or decreases the inter-class variability or both.*

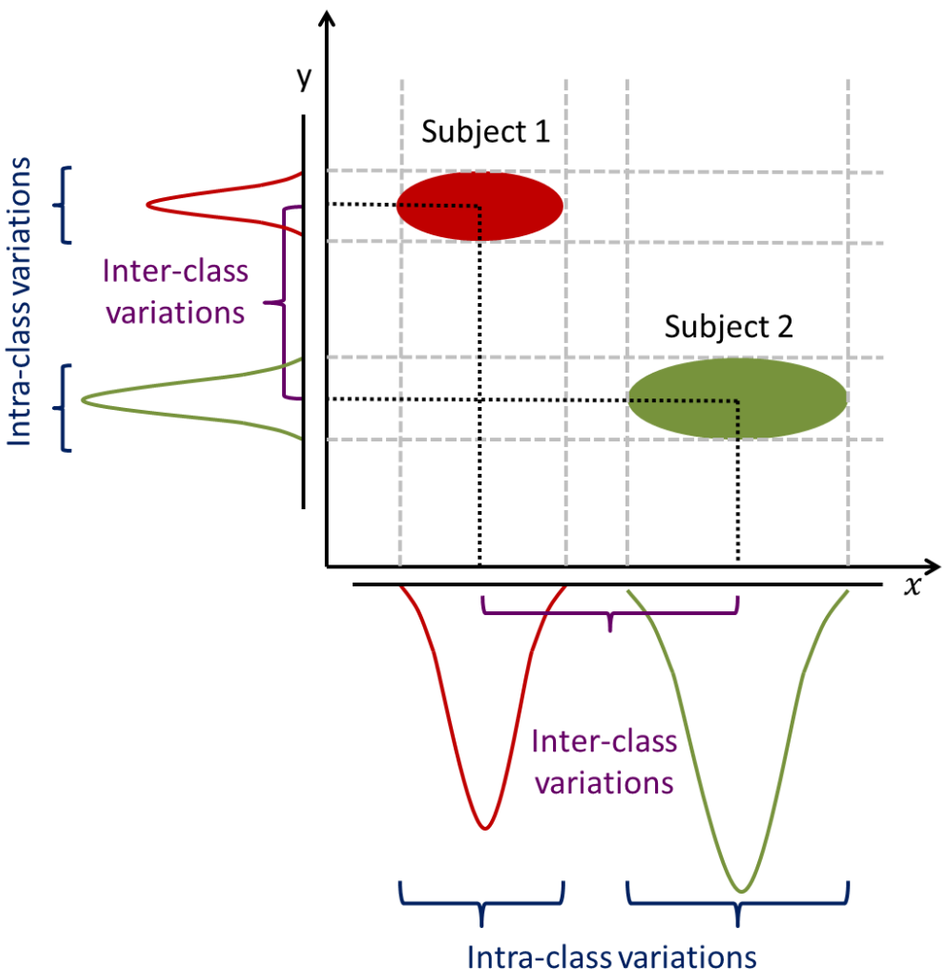

Figure 1.2: Illustrating the concepts of inter-class and intra-class variations in biometrics.

A robust face recognition system should be able to identify faces captured in uncontrolled environments such as images captured in unpredictable lighting conditions, different angles and distances from the camera, and where the subjects are non-cooperative. However, the performance of current face recognition systems significantly deteriorate for such uncontrolled but real-world conditions. Based on the applications of face recognition and how extensively different covariates have been studied in literature, the covariates of face recognition are classified into two categories: 1) existing covariates and 2) emerging covariates. Figure 1.3 shows different types of existing and emerging covariates.

1. ***Existing covariates***: In past, face recognition literature has focused on certain covariates of face recognition and a lot of research has been performed to address them [23, 25, 27, 28]. These covariates are termed as the existing covariates of face recognition and are briefly listed below.

    - Variations due to *pose* and *illumination* may camouflage some of the features and lead to incorrect recognition results.
    - Variations in *expression* can cause deformations in local facial structure and change the facial appearance and local geometry, thereby reducing the face recognition performance.
    - Many applications of face recognition require matching face images with variations in age such as matching a recent photo with image on passport or driver's license. Facial *aging* is a biological process that leads to gradual changes in the structural geometry and texture of a face.
    - *Disguise* is the process of concealing one's identity or impersonating another person by using makeup and other accessories. Law enforcement applications often require identifying individuals who try to conceal their identities using disguise accessories. Both aging and disguise can lead to large variations in intra-class and inter-class distributions and hence, degrade the face recognition performance. Many researchers are working to develop algorithms to efficiently match face images with variations in age [58, 59, 60] and disguise [61, 62].

2. ***Emerging Covariates***: With ever increasing applications of face recognition, there has emerged a need to understand and address new as well as fascinating challenges of face recognition. Since these challenges have been recently established and researchers are now developing algorithms to mitigate their effects, we term them as

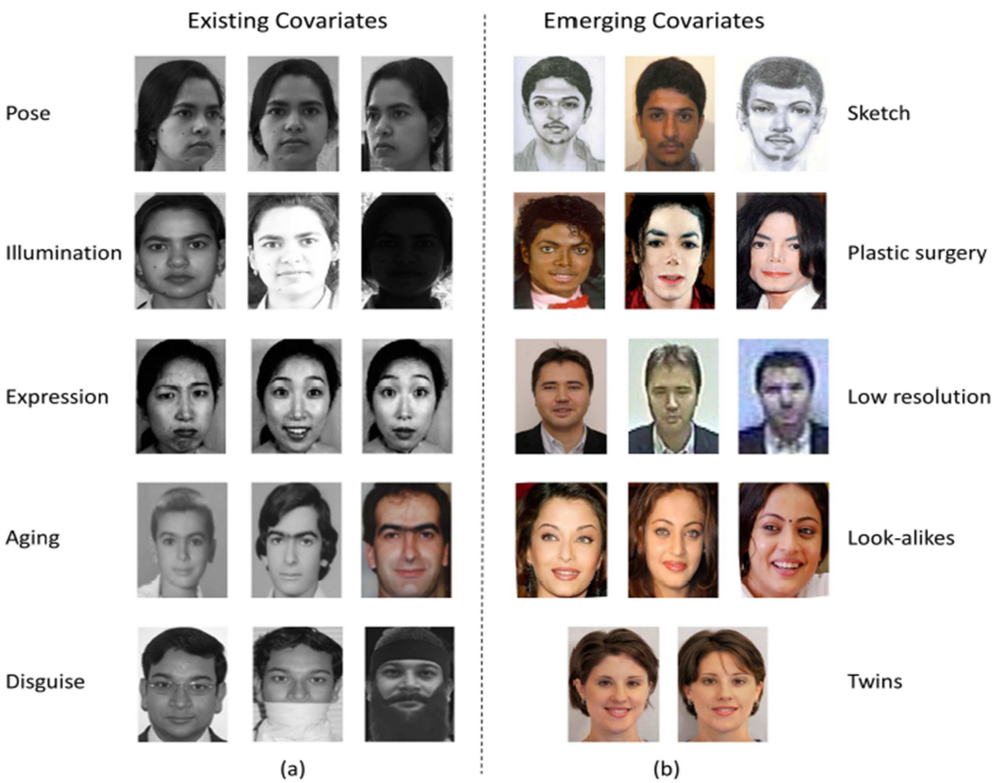

Figure 1.3: Covariates of face recognition: (a) existing covariates and (b) emerging covariates.

emerging covariates of face recognition. It has become essential for current face recognition algorithms to be robust in the presence of these emerging covariates as well.

- An important law enforcement application is *matching forensic sketches to digital face images* of known individuals. Law enforcement agencies often find the need to determine the identity of sketches obtained from the description of eye witness from the crime scene. This problem introduces a new emerging covariate of matching forensic sketches with the mugshots in the database [4, 51].
- In recent years, *facial plastic surgery* has also emerged as an important covariate of face recognition [8]. Plastic surgery is a spontaneous process and its effects are generally contrary to that of facial aging. Variations caused due to plastic surgery are long-lasting and may not be reversible. Therefore, plastic surgery poses a huge challenge for existing face recognition algorithms.
- For public safety and security, surveillance cameras are installed at public places, airport gates, security checkpoints, and government buildings primarily to monitor a large area from a single location. It is now desirable to build systems where surveillance cameras coupled with a face recognition system can be used to automatically identify individuals from a watch-list. However, due to the poor quality of face images obtained from surveillance cameras, *matching low resolution face images* [63, 64, 65, 66, 67] has emerged as an important covariate of face recognition.
- With several applications of face recognition in e-commerce and social welfare programs, *matching biological twins* [21, 54, 68], [20, 22] and *look-alikes* [19] has also instigated interest from the research community.

1.3 Research Contributions

This dissertation focuses on developing algorithms for mitigating the effects of emerging covariates of face recognition. It presents several algorithms using machine learning paradigms such as genetic and memetic algorithms, online learning, co-training, transfer learning, and clustering to make face recognition algorithms scalable and robust to the emerging covariates. Quantitative analysis with existing techniques and commercial face

recognition systems demonstrate that this research enhances the state-of-art on several publicly available databases. The major contributions of this dissertation are as follows:

1. **Memetic optimization for matching forensic sketches with digital face images**: An automated sketch recognition algorithm is developed to extract discriminating information from local regions of both sketches and digital face images. Structural information along with the minute details present in local facial regions are encoded using multi-scale circular Weber's Local descriptor. To assign optimal weights to every local facial region, an evolutionary memetic optimization is proposed to boost the identification performance. Since, forensic sketches or digital face images can be of poor quality, a pre-processing technique is proposed to enhance the quality of images and improve the identification performance. This dissertation also offers a part of the IIIT-Delhi sketch database 1) viewed and semi-forensic sketch database and 2) 61 forensic sketch-digital image pairs to the research community.

2. **Multi-objective evolutionary algorithm for matching surgically altered face images**: A multi-objective evolutionary granular algorithm is developed to match face images before and after plastic surgery. The algorithm first generates non-disjoint face granules at multiple levels of granularity. The granular information is assimilated using a multi-objective genetic approach that simultaneously optimizes the selection of feature extractor for each face granule along with the weights of individual granules.

3. **Matching cross-resolution face images using co-transfer learning**: A co-transfer learning framework is developed for matching low resolution probe images with high resolution gallery images. The proposed algorithm seamlessly combines transfer learning and co-training paradigms. The transfer learning component transfers the knowledge that is learned while matching high resolution face images during training for matching low resolution probe images with high resolution gallery during testing. On the other hand, co-training component facilitates this transfer of knowledge by assigning pseudo labels to unlabeled probe instances in the target domain. Amalgamation of these two paradigms in the proposed framework enhances the performance of cross-resolution face recognition.

4. **Recognizing Faces in Videos using Clustering Based Re-ranking and Fusion**: A video based face recognition algorithm is developed that computes a discriminative video signature as an ordered list of still face images from a large dictionary.

A three stage approach is designed for optimizing ranked lists across multiple video frames and fusing them into a single composite ordered list to compute the video signature. This signature embeds diverse intra-personal variations and facilitates in matching two videos across large variations. For matching two videos, a discounted cumulative gain measure is utilized which uses the rankings of images in a video signature as well as the usefulness of images in characterizing the individual in a video.

Chapter 2

Matching Forensic Sketches with Digital Face Images

2.1 Introduction

Face recognition is a well studied problem in many application domains. However, matching sketches with digital face images is a very important law enforcement application that has received relatively less attention. Forensic sketches are drawn based on the recollection of an eye-witness and the expertise of a sketch artist. As shown in Figure 2.1, forensic sketches include several inadequacies because of the incomplete or approximate description provided by the eye-witness. Generally, forensic sketches are manually matched with the database comprising digital face images of known individuals. Existing state-of-the-art face recognition algorithms cannot be used directly and require additional processing to address the non-linear variations present in sketches and digital face images. An automatic sketch to digital face image matching system can assist law enforcement agencies and make the recognition process efficient and relatively fast.

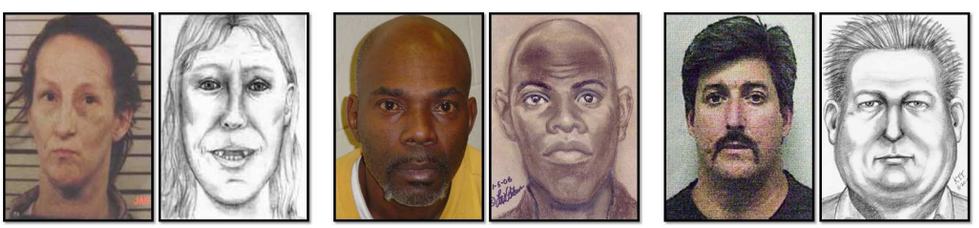

Figure 2.1: Examples showing exaggeration of facial features in forensic sketches.

2.1.1 Related Research

Sketch recognition algorithms can be classified into two categories: *generative* and *discriminative* approaches. Generative approaches model a digital image in terms of sketches and then match it with the query sketch or vice-versa. On the other hand, discriminative approaches perform feature extraction and matching using the given digital image and sketch pair and do not generate the corresponding digital image from sketches or the sketch from digital images.

Generative Approaches: Wang and Tang [69] proposed Eigen transformation based approach to transform a digital photo into sketch before matching. In another approach, they presented an algorithm with separate shape and texture information and applied Bayesian classifier for recognition [70]. Liu *et al.* [71] proposed a non-linear discriminative classifier based approach for synthesizing sketches by preserving face geometry. Li *et al.* [72] matched sketches and photos using a method similar to the Eigen-transform after converting sketches to photos. Xiao *et al.* [73] proposed to convert a sketch into photo using embedded Hidden Markov Models. The non-linearity between corresponding local patches of sketch photo pair was modeled using EHMM to generate pseudo-photo patch. These pseudo photo patches were then combined to synthesize a photo from the sketch and recognition was performed using Eigenface method. However, when synthesizing a photo from a sketch, the quality of the photo may degrade due to the exaggeration of features by artist. Wang and Tang [49] further proposed using Markov Random Fields to automatically synthesize sketches from digital face images and vice-versa. Zhang *et al.* [74] extended multiscale Markov Random Field (MRF) model to synthesize sketches under varying pose and lighting conditions. Sharma and Jacob [75] proposed a general latent space for heterogeneous face recognition using partial least squares (PLS) which projects images from two modalities to a space where they are similar. The holistic representation successfully matched sketches with digital images by maximizing the correlation in the projection of corresponding images from different modalities. However, PLS based approach cannot be expected to lead to effective recognition when such projections do not exist. Table 2.1 shows rank-1 identification accuracy of different approaches for matching sketches with digital face images.

Discriminative Approaches: Uhl and Lobo [76] proposed photometric standardization of sketches to compare it with digital photos. They further geometrically normalized sketches and photos to match them using Eigen analysis. Yuen and Man [77] used local and global feature measurements to effectively match sketches and mugshot images. Zhang

et al. [78] compared the performance of humans and PCA-based algorithm for matching sketch-photo pairs with variations in gender, age, ethnicity, and inter-artist difference. They also discussed about the quality of sketches in terms of artist's skills, experience, exposure time, and distinctiveness of features [79]. Similarly, Nizami *et al.* [80] analyzed the effect of matching sketches drawn by different artists. Nejati and Sim [81] proposed an approach for sketch photo matching using only facial component outlines and facial marks. Their analysis suggested that improved performance was achieved by comparing the abnormal features in sketches with the exaggerated digital faces. Their analysis also suggested that combining local and holistic exaggerated features led to improved face recognition performance. Further, Nejati *et al.* [82] proposed a new eye-witness testimony method where the sketches were drawn by an eyewitness thus eliminating the bias and combining additional soft information such as skin tone and ethnicity. Matching was performed by estimating shapes of the facial components and combining the relative differences using global least square optimization. However, their approach was limited to the accuracy of shape estimation from sketches and digital images. Klare and Jain [83] proposed a scale invariant feature transform (SIFT) based local feature approach where sketches and digital face images were matched using the gradient magnitude and orientation within the local region. Klare *et al.* [4] extended their approach using local feature discriminant analysis (LFDA) to match forensic sketches. In their approach, sketch and face images were first partitioned into slices. Scale-invariant feature transform (SIFT) and multiscale local binary pattern (MLBP) descriptors were computed for each slice. Next, Local-feature-based discriminant analysis (LFDA) was used to extract the most salient features for each slice and similarity between feature vectors was computed to match sketches with photos. The accuracy was further improved by incorporating subject's demographic information such as race, gender, age, and height. In another approach, Klare and Jain [84] proposed a framework for heterogeneous face recognition where both probe and gallery images were represented in terms of non-linear kernel similarities. Zhang *et al.* [85] analyzed the psychological behavior of humans for matching sketches drawn by different sketch artists. Zhang *et al.* [50] proposed an information-theoretic encoding band descriptor to capture discriminative information and random forest based matching to maximize the mutual information between the sketch and photo. Bhatt *et al.* [86] extended Uniform Local Binary Patterns to incorporate exact difference of gray level intensities to encode texture features in sketches and digital face images. Recently, Bhatt *et al.* [51] proposed to use the structural information along with the minute details present in local facial regions

Table 2.1: A comparison of some representative approaches proposed for matching sketches with digital face images.

Approach	Database	Gallery/Probe	Rank-1 accuracy
Eigen-Transformation [69]	CUHK	300/300	90.0%
LLE Transformation [71]	CUHK	300/300	87.7%
MRF Transformation [49]	CUHK	300/300	96.3%
Direct matching [83]	CUHK	300/300	97.8%
Genetic Algorithm [86]	CUHK	233/233	94.1%
LFDA+MLBP [4]	Forensic Sketch	10,100/49	32.6% (rank-50)
Memetic algorithm [51]	Forensic Sketch	7063/190	28.5% (rank-50)

using multi-scale circular Weber's local descriptor. Further, an evolutionary memetic optimization was proposed to assign optimal weights to every local facial region to boost the identification performance for matching forensic sketches. Canavan *et al.* [87] utilized a scale-space topographic feature representation to model the appearance of the sketch and a mesh adaptation approach was used to model the 3D shape. Further, a component based spatial Hidden Markov Model (HMM) for sketch recognition using the geometry of 3D face sketches. Galoogahi and Sim [88] utilized shape as a robust feature for matching cross modality images such as sketches and digital images. They encoded the shape information from sketches and digital images using LBP descriptor in Radon space for each local region.

2.1.2 Research Contributions

After discussing with several sketch artists, it is observed that *generating a sketch is an unknown psychological phenomenon, however, a sketch artist generally focusses on the local facial features and texture which he/she tries to embed in the sketch through a blend of soft and prominent edges.* Therefore, the proposed algorithm is designed based on the following observations:

- information vested in local facial regions can have high discriminating power;
- facial patterns in sketches and digital face images can be efficiently represented by local descriptors.

This chapter proposes an automatic algorithm for matching sketches with digital face images using the modified Weber's local descriptor (WLD) [1]. WLD is used for representing images at multiple scales with circular encoding. The multi-scale analysis helps

in assimilating information from minute features to the most prominent features in a face. Further, memetically optimized χ^2 distance measure is used for matching sketches with digital face images. The proposed matching algorithm improves the performance by assigning optimal weights to local facial regions. To further improve the performance, a Discrete Wavelet Transform (DWT) [89] fusion based pre-processing technique is presented to enhance forensic sketch-digital image pairs. Three different types of sketches are used for performance evaluation, 1) sketches drawn by a sketch artist while looking at the digital image of a person (viewed sketches), 2) sketches drawn by an artist based on his recollection from the digital image of a person (semi-forensic sketches), and 3) sketches drawn based on the description of an eyewitness from his recollection of the crime scene (forensic sketches). The major contributions of this chapter are summarized as follows:

1. Existing approaches for matching forensic sketches [4] manually separate *good* and *bad* forensic sketches and generally focus on *good* forensic sketches only. Such a classification is often based on the similarity between the sketch and corresponding digital face image. Since the corresponding digital face image is not available in real-time applications, selecting good and bad forensic sketches is not pragmatic for matching forensic sketches with digital face images. In this chapter, a pre-processing technique is presented for enhancing the quality of forensic sketch-digital image pairs. Pre-processing forensic sketches enhances the quality and therefore, improves the performance by at least $2 - 3\%$.

2. Multi-scale Circular WLD and memetically optimized χ^2 based algorithms are proposed for matching sketches with digital face images. The proposed algorithm outperforms existing approaches on different sketch databases.

3. To better understand the progression from viewed to forensic sketches, semi-forensic sketches are introduced to bridge the gap between viewed and forensic sketches. In the experiments, it is observed that training sketch recognition algorithms (existing as well as the proposed) on semi-forensic sketches improves the rank-1 identification performance by at least 4% compared to the traditional way, i.e. training on viewed sketches.

4. Human performance for matching sketches with digital face images is also analyzed. The information collected from individuals corroborate with our initial observation that local regions provide discriminating information.

5. The chapter also presents a part of the IIIT-Delhi database[1] (Viewed and Semi-forensic Sketch database) and 61 forensic sketch-digital image pairs to the research community to promote the research in this domain.

2.2 Pre-processing Algorithm

In sketch to digital face image matching, researchers have generally used viewed sketches where the quality of sketch-digital image pair is very good. On these good quality viewed sketches, the state-of-art is about 99% (rank-1) identification accuracy while the state-of-art in forensic sketch recognition is about 16%. One of the reasons for low recognition performance is that forensic sketches may contain distortions and noise introduced due to the excessive use of charcoal pencil, paper quality, and scanning (device noise/errors). Furthermore, in the gallery, digital images may also be noisy and of sub-optimal quality because of printing and scanning of images. As shown in Figure 2.2, forensic sketch-digital image pairs of lower visual quality may lead to reduced matching performance as compared to good quality sketch-digital image pairs.

A pre-processing technique is presented that enhances the quality of forensic sketch-digital image pairs. The steps involved in the pre-processing technique are as follows:

- Let f be the color face image to be enhanced. Let f^r and f^y be the red and luma channels[2] respectively. These two channels are processed using the multi-scale retinex (MSR) algorithm [90] with four iterations. MSR is applied on both red and luma channels to obtain f^{rm} and f^{ym}.

- f^{rm} and f^{ym} are subjected to wavelet based adaptive soft thresholding scheme [91] for image denoising. The algorithm computes generalized Gaussian distribution based soft threshold which is used in wavelet based denoising to obtain $f^{rm'}$ and $f^{ym'}$ respectively.

- Noise removal in the previous step may lead to blurring of edges. Experiments show that a symmetric low-pass filter of size 7×7 with standard deviation of 0.5 efficiently restores the genuine facial edges. Applying this (Wiener) filter on $f^{rm'}$ and $f^{ym'}$ produces f^1 and f^2.

[1] Available at http://research.iiitd.edu.in/groups/iab/sketchDatabase.html.

[2] In the watermarking literature, it is well established that red and luma channels are relatively less sensitive to the visible noise, therefore, these channels are used for enhancement.

- After computing the globally enhanced red and luma channels, DWT fusion algorithm is applied on f^1 and f^2 to compute a feature rich and enhanced face image, F. Single level DWT (with db 9/7 mother wavelet) is applied on f^1 and f^2 to obtain the detailed and approximation bands of these images. Let f^j_{LL}, f^j_{LH}, f^j_{HL}, and f^j_{HH} be the four subbands and $j = 1, 2$, where LL represents approximation band and LH, HL, and HH represent the detailed subbands. To preserve features of both the channels, coefficients from the approximation band of f^1 and f^2 are averaged.

$$f^e_{LL} = mean(f^1_{LL}, f^2_{LL}) \tag{2.1}$$

where f^e_{LL} is the approximation band of enhanced image. All three detailed subbands are divided into windows of size 3×3 and the sum of absolute pixels in each window is calculated. For the i^{th} window in HL subband of the two images, the window with maximum absolute value is selected to be used for enhanced subband f^e_{HL}. Similarly, enhanced subbands f^e_{LH} and f^e_{HH} are also obtained. Finally, inverse DWT is applied on the four subbands to generate a high quality face image.

$$F = IDWT(f^e_{LL}, f^e_{LH}, f^e_{HL}, f^e_{HH}) \tag{2.2}$$

This DWT fusion algorithm is applied on both forensic sketches and digital face images. Figure 2.3 shows quality enhanced forensic sketches and digital face images. Note that the pre-processing technique enhances the quality when there are irregularities and noise in the input image, however, it does not alter good quality face images (i.e. sketch-digital image pairs from the viewed sketch database). Sketches are scanned as three channel color images. Further, the forensic images obtained from different sources are three channel color images. If a gray scale image is obtained, multi-scale retinex and Wiener filtering are applied only on the single channel. Along with quality enhancement, face images are geometrically normalized. The eye-coordinates are detected using the OpenCV's boosted cascade of haar-like features. Using the eye-coordinates, rotation is normalized with respect to the horizontal axis and inter-eye distance is fixed to 100 pixels. Finally, the face image is resized to 192×224 pixels.

2.3 Matching Sketches with Digital Face Images

Local descriptors have received attention in face recognition due to their robustness to scale, orientation, and speed. Local Binary Patterns (LBP) is one of the widely used descriptors for face recognition [92]. In face recognition literature, several variants of LBP

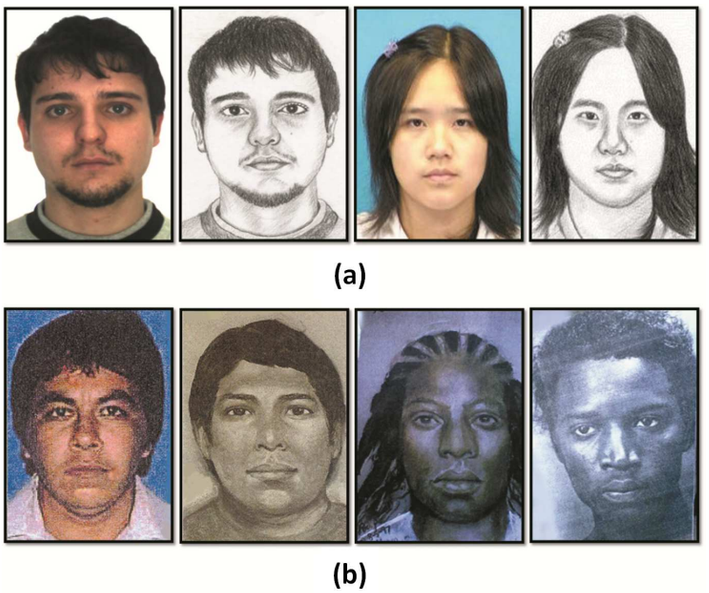

Figure 2.2: Paper quality, sensor noise, and old photographs can affect the quality of sketch-digital image pairs and hence reduce the performance of matching algorithms. (a) Good quality sketch-digital image pairs (CUHK database) and (b) poor quality sketch-digital image pairs (Forensic sketch database).

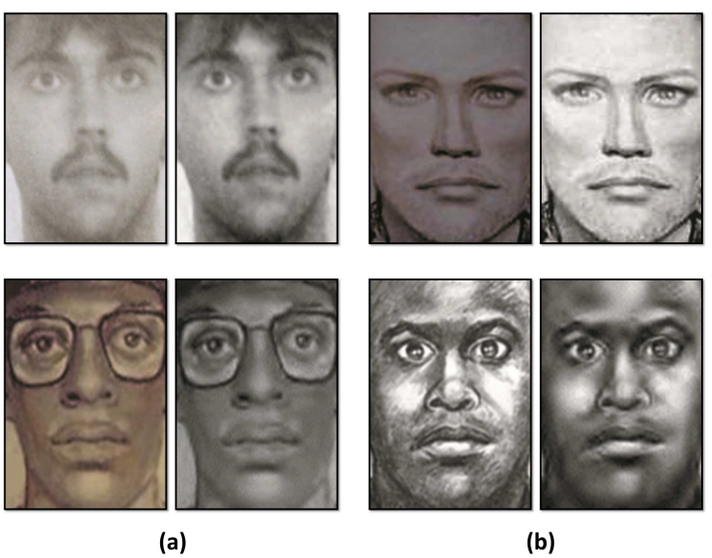

Figure 2.3: Quality enhancement using the pre-processing technique. (a) represents digital face image before and after pre-processing and (b) represents forensic sketches before and after pre-processing.

have been proposed. Bhatt *et al.* [86] extended LBP to incorporate exact difference of gray level intensities among pixel neighbors and used it for sketch recognition. Local descriptors such as LBP are generally used as dense descriptors where texture features are computed for every pixel of the input face image. On the other hand, sparse descriptor such as Scale Invariant Feature Transform (SIFT) [93] is based on interest point detection and computing the descriptor in the vicinity of detected interest points. SIFT is computed using gradient and orientation of neighboring points sampled around every detected key point. As a sparse descriptor, SIFT has been used for face recognition by Geng and Jiang [94]. Klare and Jain [83] applied SIFT in a dense manner (i.e. computing SIFT descriptor at specific pixels) for matching sketches with digital face images. It is our assertion that local descriptors can be used for representing sketches and digital face images because they can efficiently encode the discriminating information present in the local regions.

Recently, Chen *et al.* [1] proposed a new descriptor, Weber's local descriptor, which is based on Weber's law and draws its motivation from both SIFT and LBP. It is similar to SIFT in computing histogram using gradient and orientation, and analogous to LBP in being computationally efficient and considering small neighborhood regions. However, WLD has some unique features that make it more efficient and robust as compared to SIFT and LBP. WLD computes the salient micro patterns in a relatively small neighborhood region with finer granularity. This allows it to encode more discriminative local micro patterns. WLD is optimized for matching sketches with digital face images by computing multi-scale descriptor in a circular manner (in contrast to the originally proposed square neighborhood approach). Finally, two multi-scale circular WLD (MCWLD) histograms are matched using memetically optimized weighted χ^2 distance.

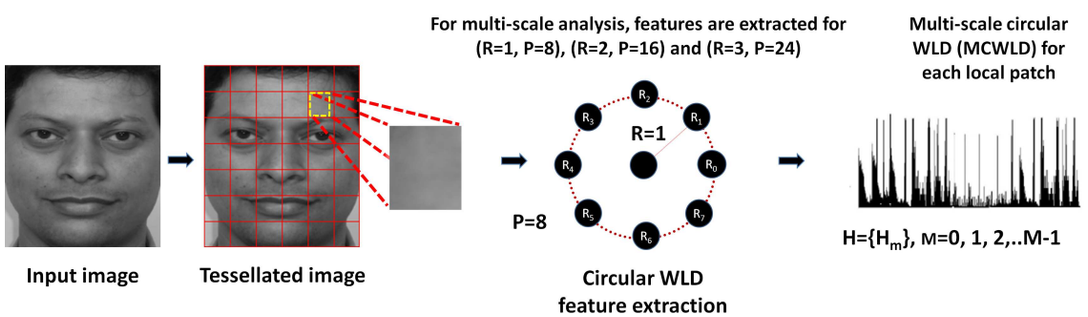

Figure 2.4: Steps involved in the proposed algorithm for matching sketches with digital face images.

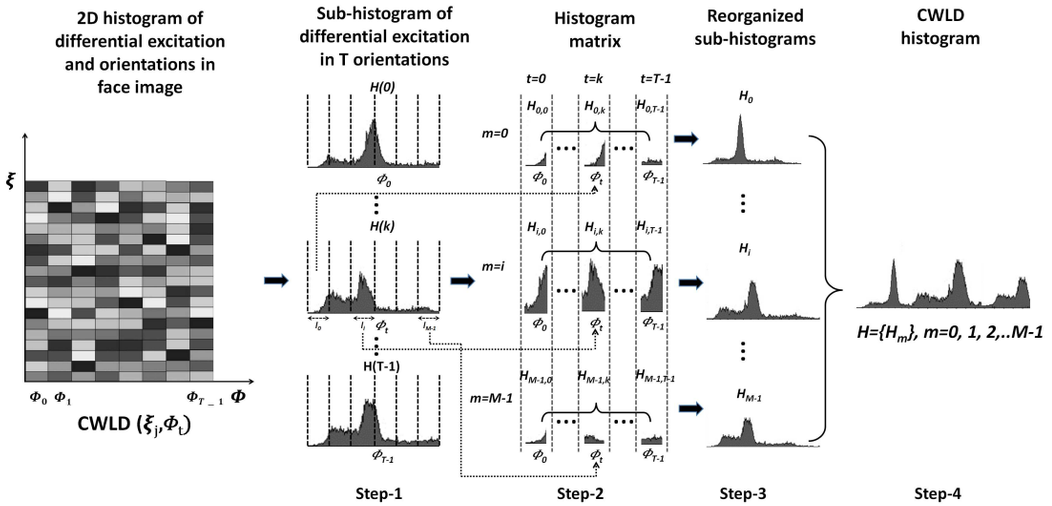

Figure 2.5: Illustrating the steps involved in computing the circular WLD histogram (adapted from [1]).

2.3.1 Feature Extraction using MCWLD

MCWLD has two components: 1) *differential excitation* and 2) *gradient orientation*. MCWLD representation for a given face image is constructed by tessellating the face image and computing a descriptor for each region. As shown in Figure 2.4, MCWLD descriptor is computed for different parameters P and R, where P is the number of neighboring pixels evenly separated on a circle of radius R centered at the current pixel. Multi-scale analysis is performed by varying radius R and number of neighbors P. Sketches and digital face images are represented using MCWLD as explained below:

2.3.1.1 Differential Excitation

Differential excitation is computed as an arctangent function of the ratio of intensity difference between central pixel and its neighbors to the intensity of central pixel. The differential excitation of central pixel $\xi(x_c)$ is computed as:

$$\xi(x_c) = arctan\left\{\sum_{i=0}^{P-1}\left(\frac{x_i - x_c}{x_c}\right)\right\} \tag{2.3}$$

where x_c is the intensity value of central pixel and P is the number of neighbors on a circle of radius R. If $\xi(x_c)$ is positive, it simulates the case that surroundings are lighter than the current pixel. In contrast, if $\xi(x_c)$ is negative, it simulates the case that surroundings are darker than the current pixel.

2.3.1.2 Orientation

The orientation component of WLD is computed as:

$$\theta(x_c) = \arctan\left\{\frac{x_{(\frac{P}{2}+R)} - x_{(R)}}{x_{(P-R)} - x_{(\frac{P}{2}-R)}}\right\} \tag{2.4}$$

The orientation is further quantized into T dominant orientation bins where T is experimentally set as eight.

2.3.1.3 Circular WLD Histogram

For every pixel, differential excitation (ξ) and orientation (θ) are computed using Eqs. 2.3 and 2.4 respectively. As shown in Figure 2.5, a 2D histogram of circular WLD feature, $CWLD(\xi_j, \theta_t)$, is constructed where $j = 0, 1, ..., N-1$, $t = 0, 1, ..., T-1$, and N is the dimension of the image. Each column in the $2D$ histogram corresponds to a dominant orientation, θ_t, and each row corresponds to a differential excitation interval. Thus, the intensity of each cell corresponds to the frequency of a certain differential excitation interval in a dominant orientation. Similar to Chen *et al.* [1], four step approach is followed to compute CWLD descriptor.

Step-1: The $2D$ histogram $CWLD(\xi_j, \theta_t)$ is further encoded into $1D$ histograms. Differential excitations, ξ, are regrouped into T orientation sub-histograms, $H(t)$, where $t = 0, 1, ..., T-1$ corresponds to each dominant orientation.

Step-2: Within each dominant orientation, range of differential excitation is evenly divided into M intervals and then reorganized into a histogram matrix. Each orientation sub-histogram in $H(t)$ is thus divided into M segments, $H_{m,t}$ where $m = 0, 1, ..., M-1$ and $M = 6$. For each differential excitation interval l_m, lower bound is computed as $\eta_{m,l} = (m/M - 1/2)\pi$ and upper bound $\eta_{m,u}$ is computed as $\eta_{m,u} = [(m+1)/M - 1/2]\pi$.

Each sub-histogram segment $H_{m,t}$ is further composed of S bins and is represented as:

$$H_{m,t} = h_{m,t,s} \tag{2.5}$$

where $s = 0, 1, ..., S-1$, $S = 3$ and $h_{m,t,s}$ is represented as:

$$h_{m,t,s} = \sum_j \delta(S_j == s), \left(S_j = \left\lfloor \frac{\xi_j - \eta_{m,l}}{(\eta_{m,n} - \eta_{m,l})/S} + \frac{1}{2} \right\rfloor\right). \tag{2.6}$$

Here $j = 0, 1, ..., N-1$, m is the interval to which differential excitation ξ_j belongs i.e. $\xi_j \in l_m$, t is the index of quantized orientation, and $\delta(\cdot)$ is defined as follows:

$$\delta(\cdot) = \begin{cases} 1, & if \;\; function\; is\; true, \\ 0, & otherwise \end{cases} \tag{2.7}$$

Step-3: Sub-histogram segments, $H_{m,t}$, across all dominant orientations are reorganized into M $1D$ histograms.

Step-4: M sub-histograms are concatenated into a single histogram represents the final $6 \times 8 \times 3$ ($M \times T \times S$) circular WLD histogram. The range of differential excitation is segmented into separate intervals to account for the variations in a given face image, and assigning optimal weights to these H_m segments further improves the performance of CWLD descriptor.

2.3.1.4 Multi-scale Circular WLD

In Multi-scale analysis, CWLD descriptor is extracted with different values of P and R and the histograms obtained at different scales are concatenated. Multi-scale analysis is performed at three different scales with parameters as $(R = 1, P = 8)$, $(R = 2, P = 16)$ and $(R = 3, p = 24)$. A face image is divided into 6×7 non-overlapping local facial regions and MCWLD histogram is computed for each region. MCWLD histograms for every region are then concatenated to form the facial representation.

2.3.2 Memetic Optimization

According to psychological studies in face recognition [95], some facial regions are more discriminating than others and hence, contribute more towards the recognition accuracy. Similarly, MCWLD histograms corresponding to different local facial regions may have varying contribution towards the recognition accuracy. Moreover, MCWLD histogram corresponding to each local facial region comprises of M sub-histogram segments (as shown in Step-3 of Figure 2.5) representing different frequency information. Generally, the regions with high variance are more discriminating as compared to flat regions, therefore, M sub-histogram segments may also have varying contribution towards the recognition accuracy. It is our assertion that while matching MCLWD histograms, different weights need to be assigned to local regions and histogram segments for better performance. Here, the weights associated with 42 local facial regions and 6 sub-histogram segments at 3 different scales have to be optimized. Optimizing such large number of weights for best performance is a very challenging problem and requires a learning based technique.

Memetic algorithm (MA) [96] can be effectively used to optimize such large search spaces. It is a form of hybrid global-local heuristic search methodology. The global search is similar to traditional evolutionary approaches such as population-based method in a Genetic Algorithm (GA), while the local search involves refining the solutions within the

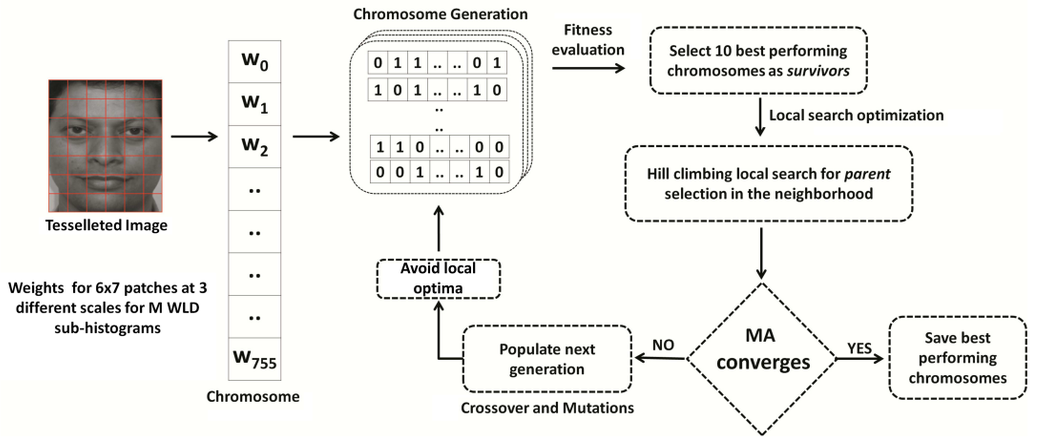

Figure 2.6: Illustrating the steps involved in memetic optimization for assigning optimal weights to each tessellated face region.

population. From an optimization perspective, MAs have been found to be more efficient (i.e. requiring fewer evaluations to find optima) and effective (i.e. identifying higher quality solutions) than traditional evolutionary approaches such as GA [97]. Memetic algorithm is used for optimizing the weights.

2.3.2.1 Weighted χ^2 Matching using Memetic Optimization

For matching two MCWLD histograms, weighted χ^2 distance measure is used.

$$\chi^2(x,y) = \sum_{i,j} \omega_j \left[\frac{(x_{i,j} - y_{i,j})^2}{(x_{i,j} + y_{i,j})} \right] \tag{2.8}$$

where x and y are the two MCWLD histograms to be matched, i and j correspond to the i^{th} bin of the j^{th} histogram segment ($j = 1, \cdots, 756$), and ω_j is the weight for the j^{th} histogram segment. As shown in Figure 2.6, a memetic search is applied to find optimal values of w_j. The steps involved in the memetic optimization process are described below:

Memetic Encoding: A chromosome is a string whose length is equal to the number of weights to be optimized i.e. $42 \times 6 \times 3 = 756$. Each unit or meme in a chromosome is a real valued number representing the corresponding weight.

Initial Population: MA is initialized with 100 chromosomes. For quick convergence, weights proportional to the rank-1 identification accuracy of each individual region are used as the initial chromosome [92]. The remaining 99 chromosomes are generated by randomly changing one or more units in the initial chromosome. Further, the weights are normalized such that the sum of all the weights in a chromosome is one.

Fitness Function: Each chromosome in a generation is a possible solution and the recognition is performed using the weights encoded by the chromosomes. The identification accuracy, used as fitness function, is computed on the training set and the 10 best performing chromosomes are selected as *survivors*. These survivors are used for crossover and mutation to populate the next generation.

Hill Climbing Local Search: MA requires a local search on *survivors* to further fine tune the solution [97]. Two *survivors* are recombined to produce two candidate *parents*. Note that in a pair of two, this process is repeated for all 10 *survivors* to find better chromosomes. If the candidate *parents* have better performance than participating *survivors*, they replace the *survivors* to become *parents* and populate the next generation. This local search is performed at each generation to find better *parents* from the competing *survivors* which leads to quick convergence and better quality of solution.

Crossover and Mutation: A set of uniform crossover operations is performed on *parents* (obtained after local search) to populate a new generation of chromosomes. After crossover, mutation is performed by changing one or more weights by a factor of its standard deviation in previous generations. After mutation and crossover, 100 chromosomes are populated in the new generation.

The MA search process is repeated till convergence and terminates when the identification performance of the chromosomes in new generation does not improve compared to the performance of chromosomes in previous five generations. At this point, weights pertaining to the best performing chromosome (i.e. chromosome giving best recognition accuracy on training data) are obtained and used for testing. Thus, for a given data set, the MA search process finds optimal weights. It also enables to discard redundant and non-discriminating regions whose contribution towards recognition accuracy is very low (i.e. the weight for that region is zero or close to zero). This leads to dimensionality reduction and better computational efficiency because MCWLD histograms for poor performing facial regions are not computed during testing.

2.3.2.2 Avoiding Local Optima

Evolutionary algorithms such as MA often fail to maintain diversity among individual solutions (chromosomes) and cause the population to converge prematurely. This leads to decrease in the quality of solution. Different techniques have been proposed to maintain certain degree of diversity in a population, without affecting the convergence. *Adaptive*

mutation rate [98] and *random offspring generation* [99] are used to prevent premature convergence to local optima.

- *Adaptive Mutation rate:* To maintain diversity in the population, mutation rate can be increased. However, higher value of mutation rate may introduce noise and affect the convergence process. Instead of using a fixed high or low mutation rate, an adaptive mutation rate, depending on population's diversity, is used. Population diversity is measured as the standard deviation of fitness values in a population as shown in Eq. 2.9:

$$stddev(P) = \sqrt{\frac{\sum_{i=1}^{N}(f_i - f_{mean})^2}{(N-1)}} \tag{2.9}$$

  where N is the population size and f_i is the fitness of the i^{th} chromosome in the population. The process starts with an initial value of mutation rate (probability 0.02), and whenever population diversity falls below the predefined threshold, mutation rate is increased.

- *Random Offspring Generation:* One of the reasons for evolutionary algorithms converging to local optima is high degree of similarity among participating chromosomes (*parents*) during crossover operation. Combination of such chromosomes is ineffective because it leads to offsprings that are exactly similar to the *parents*. If such a situation occurs where participating chromosomes (*parents*) are very similar, then crossover is not performed and offsprings are generated randomly.

The memetic optimization for computing weights is summarized in Algorithm 1.

2.3.3 Proposed Algorithm for Matching Sketches with Digital Face Images

The process of matching sketches with digital face images is as follows:

1. For a given sketch-digital image pair, the pre-processing technique is used to enhance the quality of face images.

2. Both sketches and digital face images are tessellated into non-overlapping local facial regions.

3. For each facial region, MCWLD histograms are computed at three different scales. The facial representation is obtained by concatenating MCWLD histograms for every facial region.

Algorithm 1 Memetic algorithm for weight optimization.

Step 1: Memetic Encoding: A chromosome of length $42 \times 3 \times 6 = 756$ is encoded where each unit in the chromosome is a real valued number representing the corresponding weight.
Step 2: Initial Population: A population of 100 chromosomes is generated starting with a seed chromosome.
Step 3: Fitness Function: Fitness is evaluated by performing recognition using the weights encoded by each chromosome. 10 best performing chromosomes from a population are selected as *survivors* to perform crossover and mutation.
Step 4: Hill Climbing Local Search: The *survivors* obtained in Step 3 are used to find better chromosomes in their local neighborhood and *parents* are selected.
Step 5: Crossover and Mutation: New population is generated from *parents* obtained after local search in Step 4. A set of uniform crossover operations is performed followed by mutation. To avoid local optima, adaptive mutation and random offspring generation techniques are used.
Step 6: Repeat Steps 3-5 till convergence criteria is satisfied.

4. To match two MCWLD histograms, weighted χ^2 distance measure is used where the weights are optimized using Memetic algorithm.

5. In identification mode, this procedure is applied for each gallery-probe pair and top matches are obtained.

2.4 Sketch Databases

To evaluate the performance of the proposed algorithm, three types of sketch databases are used: 1) Viewed Sketch, 2) Semi-forensic Sketch, and 3) Forensic Sketch database.

1. *Viewed Sketch Database*: It comprises a total of 549 sketch-digital image pairs from two sketch databases: the CUHK database [49] and the IIIT-Delhi Sketch database [86]. The CUHK database comprises 606 sketch-digital image pairs from CUHK students [49], the AR [41], and the XM2VTS databases. Since the XM2VTS database is not available freely, the remaining 311 sketch-digital image pairs are used. Further, the authors have prepared a database of 238 sketch-digital image pairs. The sketches are drawn by a professional sketch artist for digital images collected from different sources. This database is termed as IIIT-Delhi Viewed Sketch database.

2. *Semi-forensic Sketch Database*: As described earlier, sketches drawn based on the memory of sketch artist rather than the description of an eye-witness are termed as semi-forensic sketches. To prepare the IIIT-Delhi Semi-forensic Sketch database,

the sketch artist is allowed to view the digital image once (for about $5-10$ minutes) and is asked to draw the sketch based on his memory. The time elapsed between the artist's viewing an image and starting to draw a sketch is about 15 minutes. Sketch artist is not allowed to view the digital image while preparing the sketch. These sketches are thus drawn based on the recollection of the sketch artist, thus eliminating the effect of attrition based on how well the eyewitness remembers an individual's face and how well he/she is able to describe it to the sketch artist. 140 digital images from the IIIT-Delhi Viewed Sketch database are used to prepare the Semi-forensic Sketch database. Therefore, all images that are used to draw a semi-forensic sketch also have a corresponding viewed sketch. Figure 2.7 presents samples of viewed and semi-forensic sketches corresponding to digital face images.

3. *Forensic Sketch Database*: Forensic sketches are drawn by a sketch artist from the description of an eyewitness based on his/her recollection of the crime scene. These sketches are based on (1) how well the eyewitness can recollect and describe the face and (2) the expertise of the sketch artist. A database of 190 forensic sketches with corresponding digital face images is used. This database contains 92 forensic sketch-digital image pairs obtained from Lois Gibson [2], 37 pairs obtained from Karen Taylor (published in [3]), and 61 pairs from different source on the internet. Figure 2.8 shows sample images from the forensic sketch database.

2.5 Viewed Sketch Matching Results

To establish a baseline, the performance of the proposed and existing algorithms are first computed on the viewed sketch database. Since the application of sketch recognition is dominant with identification scenario, the performance of the proposed algorithm is evaluated in identification mode. Three sets of experiments are performed using the viewed sketch databases. In all three experiments, digital images are used as gallery and sketches are used as probe. Further, 40% of the database is used for training and the remaining 60% pairs are used for performance evaluation. The protocol for all three experiments is described in Table 2.2.

For each experiment, training is performed to compute the parameters of feature extractor and weights using the Memetic Optimization. This non-overlapping train-test partitioning is repeated five times with random sub-sampling and Cumulative Match Characteristic (CMC) curves are computed for performance comparison.

Figure 2.7: (a) Sample images from the IIIT-Delhi Sketch database. The first row represents the viewed sketches, the second row represents the corresponding digital face images and the third row represents the corresponding semi-forensic sketches. (b) Sample images from the CUHK database.

Figure 2.8: Sample images from the Forensic Sketch database. Images are obtained from different forensic artists [2], [3].

Table 2.2: Experimental protocol for matching viewed sketches.

Experiment	Number of Sketch-Digital Image Pairs	Training Database	Testing Database
Experiment 1	311 from CUHK	125	186
Experiment 2	238 from IIIT-Delhi	95	143
Experiment 3	549 from Combined	220	329

Table 2.3: Rank-1 identification accuracy of sketch to digital face image matching algorithms for matching viewed sketches. Identification accuracies are computed with five times random cross validation and standard deviations are also reported.

Database (Training/ Testing)	Algorithm	Rank-1 Identification Accuracy (%)	Standard Deviation (%)
CUHK (125/186)	COTS-1	91.25	0.83
	COTS-2	92.05	0.72
	WLD [1]	93.42	0.85
	MWLD [1]	94.14	0.82
	MCWLD	95.08	0.76
	SIFT [83]	94.36	1.03
	EUCLBP+GA [86]	95.12	0.93
	LFDA [4]	97.10	1.16
	Proposed	**97.28**	**0.68**
IIIT-Delhi Viewed Sketch (95/143)	COTS-1	71.46	0.87
	COTS-2	73.26	0.75
	WLD [1]	74.34	0.81
	MWLD [1]	75.68	0.83
	MCWLD	78.48	0.89
	SIFT [83]	76.28	1.33
	EUCLBP+GA [86]	79.36	0.87
	LFDA [4]	81.43	1.11
	Proposed	**84.24**	**0.94**
Combined (220/329)	COTS-1	80.14	0.78
	COTS-2	79.24	0.86
	WLD [1]	84.37	0.88
	MWLD [1]	85.32	0.86
	MCWLD	88.25	0.84
	SIFT [83]	85.86	1.01
	EUCLBP+GA [86]	88.75	0.87
	LFDA [4]	91.16	0.93
	Proposed	**93.16**	**0.96**

2.5.1 Experimental Analysis

The performance of the proposed approach is compared with existing algorithms designed for matching sketches with digital face images and two leading commercial face recognition systems[1]. Existing algorithms include SIFT [83], EUCLBP+GA [86] and LFDA [4]. Further, the performance gain due to multi-scale analysis and circular sampling is analyzed by comparing the performance of WLD, Multi-scale WLD (MWLD) algorithms with square sampling, and Multi-scale circular WLD (MCWLD). The same weighting scheme proposed by Chen *et al.* [1] is used in WLD, MWLD, and MCWLD algorithms. Further, to quantify improvement due to memetic optimization of weights as compared to the weighting method proposed in [1], the performance of the proposed algorithm is compared with MCWLD. The pre-processing technique enhances the quality only when there are irregularities and noise in the input image and it does not alter good quality face images (i.e. sketch-digital image pairs from the viewed sketch database). Therefore in the experiments with *Viewed Sketch* database, no pre-processing is applied on sketch-digital image pairs. Key results and observations for matching viewed sketches are summarized below:

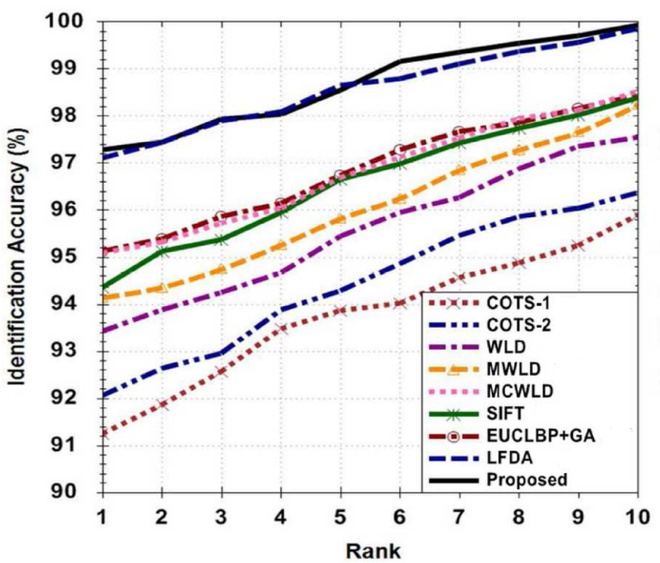

Figure 2.9: CMC curves showing the performance of sketch to digital face image matching algorithms on the CUHK database.

[1]The license agreements of these commercial face recognition systems does not allow us to name the product in any comparison. Therefore, the two products are referred to as COTS-1 and COTS-2.

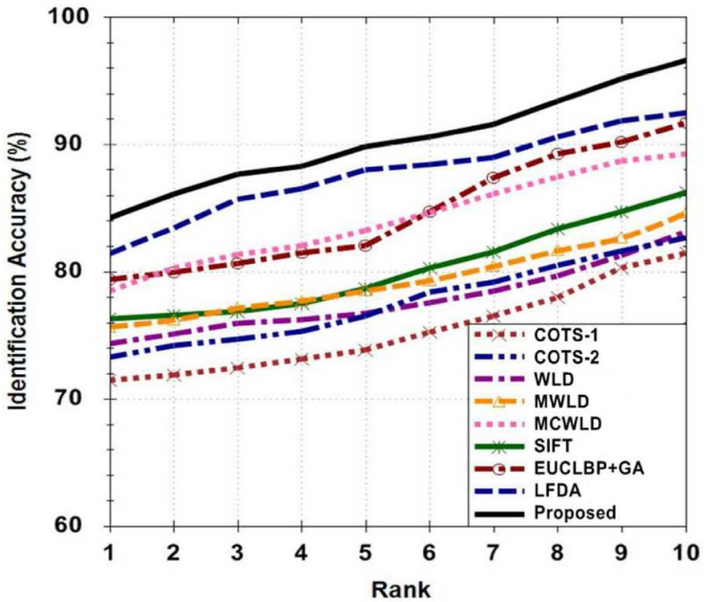

Figure 2.10: CMC curves showing the performance of sketch to digital face image matching algorithms on the IIIT-Delhi Viewed Sketch database.

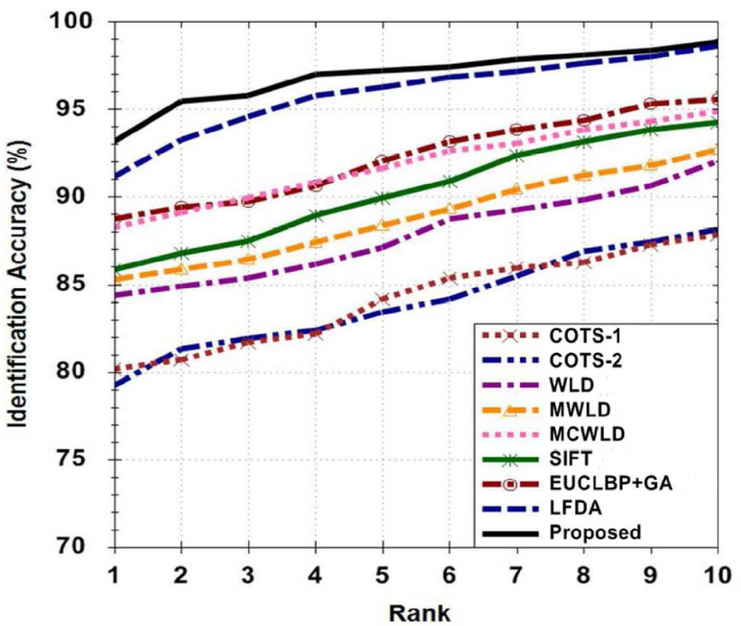

Figure 2.11: CMC curves showing the performance of sketch to digital face image matching algorithms on the Combined database.

- The CMC curves in Figures 2.9-2.11 show the rank-1 identification accuracy of sketch to digital face image matching algorithms. Table 5.2 summarizes the rank-1 identification accuracy and the standard deviation for five times random subsampling (cross validations) on all three sets of experiments. On the CUHK database, the proposed approach yields rank-1 accuracy of 97.28% which is slightly better than LFDA [4] and is at least 2% better than MWLD [1], MCWLD, SIFT [83], and EUCLBP+GA [86]. The proposed approach also outperforms the two commercial systems by at least 5%.

- On comparing WLD with MWLD, it is observed MWLD provides an improvement of about 1% on different viewed sketch databases due to multi-scale analysis. Further, compared to the MWLD algorithm [1], the proposed MCWLD algorithm improves the rank-1 identification accuracy by about 1% on the CUHK, 2.8% on the IIIT-Delhi, and 2.9% on the combined databases. It suggests that circular sampling method yields more discriminative representation of the face image as compared to square sampling. Note that both MWLD and MCWLD are applied at three different scales with parameters as $(R = 1, P = 8)$, $(R = 2, P = 16)$ and $(R = 3, p = 24)$. Parameters for WLD are $M = 6$, $T = 8$, and $S = 3$.

- Compared to the weighting scheme (proposed by Chen *et al.* [1]) used in MCWLD algorithm, the proposed memetic optimization improves the rank-1 identification accuracy by 2.2% on the CUHK, 5.7% on the IIIT-Delhi, and 4.9% on the combined databases. This improvement in rank-1 identification accuracy validates our assertion that assigning memetically optimized weights to local facial regions boosts the identification performance. This also corroborates with several psychological findings that different facial regions have varying contribution towards the recognition performance [95].

- The CUHK sketch database and the IIIT-D viewed sketch database have variations introduced by different drawing styles of artists. As discussed by Zhang *et al.* [79], drawing styles of different artists play an important role in how closely a sketch resembles the actual digital photo thus influencing the performance of different algorithms.

- As shown in Figure 2.11, the rank-1 identification accuracy of the proposed algorithm on the combined database is at least 2% better than existing approaches and

outperforms the two commercial systems by 13%. The proposed approach represents the face image by combining MCWLD histograms obtained from every local facial region. The multi-scale analysis along with memetic optimization for assigning weights corresponding to each local facial region helps in capturing the salient micro patterns from both sketches and digital face images. Further, memetic optimization helps in dimensionality reduction; i.e. at the end of memetic optimization, on an average, 32 out of 126 (42×3) local facial patches at different scales are assigned null weights. Therefore, MCWLD histogram for these patches are not computed during testing.

- Experiments are performed by reducing the dimensionality of features using PCA; however, the results are not encouraging as it does not capture the observation that information vested in local regions has varying contribution in recognition accuracy and assigning optimal weights to these regions will enhance the performance. To incorporate this observation, MA is used that leads to dimensionality reduction and better computational efficiency because MCWLD histograms for poor performing facial regions are not computed during testing.

2.6 Matching Forensic Sketches with Digital Face Images

Previous research [4] in matching forensic sketches suggests that existing sketch recognition algorithms trained on viewed sketches are not sufficient for matching forensic sketches with digital face images. Moreover, poor quality of forensic sketches further degrade the performance of sketch to digital image matching algorithms. This chapter attempts to analyze and evaluate the performance of semi-forensic sketches and use it for improving the training of the algorithms for forensic sketch matching.

2.6.1 Matching Semi-Forensic Sketches

Viewed sketches and forensic sketches are very different from each other. As shown in Figure 2.2, the level of difficulty increases from viewed to forensic sketch matching. In an attempt to bridge the gap between viewed sketches and forensic sketches, semi-forensic sketches are introduced. It is our assertion that training on semi-forensic sketches can improve modeling the variations for matching forensic sketches as compared to training on viewed sketches. Therefore, to better understand the progression from viewed to semi-forensic sketches, experiments are performed where training is done on viewed sketches and performance is evaluated on semi-forensic sketches.

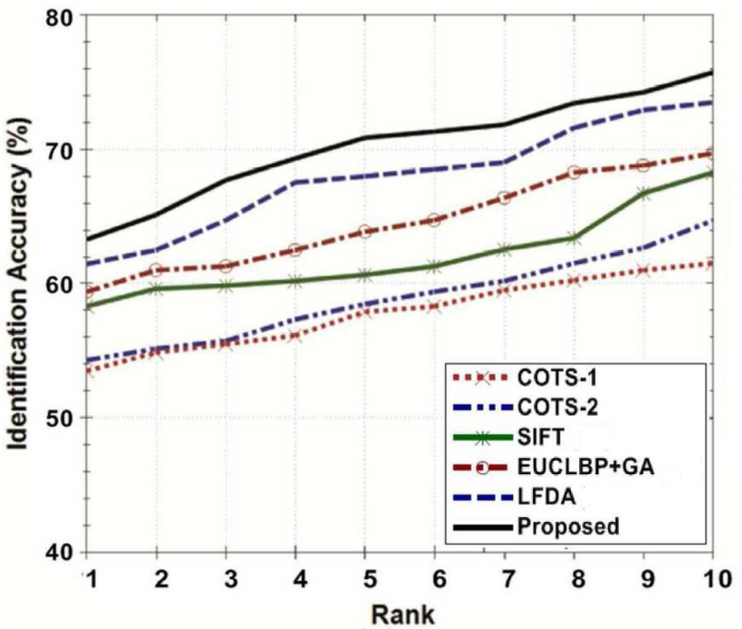

Figure 2.12: CMC curves showing the identification performance when algorithms are trained on viewed sketches and matching is performed on semi-forensic sketches.

To evaluate the performance on semi-forensic sketches, the algorithms are trained on the Viewed Sketch database. 95 sketch-digital image pairs from the IIIT-Delhi Viewed Sketch database are used for training and testing is performed with the remaining 454 digital face images as gallery and 140 semi-forensic sketches as probes. Figure 2.12 shows the rank-1 identification accuracy of sketch to digital face image matching algorithms on semi-forensic sketches. The proposed approach that uses MCWLD and memetically optimized weighted χ^2 distance yields rank-1 identification accuracy of 63.24% and outperforms existing algorithms such as SIFT [83], EUCLBP+GA [86], and LFDA [4] by $2-5\%$. The proposed approach also outperforms the two commercial face recognition systems by at least 9%.

2.6.2 Matching Forensic Sketches

Since forensic sketches are based on the recollection of an eyewitness, they are often inaccurate, incomplete, do not closely resemble the actual digital face image, and may be of poor quality. These concerns make the problem of matching forensic sketches with digital face images more challenging than matching viewed sketches. This section presents the evaluation of algorithms on the Forensic Sketch database.

2.6.2.1 Experimental Protocol

To evaluate the proposed approach for matching forensic sketches, four sets of experiments are performed. The performance of the proposed algorithm is also compared with existing algorithms and two commercial face recognition systems. The protocol for all the experiments are listed below:

1. *Training on IIIT-Delhi Viewed Sketch database*: Training is performed on 140 sketch-digital image pairs from the IIIT-Delhi Viewed Sketch database. For testing, 190 forensic sketches are used as probe. The gallery comprises of 599 digital face images (remaining 409 digital face images from the IIIT-Delhi Viewed Sketch database and 190 digital face images from the Forensic Sketch database).

2. *Training on IIIT-Delhi Semi-forensic Sketch database*: Training is performed on 140 sketch-digital image pairs from the IIIT-Delhi Semi-forensic Sketch database. For testing, 190 forensic sketches are used as probe and 599 digital face images as gallery.

3. *Enhancing Quality of Forensic Sketches*: In this experiment, the quality of Forensic Sketch database is enhanced using the pre-processing technique described in Section 2.2. Training is performed on 140 sketch-digital image pairs from the IIIT-Delhi Viewed Sketch database. 190 forensic sketches are used as probe and 599 digital face images are used as gallery.

4. *Large Scale Forensic Matching*: To replicate the real world scenario of matching forensic sketches to police mugshot database with large gallery size, 6324 digital face (frontal) images obtained from government agencies are appended to the gallery of 739 digital face images used in previous experiments. To evaluate the effect of training on semi-forensic sketches and quality enhancement using the pre-processing algorithm, two experiments are performed in large scale evaluation.

    - Training is performed on 140 sketch-digital image pairs from the IIIT-Delhi Viewed Sketch database and no pre-processing is applied on the forensic sketches.
    - Training is performed on 140 sketch-digital image pairs from the IIIT-Delhi Semi-forensic Sketch database and the forensic sketches are enhanced using the pre-processing technique.

Table 2.4: Rank-1 identification accuracy of sketch to digital face image matching algorithms for matching forensic sketches.

Experiment	**Gallery/ Probe Images**	**Algorithm**	**Rank-1 Identification Accuracy (%)**
Experiment 1 Figure 2.13	599/190	COTS-1	13.62
		COTS-2	13.92
		SIFT [83]	14.26
		EUCLBP+GA [86]	14.81
		LFDA [4]	15.26
		Proposed	**17.19**
Experiment 2 Figure 2.14	599/190	COTS-1	13.62
		COTS-2	13.92
		SIFT [83]	18.26
		EUCLBP+GA [86]	19.81
		LFDA [4]	22.78
		Proposed	**23.94**
Experiment 3 Figure 2.15	599/190	COTS-1	15.62
		COTS-2	16.01
		SIFT [83]	16.26
		EUCLBP+GA [86]	16.54
		LFDA [4]	17.78
		Proposed	**20.94**

Table 2.5: Rank-50 identification accuracy for large scale forensic sketch matching as shown in Figures 2.16 & 2.17.

Experiment 4	**Gallery /Probe Database**	**Algorithm**	**Rank-50 Identification Accuracy(%)**
Training on Viewed Sketch database without pre-processing applied on forensic sketches	6923/190	COTS-1	7.88
		COTS-2	8.46
		SIFT [83]	17.11
		EUCLBP+GA [86]	18.93
		LFDA [4]	20.81
		Proposed	**23.94**
Training on Semi-forensic database with proposed pre-processing on forensic sketches	6923/190	COTS-1	11.28
		COTS-2	12.86
		SIFT [83]	21.24
		EUCLBP+GA [86]	23.75
		LFDA [4]	24.62
		Proposed	**28.52**

2.6.2.2 Experimental Analysis

Figures 2.12-2.17 and Tables 3.2-2.5 illustrate the results of these experiments and the analysis is provided below.

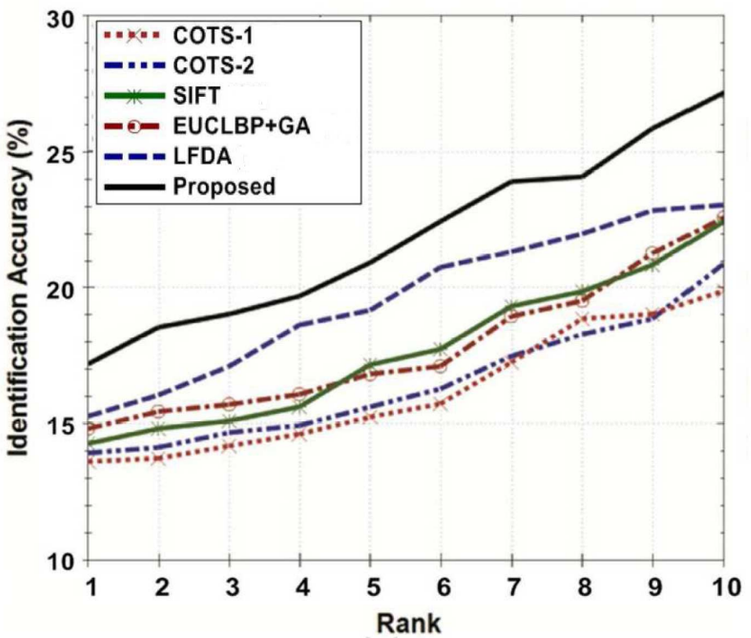

Figure 2.13: CMC curves showing the identification performance when algorithms are trained on viewed sketches and matching is performed on forensic sketches.

- Table 3.2 and Figure 2.13 show identification performance of the proposed and existing algorithms for matching forensic sketches when the algorithms are trained on the IIIT-Delhi Viewed Sketch database (Experiment 1). The proposed algorithm yields 17.19% rank-1 identification accuracy which is about 2% better than existing algorithms. The proposed approach also outperforms the two commercial face recognition systems by at least 3%.

- In Experiment 2, the training is performed on semi-forensic sketches for the same 140 subjects that are used for training in Experiment 1. The results in Figure 2.14 show that there is an improvement of about 7% in rank-1 identification accuracy of the proposed algorithm and at least 4% for existing algorithms when the algorithms are trained using the semi-forensic sketches. This improvement in accuracy validates our assertion that training sketch recognition algorithms on viewed sketches is not sufficient for matching forensic sketches. The proposed algorithm performs better

than LFDA based algorithm [4] because the proposed approach can be efficiently trained even with less number of sketch-digital image pairs whereas LFDA requires large number of training samples to compute the discriminant projection matrices.

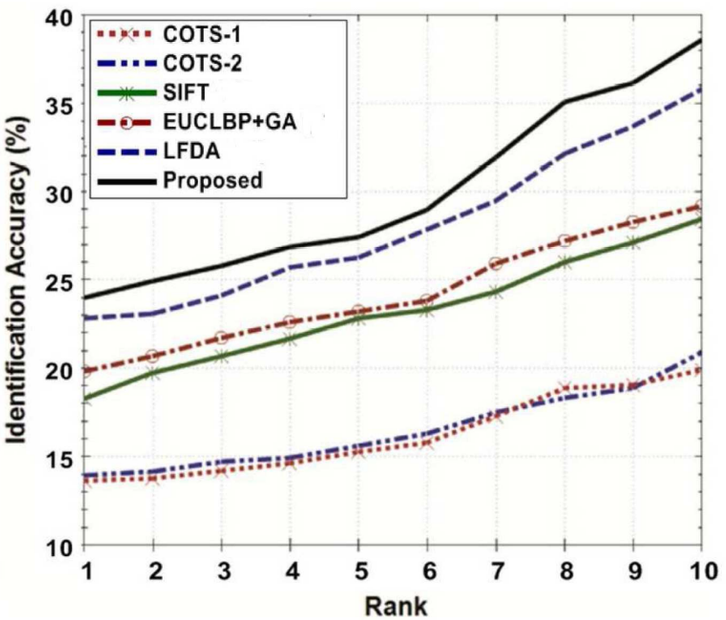

Figure 2.14: CMC curves showing the identification performance when algorithms are trained on semi-forensic sketches and matching is performed on forensic sketches.

- The forensic sketch database contains sketches and digital face images of poor quality. The pre-processing technique enhances the quality of forensic sketches by reducing noise and irregularities from the images. The CMC curves in Figure 2.15 show the results for Experiment 3 where enhancing the quality of forensic sketches leads to an improvement of $2-3\%$ in the rank-1 identification accuracy for all algorithms (compared to CMCs in Figure 2.13).

- Experiment 4 demonstrates the scenario where a forensic sketch is matched against a large mugshot database. The CMC curves in Figure 2.16 show the results for large scale forensic sketch matching when algorithms are trained using viewed sketches without any pre-processing. In this case, rank-50 identification accuracy of the proposed algorithm is 23.9% which is at least 3% better than existing algorithms.

- Comparing the CMC curves in Figure 2.17 show that the pre-processing technique along with training on semi-forensic sketches improves the identification accuracy of the proposed approach significantly (at least 4.72% improvement in rank-1 accuracy).

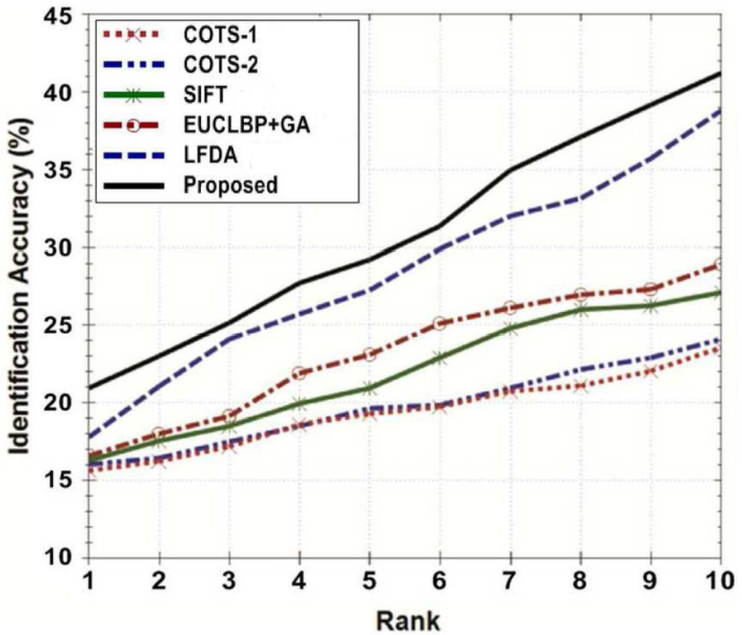

Figure 2.15: CMC curves showing the identification performance when algorithms are trained on viewed sketch-digital image pairs and testing is performed using pre-processed (enhanced) forensic sketch-digital image pairs.

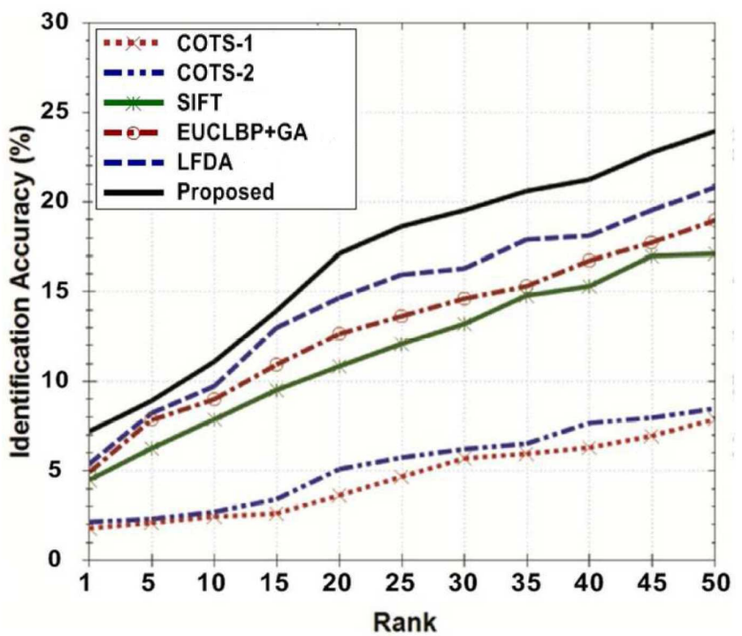

Figure 2.16: CMC curves showing the identification performance when algorithms are trained on viewed sketch-digital image pairs and tested with large scale digital gallery and forensic sketch probes.

Enhancing the quality of forensic sketch-digital image pairs improves the rank-1 identification accuracy of the two commercial face recognition systems also by at least 2%.

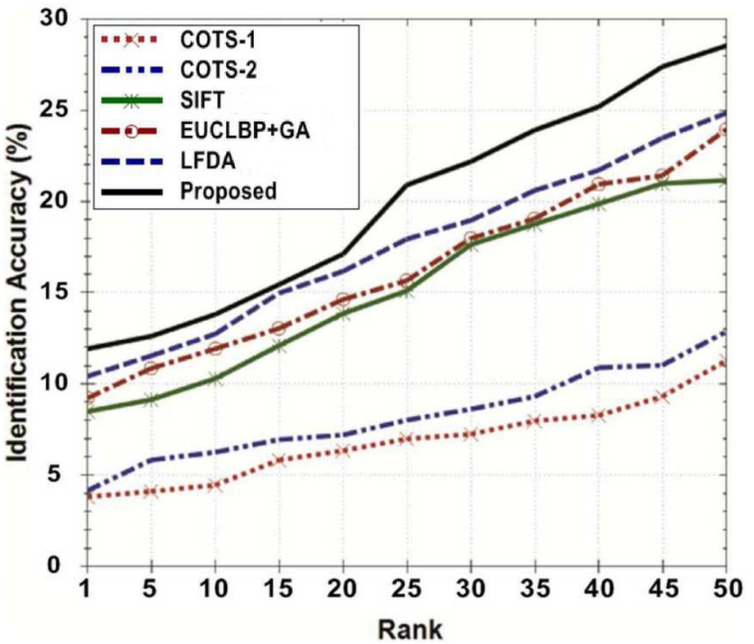

Figure 2.17: CMC curves showing the identification performance when algorithms are trained on semi-forensic sketch-digital image pairs and tested with large scale digital (enhanced) gallery and pre-processed forensic sketch probes.

- The CMC curves in Figures 2.16 and 2.17 suggest that existing algorithms for matching sketches to digital face images are still not able to achieve acceptable identification accuracy for large scale applications. However, the proposed algorithm still performs better than existing algorithms and commercial face recognition systems. As shown in Table 2.5, the proposed algorithm achieves rank-50 accuracy of 28.52% which is at least 4% better than existing algorithms and 15% better than the two commercial face recognition systems.

- It is to be noted that the performance of automated algorithms on semi-forensic sketches is better than the performance on forensic sketches. This improvement is attributed to the fact that semi-forensic sketches act like a bridge between viewed and forensic sketches. Therefore, training sketch recognition algorithms on semi-forensic sketches consistently improves the performance for all existing algorithms.

- At 95% confidence interval, non-parametric rank-ordered test (using the ranks obtained from the algorithms) and parametric t-test (using the match scores) suggest that the two top performing algorithms (i.e. the proposed and LFDA) are significantly (statistically) different.

- Finally, on a 2 GHz Intel Duo Core processor with 4 GB RAM under $C\#$ programming environment, for a given probe sketch, the proposed algorithm requires 0.096 seconds to compute the MCWLD descriptor.

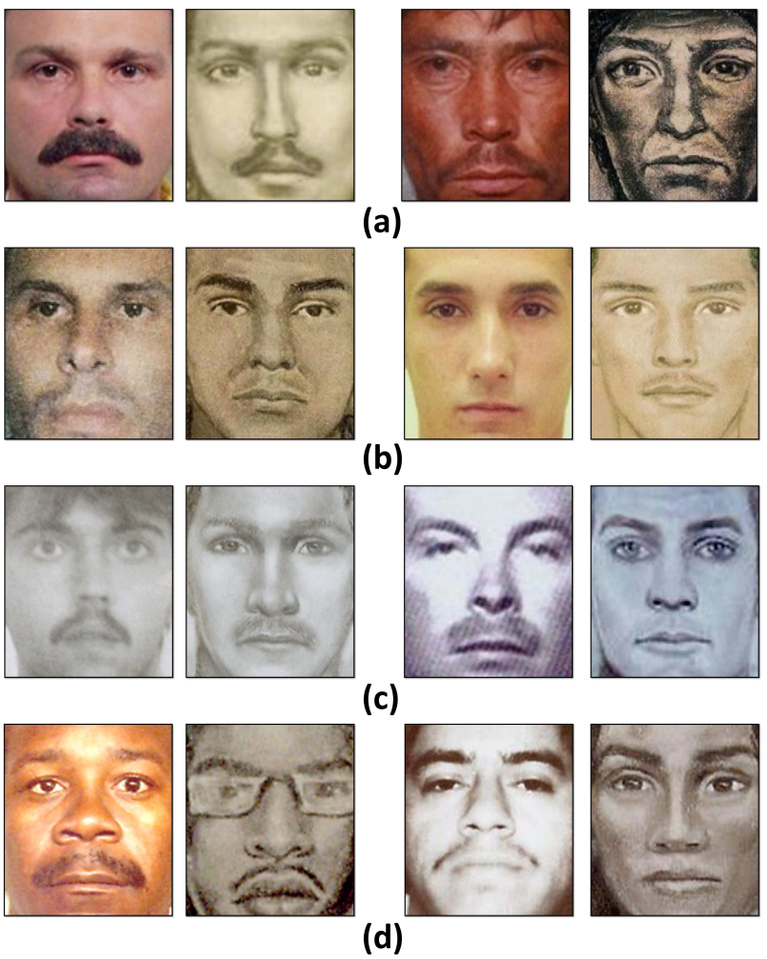

Figure 2.18: Illustrating sample cases when (a) the proposed approach and LFDA [4] correctly recognize, (b) LFDA fails while the proposed algorithm correctly recognizes, (c) the proposed algorithm fails while LFDA correctly recognizes, and (d) both the algorithms fail to recognize.

The proposed approach emphasizes on the discriminating information vested in the local regions. To capture our assertion that every local region has varying contribution, memetic algorithm assigns optimal weights to each local facial region. Assigning discriminating weights to different facial regions also supports the conclusion made by Klare *et al.* [4] that different internal, external, and individual face regions (eyes, nose, mouth, chin etc.) have significant contribution for sketch recognition. Next, Figure 2.18(a) shows some examples of sketch-digital image pairs that are correctly identified by the proposed approach as well as the LFDA [4] based approach (correctly identified in rank-50). Sketches that show high recognizability have some peculiar features such as beard, mustache, and soft marks on the face. Figure 2.18(b) shows some examples where LFDA based approach performed poorly while the proposed approach correctly identified the sketch. This is mainly because the proposed approach focuses on the structural details along with discriminating and prominent features of the face. Figure 2.18(c) shows some examples of sketch-digital image pairs where the proposed approach performed poorly, whereas, LFDA based approach correctly matched sketches with digital face images. Finally, Figure 2.18(d) shows some examples where both the proposed approach and LFDA based approach failed to match sketches with the correct digital face images. These sketches either do not resemble the actual digital face image or converge to an average face that resembles more than one digital face image in gallery because of the common features.

2.7 Human Analysis for Matching Sketches with Digital Face Images

Several studies have analyzed human capabilities to recognize faces with variations due to illumination and expression [100]. Recently, Zhang *et al.* [85] performed an extensive study to analyze the human performance in matching sketches obtained from multiple artists. This section presents a study to understand the cognitive process of matching sketches with digital face images by humans on viewed, semi-forensic and forensic sketch databases. This examination of human responses also considers local region used by each subject while matching sketches with digital face images.

2.7.1 Experimental Method

Since the validity of a psychological experiment is closely related to fatigue and interest level of the subject [85], human analysis is performed on a subset of 140 viewed, 140 semi-forensic and 190 forensic sketches.

2.7.1.1 Participants

A total of 82 subjects, largely undergraduate university students, volunteered to participate in the sketch to digital face image matching study. Some of the volunteers may be familiar with few subjects in the IIIT-Delhi Viewed and Semi-forensic Sketch database but not with any of the sketches in Forensic Sketch database.

2.7.1.2 Questions

In every question, a probe sketch must be matched to one of the 12 digital face images in the gallery. Since this is a web based application, we came up with 12 digital face images as gallery so as to properly layout the query sketch and digital face images on a computer screen. The gallery necessarily include the correct matching digital face image and the remaining images in the gallery are the top retrieved digital face images for the probe sketch obtained using the proposed MCWLD algorithm. In the interest of fairness, un-cropped images that may include hair, ears, and neck are used for human evaluation. The automatic algorithms on the other hand, do not require this additional information.

2.7.1.3 Procedure

Each volunteer interacts with a web interface, where he/she is first authenticated. It is done to ensure that the user gets different questions in every session. Subsequently, the volunteer is presented with the questions, one at a time. Each question is selected randomly from a unique unanswered question bank comprising a mixture of viewed, semi-forensic and forensic sketches. Further, the user selects one of the gallery image as a suitable match for the query sketch. Along with this selection, the user marks the local region in the digital face image that he/she finds to be the most beneficial in recognizing the query sketch. This response is indicated by the user's click on the most discriminating local facial region of the selected gallery image. A volunteer answers between 2 and 12 questions in a single session and can participate in up to four sessions.

2.7.2 Results and Analysis

A total of 1169 human responses are obtained for 470 probe sketches. Of these responses, 71.94% are found to be correct matches. Table 2.6 shows the total number of responses and individual accuracy of these responses across the three types of sketches. Further, Figure 2.19 shows human response (clicks) that the participant deemed as important in matching the sketches with digital face images. These clicks are plotted over a mean

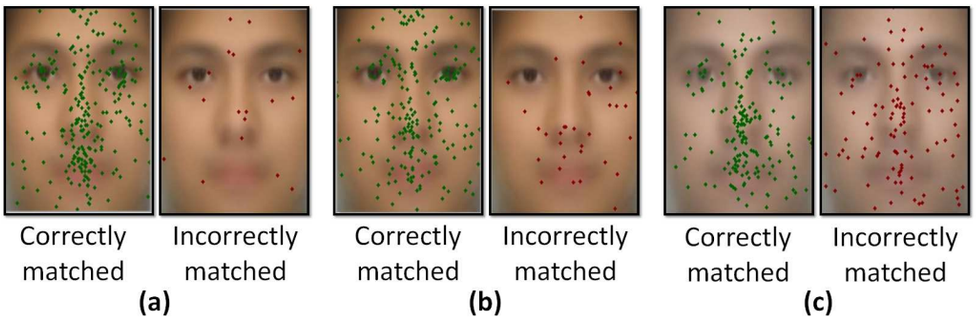

Figure 2.19: Facial regions for correctly and incorrectly matched (a) viewed sketches, (b) semi-forensic sketches, and (c) forensic sketches. Dots represents the area that user found to be most discriminating in matching the sketch with digital face images.

Table 2.6: Distribution of 1169 human responses obtained from the study.

Type	**Total Human Responses**	**% Correct**
Viewed	403	80.4
Semi-forensic	334	79.6
Forensic	432	58.1

face image to enable better visualization. The key observations from this study are listed below:

- The click-points, shown in Fig 2.19 indicate that the dominant local regions of a face image such as mouth, nose, and eyes (accurately depicted by the sketch artist), are used for matching.

- Figure 2.19(a) shows the click-plot when the user is presented viewed sketches. The high accuracy can be attributed to the correct depiction of the features by the artist. The user clicks are concentrated close to nose and mouth region.

- Figure 2.19(b) shows the click-plot when the user is presented semi-forensic sketches. The points seem to deviate towards the exaggerated features such as corners of eyes, nose and eyebrows.

Table 2.7: Distribution of user clicks between prominent facial regions.

	Viewed (%)	**Semi-forensic (%)**	**Forensic (%)**
Eyes	6.13	13.97	13.17
Nose	18.10	14.90	18.10
Mouth	10.58	10.56	14.76

- *Forensic Sketch* database contains poor quality sketches and *two-fold* exaggeration at witness description and artist depiction. The large differences in appearance, age, and high possibility of accessaries result in user preference for nose and mouth regions, as shown in Figure 2.19(c).

- As the difficulty of recognition task escalates from viewed to forensic sketches, there is a notable increase in use of the prominent facial features (eyes, nose, and mouth), as indicated in Table 2.7. This marked increase in user preference for local facial features when presented with unfamiliar sketches is a strong indication of their importance in the recognition task.

- This study supports our initial hypothesis that local regions provide discriminating information for matching sketches with digital face images. Finally, with 1169 sample size at 95% confidence level, confidence interval lies in $2-3\%$ for the three types of sketches.

The accuracy claimed by humans for different types of sketches cannot be compared with the accuracy of automatic algorithms because of different experimental protocols. This analysis is to validate our assertion that discriminating patterns in local facial regions have major contribution in recognizing sketches with digital face images.

2.8 Summary

Sketch to digital face matching is an important research challenge and is very pertinent to law enforcement agencies. This chapter presents a discriminative approach for matching sketch-digital image pairs using modified Weber's local descriptor and memetically optimized weighted χ^2 distance. The algorithm starts with the pre-processing technique to enhance sketches and digital images by removing irregularities and noise. Next, MCWLD encodes salient micro patterns from local regions to form facial signatures of both sketches and digital face images. Finally, the proposed (evolutionary) memetic optimization based weighted χ^2 distance is used to match two MCWLD histograms. Comprehensive analysis, including comparison with existing algorithms and two commercial face recognition systems, is performed using the viewed, semi-forensic, and forensic sketch databases. Semi-forensic sketches are introduced to bridge the gap between viewed and forensic sketches. It is observed that sketch recognition algorithms trained on semi-forensic sketches can better model the variations for matching forensic sketches as compared to algorithms trained on viewed sketches. Analysis of results also suggest that local regions play an important role

in matching sketch-digital image pairs and is effectively encoded in MCWLD and memetically optimized weighted χ^2 distance. The results also show that the proposed algorithm is significantly better than existing approaches and commercial systems.

Chapter 3

Recognizing Surgically Altered Face Images using Multi-objective Evolutionary Learning

3.1 Introduction

Plastic surgery procedures provide a proficient and enduring way to enhance the facial appearance by correcting feature anomalies and treating facial skin to get a younger look. Apart from cosmetic reasons, plastic surgery procedures are beneficial for patients suffering from several kinds of disorders caused due to excessive structural growth of facial features or skin tissues. Plastic surgery procedures amend the facial features and skin texture thereby providing a makeover in the appearance of face. Figure 3.1 shows an example of the effect of plastic surgery on facial appearances. With reduction in cost and time required for these procedures, the popularity of plastic surgery is increasing. Even the widespread acceptability in the society encourages individuals to undergo plastic surgery for cosmetic reasons. According to the statistics provided by the American Society for Aesthetic Plastic Surgery for year 2010 [101], there is about 9% increase in the total number of cosmetic surgery procedures, with over $500,000$ surgical procedures performed on face.

Transmuting facial geometry and texture increases the intra-class variability between the pre- and post-surgery images of the same individual. Therefore, matching post-surgery images with pre-surgery images becomes an arduous task for automatic face recognition algorithms. Further, as shown in Figure 3.2, it is our assertion that variations caused due to plastic surgery have some intersection with the variations caused due to aging and disguise. Facial aging is a biological process that leads to gradual changes in the geometry and texture of a face. Unlike aging, plastic surgery is a spontaneous process that is

Figure 3.1: Illustrating the variations in facial appearance, texture, and structural geometry caused due to plastic surgery (images taken from internet).

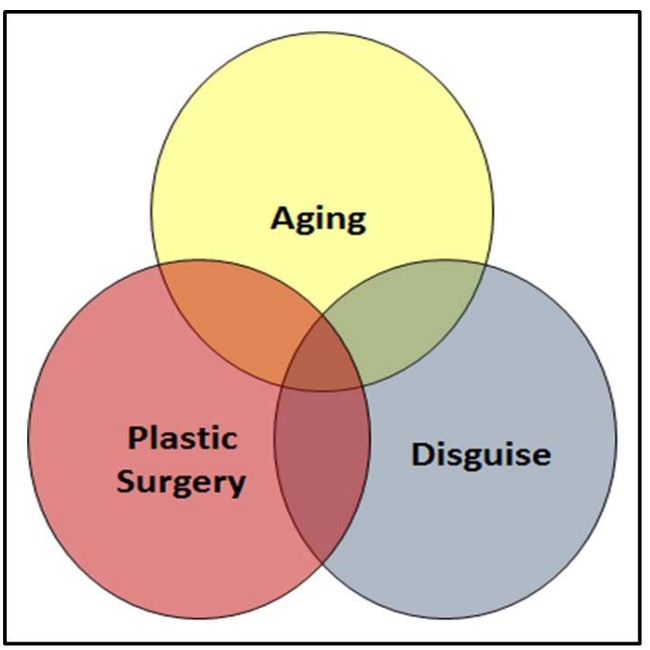

Figure 3.2: Relation among plastic surgery, aging, and disguise variations with respect to face recognition.

generally performed contrary to the effect of facial aging. Since the variations caused due to plastic surgery procedures are spontaneous, it is difficult for face recognition algorithms to model such non-uniform face transformations. On the other hand, disguise is the process of concealing one's identity by using makeup and other accessories. Both plastic surgery and disguise can be misused by individuals trying to conceal their identity and evade recognition. Variations caused due to disguise are temporary and reversible; however, variations caused due to plastic surgery are long-lasting and may not be reversible. Owing to these reasons, plastic surgery is now established as a new and challenging covariate of face recognition alongside aging and disguise.

3.1.1 Related Research

Singh *et al.* [8] analyzed several types of local and global plastic surgery procedures and their effect on different face recognition algorithms. They have experimentally shown that the non-linear variations induced by surgical procedures are difficult to address with current face recognition algorithms. Marsico *et al.* [102] also proposed an approach that integrated information derived from local regions to match pre- and post-surgery face images. Aggarwal *et al.* [103] proposed sparse representation approach on local facial fragments to match surgically altered face images. Kose *et al.* [104] proposed a block based face recognition to match face images with nose alterations in both 2D and 3D domain. The blocks which maximize the recognition performance were used in the algorithms to mitigate the effect of nose alteration both 2D and 3D. A nose alteration database was prepared from FRGC database [105] in which nose in each sample is replaced with the nose region from another randomly chosen individual. Further, Erdogmus *et al.* [106] analyzed how such changes on nose region affect the face recognition performances of several 2D and 3D algorithms and concluded that 3D algorithms were more vulnerable to variations in nose regions. Recently, Jillela and Ross [107] proposed a fusion approach that combined information from the face and ocular regions to enhance recognition performance across face images altered due to facial plastic surgery. Bhatt *et al.* [108] proposed an evolutionary granular computing based algorithm for recognizing faces altered due to plastic surgery. Their algorithm generated non-disjoint face granules where each face granule represented different information at varying size and resolution. Further, two feature extractors were used for extracting discriminative information from face granules. Finally, different responses were unified in an evolutionary manner using multi-objective genetic algorithm for improved performance.

There has been some interesting work in predicting the post-surgery faces based on the previous examples from pre- and post-surgery faces. In this direction, Liu *et al.* [109] trained a predictor on previous set of pre- and post-surgery landmark distances and used it to predict the new positions of landmark points after plastic surgery. Image morphing was used to synthesize the appearance of face after plastic surgery. Rabi and Arabi [110] proposed an approach to synthetically alter the facial features to simulate the results of plastic surgery. The proposed approach replaces the facial features of an individual with corresponding facial features of other individuals and seamlessly fuses the facial features to remove any discontinuity. This paper presents a good future direction to create large databases to study the effect of plastic surgery as creating database with plastic surgery variations is very challenging due to privacy issues. However, research in predicting the post surgery face is very naive as the proposed approach just recovers the changes in facial shape and does not incorporate the texture variations. More complex models need to be learned to simulate the effects of different local and global plastic surgery procedures.

Though results suggest that algorithms are improving towards addressing the challenge, there is a significant scope for further improvement. Plastic surgery also raises some social and ethical issues [111], being related to the medical history of an individual which is secure under law, invasion of privacy is an important aspect in this research. Apart from affecting the face recognition algorithms, plastic surgery procedures may also lead to identity theft. Identity theft can be intentional, when a person consciously attempts to resemble someone by undergoing facial plastic surgery procedures, or unintentional where he/she may resemble someone after the surgery. Plastic surgery procedures modifying the facial geometry and texture along with associated privacy issues make it an arduous research problem. These procedures can significantly modify facial regions both locally and globally. Since, existing face recognition algorithms generally rely on local and global facial features, variations in these features can affect the recognition performance. Table 3.1 summarizes the performance of different algorithms for matching pre- and post-surgery images from the plastic surgery face database [8]. Results suggest that further research is required to design optimal face recognition algorithms that can account for the challenges due to facial plastic surgery procedures.

Table 3.1: A comparison of different approaches proposed for matching pre- and post-surgery images on the Plastic Surgery face database [8].

Authors	**Approach**	**Rank-1 accuracy**
Singh *et al.* [8]	PCA	29.1%
	FDA	32.5%
	LFA	38.6%
	CLBP	47.8%
	SURF	50.9%
	GNN	54.2%
Marsico *et al.* [102]	PCA	26%
	LDA	35%
	FARO	41%
	FACE	65%
Aggarwal *et al.* [103]	Sparse representation	77.9%
Jillela and Ross [107]	Fusion of face and ocular region	87.4% (On a subset)
Bhatt *et al.* [108]	Multi-objective genetic approach	87.3%

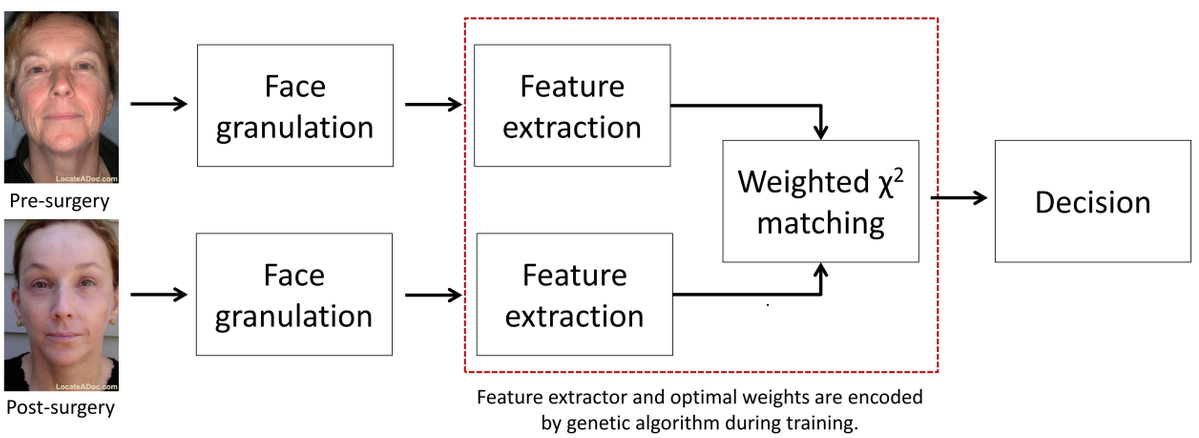

Figure 3.3: Block diagram illustrating different stages of the proposed algorithm.

3.1.2 Research Contributions

This chapter presents a multi-objective evolutionary granular computing based algorithm for recognizing faces altered due to plastic surgery procedures. As shown in Figure 3.3, the proposed algorithm starts with generating non-disjoint face granules where each granule represents different information at varying size and resolution. Further, two feature extractors, namely Extended Uniform Circular Local Binary Pattern (EUCLBP) [86] and Scale Invariant Feature Transform [93], are used for extracting discriminating information from face granules. Finally, different responses are unified in an evolutionary manner using a multi-objective genetic approach for improved performance. The performance of the proposed algorithm is compared with a commercial-off-the-shelf face recognition system (COTS) for matching surgically altered face images against large scale gallery. The chapter also analyzes the effect of plastic surgery procedures on the performance of individual granules along with periocular region. The chapter is organized as follows: Section 3.2 presents the proposed face granulation scheme along with the evolutionary optimization for selecting the feature extractor and weights of each granule. Section 3.3 presents the databases, comprehensive experimental results, and key observations.

3.2 Evolutionary Granular Computing Approach for Face Recognition

Face recognition algorithms either use facial information in a holistic way or extract features and process them in parts. In presence of variations such as pose, expression, illumination, and disguise, it is observed that local facial regions are more resilient and can therefore be used for efficient face recognition [112], [113], [114], [115]. Several part based face recognition approaches capture this observation for improved performance. Heisele *et al.* [112] proposed a component based face recognition approach using different facial components to provide robustness to pose. Weyrauch *et al.* [113] designed an algorithm in which gray-level pixel values from several facial components were concatenated and classification was performed using SVM. Similarly, Li *et al.* [114] proposed an approach where local patches were extracted from different levels of Gaussian pyramid and arranged in an exemplar manner. These exemplar based-local patches were then combined using boosting to construct strong classifiers for prediction. In another approach, a subset selection mechanism was proposed [115] where the most informative local facial locations were used in decision making.

Singh *et al.* [8] observed that with respect to plastic surgery, more than one facial region may be affected due to a procedure. For example, blepharoplasty which is primarily performed to amend forehead also affects eye-brows. Further, Singh *et al.* [8] observed that with large variations in the appearance, texture, and shape of different facial regions, it is difficult for face recognition algorithms to match a post-surgery face image with pre-surgery face images. Previous part-based face recognition approaches may not provide mechanisms to address the concurrent variations introduced in multiple features because these approaches generally emphasize on analyzing each feature independently. On the other hand, it is observed that humans solve problems using perception and knowledge represented at different levels of information granularity [95]. They recognize faces using a combination of holistic approaches together with discrete levels of information (or features). Sinha *et al.* [95] established 19 results based on the face recognition capabilities of a human mind. It is suggested that humans can efficiently recognize faces even with low resolution and noise. Moreover, high and low frequency facial information is processed both holistically and locally. Campbell *et al.* [116] reported that inner and outer facial regions represent distinct information that can be helpful for face recognition. Researchers from cognitive science also suggested that local facial fragments can provide robustness against partial occlusion and change in viewpoints [95], [117], [118]. To incorporate these observations, this chapter proposes a granular approach for facial feature extraction and matching. In the granular approach [119, 120], as shown in Figure 3.3, non-disjoint features are extracted at different granular levels. These features are then synergistically combined using multi-objective evolutionary learning to obtain the assimilated information. With granulated information, more flexibility is achieved in analyzing underlying information such as nose, ears, forehead, cheeks, and combination of two or more features. The face granulation scheme helps in analyzing multiple features simultaneously. Moreover, the face granules of different sizes and shapes (as shown in Figures 3.4-3.7) help to gain significant insights about the effect of plastic surgery procedures on different facial features and their neighboring regions.

3.2.1 Face Image Granulation

Let F be the detected frontal face image of size $n \times m$. Face granules are generated pertaining to three levels of granularity. The first level provides global information at multiple resolutions. This is analogous to a human mind processing holistic information for face recognition at varying resolutions. Next, to incorporate the findings of Campbell *et al.* [116], inner and outer facial information are extracted at the second level. Local

facial features play an important role in face recognition by human mind. Therefore, at the third level, features are extracted from the local facial regions.

3.2.1.1 First Level of Granularity

In the first level, face granules are generated by applying the Gaussian and Laplacian operators [121]. The Gaussian operator generates a sequence of low pass filtered images by iteratively convolving each of the constituent images with a 2D Gaussian kernel. The resolution and sample density of the image is reduced between successive iterations and therefore the Gaussian kernel operates on a reduced version of the original image in every iteration. Similarly, the Laplacian operator generates a series of band-pass images. Let the granules generated by Gaussian and Laplacian operators be represented by F_{Gri}, where i represents the granule number. For a face image of size 196×224, Figure 3.4 represents the face granules generated in the first level by applying Gaussian and Laplacian operators. The resultant images may be viewed as a 'pyramid' with F_{Gr1} and F_{Gr4} having the highest resolution and F_{Gr3} and F_{Gr6} having the lowest resolution. F_{Gr1} to F_{Gr3} are the granules generated by Gaussian operator and F_{Gr4} to F_{Gr6} are the granules generated by Laplacian operator. The size of the smallest granule in the first granular level is 49×56. In these six granules, facial features are segregated at different resolutions to provide edge information, noise, smoothness, and blurriness present in a face image. As shown in Figure 3.4, the effect of facial wrinkles is reduced from granule F_{Gr1} to F_{Gr3}. The first level of granularity thus compensates for the variations in facial texture, thereby providing resilience to plastic surgery procedures that alter the face texture such as facelift, skin resurfacing, and dermabrasion.

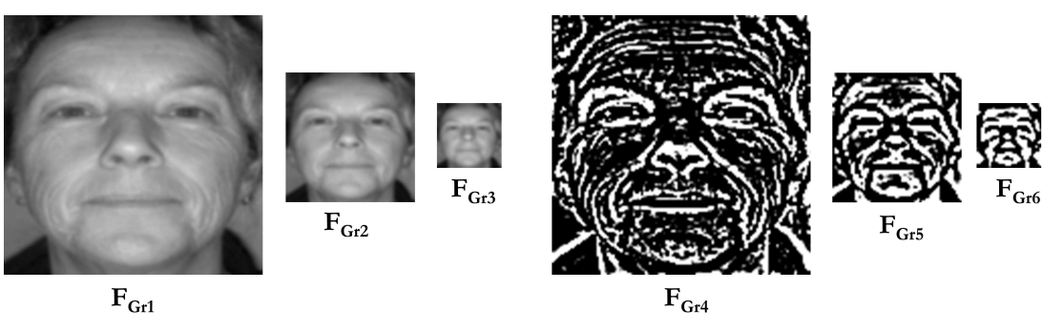

Figure 3.4: Face granules in the first level of granularity. F_{Gr1}, F_{Gr2}, and F_{Gr3} are generated by the Gaussian operator, and F_{Gr4}, F_{Gr5}, and F_{Gr6} are generated by the Laplacian operator.

3.2.1.2 Second Level of Granularity

To accommodate the observations of Campbell *et al.* [116], horizontal and vertical granules are generated by dividing the face image F into different regions as shown in Figures 3.5 and 3.6. Here, F_{Gr7} to F_{Gr15} denote the horizontal granules and F_{Gr16} to F_{Gr24} denote the vertical granules. Among the nine horizontal granules, the first three granules i.e. F_{Gr7}, F_{Gr8}, and F_{Gr9} are of size $n \times m/3$. The next three granules, i.e., F_{Gr10}, F_{Gr11}, and F_{Gr12} are generated such that the size of F_{Gr10} and F_{Gr12} is $n \times (m - \epsilon)$ and the size of F_{Gr11} is $n \times (m + 2\epsilon)$. Further, F_{Gr13}, F_{Gr14}, and F_{Gr15} are generated such that the size of F_{Gr13} and F_{Gr15} is $n \times (m + \epsilon)$ and the size of F_{Gr14} is $n \times (m - 2\epsilon)$. Nine vertical granules, F_{Gr16} to F_{Gr24}, are also generated in a similar manner. Figures 3.5 and 3.6 show horizontal and vertical granules when the size of face image is 196×224 and $\epsilon = 15$[1]. The second level of granularity provides resilience to variations in inner and outer facial regions. It utilizes the relation between horizontal and vertical granules to address the variations in chin, forehead, ears, and cheeks caused due to plastic surgery procedures.

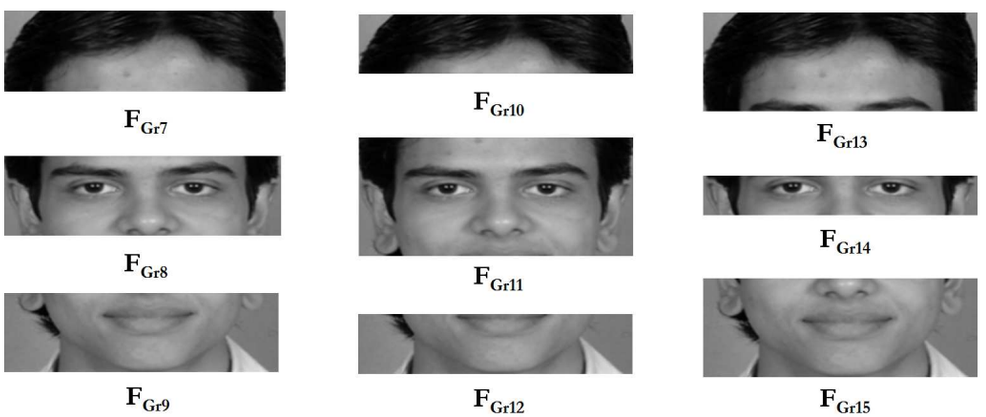

Figure 3.5: Horizontal face granules from the second level of granularity ($F_{Gr7} - F_{Gr15}$).

[1]In the experiments, it is observed that $\epsilon = 15$ yields the best recognition results when face image is of size 196×224.

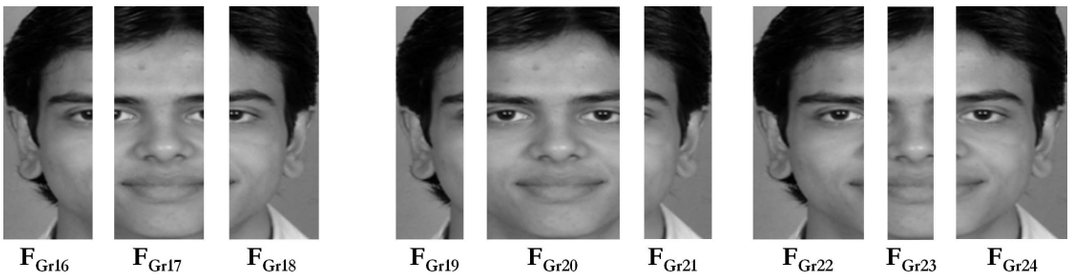

Figure 3.6: Vertical face granules from the second level of granularity ($F_{Gr16} - F_{Gr24}$).

3.2.1.3 Third Level of Granularity

As mentioned previously, human mind can distinguish and classify individuals with their local facial regions such as nose, eyes, and mouth. To incorporate this property, local facial fragments are extracted and utilized as granules in the third level of granularity. Given the eye coordinates, 16 local facial regions are extracted using the golden ratio face template [5] shown in Figure 3.7(a). Each of these regions is a granule representing local information that provides unique features for handling variations due to plastic surgery. Figure 3.7(b) shows an example of local facial fragments used as face granules in the third level of granularity.

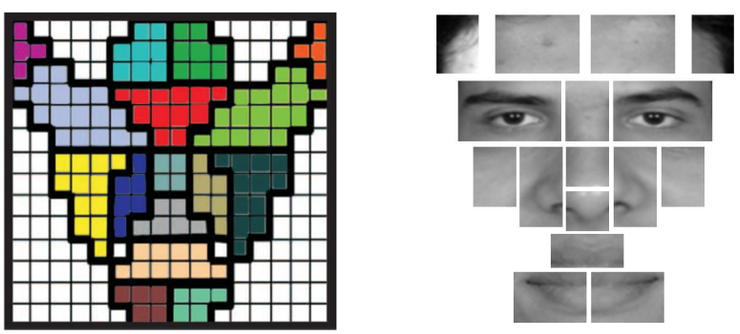

Figure 3.7: (a) Golden ratio face template [5] and (b) face granules in the third level of granularity ($F_{Gr25} - F_{Gr40}$).

The proposed granulation technique is used to generate 40 non-disjoint face granules from a face image of size 196×224. Three levels of granularity are designed to address

specific variations introduced in local facial regions by different plastic surgery procedures. For example, variations in skin texture due to dermabrasion or skin-resurfacing are more pertinent in F_{Gr1} and F_{Gr4} while texture variations are suppressed in granules F_{Gr2}, F_{Gr3}, F_{Gr5}, and F_{Gr6}. The second level of granularity (F_{Gr7} - F_{Gr24}) helps analyze different combinations of local features that provide resilience to concurrent variations introduced in multiple regions by different plastic surgery procedures (such as blepharoplasty, browlift, and rhinoplasty). The third level of granularity (F_{Gr25} - F_{Gr40}) independently analyzes each local feature to address the variations in individual facial regions. Selection of these 40 fixed structure[1] face granules is based on their capability to address specific variations.

3.2.2 Facial Feature Extraction

The proposed granulation scheme results in granules with varying information content. Some granules contain fiducial features such as eyes, nose, and mouth while some granules predominantly contain skin regions such as forehead, cheeks, and outer facial region. Therefore, different feature extractors are needed to encode diverse information from the granules. Any two (complementing) feature extractors can be used; here Extended Uniform Circular Local Binary Patterns and Scale Invariant Feature Transform are used. Both these feature extractors are fast, discriminating, rotation invariant, and robust to changes in gray level intensities due to illumination. However, the information encoded by these two feature extractors is rather diverse as one encodes the difference in intensity values while the other assimilates information from the image gradients. They efficiently use information assimilated from local regions and form a global image signature by concatenating the descriptors obtained from every local facial region. It is experimentally observed that among the 40 face granules, for some granules EUCLBP finds more discriminative features than SIFT and vice-versa (later shown in the experimental results).

3.2.2.1 Extended Uniform Circular Local Binary Patterns

Extended Uniform Circular Local Binary Pattern (EUCLBP) [86] is a texture based descriptor that encodes exact gray-level differences along with difference of sign between neighboring pixels. For computing EUCLBP descriptor, the image is first tessellated into non-overlapping uniform local patches of size 32×32. For each local patch, the EUCLBP

[1] In our approach, eye-coordinates are detected using the OpenCV's boosted cascade of Haar-like features. Since, the plastic surgery face database contains images with frontal pose and neutral expression, the OpenCV's eye detection is accurate. Using the eye-coordinates, face image is normalized with respect to the horizontal axis and inter-eye distance is fixed to 100 pixels. Finally, the detected images are resized to 196×224.

descriptor is computed based on the 8 neighboring pixels uniformly sampled on a circle (radius=2) centered at the current pixel. The concatenation of descriptors from each local patch constitutes the image signature. Two EUCLBP descriptors are matched using the weighted χ^2 distance.

3.2.2.2 Scale Invariant Feature Transform

SIFT [93] is a scale and rotation invariant descriptor that generates a compact representation of an image based on the magnitude, orientation, and spatial vicinity of image gradients. SIFT, as proposed by Lowe [93], is a sparse descriptor that is computed around the detected interest points. However, SIFT can also be used in a dense manner where the descriptor is computed around pre-defined interest points. SIFT descriptor is computed in a dense manner over a set of uniformly distributed non-overlapping local regions of size 32×32. SIFT descriptors computed at the sampled regions are then concatenated to form the image signature. Similar to EUCLBP, weighted χ^2 distance is used to compare two SIFT descriptors.

3.2.3 Multi-objective Evolutionary Approach for Selection of Feature Extractor and Weight Optimization

Every face granule has useful but diverse information, which if combined together can provide discriminating information for face recognition. Moreover, psychological studies in face recognition [95] have also shown that some facial regions are more discriminating than others and hence, contribute more towards the recognition accuracy. Feature selection methods are used for selective combination of features to assimilate diverse information for improved performance. Sequential feature selection (SFS) [122] and sequential floating forward selection (SFFS) [122] are widely used feature selection methods that evaluate the growing feature set by sequentially adding (or removing) features one-at-a-time. On the other hand, a definitive feature selection approach concatenates different features (for example, EUCLBP and SIFT) and performs dimensionality reduction using PCA to yield the final feature set. Other approaches such as genetic search [115] and conditional mutual information (CMI) [123] are also used to find the most informative features. These existing feature selection techniques are single objective functions and may not be sufficient for improving the performance with single gallery evaluations.

Feature selection problem encompasses around two objectives: 1) select an optimal feature extractor for each granule, and 2) assign proper weight for each face granule. The problem of finding optimal feature extractor and weight for each granule involves

searching very large space and finding several suboptimal solutions. Genetic algorithms (GA) are well proven in searching very large spaces to quickly converge to the near optimal solution [124]. Therefore, a multi-objective genetic algorithm is proposed to incorporate feature selection and weight optimization for each face granule. Figure 3.8 represents the multi-objective genetic search process and the steps involved are described below.

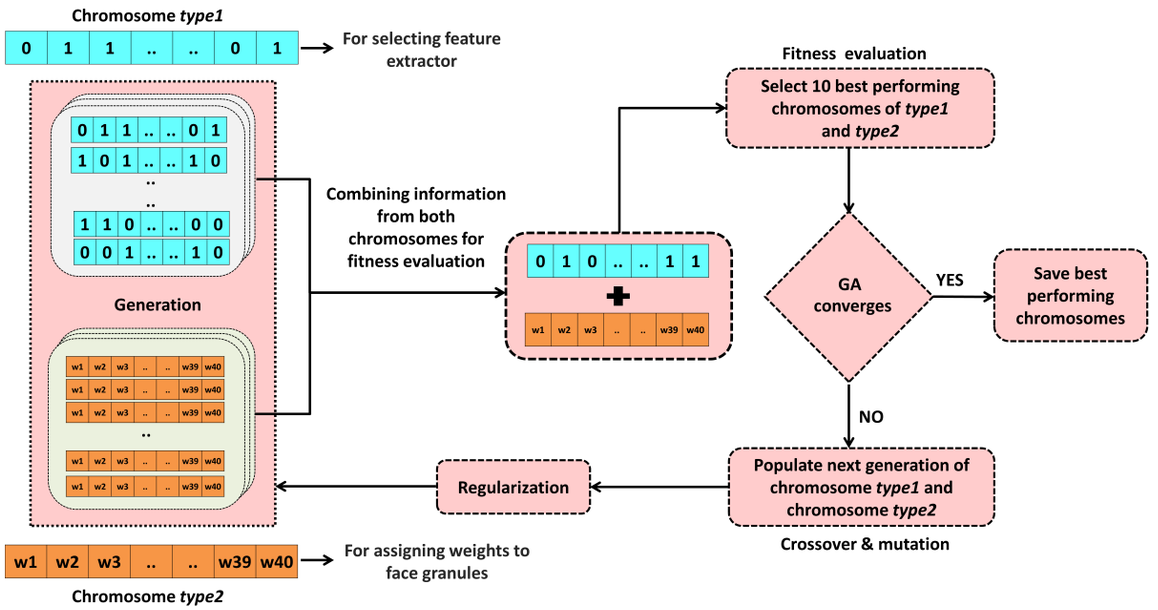

Figure 3.8: Genetic optimization process for selecting feature extractor and weight for each face granule.

Genetic Encoding: A chromosome is a string whose length is equal to the number of face granules i.e. 40 in our case. For simultaneous optimization of two functions, two types of chromosomes are encoded: (i) for selecting feature extractor (referred to as chromosome *type*1) and (ii) for assigning weights to each face granule (referred to as chromosome *type*2). Each gene (unit) in chromosome *type*1 is a binary bit 0 or 1 where 0 represents the SIFT feature extractor and 1 represents the EUCLBP feature extractor. Genes in chromosome *type*2 have real valued numbers associated with corresponding weights of the 40 face granules.

Initial Population: Two generations with 100 chromosomes are populated. One generation has all *type*1 chromosomes while the other generation has all *type*2 chromosomes.

1. For selecting feature extractors (*type*1 chromosome), half of the initial generation (i.e. 50 chromosomes) is set with all the genes (units) as 1, which represents EUCLBP

as the feature extractor for all 40 face granules. The remaining 50 chromosomes in the initial generation have all genes as 0 representing SIFT as the feature extractor for all 40 face granules.

2. For assigning weights to face granules (*type*2 chromosome), a chromosome with weights proportional to the identification accuracy of individual face granules (as proposed by Ahonen [92]) is used as the seed chromosome. The remaining 99 chromosomes are generated by randomly changing one or more genes in the seed chromosome. Further, the weights are normalized such that the sum of all the weights in a chromosome is 1.

Fitness Function: Both *type*1 and *type*2 chromosomes are combined and evaluated simultaneously. Recognition is performed using the feature extractor selected by chromosome *type*1 and weight encoded by chromosome *type*2 for each face granule. Identification accuracy, used as the fitness function, is computed on the training set and 10 best performing chromosomes are selected as *parents* to populate the next generation.

Crossover: A set of uniform crossover operations is performed on *parents* to populate a new generation of 100 chromosomes. Crossover operation is same for both *type*1 and *type*2 chromosomes.

Mutation: After crossover, mutation is performed for *type*2 chromosomes by changing one or more weights by a factor of its standard deviation in the previous generation. For *type*1 chromosome, mutation is performed by randomly inverting the genes in the chromosome.

The search process is repeated till convergence and terminated when the identification performance of the chromosomes in new generation do not improve compared to the performance of chromosomes in previous five generations. At this point, the feature extractor and optimal weights for each face granule (i.e. chromosomes giving best recognition accuracy on the training data) are obtained. Genetic optimization also enables discarding redundant and non-discriminating face granules that do not contribute much towards the recognition accuracy (i.e. the weight for that face granule is close to zero). This optimization process leads to both dimensionality reduction and better computational efficiency.

Evolutionary algorithms such as genetic algorithms often fail to maintain diversity among individual solutions (chromosomes) and cause the population to converge prematurely. This problem is attributed to loss of diversity in a population that decreases the quality of solution. *Adaptive mutation rate* [98] and *random offspring generation* [99] are used to prevent premature convergence to local optima by ensuring sufficient diversity in

a population. Depending on population diversity, mutation is performed with an adaptive rate that increases if diversity decreases and vice-versa. Population diversity is measured as the standard deviation of fitness values in a population. Further, random offspring generation is used to produce random offsprings if there is a high degree of similarity among participating chromosomes (parents) during the crossover operation. Combination of such similar chromosomes is ineffective because it leads to offsprings that are exactly similar to parents. Therefore, under such conditions, crossover is not performed and offsprings are generated randomly.

3.2.4 Combining Face Granules with Multi-objective Evolutionary Learning for Recognition

The granular approach for matching faces altered due to plastic surgery is summarized below.

1. For a given gallery-probe pair, 40 face granules are extracted from each image.

2. EUCLBP or SIFT features are computed for each face granule according to the evolutionary model learned using the training data.

3. The descriptors extracted from gallery and probe images are matched using weighted χ^2 distance measure.

$$\chi^2(a,b) = \sum_{i,j} \omega_j \frac{(a_{i,j} - b_{i,j})^2}{a_{i,j} + b_{i,j}} \tag{3.1}$$

   where a and b are the descriptors computed from face granules pertaining to a gallery-probe pair, i and j correspond to the i^{th} bin of the j^{th} face granule, and ω_j is the weight of the j^{th} face granule. Here, the weights of each face granule are learned using the genetic algorithm.

4. In identification mode $(1 : N)$, this procedure is repeated for all the gallery-probe pairs and top matches are obtained based on the match scores.

3.3 Experimental Results

Several experiments are performed to evaluate the performance of the proposed algorithm. The performance of the algorithm is also compared with SIFT and EUCLBP applied on full face image, SIFT and EUCLBP applied on the 40 face granules, sum-rule fusion [125] of SIFT and EUCLBP on face granules, and a commercial-off-the-shelf face recognition

system (COTS)[1]. Further, to evaluate the effectiveness of the proposed multi-objective genetic approach for feature selection and assimilation, the performance is compared with other feature selection methods, namely, definitive feature selection (referred to as "EU-CLBP+SIFT+PCA"), SFS [122], and SFFS [122].

3.3.1 Database

Experiments are performed on two databases: (a) plastic surgery face database [8] and (b) combined heterogeneous face database. The plastic surgery face database comprises 1800 pre- and post-surgery images corresponding to 900 subjects with frontal pose, proper illumination, and neutral expression. The database consists of different types of facial plastic surgery cases such as rhinoplasty (nose surgery), blepharoplasty (eyelid surgery), brow lift, skin peeling, and rhytidectomy (face lift). It is difficult to isolate individuals who have undergone plastic surgery and use special mechanism to recognize them. Therefore, face recognition algorithms should be robust to variations introduced by plastic surgery even in general operating environments. Considering such generality of face recognition, the second database is prepared by appending the plastic surgery face database with 1800 non-surgery images pertaining to 900 subjects from other publicly available face databases. This database is termed as the *combined heterogeneous face database* and comprises 3600 images pertaining to 1800 subjects. The non-surgery images are from the same databases used by Singh *et al.* [8] and consists of two frontal images per subject with proper illumination and neutral expression.

Images in the plastic surgery face database are collected from different sources on internet and have noise and irregularities. The detected images in the database are first preprocessed to zero mean and unit variance followed by applying histogram equalization to maximize image contrast. Further, Wiener filtering is applied to restore the blurred edges. As mentioned previously, the face images are geometrically normalized and the size of each detected face image is 196×224 pixels.

3.3.2 Experimental Protocol

To evaluate the efficacy of the proposed algorithm, experiments are performed with 10 times repeated random sub-sampling (cross validations). In each experiment, 40% of the database is used for training and the remaining 60% is used for testing. The training data is used for learning EUCLBP/SIFT feature selection and weights of each face granule,

[1]COTS used in our experiments is one of the highly accurate and widely used face recognition software; however the academic license agreement does not allow us to name it in any comparison.

while the unseen testing data is used for performance evaluation. Experimental protocol for all the experiments are described here:

- *Experiment 1:* 1800 pre- and post-surgery images pertaining to 900 subjects from the plastic surgery face database [8] are used in this experiment. Images of 360 subjects are used for training and the performance is evaluated on the remaining 540 subjects. Pre-surgery images are used as the gallery and post-surgery images are used as the probe.

- *Experiment 2:* Out of 1800 subjects from the combined heterogeneous face database, 720 subjects are used for training and the remaining 1080 subjects are used for testing. The training subjects are randomly selected and there is no regulation on the number of training subjects with plastic surgery. This experiment resembles real world scenario of training-testing where the system is unaware of any plastic surgery cases.

- *Experiment 3:* To evaluate the effectiveness of the proposed algorithm for matching individuals against large size gallery, two different experiments are performed. In both the experiments, 6324 frontal face images obtained from government agencies are appended to the gallery of 1800 face images used in *Experiment* 2.

  - Case 1: Training is performed with images of 360 subjects from the plastic surgery face database. The performance is evaluated on post-surgery images from the remaining 540 subjects as probes against the large scale gallery of 7764 subjects.
  - Case 2: Training is performed with images of 720 subjects from the combined heterogeneous face database. The performance is evaluated on images from the remaining 1080 subjects as probes against the large scale gallery of 7404 subjects.

3.3.3 Analysis

The proposed algorithm utilizes the observation that human mind recognizes face images by analyzing the relation among non-disjoint spatial features extracted at multiple granular levels. Further, simultaneously optimizing the feature selection and weight computation pertaining to each face granule allows for addressing the non-linear and spontaneous variations introduced by plastic surgery. Key results and observations from the experiments are summarized below.

- The CMC curves in Figures 3.9 and 3.10 and Table 3.2 show rank-1 identification accuracy for Experiments 1 and 2. The proposed algorithm outperforms other algorithms by at least 4.22% on the plastic surgery face database and 4.86% on the combined heterogeneous face database. The proposed algorithm also outperforms the commercial system by 2.66% and 1.93% on the plastic surgery face database and the combined heterogeneous face database respectively.

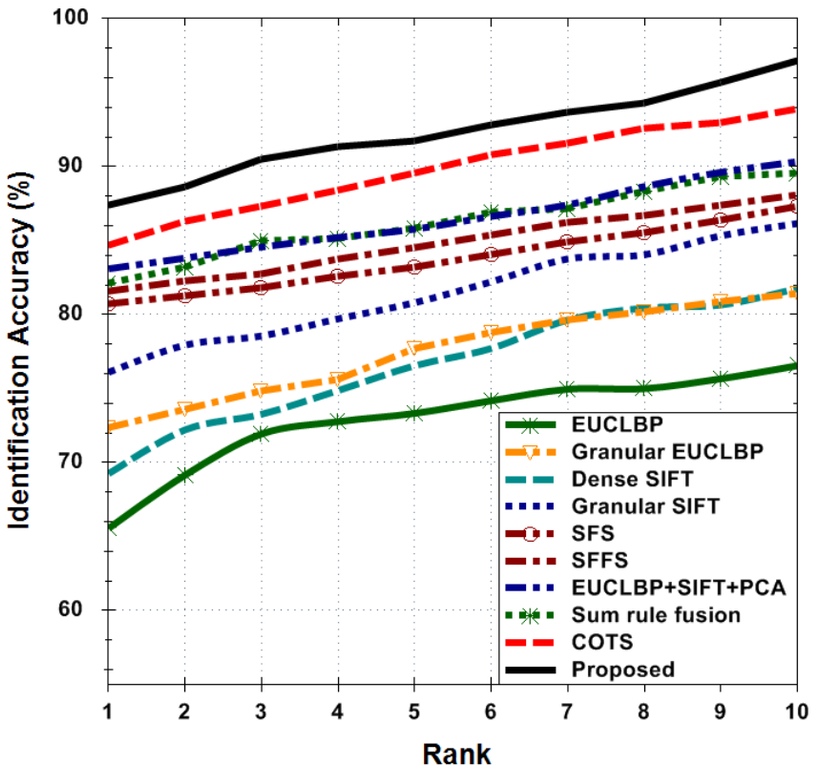

Figure 3.9: CMC curves for the proposed and existing algorithms on the plastic surgery face database.

- In Experiment 2, the training-testing partitions have plastic surgery as well as non-surgery images. It closely resembles the condition which a real world face recognition system encounters in general operating environment. Without the knowledge of specific plastic surgery cases, face recognition system has to be robust in matching surgically altered face images in addition to matching regular face images. Different types of plastic surgery procedures have varying effect on one or more facial regions. The proposed algorithm inherently provides the benefit of addressing the non-linear variations introduced by different types of plastic surgery procedures.

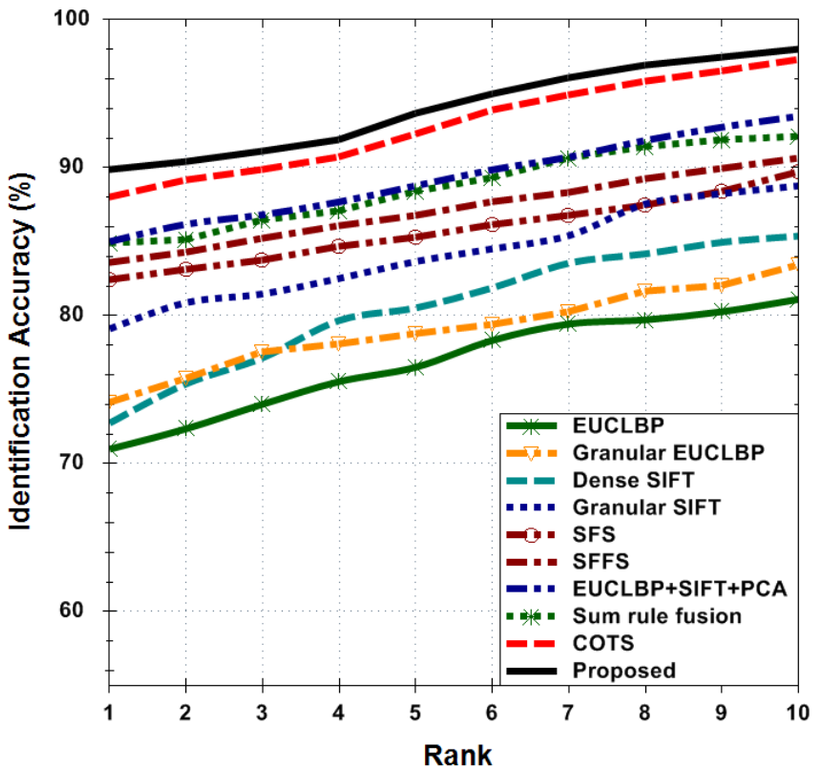

Figure 3.10: CMC curves for the proposed and existing algorithms on the combined heterogeneous face database.

Table 3.2: Rank-1 identification accuracy of the proposed multi-objective evolutionary granular approach and comparison with existing approaches. Identification accuracies and standard deviations are computed with 10 times cross validation.

Database (#Train/#Test)	**Algorithm**	**Rank-1 Accuracy (%)**	**Standard Deviation**
Plastic surgery face database (360/540)	**Single Algorithm**		
	EUCLBP [86]	65.56	1.02
	SIFT [93]	69.26	1.13
	Granular EUCLBP	72.35	0.85
	Granular SIFT	76.11	0.86
	COTS	84.66	0.76
	Match Score Fusion		
	Sum Rule Fusion	82.05	0.90
	Feature Selection Approach		
	SFS [122]	80.66	0.94
	SFFS [122]	81.58	0.96
	EUCLBP+SIFT+PCA	83.10	0.71
	Proposed	**87.32**	**0.64**
Combined heterogeneous face database (720/1080)	**Single Algorithm**		
	EUCLBP [86]	70.98	0.92
	SIFT [93]	72.75	0.98
	Granular EUCLBP	74.08	0.78
	Granular SIFT	79.12	0.82
	COTS	87.94	0.80
	Match Score Fusion		
	Sum Rule Fusion	84.85	1.16
	Feature Selection Approach		
	SFS [122]	82.43	0.76
	SFFS [122]	83.59	0.88
	EUCLBP+SIFT+PCA	85.01	0.68
	Proposed	**89.87**	**0.70**

- CMC curves in Figures 3.11 and 3.12 show the performance of the proposed algorithm and commercial system for matching probes against a large gallery (Experiment 3). The proposed algorithm outperforms the commercial system by 4.6% and 2.21% on Case 1 and Case 2 of Experiment 3 respectively. Assimilating discriminating information from different levels of granulation and combining them in an evolutionary manner helps to mitigate the effect of plastic surgery procedures and leads to improved performance.

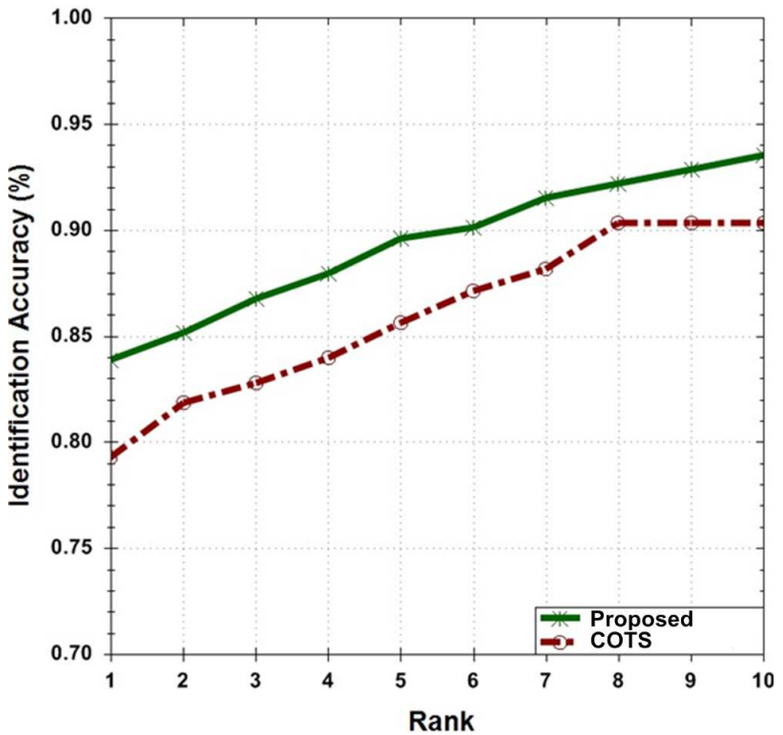

Figure 3.11: CMC curves for the proposed and commercial algorithms for large scale evaluation on probe images from (a) Case 1 of Experiment 3 and (b) Case 2 of Experiment 3.

- Table 3.3 shows individual rank-1 identification accuracy of all 40 face granules using EUCLBP and SIFT on the plastic surgery face database. Face granules $4, 7, 19, 21, 29,$ and 31 yield significantly better recognition performance with EUCLBP as compared to SIFT. On the other hand, face granules $2, 3, 8, 11, 14, 26, 39,$ and 40 provide better recognition performance with SIFT as compared to EUCLBP. SIFT generally performs better on granules containing fiducial features such as eyes, nose, and mouth, however its performance on predominant skin regions such as forehead, cheeks, and outer facial region is not optimal. Since EUCLBP is based on

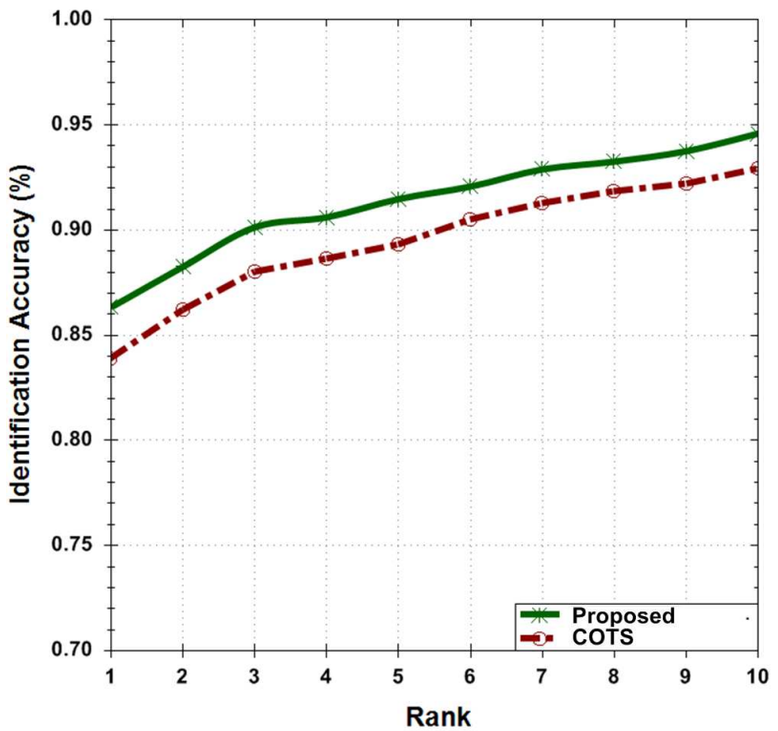

Figure 3.12: CMC curves for the proposed and commercial algorithms for large scale evaluation on probe images from (a) Case 1 of Experiment 3 and (b) Case 2 of Experiment 3.

Table 3.3: Rank-1 identification accuracy of face granules using SIFT and EUCLBP.

Granule	SIFT (%)	EUCLBP (%)	Granule	SIFT (%)	EUCLBP (%)
F_{Gr1}	69.26	65.56	F_{Gr21}	14.12	22.08
F_{Gr2}	51.42	42.26	F_{Gr22}	19.25	23.96
F_{Gr3}	46.18	21.32	F_{Gr23}	23.64	19.25
F_{Gr4}	22.86	36.20	F_{Gr24}	20.88	23.94
F_{Gr5}	20.15	25.75	F_{Gr25}	9.72	5.50
F_{Gr6}	16.26	19.50	F_{Gr26}	19.36	8.85
F_{Gr7}	10.46	19.38	F_{Gr27}	18.12	12.50
F_{Gr8}	39.06	28.64	F_{Gr28}	9.22	7.25
F_{Gr9}	17.85	23.42	F_{Gr29}	17.36	22.50
F_{Gr10}	13.14	19.64	F_{Gr30}	8.54	6.48
F_{Gr11}	41.43	32.38	F_{Gr31}	18.52	22.86
F_{Gr12}	28.20	24.44	F_{Gr32}	14.24	6.48
F_{Gr13}	16.88	22.02	F_{Gr33}	13.16	11.24
F_{Gr14}	33.06	23.84	F_{Gr34}	11.35	5.65
F_{Gr15}	30.56	24.68	F_{Gr35}	10.75	7.94
F_{Gr16}	15.76	21.84	F_{Gr36}	15.10	13.54
F_{Gr17}	33.12	25.50	F_{Gr37}	12.64	6.28
F_{Gr18}	15.64	21.28	F_{Gr38}	12.20	10.38
F_{Gr19}	11.82	20.10	F_{Gr39}	22.86	12.82
F_{Gr20}	51.60	44.40	F_{Gr40}	24.92	11.18

exact difference of gray level intensities, it can better encode discriminating micro patterns even from predominant skin regions.

- Multi-objective evolutionary approach for selecting feature extractor using genetic algorithm provides the advantage of choosing better performing feature extractor for each face granule. It is observed in our experiments that on average, SIFT is selected for 22 face granules whereas EUCLBP is selected for 18 face granules. This process in performed only during training and requires around 3 hours and 16 minutes to converge when we optimize it for 340 subjects in the plastic surgery face database. During testing, the weights learned during training are used to match the histogram features corresponding to different face granules using weighted ξ^2 distance.

- To show the improvement due to face granulation, Table 3.2 compares the rank-1 identification accuracy of granular EUCLBP and granular SIFT with EUCLBP and SIFT applied on the full face. The results show that applying EUCLBP and SIFT on face granules improves the rank-1 accuracy by at least 3% as compared to a full face

image. The ability to encode local features at different resolutions and sizes (face granules) allows the proposed algorithm to be resilient to the non-linear variations introduced by plastic surgery procedures.

- To show the efficacy of the multi-objective evolutionary approach, the performance is compared with sum-rule fusion [125] of SIFT and EUCLBP on face granules. Table 3.2 shows that the proposed algorithm outperforms sum-rule fusion by at least 5% on both the databases.

- While comparing with existing feature selection approaches, SFS and SFFS algorithms are used to select either EUCLBP or SIFT features for each face granule based on the identification accuracy (optimization function). The dimension of selected features is empirically decided based on the best performance achieved during training. In SFS, the best performance is achieved with all 40 face granules. However for feature selection using SFFS, the best performance is obtained with 32 face granules. Unlike SFS and SFFS algorithms, the proposed multi-objective evolutionary granular algorithm allows for simultaneous optimization of feature selection and weights for each face granule. As shown in Table 3.2, the proposed algorithm outperforms SFS by at least 6.66% and SFFS by at least 5.74% in rank-1 identification accuracy on both the databases.

- In definitive feature selection (EUCLBP+SIFT+PCA), PCA is used for dimensionality reduction in which top eigen vectors are retained to preserve 95% of the total energy of the distribution. Unlike the multi-objective genetic optimization, PCA based dimensionality reduction does not allow assigning distinct weights to different face granules and therefore the proposed algorithm outperforms definitive feature selection by at least 4.22%.

- Recently, Aggarwal *et al.* [103] proposed a sparse representation based approach to match surgically altered face images in a part-wise manner. The proposed granular algorithm outperforms the sparse representation based approach [103] by 9.4% on the plastic surgery face database under the same experimental protocol.

- From non-parametric rank-ordered test (Mann-Whitney test on the ranks obtained from the algorithms), it can be concluded that there is a statistically significant difference between the proposed algorithm and COTS. Further, at 95% confidence level, parametric t-test (using the match scores) also suggests that the proposed algorithm and COTS are statistically different.

3.3.4 Identification Performance with Different Plastic Surgery Procedures

According to Singh *et al.* [8], plastic surgery procedures can be categorized into *global* and *local* plastic surgery. **Global plastic surgery** completely transforms the face and is recommended in cases where functional damage is to be cured such as patients with fatal burns or trauma. In these kind of surgeries, facial appearance, skin texture, and feature shapes vary drastically thus making it arduous for any face recognition system to recognize pre- and post-surgery faces. Rhytidectomy (full facelift) is used to treat patients with severe burns on face and neck. It can also be used to reverse the effect of aging and get a younger look, thus modifying the appearance and texture of the whole face. Analogous to rhytidectomy, skin peeling procedures such as laser resurfacing and chemical peel alter the texture information thus affecting the performance of face recognition algorithms. These procedures are used to treat wrinkles, stretch marks, acne, and other skin damages caused due to aging and sunburns. These two global plastic surgery procedures severely impact the performance of the proposed algorithm that yields rank-1 identification accuracy of 71.76% and 85.09% for cases with rhytidectomy and skin peeling respectively, as shown in Figure 3.13 and Table 3.4.

On the other hand, **local plastic surgery** is meant for reshaping and restructuring facial features to improve the aesthetics. These surgical procedures result in varying amounts of change in the geometric distance between facial features but the overall texture and appearance of the face remains similar to the original face. Dermabrasion is used to give a smooth finish to face skin by correcting the skin damaged by sunburns or scars (developed as a post-surgery effect), dark irregular patches (melasma) that grow over the face skin, and mole removal. Among all the local plastic surgery procedures listed in [8], dermabrasion has the most prominent effect on the performance of the proposed algorithm as it drastically changes the face texture. As shown in Figure 3.13 and Table 3.4, the proposed approach yields rank-1 identification accuracy of 77.89% for dermabrasion cases. Other local plastic surgery procedures also affect the performance of the proposed algorithm to varying degrees. Since plastic surgery procedures increase the difference between pre- and post-surgery images of the same individual (intra-class variations), they drastically reduce the performance of existing face recognition algorithms. The performance of the proposed algorithm with various global and local plastic surgery procedures is also shown in Table 3.4 and CMC curves in Figure 3.13. These results show that the proposed

algorithm provides improvement of at least 21.7% compared to existing algorithms[1].

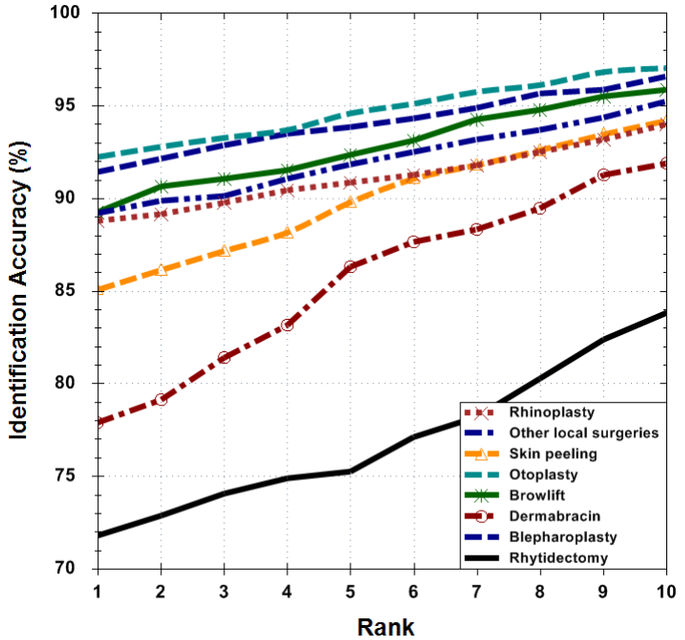

Figure 3.13: CMC curves on different types of local and global plastic surgery procedures for the proposed algorithm.

3.3.5 Analysis of Different Granularity Levels

To understand the contribution of different granularity levels for recognizing face images altered due to plastic surgery, a detailed experimental study of individual granular levels is performed. The correlation analysis of all three levels of granularity is reported in Table 3.5. The complementary information vested in different levels is utilized by the proposed algorithm for efficiently matching surgically altered face images. Table 3.6 shows the identification accuracy of individual levels of granularity for the two databases. The first level of granularity has different Gaussian and Laplacian pyramids that assimilate discriminating information across multiple resolutions. Pyramids at level-0 contain minute features whereas the pyramids at level-1 and level-2 provide high level prominent features of a face. Several psychological studies have shown that humans use different inner and outer facial features to identify individuals [126]. The inner facial features include nose, eyes, eyebrows, and mouth while the outer facial region comprises face outline, structure of jaw/chin, and forehead. The second level of granularity therefore extracts information

[1]Since the experimental protocol and sample distribution across different validation trials in [8] and the current research are same, the results of PCA, FDA, LFA, CLBP, SURF, and GNN are directly compared.

Table 3.4: Rank-1 identification accuracy on different types of local and global plastic surgery procedures.

Type	Surgery	#	PCA	FDA	LFA	CLBP	SURF	GNN	Periocular	Proposed
Local	Browlift	60	28.5	31.8	39.6	49.1	51.1	57.2	34.42	89.22
	Dermabrasion	32	20.2	23.4	25.5	42.1	42.6	43.8	44.56	77.89
	Otoplasty	74	56.4	58.1	60.7	68.8	66.4	70.5	47.25	92.25
	Blepharoplasty	105	28.3	35.0	40.2	52.1	53.9	61.4	30.96	91.42
	Rhinoplasty	192	23.1	24.1	35.4	44.8	51.5	54.3	40.71	88.85
	Other	56	26.4	33.1	41.4	52.4	62.6	58.9	35.81	89.17
Global	Rhytidectomy	308	18.6	20.0	21.6	40.9	40.3	42.1	37.27	71.76
	Skin peeling	73	25.2	31.5	40.3	53.7	51.1	53.9	45.83	85.09
	Overall	**900**	**27.2**	**31.4**	**37.8**	**47.8**	**50.9**	**53.7**	**40.11**	**87.32**

Table 3.5: Pearson correlation coefficient between different granular levels on the plastic surgery face database.

Database	Granules	Genuine correlation	Impostor correlation
Plastic surgery face database	Level 1 - Level 2	0.67	0.59
	Level 1 - Level 3	0.43	0.21
	Level 2 - Level 3	0.63	0.55
Combined heterogeneous face database	Level 1 - Level 2	0.81	0.78
	Level 1 - Level 3	0.38	0.20
	Level 2 - Level 3	0.42	0.26

Table 3.6: Performance of different levels of granules and their combinations on the plastic surgery and the combined heterogeneous face database.

Database	Granular level	Accuracy (%)
Plastic surgery face database	Granular level 1	78.3
	Granular level 2	82.7
	Granular level 3	58.4
	Granular level 1+2	80.1
	Granular level 2+3	82.8
	Granular level 1+3	85.0
	Proposed	**87.3**
Combines Heterogeneous face database	Granular level 1	80.7
	Granular level 2	84.1
	Granular level 3	61.5
	Granular level 1+2	83.2
	Granular level 2+3	84.4
	Granular level 1+3	86.9
	Proposed	**89.8**

from different inner and outer facial regions representing discriminative information that is useful for face recognition. Local facial fragments such as nose, eyes, and mouth provide robustness to variations in local regions caused due to plastic surgery procedures. Human mind can efficiently distinguish and classify individuals with their local facial fragments. Therefore, the third level of granularity assimilates discriminating information from these regions. The proposed granular approach unifies diverse information from levels that are useful for recognizing faces altered due to plastic surgery.

To analyze the complementary information provided by different granularity levels, the performance is evaluated for different combinations. The performance of the proposed multi-objective evolutionary granular approach is optimized for a particular level of granulation or their combination by assigning null weights to the face granules corresponding to other levels of granulation during genetic optimization. Table 3.6 also shows the results for different combinations of granular levels on the two databases.

According to the statistics provided by American Society for Aesthetic Plastic Surgery [101], blepharoplasty (eyelid surgery) is identified as one of the top five surgical procedures performed in 2010. Eyelid is the thin skin that covers and protects our eyes and is a major feature in periocular recognition algorithms. Blepharoplasty is used to reshape upper and lower eyelids to treat excessive growth of skin tissues obstructing vision. Some other global plastic surgery procedures such as rhytidectomy or skin peeling may also affect the periocular region. Periocular region can be used as a biometric when the face

is occluded [127] and/or the iris cannot be captured [128]. Recently, Juefei-Xu *et al.* [129] proposed using periocular region for age invariant face recognition and reported substantial improvements in both verification and identification performance. Driven by the robustness of periocular biometrics against occlusion and aging, this chapter also evaluates the performance of periocular biometrics for recognizing surgically altered face images. In the proposed granulation scheme, F_{Gr29} and F_{Gr31} represent the right and left periocular regions as shown in Figure 3.14. Experiments are performed using the protocol of Experiment 1 in Section 3.3.2. CMC curves in Figure 3.15 show the performance of periocular region for matching surgically altered faces from the plastic surgery face database. The performance is computed individually for the left and right periocular regions using SIFT and EUCLBP. Sum-rule fusion [125] of SIFT on the left and right periocular regions (fusion of SIFT) and sum-rule fusion of EUCLBP on the left and right periocular regions (fusion of EUCLBP) is also reported. Finally, the overall performance of periocular region is computed based on the sum-rule fusion of SIFT and EUCLBP on left and right periocular regions (fusion of SIFT and EUCLBP). The performance is also compared with an existing periocular based recognition algorithm, referred to as Bharadwaj *et al.* [128].

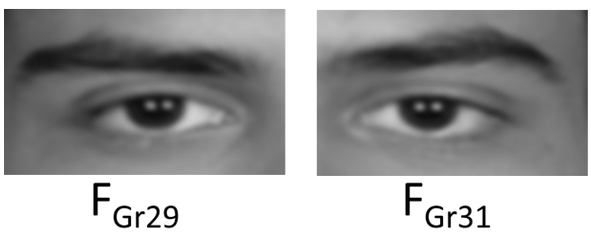

Figure 3.14: F_{Gr29} represents the right periocular region and F_{Gr31} represents the left periocular region.

Experiments are also performed to analyze the effect of different global and local plastic surgery procedures (especially blepharoplasty) on periocular region. Table 3.4 reports rank-1 identification accuracy of periocular region for matching faces altered due to specific types of plastic surgery. Blepharoplasty alters the periocular region thereby affecting the performance of periocular biometrics. It is also observed that the performance of periocular biometrics is reduced when a local region neighboring the periocular region (such as nose and forehead) is transformed due to plastic surgery. This is mainly because modifying local features also transmits some vicissitudes in the adjacent facial regions. The

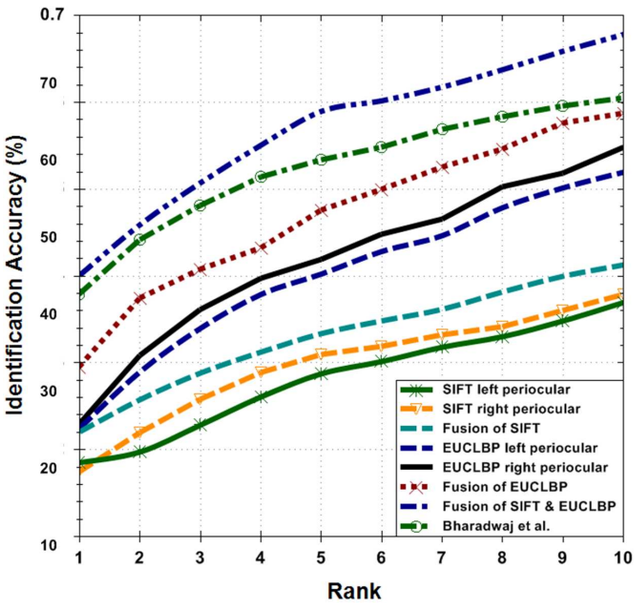

Figure 3.15: CMC curves comparing the performance of different algorithms for matching periocular region on the plastic surgery face database.

results suggest that although, periocular biometrics has shown robustness to aging and occlusion, plastic surgery is an important challenge for periocular recognition algorithms.

3.4 Summary

Plastic surgery has emerged as a new covariate of face recognition and its allure has made it indispensable for face recognition algorithms to be robust in matching surgically altered face images. This chapter presents a multi-objective evolutionary granular algorithm that operates on several granules extracted from a face image. The first level of granularity processes the image with Gaussian and Laplacian operators to assimilate information from multi-resolution image pyramids. The second level of granularity tessellates the image into horizontal and vertical face granules of varying size and information content. The third level of granularity extracts discriminating information from local facial regions. Further, a multi-objective evolutionary genetic algorithm is proposed for feature selection and weight optimization for each face granule. The evolutionary selection of feature extractor allows switching between two feature extractors (SIFT and EUCLBP) and helps in encoding discriminatory information for each face granule. The proposed algorithm utilizes the observation that human mind recognizes faces by analyzing the relation among non-disjoint spatial features extracted at different granularity levels. Experiments under

different protocols, including large scale matching, show that the proposed algorithm outperforms existing algorithms including a commercial system for matching surgically altered face images. Further, experiments on several local and global plastic surgery procedures also show that the proposed algorithm consistently outperforms other existing algorithms. Detailed analysis on the contribution of three granular levels and individual face granules corroborates the hypothesis that the proposed algorithm unifies diverse information from all granules to address the non-linear variations in pre-and post-surgery images.

Chapter 4

Matching Cross-resolution Face Images using Co-transfer Learning

4.1 Introduction

With advancements in technology, surveillance cameras now have a profound presence and are widely used in security and law enforcement applications. There are several instances where surveillance videos have helped agencies in apprehending individuals who have committed crime or identify individuals with the intent to commit crime. For example, in 2005 subway bomb blasts in London [130], CCTV footage helped law enforcement officers in identifying the bombers. In 2008 Mumbai terrorist attacks [131], surveillance cameras installed at different locations (CST railway station, Taj Palace, and Trident hotels) helped the agencies to track the activities of terrorists and later identify them. In the 2010 car bomb case at Times Square [132], the surveillance footage captured an unidentified individual leaving the car with explosives. Later, widespread distribution and manual investigation of the video helped the investigating agencies to apprehend the individual.

In all these cases, surveillance cameras could not foil the terrorist attacks, however, they served as the primary evidence in leading the investigation and also recognizing the individuals at the end. It is therefore desirable to build a system where surveillance cameras coupled with a face recognition algorithm can be used to automatically identify individuals from a watch-list. Along with the challenges of pose, expression, illumination, aging, and disguise in face recognition, matching a watch-list photograph to an image obtained from surveillance camera also requires the capability of matching across resolution. For example, the watch-list photograph could be a high resolution image whereas the surveillance camera images are generally low resolution images. As shown in Figure 4.1, even if both the images are frontal, the information content in both the images could

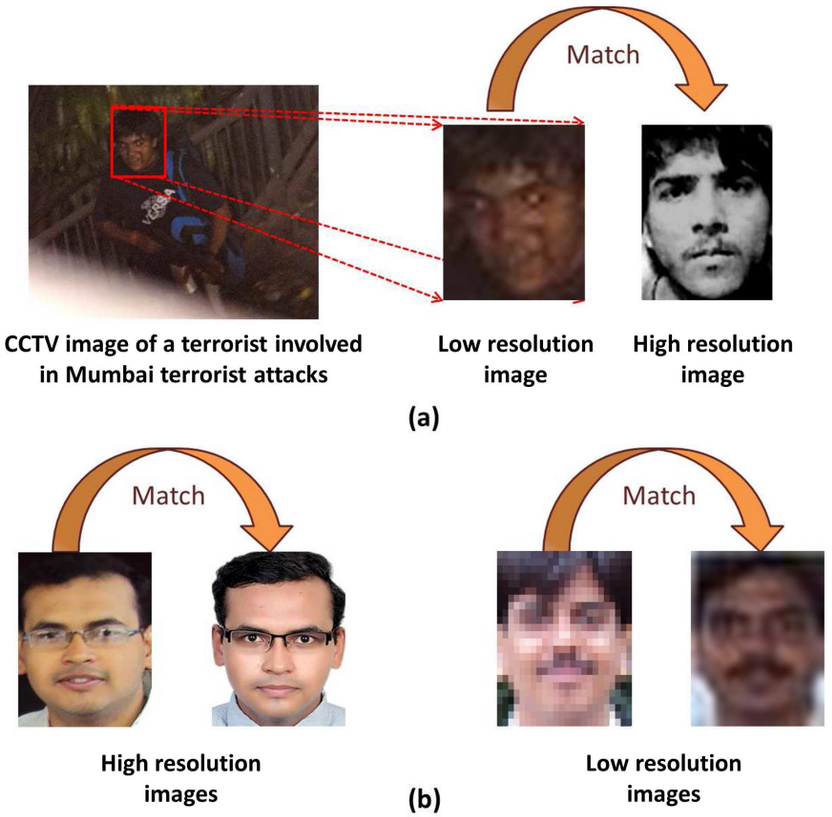

Figure 4.1: Illustrating the difference in matching (a) low resolution and high resolution images, (b) two high resolution images, and (c) two low resolution images.

be significantly different. The presence of pose, illumination, and expression along with different resolution could further exacerbate the problem, as shown in Figure 4.2.

4.1.1 Related Research

In literature, several approaches have been proposed to match cross-resolution face images. As shown in Table 4.1, these algorithms can be classified into two categories: *super-resolution* and *transformation* based approaches. Figure 4.3 shows the steps involved in super-resolution based approaches for cross-resolution matching which enhance the low quality probe image before recognition. On the other hand, Figure 4.4 shows transformation based approaches which extract features that are resilient to resolution changes and match cross-resolution face images. Some of the transformation based approaches also perform resolution invariant transformations either in the image space or the feature

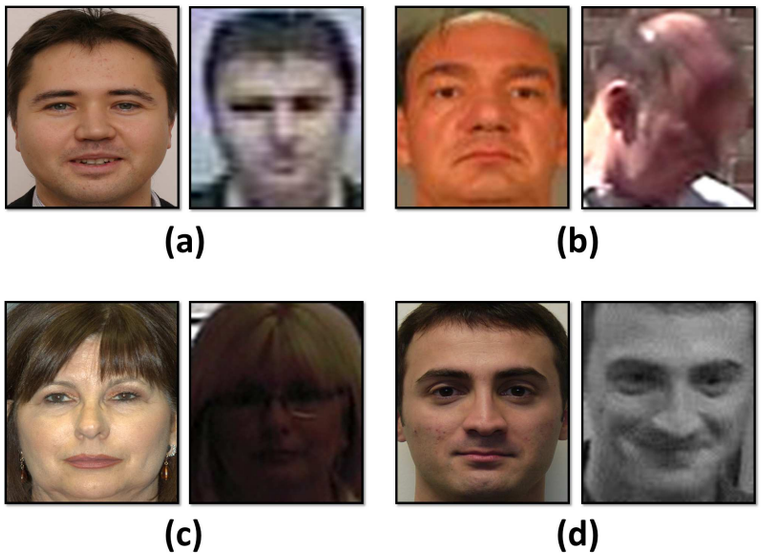

Figure 4.2: Illustrates the challenge in matching low resolution images when coupled with other covariates. Low resolution challenge (a) alone, (b) with pose, (c) with illumination, and (d) with expression.

space for matching.

1. *Super-resolution based approaches:* Huang and He [133] proposed to build a coherent subspace between the PCA features of high resolution (HR) and low resolution (LR) images mapped using the radial basis functions for recognition. Baker and Kanade [138] proposed an algorithm to *apriori* learn the spatial distribution of image gradients to enhance the resolution of local features before matching. Chakrabarty *et al.* [139] proposed a learning based method to super-resolve face images with kernel principal component analysis-based prior model. Chang *et al.* [140], [141] formed geometrically similar manifolds using local facial patches in the low and high resolution images. They used training images to estimate the high-resolution embedding and construct a smooth super-resolved image. Yang *et al.* [142] proposed a super resolution approach by representing local patches as a sparse linear combination of elements from high resolution images. In addition to these local models, Liu *et al.* [143] integrated a holistic parametric and a local nonparametric model using two-step statistical modeling for face hallucination. It was observed that super-resolution approaches, due to environmental variations and distortions, failed to significantly

Table 4.1: Existing algorithms for cross-resolution face image matching.

Approach	Technique	Database	Gallery/probe resolution
Super-resolution	Coherent features [133]	FERET	72×72/ 12×12
		UMIST	
		ORL	
	Multi-modal tensor face [134]	AR	56×36/ 14×9
		Yale	
		FERET	
	S2R2 [66]	Multi-PIE	24×24/6×6
		FERET	
		FRGC v.2	
	Relationship learning [135]	FRGC v.2	64×48/ 28×24
Transformation	LFD [136]	FERET	88×80/33×30
	Coupled locality preserving mapping (CLPM) [137]	FERET	72×72/12×12
	Synthesis based LR face recognition[67]	CMU-PIE	48×40/19×16
		FRGC v.2	
	MDS [63]	Multi-PIE	48×40/12×10

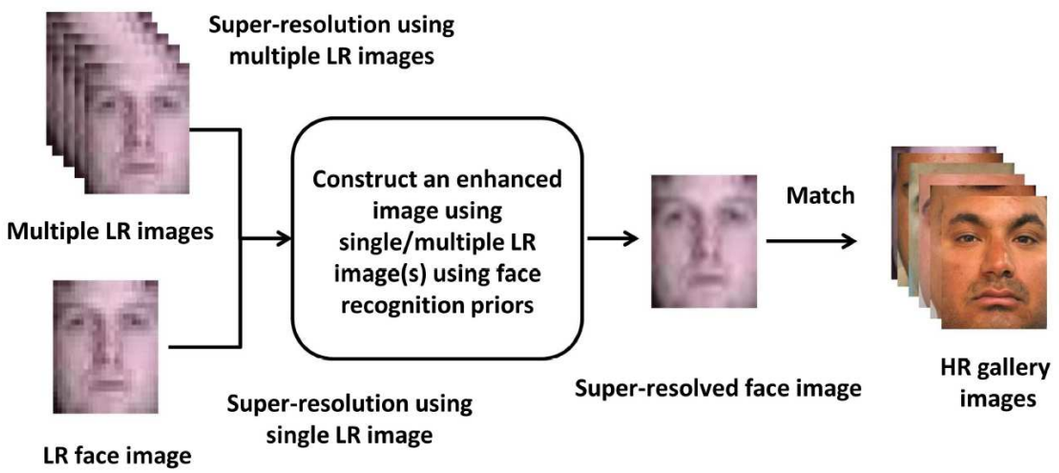

Figure 4.3: Broad view of super resolution based approaches for cross-resolution face matching.

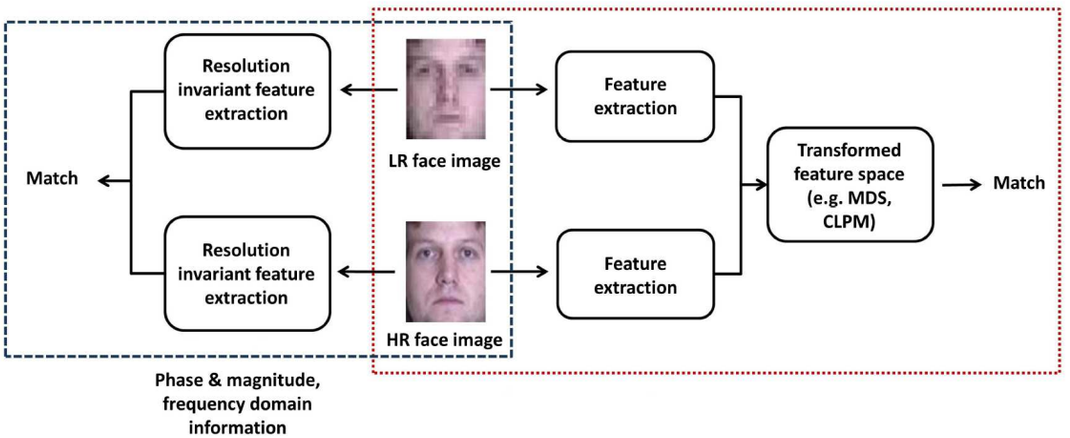

Figure 4.4: Broad view of transformation based approaches for cross-resolution face matching.

improve the recognition performance. It is our assertion that the primary objective of super-resolution is to obtain a good visual reconstruction from low resolution face(s), and these algorithms are generally not intended for recognition. However, there are some approaches that simultaneously optimize both super resolution and face recognition. Jia and Gong [134] combined super-resolution and face recognition by computing a maximum likelihood identity parameter vector in high-resolution tensor space for recognition. Further, Hennings-Yeomans *et al.* [66] proposed an approach where facial features were included in a super-resolution method as the prior information for simultaneous reconstruction of super-resolved images. Recently, Zou and Yuen [135] proposed a super-resolution technique based on the relationship between the high-resolution image space and the very low resolution image space. Their technique allowed for better visual appearance as well as improved face recognition performance for the very low resolution problem.

2. *Transformation based approaches:* Unlike super-resolution, another method to match cross-resolution images is to downsample high resolution images to the level of low resolution images before matching. However, information useful for face recognition such as texture, edges, and other high frequency information is compromised while downsampling the images. To address this problem, Li *et al.* [137] proposed to project both high resolution and low resolution images to a feature space using coupled mappings. Biswas *et al.* [144] proposed a multidimensional scaling approach to

simultaneously transform the features from high resolution gallery and low resolution probe images. The Euclidean distance between the transformed feature vectors approximates the distance computed when the probe images were captured at similar resolution as that of the gallery images. Researchers have also studied that the phase and magnitude in frequency domain can be used as a resolution invariant representation for efficiently matching cross-resolution face images. Lei *et al.* [136] proposed a local frequency descriptor based on the magnitude and phase information to match cross-resolution face images in the frequency domain. Shekhar *et al.* [67] proposed a generative approach using the information from high resolution gallery to match low resolution probe images with illumination variations. Lei *et al.* [65] proposed a coupled discriminant analysis for heterogeneous face recognition (matching high vs. low resolution images). To maintain the discriminative power and generalizability of their approach, they utilized multiple samples from different resolutions along with locality information in the kernel space.

4.1.2 Research Contribution

The conditions in which a face recognition algorithm is trained are referred to as the *source domain* where the availability of large training data helps the algorithm to efficiently learn the task. In the source domain, face recognition algorithms are trained to match high resolution images. However, for surveillance applications, the probe data i.e., the *target domain*, comprises low resolution face images and the gallery contains high resolution face images. Under these variations, the performance of a biometric system degrades because it is unable to efficiently utilize the knowledge learned in the source domain and there is a scarcity of labeled low resolution data that can be used for training the algorithms. Obtaining sufficient labels for the target data is time consuming, requires human effort, and very expensive. However, there is an abundance of *unlabeled* low resolution data in target domain during testing. This observation motivates us to formulate the problem of cross-resolution face matching where sufficient labeled data is available in source domain and only a few labeled instances are available from the target domain. In this chapter, we propose a co-transfer learning (CTL) framework which is a cross-pollination of transfer learning [145] and co-training [146]. The framework integrates transfer learning and co-training in a non-separable manner to efficiently transfer the knowledge from the source domain to the target domain with sequentially available unlabeled instances from the target domain.

- *transfer learning* is used to leverage the knowledge learned in the source domain for efficiently matching LR probes with HR gallery in the target domain.

- *co-training* is used to enable transfer learning with unlabeled probe instances from the target domain by assigning pseudo-labels to probes.

In face recognition literature, to the best of our knowledge, this is the first work that leverages unlabeled probe instances to facilitate knowledge transfer. The performance of the proposed framework is evaluated in a cross-resolution face recognition application and the experiments are performed on four face databases, namely, the CMU Multi-PIE [9], SCface [10], ChokePoint [11], and MBGC v2 video challenge [7] databases. The results show that the proposed algorithm outperforms existing algorithms including FaceVACS which is a commercial face recognition system.

4.2 Co-transfer Learning Framework

We, humans, have innate abilities of transferring knowledge between related tasks. It is observed that if the new task is closely related to the previous learning, humans can quickly transfer this knowledge to perform the new task. However, given some prior knowledge in a related task, traditional algorithms are unable to adapt to a new task and have to learn the new task from the beginning. Generally, they do not consider that the two tasks may be related and the knowledge gained in one may be used to learn the new task efficiently in lesser time. *Transfer learning* attempts to mimic this human behavior by transferring the knowledge learned in one or more source tasks and use it for learning the related target task. Several approaches have been proposed for transfer learning and they can be categorized as 1) inductive, 2) transductive, and 3) unsupervised transfer learning. Based on the domain representation, transfer learning approaches can be further categorized into homogeneous and heterogeneous transfer learning. The source and target domains share same feature space in the former whereas feature space is different in the later one. For a more detailed discussion on different transfer learning approaches, readers are directed to [145].

Transfer learning has been explored in many computer vision applications. Zhu *et al.* [147] proposed a heterogeneous transfer learning framework that utilized annotated images from the web as a bridge to transfer knowledge between text and images using a matrix factorization approach. Quattoni *et al.* [148] proposed a method for learning a sparse prototype image representation for transfer across visual categories. Their approach used

a large set of unlabeled data and a kernel function to form a representation. Ahmed *et al.* [149] proposed a hierarchical feed-forward model for visual recognition using transfer learning from pseudo tasks which include a set of pattern matching operations constructed from the data. Geng *et al.* [150] proposed a domain adaptation metric learning by introducing a data dependent regularization to conventional metric learning in the reproducing kernel Hilbert space. This minimized the empirical maximum mean discrepancy between different domains. Wang *et al.* [151] proposed dyadic knowledge transfer which is a non-negative matrix tri-factorization based approach to transfer cross-domain image knowledge for the new computer vision tasks. In face recognition or related domains, transfer learning has been applied to verify kinship using face images through subspace transfer learning [152]. Chen *et al.* [153] also proposed to learn a person-specific facial expression model by transferring the informative knowledge from other people. Their approach allows to learn an accurate person-specific model for a new subject with only a small amount of person specific data. Most of the transfer learning techniques work in offline manner and assumes that the data from the target domain is available upfront.

Generally, labeled data in target domain is scarce and obtaining labels for the target data is time consuming and expensive in most real world scenarios; therefore, it is difficult to learn a model for the target data. On the other hand, large amount of unlabeled data, available in the form of probe, can be leveraged to learn the model. There are some existing semi-supervised approaches for face recognition [154, 155, 156, 157] that utilize few labeled and ample amount of unlabeled data for enhancing face recognition performance. Many of these semi-supervised approaches are used for template update such as semi-supervised PCA [158, 159] or LDA [160]. There are few approaches [154, 161] that update/retrain the model with few labeled and large unlabeled data. Mostly, existing semi-supervised algorithms require entire unlabeled data upfront and do not perform well for single sample per subject.

The proposed co-transfer learning algorithm builds on the limitations of existing approaches to address the challenge of single sample per subject and performs transfer learning in online manner with sequential unlabeled data available from the target domain. Transfer learning and co-training are jointly used to *transfer* the knowledge learned in the source domain to the target domain with unlabeled instances, as shown in Figure 4.5. Co-training to update the classifiers has been explored by Bhatt *et al.* [161] where biometric classifiers are updated using labeled as well as unlabeled instances. However, to the best of our knowledge, it is the first algorithm that uses transfer learning for face recognition

as a semi-supervised approach using few labeled and a large number of unlabeled probe instances.

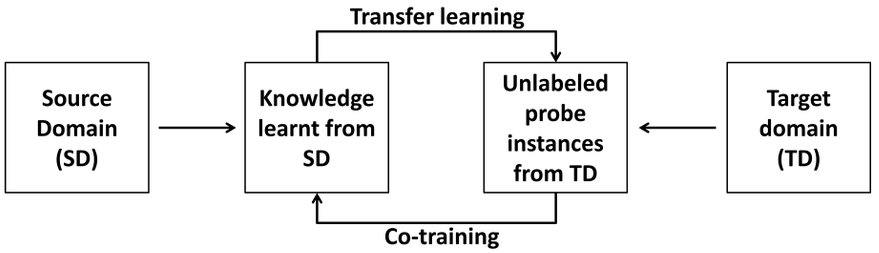

Figure 4.5: Illustrating the cross-pollination of transfer learning and co-training for transferring knowledge from source domain to target domain.

The proposed framework is a generalized framework that can be applied to any classifier which allows re-training with incremental data. In this chapter, we have applied the concept of co-transfer learning to support vector machine (SVM). Re-training the SVM classifier in batch mode is computationally expensive [162] and may not be feasible in real-wold applications. Some approaches have been proposed that allow re-training the SVM classifier using only previous support vectors and the new incremental data points. A method to add or remove one sample at a time to update SVM is proposed in [163] where a solution for $N \pm 1$ samples can be obtained using the N old samples and the sample to be added or removed. In the proposed approach, SVM is first trained using an initial training set and a decision hyperplane is obtained. The SVM classifiers are then updated using the new available instances and the previous support vectors. For more details on updating SVM classifiers with new incremental data, readers are directed to [161], [162], [163].

Transfer Learning: In face recognition, the classifiers such as SVM, are learned using training data (from the source domain) while the performance is evaluated on a separate unseen test data (the target domain) which may have different properties and follow a different distribution compared to the training data. Consider a scenario where there are two classifiers, one trained using the source and an other trained using the target domain data. During training, there is a large labeled data in the source domain i.e., for matching HR probe with HR gallery images (source domain) but only a few labeled instances are available in the target domain, i.e., for matching LR probes with HR gallery images. In such a case, the source domain classifier alone may not efficiently classify the test instances

because of the variations in data distribution of source and target domains. Since, the classifier in target domain is trained using only a few labeled samples, it is not able to efficiently classify the test instances. It has to learn/update its decision boundary with the incremental data available in the target domain. Both the classifiers are individually insufficient to classify the test data from the target domain. Therefore, in the proposed algorithm, an ensemble is built as a weighted combination of the source and target domain classifiers. It efficiently classifies test instances and subsequently transfers the knowledge from the source domain to the target domain as and when the data from the target domain is available. For this, the two classifiers trained on the source and target domains are combined to efficiently classify the unlabeled probe instances.

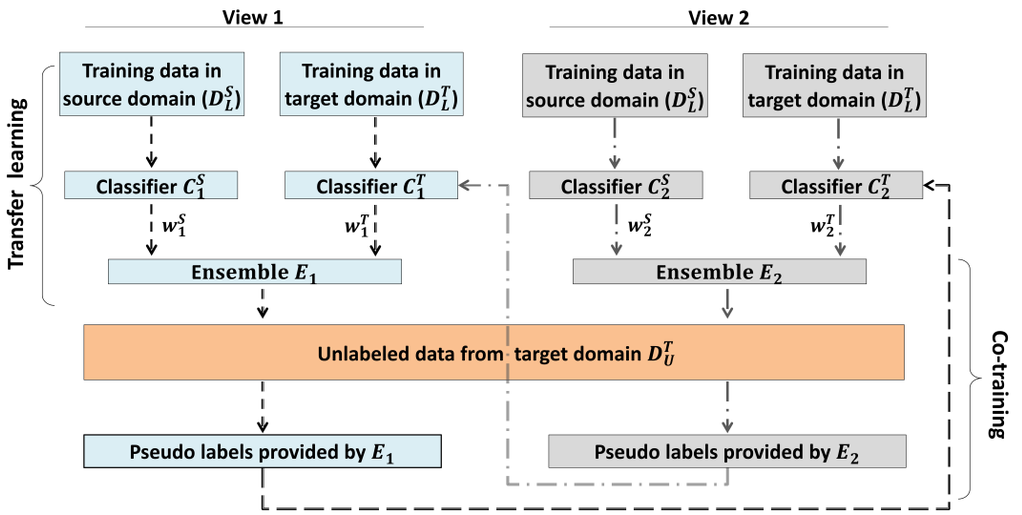

Figure 4.6: Block diagram illustrating the steps involved in the proposed co-transfer learning framework.

As shown in Figure 4.6, the source domain classifiers (C_j^S) are trained using sufficient HR labeled training data in the source domain denoted by $D_L^S = \{(\mathbf{u}_1^S, z_1), (\mathbf{u}_2^S, z_2), ..., (\mathbf{u}_n^S, z_n)\}$. Every i^{th} instance, $\mathbf{u}_i$ has two views $\{x_{i,1}, x_{i,2}\}$ for the training label $z_i \in \{-1, +1\}$; here $x_{i,1}$ and $x_{i,2}$ represent the input vectors obtained from two separate views (features). The two views are utilized for co-training (explained later). The target domain classifiers (C_j^T) are initially trained on a few labeled training instances from the target domain represented as D_L^T=$\{(\mathbf{u}_1^T, z_1), (\mathbf{u}_2^T, z_2), ..., (\mathbf{u}_m^T, z_m)\}$. Here, n and m are the number of training instances in the source and target domains respectively, such that n>m and $j = 1, 2$ represents the view (feature). Let a set of r unlabeled probe instances in the target domain

be represented as D_U^T=$\{(\mathbf{u'}_1^T), (\mathbf{u'}_2^T), ..., (\mathbf{u'}_r^T)\}$. An ensemble prediction function, denoted as E_j, is constructed for each view. E_j is a weighted combination of the source domain classifier, C_j^S, and the target domain classifier, C_j^T, with w_j^S and w_j^T as the weights of the source domain classifier and the target domain classifier for the j^{th} view respectively. For the i^{th} unlabeled probe instance in the j^{th} view, the ensemble function E_j predicts the label, $E_j(x_{i,j}) \rightarrow y_{i,j}$. For the i^{th} instance in the target domain $\mathbf{u'}_i$, class label is predicted by the ensemble as given in Eq. 4.1.

$$y_{i,j} = sign(w_{i,j}^S \ \Pi(C_j^S(\mathbf{u'}_i)) + w_{i,j}^T \ \Pi(C_j^T(\mathbf{u'}_i)) - \frac{1}{2}) \tag{4.1}$$

where Π is a normalization function such that $\Pi(x) = max(0, min(1, \frac{x+1}{2}))$, $w_{i,j}^S$ and $w_{i,j}^T$ are the weights for the source and target domain classifiers at the i^{th} instance respectively. Initially, both the weights are set to 0.5 so that each classifier contributes equally within an ensemble and gradually, they are automatically adjusted to emphasize the contribution from the updated target domain classifiers in an ensemble. As proposed by Zhao and Hoi [164], the two weights are updated dynamically as shown in Eqs. 4.2 and 4.3.

$$w_{i+1,j}^S = \frac{w_{i,j}^S h_i(C^S)}{w_{i,j}^S h_i(C^S) + w_{i,j}^T h_i(C_j^T)} \tag{4.2}$$

$$w_{i+1,j}^T = \frac{w_{i,j}^T h_i(C^T)}{w_{i,j}^S h_i(C^S) + w_{i,j}^T h_i(C^T)} \tag{4.3}$$

where $w_{i+1,j}^S$ and $w_{i+1,j}^T$ are the updated weights and h_i is defined as:

$$h_i(C) = \exp\{-\eta \ l(\Pi(C_i), \Pi(\hat{y}_i))\}, \tag{4.4}$$

$\eta = 0.5$, $l(y, \hat{y}) = (y - \hat{y})^2$ is the square loss function, y is the predicted label and $\hat{y}$ is the pseudo label provided by co-training (explained later).

Co-training: As mentioned previously, unlabeled probe instances are available in abundance and can be utilized to update/learn the classifiers in the target domain. However, it is required to obtain the labeled target data. Obtaining labeled training instances from the target domain is difficult, expensive, and requires human effort. In biometrics, there are situations when only a small set of labeled data is available for training while a huge amount of unlabeled data is readily available as probe. This situation is similar to a semi-supervised learning scenario, where co-training [146], [161] has proven beneficial as it can be used to transform unlabeled probe instances into pseudo-labeled training instances.

In the proposed co-training approach, a small initial labeled set is available from the target domain for training the classifiers and a large number of unlabeled instances are available as probe. It assumes the availability of two ensemble functions (classifiers), E_1 and E_2, trained on separate views (features) where each ensemble function has sufficient (better than random) accuracy. If the first ensemble confidently predicts genuine label for an instance while the second ensemble predicts impostor label with low confidence, then this particular instance (with pseudo label provided by the first ensemble) is utilized for updating the second ensemble and vice-versa. In this chapter, the confidence of prediction for an instance on the j^{th} view, denoted by α_j, is measured as the distance of that instance from the decision boundary. For confidently predicting an instance to belong to genuine class, the distance from the decision hyperplane should be greater than the genuine threshold (P_j). Similarly, an instance is confidently predicted as impostor if the distance from the hyperplane is greater than the impostor threshold (P_j). Note, here P_j refers to genuine threshold when comparing for genuine class and to impostor threshold when comparing for impostor class. Since SVM is used for classification, a genuine threshold is computed as the distance of the farthest support vector of *genuine* class. Similarly, an impostor threshold is computed as the distance of the farthest support vector of *impostor* class. Varying the thresholds will change the number of instances on which the co-training is performed. High threshold value implies conservative co-training while smaller value of the threshold leads to aggressive co-training. In this manner, unlabeled probe instances are transformed into pseudo-labeled training instances which are then used to update the ensembles. In an ensemble, knowledge is transferred by updating the decision boundary of the target domain classifier C_j^T using only the new incremental data as proposed in [161].

Co-transfer: In the proposed framework, transfer learning and co-training work concurrently to improve the target domain task with pseudo labels provided by co-training that lead to transfer of knowledge from the source to the target domain. Within each ensemble, the target domain classifier updates its decision boundary [161] with every pseudo-labeled instance obtained during testing. Moreover, the weights corresponding to the source and target domain classifiers are also adjusted dynamically using Eqs. 4.2 and 4.3. This scheme avoids the need to learn the target domain classifiers from the beginning and hence, makes the system scalable and computationally efficient. Note that in the co-transfer learning framework, only target domain classifiers are updated with pseudo-labeled instances. The source domain classifiers do not need any update because they are

Algorithm 2 Co-transfer learning

Input: Initial labeled training data D_L^S in the source domain, a few labeled instances D_L^T from the target domain. Unlabeled probe instances D_U^T from target domain (available sequentially).
Iterate: j= 1 **to** 2 (number of views)
Process: Train classifiers C_j^S and C_j^T on j^{th} view of D_L^S and D_L^T respectively to construct ensemble E_j. Compute confidence thresholds P_j for each view.
for $i = 1$ **to** r (number of probe instances) **do**
  Predict labels: $E_j(x_{i,j}) \rightarrow y_{i,j}$; calculate α_j: confidence of prediction
  if $\alpha_1 > P_1$ & $\alpha_2 < P_2$ **then**
    Update C_2^T with pseudo-labeled instance $\{x_{i,2}, y_{i,1})\}$ & recompute w_2^S and w_2^T.
  end if.
  if $\alpha_1 < P_1$ & $\alpha_2 > P_2$ **then**
    Update C_1^T with pseudo-labeled instance $\{x_{i,1}, y_{i,2})\}$ & recompute w_1^S and w_1^T.
  end if.
end for.
end iterate.
Output: Updated classifiers C_1^T, C_2^T and weights w_1^S, w_1^T, w_2^S and w_2^T.

well trained using large amount of labeled data available upfront in the source domain. The proposed *co-transfer learning* framework is summarized in Algorithm 2.

Error bounds: To analyze the effectiveness of the proposed co-transfer learning algorithm, we compute the error bounds. For an ensemble E, let M_E denote the number of errors by the ensemble, then as shown in [164], M_E is bounded by:

$$M_E \leq 4min\left(\sum C^S, \sum C^T\right) + 8ln(2). \tag{4.5}$$

where $\sum C^S = \sum_{i=1}^{I} l^*(\Pi(C_i^S), \Pi(\hat{y_i}))$, $\sum C^T = \sum_{i=1}^{I} l^*(\Pi(C_i^T), \Pi(\hat{y_i}))$, I is the number of instances, and $\hat{y_i}$ is the pseudo label for the i^{th} instance. Proof of the error bound of an ensemble is provided in [164] and also provided as Appendix B. For two ensembles when the final decision classification decision is based on their combination, the error bounds M for the co-transfer learning algorithm are given as:

$$min(M_{E1}, M_{E2}) \leq M \leq max(M_{E1}, M_{E2}) \tag{4.6}$$

The primary objective of selecting two ensembles is to facilitate co-transfer learning as one ensemble provides pseudo labeled training instances to the other. Therefore, the error bounds of the proposed algorithm will lie between the error bounds of the two participating ensembles as shown in Eq. 4.6. It follows the concept of lifelong learning

Table 4.2: Experimental protocol on different databases for cross-resolution face matching. Training subjects in the source domain specifies the total number of subjects used for training different algorithms. * For ChokePoint database, training of source and target domain classifiers is performed using the CMU Multi-PIE [9] database.

Database	Training		Testing/ Co-transfer learning (# subjects)	Resolution range (pixels)	Covariates (apart from low resolution)
	Source domain (# subjects)	Target domain (# subjects)			
CMU Multi-PIE [9]	100	40	237	216×216 - 16× 16	Illumination
SCface [10]	50	20	80	72×72 - 24×24	Variation in camera & distance, pose illumination
ChokePoint* [11]	50	20	29	216×216 - 16× 16	Pose, illumination, and expression
MBGC v.2 [7]	60	30	87	216×216 - 16× 16	Pose, illumination, and activity (walking/talking)

where the classifiers continue to learn as and when additional training data is available. However, as more and more pseudo labeled instances are available, the weights for the source and target domain classifiers saturate. Co-transfer learning can be stopped when the saturation occurs and the emphasis is shifted towards the target domain classifiers.

4.3 Co-transfer Learning for Cross-resolution Face Recognition

In an operational scenario, training is performed in a controlled environment; whereas during testing, a biometric system encounters data from uncontrolled environment. Co-training is particularly useful for recognizing cross-resolution face images. The source and target domain classifiers are trained on two views (features) and two ensemble functions (E_1 and E_2) are built. One view is the local phase quantization (LPQ) [165] and the second view is the scale invariant feature transform [93]. Both these views are resilient to scale changes and can be effectively used for matching face images with different resolutions. The two features provide diverse information, one encodes the discriminative phase information whereas the other encodes information from the image gradients. The two feature extractors used are briefly described below:

- **Local Phase Quantization [165]** operates on the Fourier phase computed locally for a window in every image position. It uses the local phase information extracted using a short-term Fourier transform. The phases of the multiple low-frequency coefficients are uniformly quantized into one of the 256 bins. These LPQ codes for

all image pixel neighborhoods are concatenated to form a histogram and is used for recognition. Since only phase information is used, the method is also invariant to uniform illumination changes. In our experiments, same parameters as proposed by Ahonen *et al.* [165] are used. Finally, χ^2 distance is used to compare two LPQ descriptors.

- **Scale Invariant Feature Transform [93]** is a scale and rotation invariant descriptor that generates a compact representation of an image based on the magnitude, orientation, and spatial vicinity of image gradients. SIFT, as proposed by Lowe [93], is a sparse descriptor that is computed around detected interest points. However, it can also be used in a dense manner where the descriptor is computed around pre-defined interest points. SIFT descriptor is computed in a dense manner. SIFT descriptors computed at the sampled regions are concatenated to form the image signature and χ^2 distance is used to compare two SIFT descriptors.

Initial training on labeled data from the source and target domains: The co-transfer learning framework assumes that during training, each subject has high resolution gallery-probe pairs and a few subjects have corresponding low resolution images from the target domain. As shown in Figure 4.7, face images are tessellated into non-overlapping facial patches[1]. LPQ and SIFT descriptors are computed for each local patch and matched using the χ^2 distance measure. Distance scores corresponding to each local patch are vectorized to an input vector $\{\mathbf{u}_i, z_i\}$, where $z_i \in \{-1, +1\}$ is the associated label. $\{+1\}$ signifies that the gallery-probe pair belongs to the same individual (i.e. genuine pair) whereas $\{-1\}$ signifies that the gallery-probe pair belongs to images corresponding to different individuals (i.e. impostor pair).

Input vectors obtained by matching LPQ descriptors of two high resolution images are utilized for training the source domain SVM classifier (C_1^S) on view 1. On the contrary, the target domain SVM classifiers for view 1 are trained using one high resolution and one low resolution images. The source domain and the target domain SVM classifiers are then combined to form an ensemble, E_1. Similarly, the SVM classifiers for view 2 (SIFT) are trained and the ensemble function E_2 is learned.

Co-transfer learning with unlabeled probes from the target domain: As shown in Figure 4.8, for matching a LR probe with a HR gallery image, the images are tessellated

[1]It is empirically determined that the best performance is obtained when a face is tesselated into 3×3 non-overlapping patches.

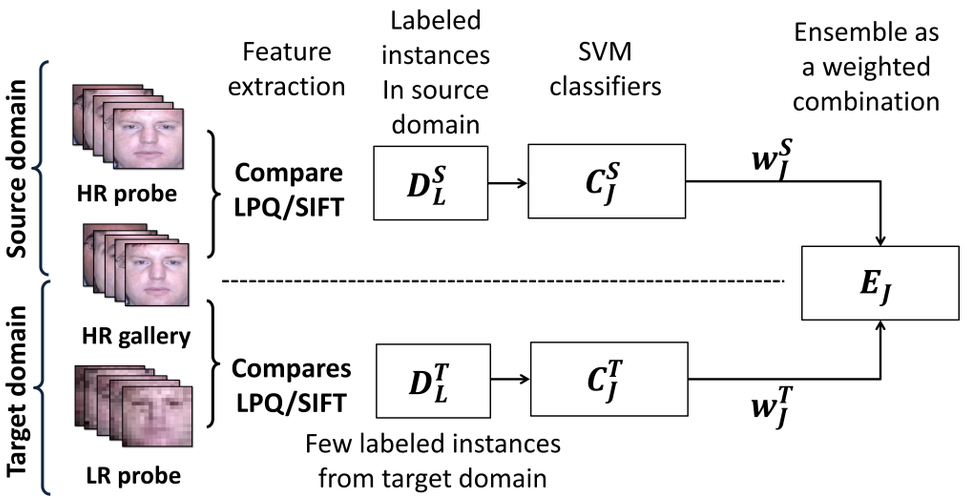

Figure 4.7: Block diagram illustrating the training process of the source and target domain classifiers to build the ensembles.

into non-overlapping local patches and LPQ and SIFT descriptors are computed for each local patch. LPQ descriptors from the corresponding local patches on the gallery and probe images are matched using χ^2 distance and the distance scores from these local patches are vectorized to form an input vector **u**' for view 1. Similarly, an input vector corresponding to SIFT (view 2) is computed using the χ^2 distance measure. Unlike training, the instances obtained during testing are unlabeled. For every query given to the biometric system, both the ensembles, E_1 and E_2, are used to classify the instance. If one ensemble confidently predicts genuine label for an instance while the other ensemble predicts impostor label with low confidence, then this instance is added as a *labeled* re-training sample for the second ensemble and vice-versa. The target domain SVM classifiers (C^T) in the ensembles are updated with pseudo-labeled probe instances obtained during testing. Further, the weights for both source domain and target domain SVM classifiers are also updated with each pseudo-labeled probe instance, as shown in Eqs. 4.2 and 4.3. Thus each ensemble updates the target domain classifier of the other ensemble. The final decision is computed by combining responses from both the ensembles.

4.4 Database and Experimental Protocol

The performance of the proposed co-transfer learning framework is evaluated on four different databases, (1) CMU Multi-PIE [9], (2) SCface [10], (3) ChokePoint [11], and (4)

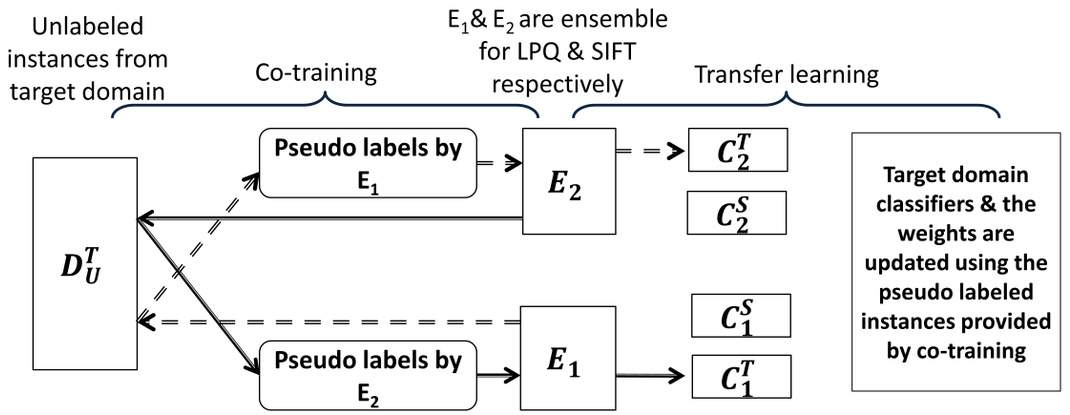

Figure 4.8: Block diagram illustrating the co-transfer learning in the target domain with unlabeled probe instances.

Multiple Biometric Grand Challenge (MBGC) v.2 video challenge database [7]. To evaluate the efficacy of the proposed framework, a joint transfer-and-test [164] strategy is used which allows the data used in model adaptation to be concurrently used for performance evaluation. The experiments are designed to resemble real world scenario where ample training data is available in the source domain to train the classifiers for classifying the high resolution gallery-probe pairs as genuine or impostor. However, only a few low resolution probe and corresponding high resolution gallery images are available for training the classifiers in the target domain. To emulate such conditions, Table 4.2 lists the number of high resolution gallery-probe pairs that are used for training the classifiers in the source domain and the number of low resolution probe and corresponding high resolution gallery images used for training classifiers in the target domain. The training subjects in the target domain are a subset of the training subjects in the source domain. Sample images from all the databases are shown in Figure 4.9. Details about the databases are further described below:

1. CMU Multi-PIE [9] database comprises images from 337 individuals captured in four different sessions with varying pose, expression, and illumination. For experiments, a subset pertaining to 337 individuals with frontal pose and neutral expression are selected; however, the gallery and probe images vary in illumination conditions. For each subject, one high resolution image is kept in the gallery and one low resolution image is used as probe.

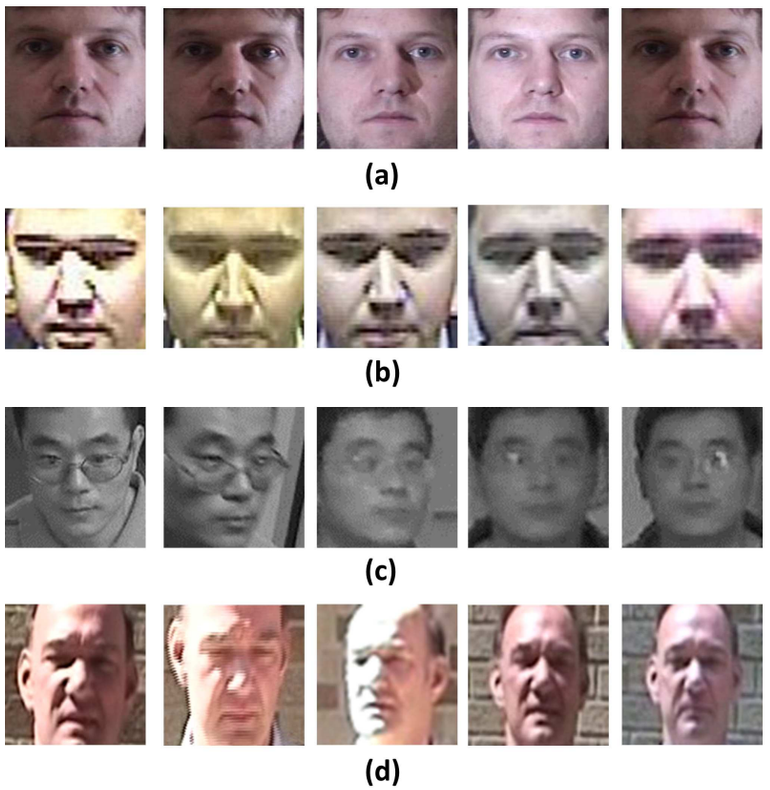

Figure 4.9: Sample images from the (a) CMU Multi-PIE, (b) SCface, (c) ChokePoint, and (d) MBGC v.2 video challenge databases.

2. The SCface database is a real-world surveillance database comprising images of 130 individuals captured in uncontrolled indoor environment using multiple surveillance cameras placed at different distances. For each subject, one high resolution image is kept in gallery and five images captured from different cameras are used as probe. SCface database contains low resolution images ranging from 48×48 - 24×24 pixels and experiments are performed without interpolating these images. Therefore in the experimental protocol for SCface database, gallery and probe images vary from 72×72 to 24×24 pixels.

3. The ChokePoint database is a video database captured under real-world surveillance conditions. Three cameras placed above the portals are used to capture individuals walking through the portal. Images are captured with surveillance cameras in unconstrained environment and include illumination, expression, and pose variations. The database consists of 29 unique subjects captured in two portals with a time gap

of about one month. Since there are only 29 subjects in the database, training of both source and target domain classifiers is performed using the CMU Multi-PIE database. For each subject in the ChokePoint database, one high resolution image is kept in the gallery and five images are used as the probe.

4. From the MBGC v.2 video challenge database, multiple videos in standard definition (720×480 pixels) and high definition (1440×1080 pixels) format corresponding to 147 subjects are used. The database includes videos where the user is walking or performing some activity. Faces present in these videos have variations due to pose, illumination, and expression. The faces extracted from video frames are partitioned into the gallery and probe data sets (here we ensure that gallery and probe images are from different sessions i.e. from different videos of the person). Gallery consists of single image per user and probe set comprises five images from different sessions.

To emulate the conditions that the gallery is generally captured under controlled conditions, the experiments are performed with settings such that the resolution of gallery images is always higher than the probe images. Generally in a practical surveillance scenario, a single image is available in the gallery watch-list. To match the complexity of such surveillance scenario, experiments are performed with single image per subject in the gallery. The performance is reported in identification mode with 10 times repeated random sub-sampling (cross-validations) for non-overlapping training-testing partitions. Experiments are performed at different resolutions of gallery and probe images ranging from 216×216 pixels to 16×16 pixels. Face images in the databases are available at different resolutions and are interpolated[1] to the nearest resolution in the experimental protocol.

4.5 Experimental Results and Analysis

For cross resolution face matching, the performance of algorithms degrade mainly due to the 1) difference in information content between the high resolution gallery and low resolution probes and 2) limited biometric information in face images at low resolution. The proposed algorithm attempts to address these issues by using the knowledge learned for matching high resolution images from the source domain to efficiently match low resolution images from the target domain. The objective of the experiments is to determine the effectiveness of the proposed algorithm in transferring knowledge from the source domain to target domain for cross resolution face matching.

[1]Images are interpolated to the required resolution using bi-cubic interpolation.

1. SIFT with SVM classifier and LPQ with SVM classifier, referred to as SIFT and LPQ in the results.

2. Sum-rule score level fusion [125] of two ensembles trained on the initial labeled data from the source and target domains (referred to as *'fusion'*).

3. Multidimensional scaling algorithm (MDS) proposed by Biswas *et al.* [63] for matching low resolution face images.

4. A widely used commercial-off-the-shelf face recognition algorithm, FaceVACS, referred to as COTS.

5. Three super-resolution techniques. Super-resolution-1 (SR-1) is the standard bicubic interpolation, super-resolution-2 (SR-2)[1] is a regression based technique proposed by Kim and Kwon [166], and super-resolution-3 (SR-3)[2] is a sparse representation based approach proposed by Yang *et al.* [142].

6. Match score fusion of the proposed algorithm with MDS [63] and COTS using sum-rule [125].

4.5.1 Analysis

The experimental results suggest that the proposed approach efficiently matches cross-resolution face images by leveraging knowledge learned in the source domain. It also validates our assertion that co-training enables updating the decision boundary of the target domain classifiers with unlabeled probe instances as and when they arrive.

- Cross-pollination of transfer learning and co-training seamlessly transfers the knowledge learned in the source domain for matching cross-resolution face images. Co-training and transfer learning go hand-in-hand as co-training provides pseudo labels for unlabeled test instances which in-turn are used to update the target domain classifiers within each ensemble and thus transfer the knowledge.

- Updating the weights of the source and target domain classifiers allows to dynamically adjust the contribution from the constituent source and target domain classifiers in an ensemble. Initially, equal weights are assigned to both the classifiers; however with knowledge transfer, weights of classifiers in the target domain become more

[1]Source code is available at authors webpage "http://www.mpi-inf.mpg.de/kkim/."
[2]Source code is obtained from "http://www.ifp.illinois.edu/jyang29/."

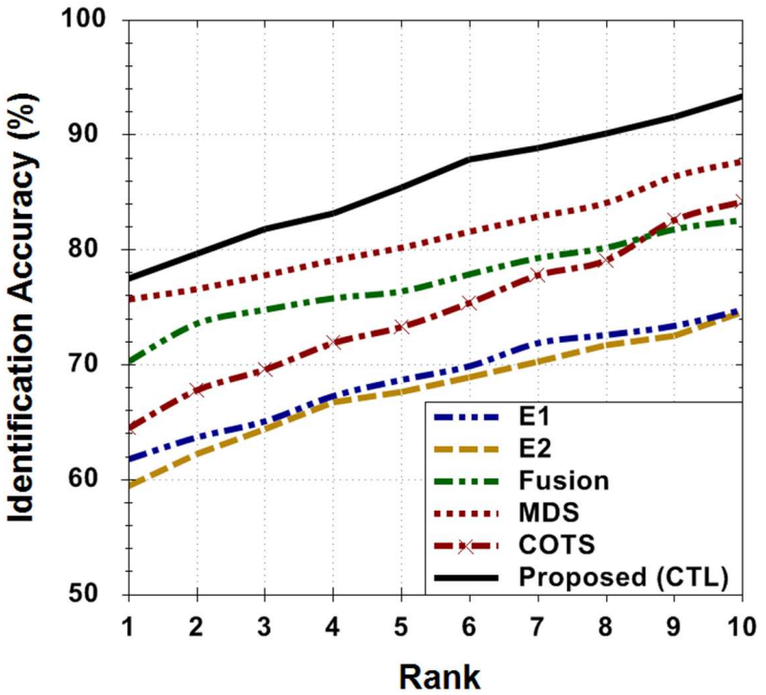

Figure 4.10: CMC curves showing the performance for matching 24×24 probe images with 72×72 gallery images on the CMU Multi-PIE database.

prominent. Table 4.3 shows the number of instances on which co-transfer learning is performed for different databases. It also shows how the co-transfer learning on unlabeled instances changes the weights of an ensemble so as to better classify the target domain samples. The experiments show that on all the four databases combined, co-training provides correct pseudo labels for about 98% of the total instances.

- The behavior of the proposed algorithm is further analyzed and Figure 4.14(a) illustrates sample cases where the proposed co-transfer learning algorithm correctly recognizes the low resolution probe images. Examples in Figure 4.14(b) illustrate cases where the proposed algorithm performs poorly. The poor performance can be attributed to the fact that some of the pseudo labels assigned to unlabeled probe instances may be incorrect leading to *negative transfer*. However, the effect of *negative-transfer* can be minimized by optimally selecting the confidence threshold for co-training. High threshold value implies conservative transfer while smaller value of the threshold leads to aggressive transfer.

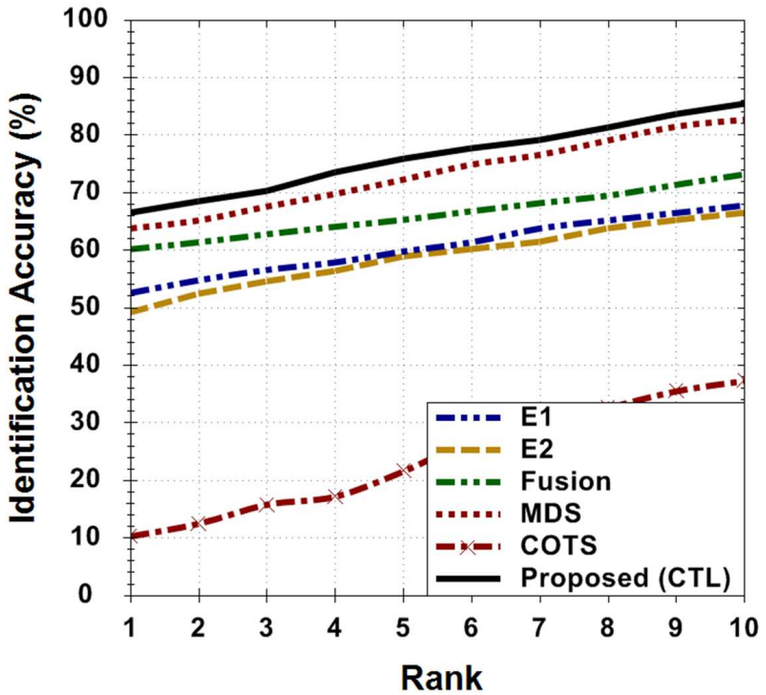

Figure 4.11: CMC curves showing the performance for matching 24×24 probe images with 72×72 gallery images on the SCface database.

As discussed before, existing techniques for matching cross-resolution face images can be divided into transformation and super-resolution approaches. The subsections below compare the performance of the proposed algorithm with both kinds of approaches.

4.5.1.1 Comparison with COTS and Transformation based Approaches

The performance of the proposed co-transfer learning (CTL) algorithm is compared with MDS [63], COTS, individual ensembles of SIFT [93], and LPQ [165], and their fusion. The results are also evaluated by fusing the proposed CTL algorithm with other techniques such as MDS [63] and COTS. Tables 4.4-4.7 show the results of the proposed and existing algorithms different combinations of gallery-probe resolution on the four databases.

- The results show that independently, SIFT and LPQ are not efficient for matching cross resolution face images. However, the ensembles, E1 (for LPQ) and E2 (for SIFT), developed by combining the classifiers trained on both the source and target domains, improves the performance. It is also observed that fusion [125] of two ensembles further improves the performance for cross-resolution face matching.

Table 4.3: Illustrates the number of instances on which co-transfer learning is performed and how the weights within an ensemble shift to emphasize the contribution of the target domain classifier.

Database	# pseudo labels		Weights after co-transfer			
	C_1^T	C_2^T	w_1^S	w_1^T	w_2^S	w_2^T
CMU Multi-PIE [9]	5184	4210	0.18	0.82	0.23	0.77
SCface [10]	7346	5268	0.21	0.79	0.27	0.73
ChokePoint [11]	456	540	0.33	0.67	0.36	0.64
MBGC v2 [7]	8136	6874	0.22	0.78	0.24	0.76

Table 4.4: Rank-1 identification accuracy of the proposed CTL algorithm and comparison with existing algorithms and commercial system on the CMU Multi-PIE database [9].

Resolution		Algorithm									
Gallery	**Probe**	**LPQ**	**SIFT**	**E1**	**E2**	**Fusion**	**MDS**	**CTL**	**COTS**	**CTL+ MDS**	**CTL+ COTS**
216×216	72×72	66.3	61.7	72.4	68.1	76.2	77.8	81.0	99.5	80.2	99.8
	48×48	63.6	58.2	70.6	67.3	74.5	75.2	79.7	98.1	79.4	99.3
	32×32	45.4	41.8	53.2	47.4	58.7	61.3	65.3	97.4	63.7	98.5
	24×24	22.2	21.4	29.5	26.8	32.9	33.4	37.7	54.5	35.6	58.2
	16×16	10.8	9.6	16.7	13.3	18.1	20.2	23.6	10.9	22.1	24.8
72×72	48×48	73.8	71.4	79.4	76.3	86.1	89.2	92.3	98.2	92.7	99.1
	32×32	62.8	49.8	69.1	55.2	79.4	81.5	84.1	96.3	84.3	97.4
	24×24	56.8	52.6	61.8	59.4	70.3	75.7	77.4	64.5	78.5	80.1
	16×16	50.2	47.4	56.7	52.1	66.2	68.9	72.4	11.5	72.8	76.1
48×48	32×32	44.2	42.5	50.3	47.8	55.2	58.7	61.8	96.8	60.5	97.1
	24×24	42.6	39.8	48.6	44.5	51.7	54.9	57.1	75.9	55.8	78.5
	16×16	20.6	18.2	26.2	22.3	29.9	31.3	32.9	6.4	39.4	43.2
32×32	24×24	37.6	30.1	41.2	30.4	44.8	40.9	45.7	78.4	45.4	80.6
	16×16	22.1	16.8	24.3	17.2	27.0	25.1	28.1	5.4	29.8	30.0
24×24	16×16	30.8	26.4	35.6	30.2	42.1	38.1	43.2	16.3	44.6	47.8

Table 4.5: Rank-1 identification accuracy of the proposed CTL algorithm and comparison with existing algorithms and commercial system on the SCface database [10].

Resolution		Algorithm									
Gallery	**Probe**	**LPQ**	**SIFT**	**E1**	**E2**	**Fusion**	**MDS**	**CTL**	**COTS**	**CTL+ MDS**	**CTL+ COTS**
72×72	48×48	58.4	55.8	63.2	60.4	74.4	76.1	79.4	35.7	80.4	83.4
	32×32	53.4	52.3	58.1	57.8	67.4	70.4	72.8	18.5	73.7	76.2
	24×24	48.1	43.5	52.6	49.1	60.2	64.8	66.4	10.3	67.6	70.1
48×48	32×32	36.2	32.6	40.2	36.5	45.8	47.9	50.0	23.8	50.6	54.3
	24×24	25.6	24.2	30.2	28.3	35.6	38.1	40.3	14.5	39.5	45.1
32×32	24×24	22.5	17.3	26.4	21.3	29.7	31.2	33.1	8.4	33.9	36.2

Table 4.6: Rank-1 identification accuracy of the proposed CTL algorithm and comparison with existing algorithms and commercial system on the ChokePoint database [11].

Resolution		Algorithm									
Gallery	**Probe**	**LPQ**	**SIFT**	**E1**	**E2**	**Fusion**	**MDS**	**CTL**	**COTS**	**CTL+ MDS**	**CTL+ COTS**
216×216	72×72	32.2	28.6	36.3	32.5	39.8	41.6	44.6	46.2	43.2	50.9
	48×48	23.1	22.1	29.6	28.1	31.5	33.8	38.4	33.7	36.8	42.3
	32×32	21.8	21.8	27.3	25.7	30.6	32.5	35.5	20.4	34.1	39.5
	24×24	18.4	16.2	23.2	20.7	28.4	29.1	32.4	10.3	31.7	35.1
	16×16	9.6	8.2	14.7	11.2	15.6	17.8	20.2	6.04	19.3	23.4
72×72	48×48	42.4	36.1	48.4	42.6	50.5	50.9	53.7	22.7	53.1	56.4
	32×32	32.6	31.8	37.6	35.7	39.5	41.6	43.8	12.7	42.6	47.2
	24×24	25.4	23.6	30.5	28.9	31.6	32.4	36.1	9.5	34.8	39.5
	16×16	21.4	19.6	26.2	23.8	28.1	28.7	31.6	7.6	30.4	35.2
48×48	32×32	35.4	32.6	41.2	37.6	44.7	45.4	48.2	18.5	47.8	50.9
	24×24	23.2	20.4	27.4	24.8	29.5	30.2	33.1	11.8	32.6	37.2
	16×16	17.6	14.5	21.8	19.6	24.1	26.3	28.3	4.7	27.5	31.6
32×32	24×24	20.4	14.8	23.4	18.7	24.3	28.6	31.6	16.4	30.8	35.4
	16×16	14.6	9.6	17.3	13.4	19.6	21.9	23.1	3.5	22.5	26.0
24×24	16×16	19.4	15.6	22.7	18.6	25.8	28.7	30.5	13.5	31.4	35.8

Table 4.7: Rank-1 identification accuracy of the proposed CTL algorithm and comparison with existing algorithms and commercial system on the MBGC v.2 video challenge database [7].

Resolution		**Algorithm**									
Gallery	**Probe**	**LPQ**	**SIFT**	**E1**	**E2**	**Fusion**	**MDS**	**CTL**	**COTS**	**CTL+ MDS**	**CTL+ COTS**
216×216	72×72	27.2	25.4	30.8	28.2	33.4	36.5	40.7	44.3	39.2	47.3
	48×48	22.6	24.8	26.2	23.7	29.3	30.8	33.5	31.4	32.7	36.8
	32×32	20.8	17.2	23.6	20.9	26.1	28.4	32.6	18.5	31.4	35.2
	24×24	17.6	15.4	21.5	18.6	23.7	25.3	28.1	9.8	26.9	29.5
	16×16	9.6	8.8	12.5	10.1	14.8	16.8	19.5	5.7	18.7	20.8
72×72	48×48	38.2	33.6	43.1	39.7	45.3	46.8	49.3	21.2	48.6	50.7
	32×32	29.2	26.4	33.4	29.1	35.9	38.3	41.9	11.4	40.5	45.2
	24×24	23.6	20.4	26.3	22.5	28.7	29.8	33.2	8.7	31.5	36.9
	16×16	19.6	16.8	22.7	19.4	24.8	26.5	29.5	6.2	28.1	33.5
48×48	32×32	33.6	31.2	38.4	33.1	40.5	44.3	47.4	17.2	46.8	48.7
	24×24	21.4	20.2	24.6	21.3	25.8	27.6	30.3	10.2	29.4	33.5
	16×16	16.5	14.8	18.2	15.7	20.3	24.1	26.5	4.2	26.1	27.9
32×32	24×24	17.8	12.6	20.7	15.6	22.3	27.1	28.6	14.8	29.2	33.2
	16×16	12.6	8.6	14.6	10.9	16.3	19.8	21.3	3.1	20.6	23.9
24×24	16×16	18.4	14.8	20.4	16.1	23.5	27.3	28.2	11.9	29.4	31.8

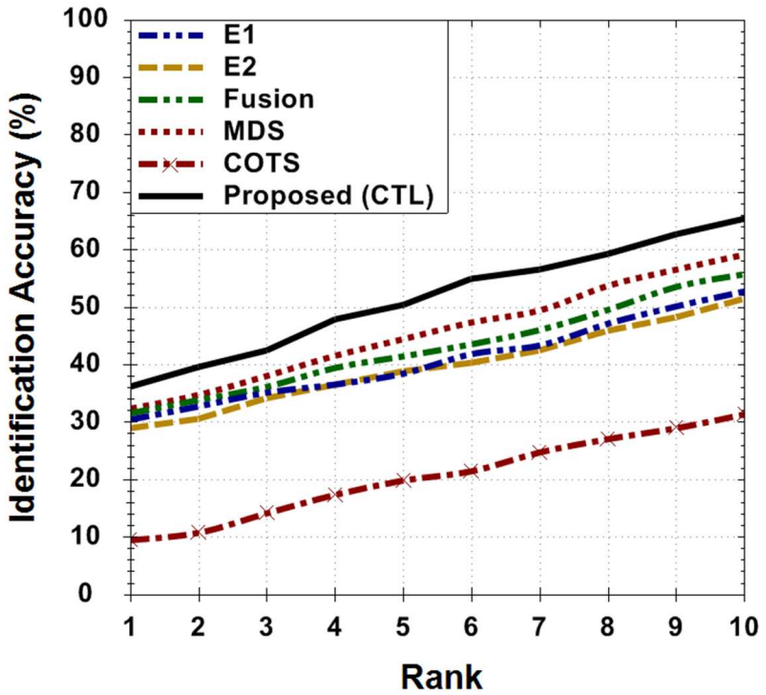

Figure 4.12: CMC curves showing the performance for matching 24×24 probe images with 72×72 gallery images on the ChokePoint database.

- The Cumulative Match Characteristics curves in Figures 4.10, 4.11, 4.12, and 4.13 show the performance of different algorithms for matching probe images of resolution 24×24 with gallery images of resolution 72×72. As compared to the fusion of two ensembles, the knowledge transfer from the source to target domain improves the accuracy by at least 4-5%. During initial training, since the source and target domain classifiers are trained independently, the knowledge transfer is not available in an ensemble. It is feasible only with pseudo labeled probe instances available in the target domain during testing.

- Table 4.4 shows the results on the CMU Multi-PIE database. The images in the CMU Multi-PIE database are of very high quality and therefore the results on this database may not be representative of cross resolution face matching with surveillance quality databases. However, previous research on low resolution face recognition has shown results on the CMU Multi-PIE database, therefore, we used this database (along with three surveillance databases) to establish the baseline comparison with MDS. The results show that for high resolutions, COTS performs better than the proposed

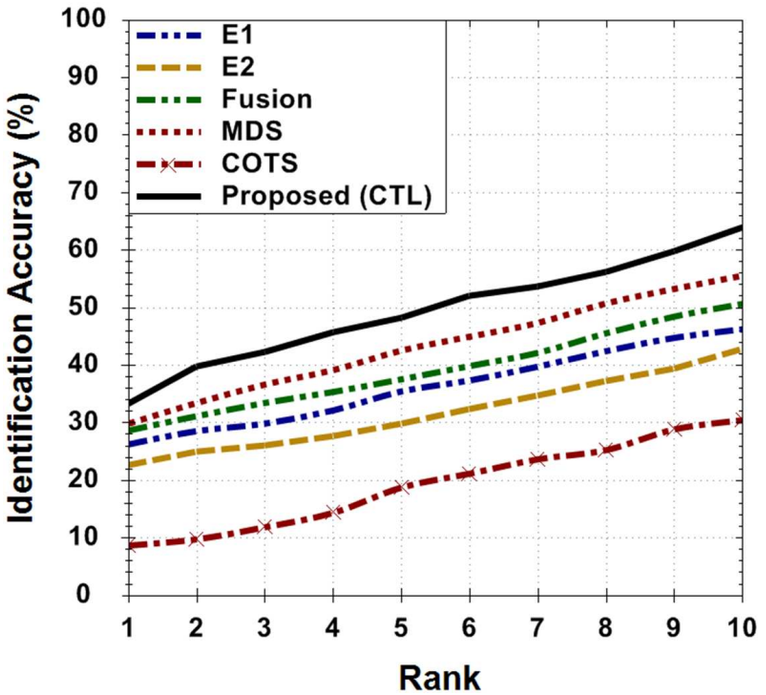

Figure 4.13: CMC curves showing the performance for matching 24×24 probe images with 72×72 gallery images on the MBGC v.2 video challenge database.

CTL and MDS algorithms. However, the performance of the commercial system reduces significantly on reducing the resolution of probe images. On the contrary, the performance of CTL reduces at a lower rate and it yields better results that COTS when the probe image is of resolution 16×16.

- Table 4.5 shows the results on the SCface database [10]. The proposed algorithm yields promising results on the real-world surveillance database and even outperforms COTS by at least 24% on all combinations of gallery and probe resolutions. Since the proposed algorithm uses SIFT and LPQ features that are resilient to pose variations and changes in gray-level intensities due to illumination variations, it inherently addresses the problem of head-pose and illumination variations in the SCface database. Moreover, the knowledge transfer with unlabeled probe instances in the target domain facilitates to efficiently classify the low resolution probes.

- Tables 4.6 and 4.7 illustrate the performance on the ChokePoint [11] and MBGC v.2 video challenge [7] databases respectively. The results that on both the databases,

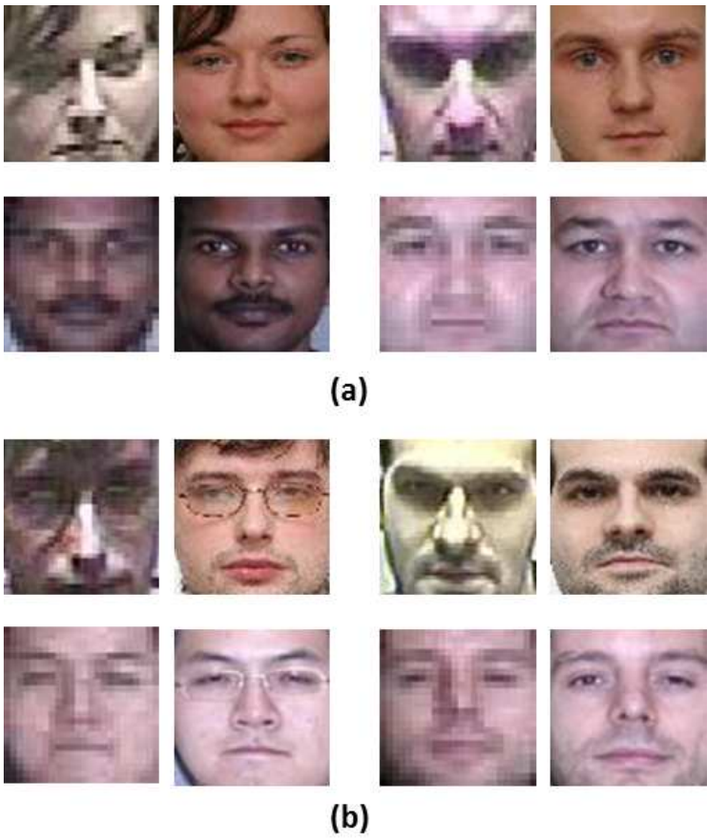

Figure 4.14: Illustrating sample cases when the proposed approach (a) correctly recognizes and (b) fails to recognize. All the examples are with probe (left image) size 24×24 and gallery (right image) size 72×72.

the proposed algorithm performs better than the existing algorithms and COTS for all combinations of gallery-probe resolutions (except for gallery 216×216 and probe 72×72, where COTS gives better performance).

- From the results shown on the three surveillance databases, it can be inferred that for high resolution gallery-probe pairs, COTS performs better than the proposed algorithm. However, for lower resolutions, the proposed algorithm yields better results. The performance of transformation based approaches such as MDS [63] degrade when the difference in resolution of gallery and probe images increases (i.e. matching gallery images of 216×216 with probe image of resolution 32×32 or lower). The transformations learned for such wide variations in gallery-probe resolution may not be precise and thus degrade the performance.

- Experimental results in Tables 4.4-4.7 also show that sum-rule fusion [125] of the proposed algorithm with COTS further enhances the performance of cross-resolution face matching. This improvement in performance may be attributed to the combined effect of COTS and CTL. COTS efficiently addresses the difference in the information content at higher resolutions, while CTL addresses the problem of limited biometric information at low resolution images. On the contrary, sum-rule fusion of the proposed CTL with MDS [63] slightly degrades the performance as it may not efficiently accommodate for large difference in information content between the gallery-probe pairs.

- Figures 4.15-4.18 show the confidence interval for the proposed algorithm, COTS and MDS on the four databases used for measuring the efficacy of the proposed algorithm for matching cross-resolution face images.

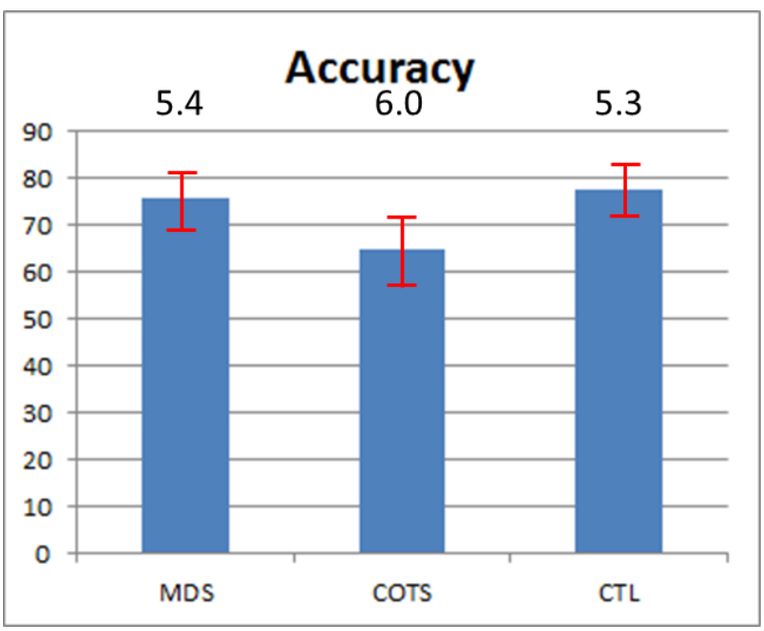

Figure 4.15: Illustrates the confidence interval for different algorithms for matching cross-resolution face images on the CMU Multi-PIE database.

4.5.1.2 Comparison with Super-resolution based Approaches

In this section, the performance of the proposed co-transfer learning algorithm is compared with three super-resolution techniques proposed in literature. For evaluating the effective-

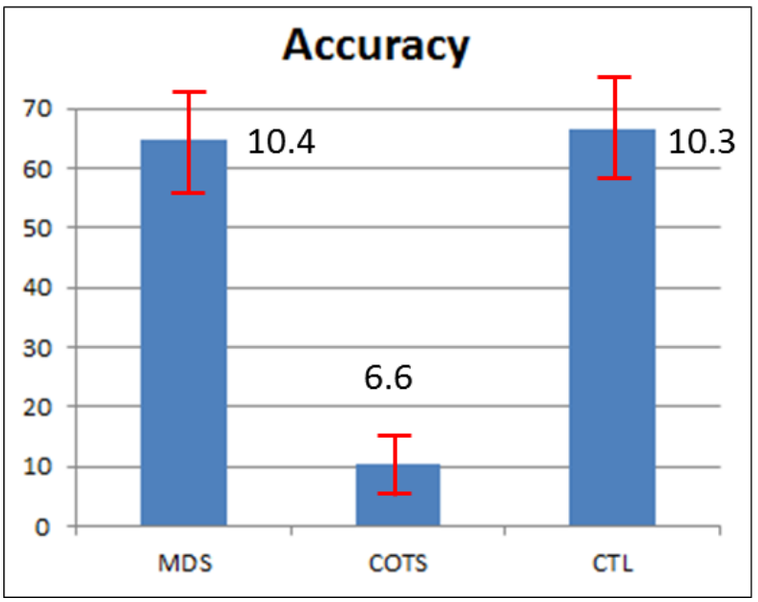

Figure 4.16: Illustrates the confidence interval for different algorithms for matching cross-resolution face images on the SCface database.

ness of super-resolution techniques for matching low and high resolution face images, it is used as a pre-processing step to enhance the quality of low resolution face images before matching. The enhanced image is matched with the high resolution gallery image using different algorithms. The LPQ and SIFT features are extracted from the super-resolution images and the performance is computed after sum-rule fusion [125] of LPQ and SIFT match scores computed using the χ^2 distance metric. For evaluating the performance with the proposed technique and COTS, super-resolution based on sparse representation (SR-3) is applied on the probe images and then feature extracting and matching are performed using the CTL algorithm (referred to as "CTL+SR") and COTS (referred to as "COTS+SR"). The target domain thus includes enhanced images obtained using super-resolution. It is to be noted that transfer learning is still applicable as super-resolution introduces several artifacts that may affect the biometric information in a face image and leads to variations in data distribution (of features or match scores) between the source and target domains. The classifiers in target domain are now trained to match the enhanced probe images with HR gallery. For the experiments, super-resolution is performed with a magnification factor of three to match probe images of size 24×24 with 72×72 gallery images. Figure 4.19 shows examples of enhanced images obtained using the three

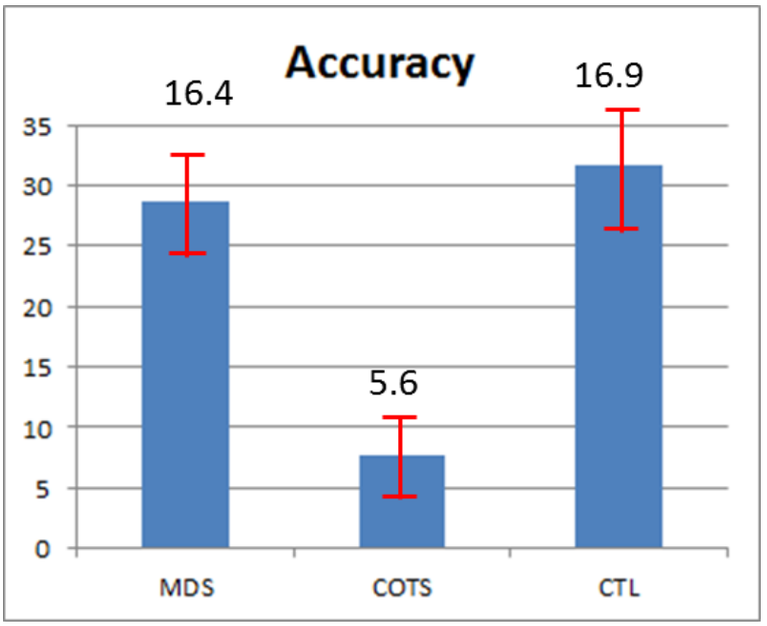

Figure 4.17: Illustrates the confidence interval for different algorithms for matching cross-resolution face images on the ChokePoint database.

super-resolution techniques and Figures 4.20, 4.21, 4.22, and 4.23 show the CMC curves. The key analysis and observations from the experiments are listed below:

- CMC curves in Figures 4.20, 4.21, 4.22, and 4.23 show that the proposed co-transfer learning algorithm outperforms all three super-resolution techniques by at least ~11% on the CMU Multi-PIE database, ~10% on the SCface database, and ~4% on the ChokePoint and MBGC v.2 video challenge databases.

- As shown in Figures 4.20, 4.21, 4.22, and 4.23, enhancing probe images using super-resolution boosts the performance of both CTL and COTS. It is observed that super-resolution minimizes the difference in the resolutions of gallery and probe images. However, it does not enhance the biometric information in the face images. Therefore, the performance gain is constrained by limited biometric information in low resolution face images.

4.5.1.3 Performance on Real World Cases

Recently, Klontz and Jain [167] have investigated the opportunity for face recognition algorithms to facilitate law enforcement agencies in identifying individuals from the crime

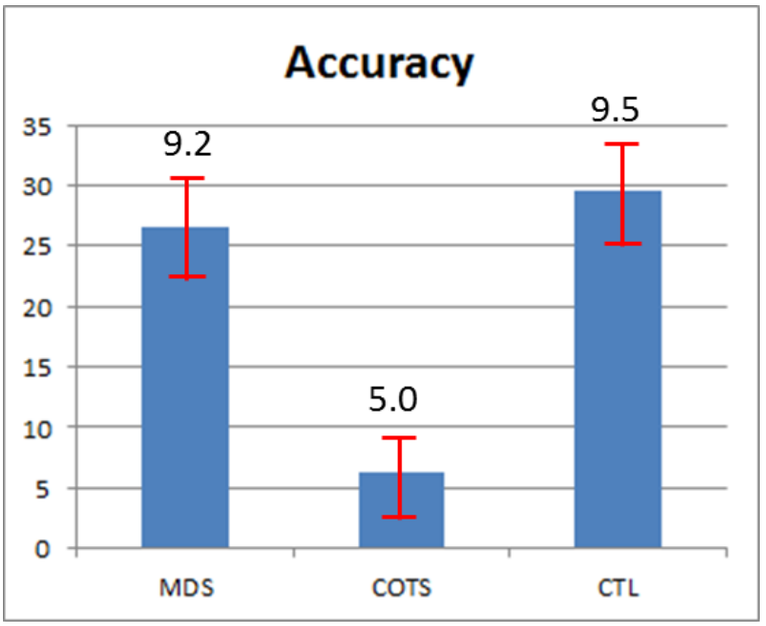

Figure 4.18: Illustrates the confidence interval for different algorithms for matching cross-resolution face images on the MBGC v2 video challenge database.

scene CCTV images during the Boston bombings incident. Inspired by their study, the performance of the proposed co-transfer learning algorithm is also evaluated on some real world examples pertaining to cross-resolution face matching. In our experiments, some real world examples are collected from different sources on the internet which includes two individuals from Boston bombing [167, 168], four individuals from London bombing [130] and one individual from Mumbai terrorist attack [131]. Figure 4.24 shows the low resolution probes and corresponding gallery images considered in the experiment. In this additional experiment for evaluating the performance with these seven real world examples, we appended these images to the SCface database for co-transfer learning. The experiments are performed with gallery image resolution of 72×72 pixels and query image resolution of 32×32 pixels. Each individual has one image in the gallery and one or more low resolution images as probe. Further, an extended gallery of 6534 individuals is created by using frontal images acquired from a law enforcement agency and appending it to the gallery of the SCface database. The performance of the proposed co-transfer learning algorithm is also compared with COTS for matching 15 probe images corresponding to these 7 real world cases. The results in Table 4.8 show that the proposed algorithm

Figure 4.19: Enhanced images obtained using three super-resolution techniques (SR-1,SR-2, and SR-3). The leftmost column represents low resolution (24×24) images and the rightmost column represents the original high resolution images (72×72) from the (a) CMU Multi-PIE, (b) SCface, (c) ChokePoint, and (d) MBGC v.2 video challenge databases.

consistently retrieves the correct match at a lower rank than COTS[1] on all the cases. The results validate our initial assertion that the proposed co-transfer learning algorithm can efficiently be coupled with surveillance systems to assist law enforcement agencies.

4.6 Summary

The chapter introduces a co-transfer learning framework which seamlessly combines the co-training and transfer learning paradigms for efficient cross-resolution face matching. During training, the proposed framework learns to match high resolution face images in the source domain. This knowledge is then transferred from the source domain to the target domain to match low resolution probes with high resolution gallery. The proposed

[1]Since, the eye region is occluded in some of the probe images, COTS is not able to process such cases (represented as NP - Not Processed).

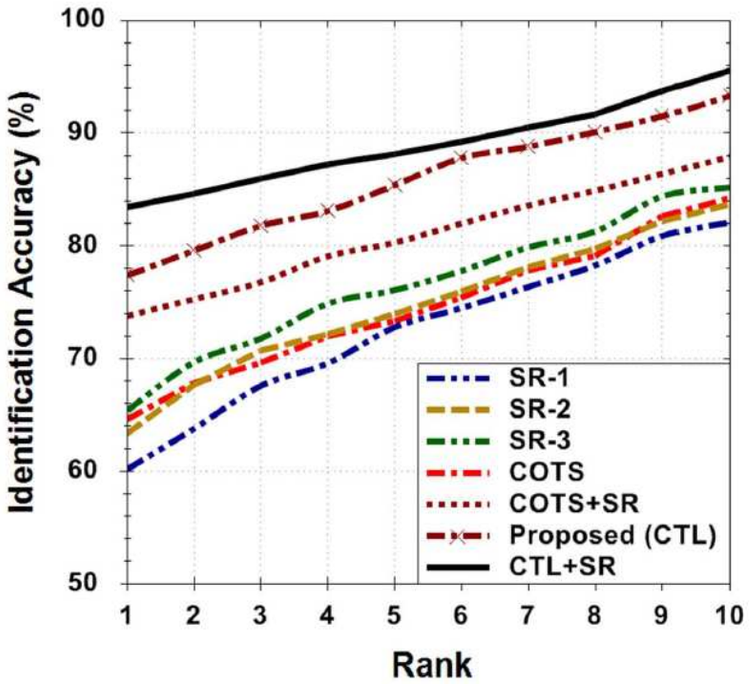

Figure 4.20: CMC curves comparing the performance of the proposed algorithm with three super resolution techniques on the CMU Multi-PIE database. Probe images of 24×24 pixels are super-resolved by a magnification factor of 3 to match the gallery resolution of 72×72 pixels.

framework builds ensembles from the weighted combination of source and target domain classifiers on two separate views. Two ensembles trained on separate views transform the unlabeled probe instances into pseudo-labeled instances using co-training. These pseudo-labeled instances are utilized for updating the decision boundary of the target domain classifier, thus, transferring knowledge from the source domain to the target domain. Further, dynamically updating the weights assigned to each classifier facilitates gradual shift of knowledge from the source to target domain. The amalgamation of transfer learning and co-training helps to transfer the knowledge from the source to target domain with probe instances as and when they arrive. Comprehensive analysis, including comparison with existing cross-resolution face matching algorithms, super-resolution techniques, and a commercial face recognition system, is performed for different gallery-probe resolutions ranging from 216×216 to 16×16 pixels. The proposed co-transfer learning framework provides significant improvement for cross-resolution face matching on different surveillance quality face databases.

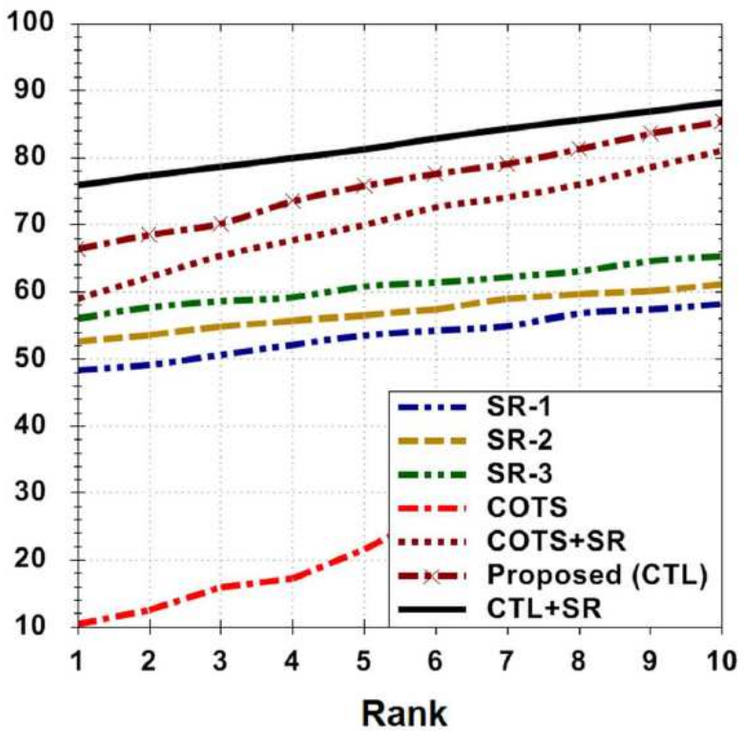

Figure 4.21: CMC curves comparing the performance of the proposed algorithm with three super resolution techniques on the SCface database. Probe images of 24×24 pixels are super-resolved by a magnification factor of 3 to match the gallery resolution of 72×72 pixels.

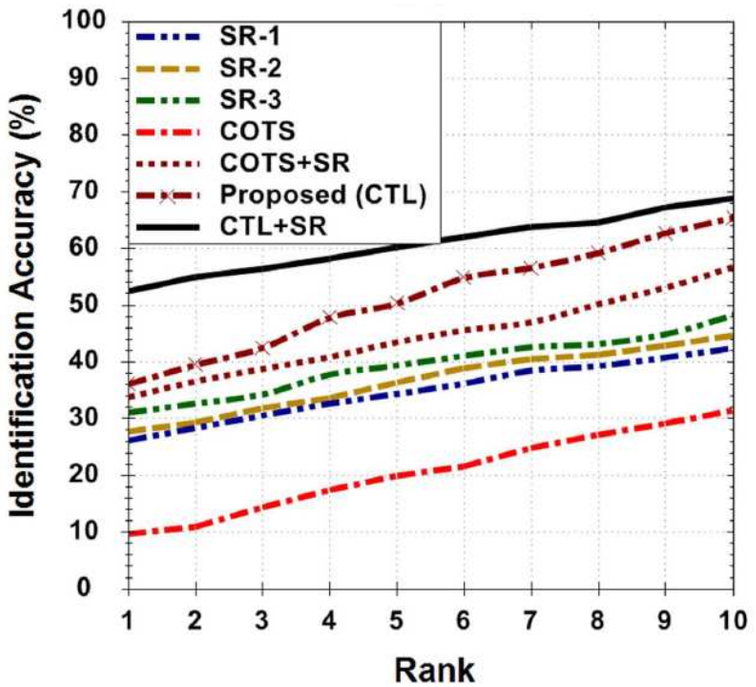

Figure 4.22: CMC curves comparing the performance of the proposed algorithm with three super resolution techniques on the ChokePoint database. Probe images of 24×24 pixels are super-resolved by a magnification factor of 3 to match the gallery resolution of 72×72 pixels.

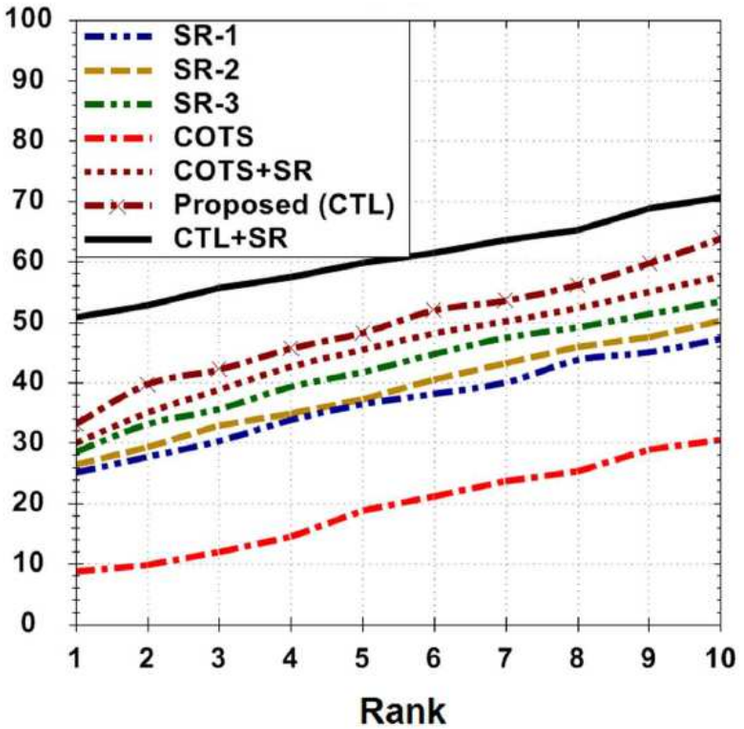

Figure 4.23: CMC curves comparing the performance of the proposed algorithm with three super resolution techniques on the MBGC v.2 video challenge database. Probe images of 24×24 pixels are super-resolved by a magnification factor of 3 to match the gallery resolution of 72×72 pixels.

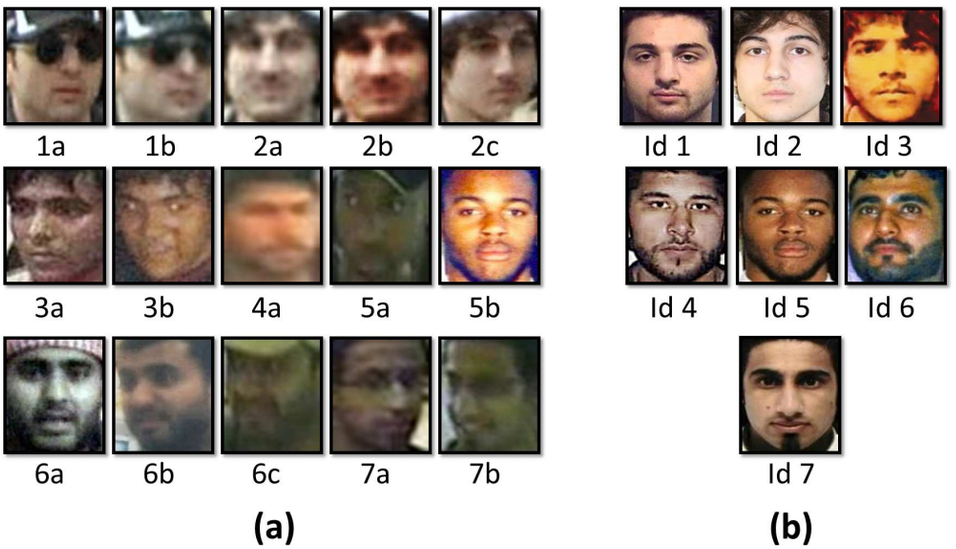

Figure 4.24: Real world cases for cross-resolution face matching: (a) low resolution probe images and (b) corresponding gallery images.

Table 4.8: Results for matching real world examples against a large scale gallery of 6534 individuals. Values in the table represents the rank at which the correct identity is retrieved. NP represents the cases which are not processed by the COTS.

Probe	CTL	COTS
1a	7	NP
1b	29	NP
2a	8	11
2b	17	26
2c	1	3
3a	15	28
3b	5	9
4a	19	22
5a	17	NP
5b	1	1
6a	1	1
6b	10	14
6c	18	NP
7a	2	4
7b	4	NP

Chapter 5

Recognizing Faces in Videos using Clustering Based Re-ranking and Fusion

5.1 Introduction

With the increase in usage of camera technology in both surveillance and personal applications, enormous amount of video feed is being captured everyday. For instance, almost 100 hours of video are being uploaded every minute on Youtube alone[1] and it is increasing rapidly. Surveillance cameras are also capturing significant amount of data across the globe. In terms of face recognition, the amount of data collected by surveillance cameras every day is probably more than the size of all the publicly available face image databases combined. One primary purpose of collecting these data from surveillance cameras is to detect any unwanted activity during the act or at least enable to analyze the events and may be determine the persons of interest after the act. Therefore, widespread use of video cameras for surveillance and security applications have stirred extensive research interest in video based face recognition.

While face recognition is a well-studied problem and several algorithms have been proposed [13, 169], a majority of the literature is on matching still images and face recognition from videos is relatively less explored. Recognizing the individuals appearing in videos has both advantages and disadvantages compared to still face matching. Since the acquisition in videos is unconstrained, the presence of covariates such as pose, illumination, and expression is significantly more but at the same time, the information available in a video is generally more than the information available for matching two still images.

[1] http://www.youtube.com/yt/press/statistics.html

As shown in Figure 5.1, videos provide several cues in the form of multiple frames and temporal information as compared to still images. These cues can be used for improving the performance of face recognition and provide robustness to large variations in facial pose, expression, and lighting conditions.

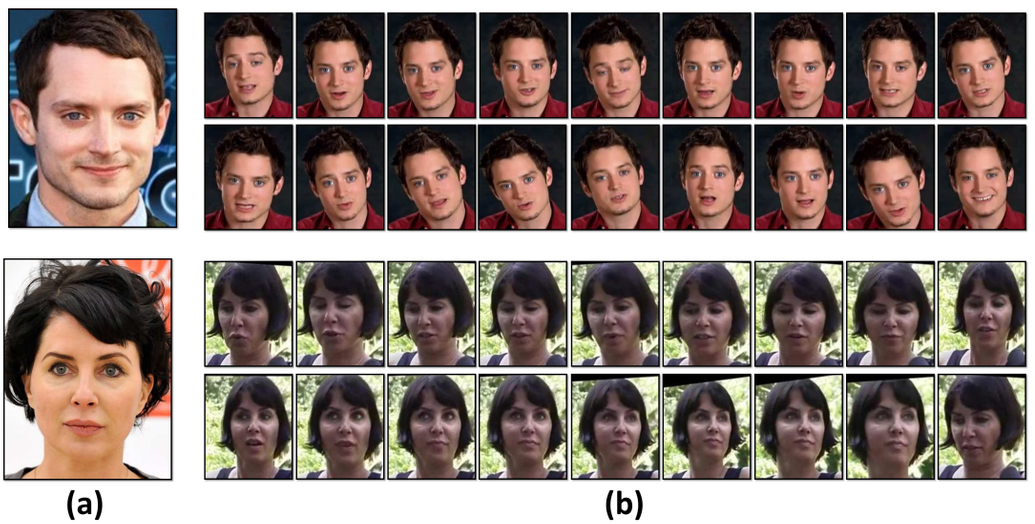

Figure 5.1: Illustrates the abundant information present in videos. Compared to (a) still face images, (b) video frames represent large intra-personal and temporal variations useful for face recognition.

Video based face recognition includes (1) matching video-to-still face images (or still-to-videos) and (2) matching two videos. In video-to-still face recognition, the probe (query) is a video sequence and the gallery is composed of still face images whereas in still-to-video face matching, the gallery and probe are switched. As proposed by Zhang *et al.* [170], video-to-still/still-to-video face recognition techniques can be broadly categorized into frame selection and multi-frame fusion approaches. In frame selection, one or more optimal frames are selected from a video sequence and used to compute the similarity between the video and still images. On the other hand, in multi-frame fusion approaches, recognition results of multiple frames are fused together. In video-to-video face recognition, both gallery and probe (query) are videos of individuals to be matched. Poh *et al.* [171] evaluated several existing approaches for video-to-video face recognition and their analysis suggests that existing techniques do not efficiently utilize the abundant information in videos for enhancing face recognition performance. They also suggest that (1) part-based approaches generally out-perform holistic approaches and (2) selecting frames based on

the image quality boosts the recognition performance. To further evaluate existing algorithms for video-to-video face recognition, the Multiple Biometric Grand Challenge [7] also featured a problem on face recognition from unconstrained videos. The results from the challenge suggest that there is a huge gap in the performance of state-of-the-art algorithms from still image to video based face recognition. Observations and analysis from these evaluations elicit further research in video based face recognition.

5.1.1 Related Research

The survey on video based face recognition by Barr *et al.* [190] categorizes existing approaches as set based and sequence based approaches. Table 5.1 summarizes the existing video based face recognition algorithms. Set based approaches [6, 191] utilize the abundance and variety of observations in a video to achieve resilience to sub-optimal capture conditions. The approaches [172, 192] that model image sets as distributions use the between-distribution similarity to match two image sets. However, the performance of such approaches depend on the parameter estimation of the underlying distribution. Modeling image sets as linear sub-spaces [176, 177, 178] and manifolds [172, 173, 179, 193] is also proposed where matching between two image sets is performed by measuring similarity between the input and reference subspaces/manifolds. However, the performance of a subspace/manifold based approach depends on maintaining the image set correspondences. To address these limitations, Cui *et al.* [181] proposed to align two image sets using a common reference set before matching. Lee *et al.* [183] proposed a connected manifold approach that utilizes the likelihood and transition probability of the nearest previous manifold for recognition. Hu *et al.* [184] proposed to represent an image set using sample images, their mean, and an affine hull model. A sparse approximated nearest point method was proposed to compute the between-set distance as a pair of nearest points on the sets that are sparsely approximated by sample images. On the other hand, sequence based approaches explicitly utilize the temporal information for improved face recognition. To utilize the temporal information, Zhou *et al.* [185] proposed to use a joint posterior probability distribution of motion vector and identity variable estimated using sequence importance sampling. Several approaches that model the temporal information with Hidden Markov Models (HMM) [182, 186] are also proposed for improved video based face recognition.

Recently, the research focus has shifted and advancements in face recognition have led to a new paradigm of matching face images using a large dictionary. Patel *et al.* [194] proposed a sparse approximation based approach where test images were projected

Category	Authors	Technique	Database	Recognition Rate (%)
Set Based	Arandjelovic *et al.* [172]	Manifold density divergence	Private	93.6 (avg)
	Wang *et al.* [173]	Manifold-manifold distance	Honda/UCSD [174]	96.9
			CMU MoBo [175]	93.6
	Aggarwal *et al.* [176]	Linear dynamic modeling	Private	93.7
			Honda/UCSD [174]	90.0
	Fukui & Yamaguchi [177]	Kernel orthogonal mutual subspace	Private	97.42/EER=3.5
	Nishiyama *et al.* [178]	Hierarchical image-set matching	Private	97.4/EER=2.3
	Harandi *et al.* [179]	Grassmannian manifolds	CMU PIE [31]	65.2
			BANCA [180]	64.5
			CMU MoBo [175]	64.9
	Cui *et al.* [181]	Image set alignmnet	Honda/UCSD [174]	98.9
			CMU MoBo [175]	95.0
			YouTube celebrity [182]	74.6
	Lee *et al.* [183]	Probabilistic appearance manifolds	Private	93.2
	Hu *et al.* [184]	Sparse Approximated Nearest Point	Honda UCSD [174]	92.3
			CMU MoBo [175]	97
			YouTube Celebrity [182]	65.0
	Wolf *et al.* [6]	Set-to-set similarity	YouTube Faces [6]	72.6 at EER
Sequence Based	Zhou *et al.* [185]	Sequential importance sampling	Private	100
			Private	~93
			CMU MoBo [175]	~56
	Liu & Chen [186]	Adaptive Hidden Markov models	Private	1.2 EER
			CMU MoBo [175]	4.0 EER
	Kim *et al.* [182]	Visual constraints using generative & discriminative models	Honda/UCSD [174]	100
			YouTube celebrity [182]	~70
Dictionary Based	Chen *et al.* [187]	Video-dictionaries	MBGC v1 [188]	~59 at EER (WW)
				~55 at EER (AW)
				~51 at EER (AA)
	Bhatt *et al.* [189]	Rank aggregation	YouTube Faces [6]	78.3 at EER
	Proposed	Clustering based re-ranking and fusion	YouTube Faces [6]	80.7 at EER
			MBGC v2 [7]	62.2 at EER (WW)
				57.3 at EER (AW)
				54.1 at EER (AA)

Table 5.1: Categorization of existing approaches of video based face recognition.

onto a span of elements in learned dictionaries and the resulting residual vectors were used for classification. Chen *et al.* [187] proposed a generative approach for video based face recognition where a video sequence was first partitioned into sub-sequences and then sequence-specific dictionaries were learned. The frames from every query video were projected onto the span of atoms in every sequence-specific dictionary and the residuals were utilized for recognition. Their approach has a computational overhead of creating multiple sequence-specific dictionaries for specific pose and illumination variations. Chen *et al.* [195] proposed a multi-variate sparse representation that simultaneously takes correlation as well as coupling information between frames. Different sub-directories were trained for multiple partitions which represents a particular viewing condition and a joint sparse representation was used for face recognition using minimum class reconstruction error criteria. Recently, Bhatt *et al.* [189] proposed to compute a video signature as an ordered list of still face images from a large dictionary. In their approach, temporal and wide intra-personal variations from multiple frames were combined using Markov chain based rank aggregation approach.

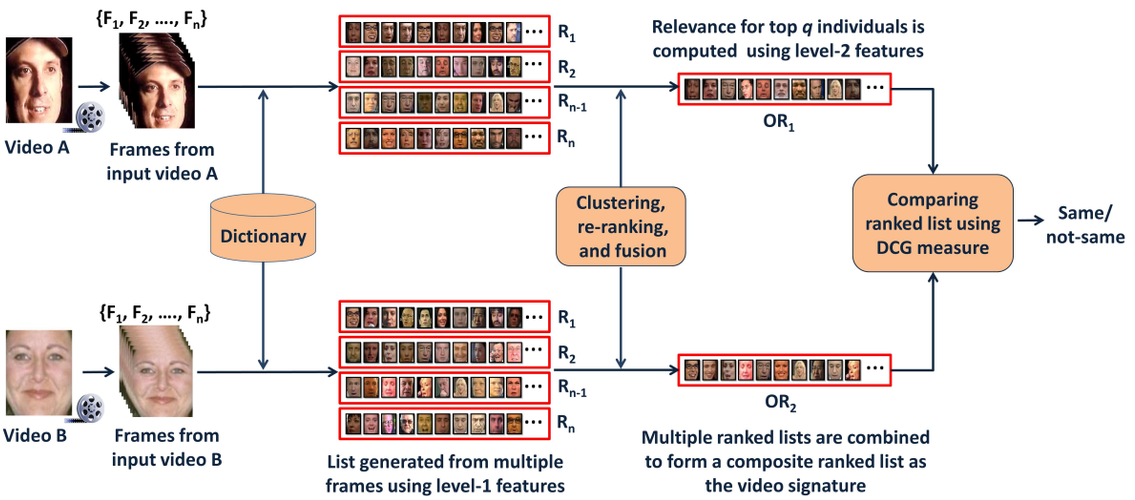

Figure 5.2: Illustrates the block diagram of the proposed algorithm for matching two videos.

5.1.2 Research Contributions

This chapter proposes a novel algorithm for video based face recognition that computes the signature of a video as an ordered list of still face images from a dictionary. Figure 5.2 shows the outline of the proposed algorithm which starts by computing a ranked list

for every frame in the video to utilize the abundant information and capture the wide intra-personal variations. It utilizes the taxonomy of facial features [196] to efficiently compute the video signature. Taxonomy of facial features [196] groups the salient information available in face images into different feature categories: level-1, level-2, and level-3. Out of these three, level-1 facial features capture the holistic nature of face such as skin color, gender, and appearance of the face. These features are highly discriminative in differentiating an image from other images that have largely different facial appearances. These features being computationally efficient are generally used for indexing or reducing the search space. Therefore, level-1 features are used to generate a ranked list by congregating images from the dictionary that are similar to the input frame. A ranked list is an ordered list of face images retrieved from the dictionary where the face image with the highest similarity is positioned at the top of the list. To characterize an individual in a video, ranked lists from multiple frames are combined using a three stage process that involves clustering, re-ranking, and fusion. It produces the final composite ranked list for a video which represents the discriminative video signature. Combining multiple ranked lists into a single optimized ranked list that minimizes the overall distance from all ranked lists is a well studied problem in information retrieval domain. However, to the best of our knowledge, this dissertation presents the first approach to combine ranked lists pertaining to individual frames to generate a composite video signature. It transforms the problem of video based face recognition into matching two ordered lists (ranked lists). Further, a relevance score is computed for images in the final composite ranked list using the discriminative level-2 features. These are locally derived features and describe structures in a face that are pertinent for face recognition. As compared to level-1 features, these features are more discriminative and are predominantly used for face recognition. Relevance score computed using level-2 features represent the usefulness of an image in characterizing the individual in a video. Finally, to match two videos, their composite ranked lists (video signatures) are compared using a discounted cumulative gain (DCG) measure [197]. The major contributions of this chapter can be summarized as follows:

- It utilizes the taxonomy of facial features for efficient video based face recognition. Computationally efficient level-1 features are used for computing multiple ranked lists pertaining to multiple video frames and discriminative level-2 features are used to compute the relevance of images in the final composite ranked list.

- Existing dictionary based face recognition algorithms [198] compute the signature of a still face image as an ordered list of images from dictionary. In this chapter, a new

paradigm is introduced using a three-stage technique for generating video signatures as an ordered list of still face images from the dictionary.

- Existing approaches discard the characteristics embedded in the ranked lists and only consider the overlap between two lists as the final similarity. In this chapter, the DCG measure seamlessly utilizes rank and relevance scores of images to compute the final similarity between two lists.

5.2 Dictionary Based Video Face Recognition Algorithm

Recent studies in face recognition [194, 198, 199] have shown that generating image signatures based on a dictionary is more efficient for matching images across large variations than direct comparison between two images or some of its features. In this chapter, video based face recognition is addressed by computing a discriminative video signature using a dictionary of still face images. The proposed algorithm congregates abundant information present in multiple video frames to generate a discriminative video signature. It facilitates in characterizing an individual as it embeds the information in the form of a ranked list of images under similar intra-personal settings from the dictionary. Figure 5.2 shows different stages of the proposed algorithm which are elaborated in the following subsections.

5.2.1 Dictionary

Dictionary is a large collection of still face images where every individual has multiple images capturing a wide range of intra-personal variations i.e. pose, illumination, and expression variations. Our definition of dictionary is different from the dictionary in sparse representation based approaches [184, 187]. They represent a dictionary as a collection of atoms such that the number of atoms exceeds the dimension of the signal space, so that any signal can be represented by more than on combination of different atoms. In this chapter, the dictionary comprises $38,488$ face images pertaining to 337 individuals from the CMU Multi-PIE [9] database captured in multiple sessions. OpenCV's boosted cascade of Haar-like features provide the face boundaries and eye-coordinates. These boundaries are used to detect and crop faces from the dictionary images and eye-coordinates are used to normalize the detected image with respect to rotation. The normalized face images are resized to 196×224 pixels with inter-eye distance of 100 pixels.

5.2.2 Computing Ranked List

Let V be the video of an individual comprising n frames where each frame depicts the temporal variations of the individual. Face region from each frame is detected[1] and pre-processed[2]. Face regions corresponding to different frames across a video are represented as $\{F_1, F_2, ..., F_n\}$. To generate ranked lists, each frame is compared with all the images in the dictionary. Since the dictionary consists of a large number of images and each video has multiple frames; it is essential to compute the ranked list in a computationally efficient manner. Linear discriminant analysis (LDA), level-1 feature, is therefore used to generate a ranked list by congregating images from the dictionary that are similar to the input frame. A linear discriminant function [202] is learned from the dictionary images that captures the variations in pose, illumination, and expression. The linear discriminant function learns these variations and retrieves images from the dictionary that are similar to the input video frame i.e. images with similar pose, illumination, and expression. The ranking of retrieved images from such a dictionary is found to be more discriminative for face recognition than that of a signature based on the pixel intensities or some image features [198]. Each column of the projection matrix W represents a projection direction in the subspace and the projection of an image onto the subspace is computed as:

$$Y = W^T X \tag{5.1}$$

where X is an input image and Y is its subspace representation. The input frame F_i and all images in the dictionary are projected onto the subspace. The Euclidean distance is computed between the subspace representations of the input frame F_i and each of the dictionary images. An ordered list of images is retrieved from the dictionary based on their similarity[3] to the input frame. To generate a ranked list $\mathbf{R_i}$ corresponding to the input frame F_i, retrieved dictionary images are positioned based on their similarity to F_i with the most similar image positioned at the top of the list. For a video V, the proposed algorithm computes a set of ranked list $\{\mathbf{R_1}, \mathbf{R_2}, ..., \mathbf{R_n}\}$ corresponding to the n frames of the video.

[1] OpenCV's boosted cascade of haar-like features is used for face detection in near-frontal videos. For profile-face videos, a tracking technique [200] is used to track and crop faces. Obtaining cropped faces from videos was a combination of automatic and manual tasks where we located the tracker for the face region in the first frame each time.

[2] A multi-scale retinex with wavelet based de-noising technique [201] is utilized to enhance the quality of poor quality video frames before computing the ranked list.

[3] The distance scores computed using level-1 features are normalized in range {0-1} using min-max normalization and then converted into similarity scores.

5.2.3 Clustering, Re-ranking, and Fusion

Multiple ranked lists computed across n frames of a video have significant amount of overlap in terms of positioning of the dictionary images. Due to this redundant information, it is computationally expensive and inefficient to compare multiple ranked lists across two videos. Therefore, multiple ranked lists of a video are combined to generate a single composite ranked list, denoted as $\mathbf{R}'$. As shown in Figure 5.3, the proposed algorithm generates a composite ranked list in three steps. First, each ranked list corresponding to a video frame is partitioned into different clusters and reliability of each cluster is calculated. Secondly, the similarity scores of images within a cluster are adjusted based on the reliability of that cluster [203]. Finally, multiple ranked lists of a video are fused based on the adjusted similarity scores of images to generate a composite ranked list as the video signature. The video signature thus obtained minimizes the distance from all the constituent ranked lists. These stages are described in Algorithm 3 and are elaborated below.

Figure 5.3: Illustrates clustering based re-ranking and fusion to form the video signature. Clustering based re-ranking associates dictionary images to different clusters and adjusts their similarity scores. It facilitates to bring images similar to the query frame towards the top of the ranked list. The lists are then re-ranked using the adjusted scores and are finally combined to generate the video signature.

Algorithm 3 Fusing ranked lists with clustering and re-ranking.

Input: A set of ranked lists $\mathbf{R_1}, \mathbf{R_2}, ..., \mathbf{R_n}$ from multiple frames in a video V.
Iterate: i= 1 **to** n (number of ranked lists)
Clustering: Partition ranked list $\mathbf{R_i}$ into different clusters $C_{i,1}, C_{i,2}, ..., C_{i,k}$, where k is the number of clusters.
end iterate.
Iterate: i= 1 **to** n, j= 1 **to** k.
Reliability: Compute reliability of cluster $r(C_{i,j})$.
Re-ranking: Adjust the similarity score of each image d based on the reliability of the cluster it belongs.
$Sim_i^*(d)$=$Sim_i(d) \times (1 + r(C_{i,j}))$, $d \in C_{i,j}$.
end iterate.
Fusion: Compute an ordered composite ranked list $\mathbf{R}'$ where similarity score of an image d is given as:
$SS_d = \frac{\sum_{i=1}^{n} Sim_i^*(d)}{n}$.
Output: Final composite ranked list $\mathbf{R}'$ for video V.

5.2.3.1 Clustering

Multiple frames in a video exhibit different intra-personal variations; therefore, each ranked list positions dictionary images based on the similarity to the input frame. Images in the ranked list are further partitioned into different clusters such that if an image in a cluster has high similarity to the input frame, then all images in that cluster tend to be more similar to the input frame. The main idea behind clustering is to congregate images in a ranked list into different clusters where each cluster represents a particular viewing condition i.e. a specific pose, illumination or expression. Let $\mathbf{R_i}$ be the i^{th} ranked list of a video corresponding to frame F_i, then $\{C_{i,1}, C_{i,2}, ..., C_{i,k}\}$ form k clusters of $\mathbf{R_i}$. In this chapter, K-means clustering [204] which is an unsupervised, non-deterministic technique for generating a number of disjoint and flat (non-hierarchical) clusters is used to cluster similar images with an equal cardinality constraint. To guarantee that all clusters have equal number of data points, k centroids are initially selected at random. For each point, similarity to the nearest cluster is computed and a heap is build. Similarity is measured using the Euclidean distance in LDA projection space, as described in Eq. 5.1. A data point is drawn from the heap and assigned to the nearest cluster, unless that cluster is already full. If the nearest cluster is full, distance to the next nearest cluster is computed and the data is re-inserted into the heap. The process is repeated till the heap is empty i.e. all the data points are assigned to a cluster. It guarantees that all the clusters contain equal number of data points ($\pm$1 data points per cluster). K-means clustering is used as it is computationally faster and produces tighter clusters than hierarchical clustering

techniques. After clustering, each ranked list $\mathbf{R_i}$ has a set of clusters $C_{i,1}$, $C_{i,2}$,..., $C_{i,k}$, where k is the number of clusters. K-means clustering is affected by the initialization of initial centroid points; however, we start with five different random initializations of k clusters. Finally, clusters which minimize the overall sum of square distances are selected.

5.2.3.2 Re-ranking

Clusters across multiple ranked lists overlap in terms of common dictionary images. Since the overlap between the clusters depends on the size of each cluster, it is required that all the clusters should be of equal size. Higher the overlap between the clusters, more likely that they contain images with similar appearances (i.e. with similar pose, illumination, and expression). Based on this hypothesis, the reliability of each cluster is computed as the weighted sum of similarities between the cluster and other clusters across multiple ranked lists [203]. The *reliability* $r(C_{A,j})$ of a cluster $C_{A,j}$ in ranked list A is computed as shown in Eq. 5.2.

$$r(C_{A,j}) = \sum_{i=1,i\neq A}^{n} \sum_{p=1}^{k} \left[\frac{Sim_F C(F_i, C_{i,p})}{norm_A} Sim(C_{A,j}, C_{i,p}) \right] \tag{5.2}$$

where

$$norm_A = \sum_{i=1,i\neq A}^{n} \sum_{p=1}^{k} [Sim_F C(F_i, C_{i,p})] \tag{5.3}$$

$$Sim_F C(F_A, C_{A,j}) = \frac{\sum_{d\in C_{A,j}} ||F_A - d||^2}{|C_{A,j}|} \tag{5.4}$$

$$Sim(C_{A,j}, C_{B,j}) = |C_{A,j} \cap C_{B,j}| \tag{5.5}$$

where d is an image from the dictionary, $norm_A$ is a normalization factor for clusters in the ranked list $\mathbf{R_A}$, $|C_{A,j}|$ is the number of images in cluster $C_{A,j}$, F_A is the current frame of the video, and $||F_A - d||^2$ represents the similarity between the input frame and a dictionary image computed using the Euclidean distance[1] between their subspace representations. The similarity between frame F_i and cluster $C_{i,j}$ is measured as the average similarity score of all images in that cluster to the input frame F_i, as shown in Eq. 5.4. The similarity between two clusters is estimated in terms of the number of common images as shown in Eq. 5.5. Higher the reliability of a cluster, higher is the contribution of its constituent images. The similarity scores of images in a cluster are adjusted based on the reliability of the cluster. It enhances the similarity scores of images from a cluster

[1]The distance scores computed using level-1 features are normalized in range {0-1} using min-max normalization and then converted into similarity scores.

that exhibits similar settings as the input video frame and reduces the similarity scores of images from clusters which exhibit different settings i.e. pose, illumination, and expression variation. The reliability score of a cluster is then used to adjust the similarity scores of all images belonging to that cluster, as shown in Eq. 5.6:

$$Sim_i^*(d) = Sim_i(d) \times [1 + r(C_{i,j})], \forall\ d \in C_{i,j} \tag{5.6}$$

where $Sim_i(d)$ is the similarity score of an image d in ranked list $\mathbf{R_i}$ computed using level-1 features and $r(C_{i,j})$ is the reliability of the j^{th} cluster of the i^{th} ranked list, $C_{i,j}$, such that $d \in C_{i,j}$.

5.2.3.3 Fusion

The ranked lists across multiple frames have redundant information and matching such ranked lists across two videos can be computationally inefficient. Therefore, it is imperative to compute a composite ranked list as the video signature. Once the similarity scores of images are adjusted across all the ranked lists, multiple ranked lists are fused into a final composite ranked list, R'. The final similarity score of an image d (denoted as SS_d) is the average of adjusted similarity scores of image d across all the ranked lists, as shown in Eq. 5.7.

$$SS_d = \frac{\sum_{i=1}^{n} Sim_i^*(d)}{n} \tag{5.7}$$

where n is the number of frames in a video. There are different types of fusion methods proposed in the literature [125, 205] such as sensor level, feature level, score level, and decision level fusion. Chen *et al.* [195] proposed to concatenate n sub-dictionaries using a use joint sparsity coefficient approach to make a combined decision. However, in the proposed algorithm, adjusted similarity scores of all images in the dictionary are averaged across multiple ranked lists. The final composite ranked list $\mathbf{R'}$ of a video is generated by ordering all images from dictionary such that the image with maximum adjusted similarity score (SS) is positioned at the top of the list.

5.2.4 Matching the Composite Ranked Lists

To match two videos, their composite ranked lists obtained after clustering based re-ranking and fusion are compared. The discounted cumulative gain (DCG) [197] measure is used to compare two ranked lists. DCG measure is widely used in information retrieval domain [206] to compare the lists of documents. Each document in the ranked list is

arranged based on its similarity to the input query and also has a relevance score provided by a domain expert (or the user). It uses both these attributes (i.e. rank and relevance) to compare two ranked lists. The relevance in our context is the usefulness of a dictionary image in characterizing the individual in a video. The relevance rel_d of a dictionary image d is computed as the maximum similarity score of the image across multiple frames of the video, as shown in Eq. 5.8.

$$rel_d = \arg\max_{1 \leq i \leq n} \{Sim_{level2}(d, F_i)\} \tag{5.8}$$

where n is the number of frames in a video, $Sim_{level2}(d, F_i)$ is the similarity score of a dictionary image d with the frame F_i computed using level-2 features (LBP) and χ^2 distance measure. It is observed that the similarity between a video frame and images in the ranked list drop after a particular rank and the order of images is less discriminative beyond that point. Therefore, images retrieved till rank q are considered in the video signature and their relevance is computed. Now, the images in the composite ranked list $\mathbf{R}'$ are positioned based on level-1 features and have a relevance score computed using level-2 features.

The DCG measure captures the observation that relevant images are more useful when appearing earlier in the ranked list. Greater the rank of an image, smaller is the contribution of its relevance to the final decision. Similarity is accumulated from top of the ranked list to the bottom by discounting the relevance score of an image by its position. Therefore, DCG measure is more efficient in matching two ranked list than just comparing the overlap between two lists (later shown in results). As shown in Eq. 5.9, DCG measure discounts the relevance of an image by the logarithm of its rank.

$$DCG_q = \sum_{i=1}^{<b} rel_i + \sum_{i=b}^{q} \frac{rel_i}{log_b(i)} \tag{5.9}$$

where rel_i is the relevance score of an image at rank i and the DCG is computed till rank q. In our experiments, $q = 100$ and logarithm to the base $b = 2$ are empirically set to yield the best performance. Further, the DCG value is normalized by dividing it with ideal discounted cumulative gain ($IDCG$) to obtain normalized discounted cumulative gain $nDCG$, as shown in Eq. 5.10.

$$nDCG_q = \frac{DCG_q}{IDCG_q} \tag{5.10}$$

$IDCG$ at rank q is obtained by calculating DCG values when the images in the ranked list are positioned based on their relevance instead of similarity scores computed using level-1

features (i.e. the image with maximum relevance is positioned at the top of the list). To compute the similarity between two ranked lists, a two sided $nDCG$ measure is used. For two ranked lists $\mathbf{R'_1}$ and $\mathbf{R'_2}$, $nDCG_q$ for $\mathbf{R'_1}$ with respect to $\mathbf{R'_2}$ at rank q is computed by considering $\mathbf{R'_2}$ as the ideal ranking of images. Similarly, $nDCG_q$ for $\mathbf{R'_2}$ with respect to $\mathbf{R'_1}$ is computed by considering $\mathbf{R'_1}$ as the ideal ranking of images. The final similarity K_{sim} between two lists $\mathbf{R'_1}$ and $\mathbf{R'_2}$ is the average of the two $nDCG$ values.

$$\begin{aligned} K_{sim}(R'_1, R'_2) = \quad & \frac{1}{2}\{nDCG_q(\mathbf{R'_1}, \mathbf{R'_2}) \\ + \quad & nDCG_q(\mathbf{R'_2}, \mathbf{R'_1}))\} \end{aligned} \tag{5.11}$$

5.2.5 Dictionary Based Video Face Recognition Algorithm

The proposed algorithm for computing the video signatures and matching is summarized below:

1. For a given video pair, frames from each video are extracted and pre-processed. Face region from each frame is detected and resized to 196×224 pixels.

2. For each frame in a video, a ranked list of still face images from the dictionary is computed using level-1 features. The retrieved dictionary images are arranged in a ranked list such that the image with the maximum similarity score is positioned at the top of the list.

3. Ranked list across multiple frames of a video are combined to form a video signature using clustering based re-ranking and fusion as elaborated in Algorithm 3.

4. To match two videos, their video signatures are compared using the $nDCG$ measure that incorporates scores computed using both level-1 (rank) and level-2 (relevance) features.

The proposed video based face recognition algorithm efficiently computes the video signature and transforms the problem of video based face recognition into matching two ranked lists. Generally, in face recognition applications, level-1 and level-2 features are sufficient for efficiently matching face images. In some law enforcement applications such as matching identical twins or look-alikes, level-3 features are widely used as an additional layer of discrimination over level-1 and level-2 features. However, level-3 features are extracted from good quality high resolution face images which are generally not available in the application focussed in this research i.e. face recognition from unconstrained

videos. Therefore, only level-1 and level-2 features are used in this chapter for computing a discriminative video signature.

5.3 Experimental Results

The efficacy of the proposed algorithm is evaluated on multiple databases with different scenarios such as video-to-still, still-to-video, and video-to-video. For a thorough analysis, the performance of individual components of the proposed algorithm is evaluated along with comparing it with the min-max normalization and sum rule fusion [125], referred to as MNF, for combining multiple ranked lists across the video frames. The performance is also compared with FaceVACS which is a commercial off-the-shelf face recognition system (denoted as COTS). Section 5.3.1 explains the databases used in this chapter, Section 5.3.2 elaborates the experimental protocol, and finally Section 5.3.3 lists the key observations and analysis.

5.3.1 Databases

The experiments are performed on two publicly available video databases: The YouTube faces database [6] and MBGC v2 video challenge database [7]. The YouTube faces database [6] is the largest available unconstrained video database comprising $3,425$ videos of $1,595$ different individuals downloaded from YouTube where each video has $\sim$180 frames on average. The database provides 10-fold pair-wise matching ('same'/'not-same') test benchmark protocol for comparison with existing algorithms. $5,000$ video pairs are randomly selected from the database, half of which are pairs of videos of the same individual, and half of different individuals. As per the given protocol [6], these pairs are further divided into 10 splits where each split contains 250 'same' and 250 'not-same' pairs. Further details about the database are available in [6].

The MBGC v2 video challenge database [7] comprises videos in standard (720×480) and high definition (1440×1080) formats pertaining to individuals either walking or performing some activity. From the MBGC v2 video challenge database, experiments are performed on the data collected from the University of Notre Dame. The experiments are performed for matching videos of individuals under three settings, 1) walking vs walking (WW), 2) walking vs activity (WA), and 3) activity vs activity (AA). Further, to evaluate the performance of the proposed algorithm for still-to-video and video-to-still matching, face images pertaining to 147 individuals from the MBGC v2 still portal are utilized. These

individuals have good quality still face images and their corresponding videos. Figure 5.4 shows still face images along with samples from activity and walking video frames.

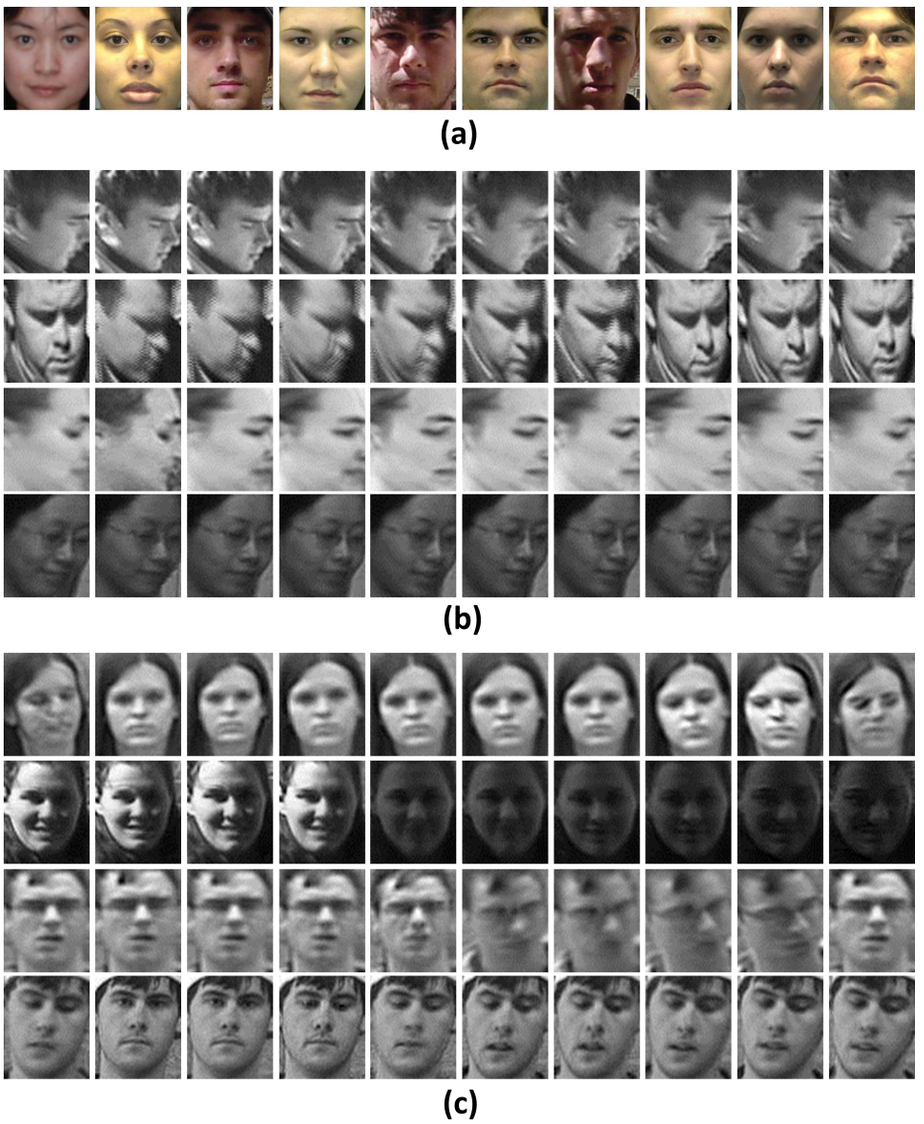

Figure 5.4: Sample images from the MBGC v2 database (a) still face images, (b) frames from activity video, and (c) frames from walking video.

5.3.2 Protocol

The efficacy of the proposed algorithm for video based face recognition is evaluated in verification mode (1:1 matching). The performance of the proposed algorithm is compared with existing video based face recognition algorithms using the experimental protocol de-

fined in [6] where the verification accuracy is reported at equal error rate (EER) along with area under the curve (AUC). For matching two videos using COTS, set-to-set matching is used where each frame in the first video is matched to all the frames in the second video. The mean score obtained corresponding to all frames of the second video is assigned as the similarity score of the frame in the first video. The final similarity score of the first video is the average score of all the frames in that video. In MNF, similarity scores across multiple ranked lists are normalized using min-max score normalization [207]. The score for each dictionary image is then re-computed as the average score across all the ranked lists. Finally, the combined ranked list is generated based on the averaged similarity scores where the dictionary image with the largest similarity score is positioned at the top of the list. The experimental protocol for the two databases are further elaborated below:

5.3.2.1 YouTube Faces Database

The performance of the proposed algorithm is evaluated using the experimental protocol defined by Wolf *et al.* [6]. In this experiment both gallery and probe consist of videos and training is performed as two class problem with 'same'/'not-same' labels. In our experiments, ten splits provided along with the database are used. Training is performed on nine splits and the performance is computed on the tenth split. The final performance is reported as an average of 10 folds. In this protocol, the information about the subject's label associated with the video is discarded and only the information about whether a pair is 'same' or 'not-same' is retained.

On the YouTube faces database, the performance of the proposed algorithm is compared with the benchmark test results provided with the database [6]. For performance comparison LBP descriptor with matched background similarity (MBGC (mean) LBP), minimum distance (mindst LBP), maximum correlation measures ($||U1'U2||$ LBP) and FPLBP descriptor with matched background similarity (MBGC (mean) FPLBP), minimum distance (mindst FPLBP), maximum correlation measures ($||U1'U2||$ FPLBP), APEM+FUSION [208], STFRD+PMML [209], VSOF+OSS [210] and one recently proposed algorithm, referred to as Bhatt *et al.* [189], are used. In the proposed approach, videos from the YouTube faces database are pre-processed using multi-scale retinex with wavelet based denoising. Experiments are performed to evaluate the performance enhancement due to different stages of the proposed algorithm on the YouTube faces database. First, to evaluate the performance gain due to clustering based re-ranking and fusion steps, the performance is compared when ranked list across multiple frames are combined using the MNF approach. Secondly, to evaluate the gain in the performance due to $nDCG$

measure, the performance is evaluated when two ranked lists are compared using the distance measure proposed by Schroff *et al.* [198]. Their distance measure only considers the overlap between two ranked lists and ignores other information such as relevance of images in the ranked list. It should be noted that while evaluating the gain in performance due to an individual step, all other steps in the proposed algorithm remain the same.

5.3.2.2 Multi Biometric Grand Challenge v2 Database

Multiple experiments are performed on this database to evaluate the efficacy of the proposed algorithm. Specifically, the algorithm is evaluated for two different scenarios: (1) matching still face images with videos and (2) matching videos with videos.

Matching still face images with videos: In many real world applications, such as surveillance, it is required to match still face images with videos for authenticating the identity of individuals. In this experiment, still face images from the MBGC v2 still portal and videos (comprising both walking and activity videos) from the MBGC v2 video challenge database [7] pertaining to 147 subjects are used. To evaluate the efficacy of the proposed algorithm, experiments are performed with 10 times repeated random subsampling (cross validations). In each experiment, training is performed on 47 subjects and the performance in reported on the remaining 100 subjects. This experiment further comprises two different subsets:

- *Matching video probe with still gallery images*: In this experiment, the probe is a video of an individual whose identity is to be matched against a gallery of still face images. In this experiment, the composite ranked list of a probe video is compared with the ranked list computed for each of the gallery images. The ranked list of an image in the gallery is computed by positioning the images retrieved from the dictionary based on their level-1 similarity scores. The experiment is further divided as: 1) probe comprises 618 walking videos pertaining to 100 subjects and 2) probe comprises 513 activity videos pertaining to 100 subjects. In both the cases the gallery consists of 100 still face images, one image per subject.

- *Matching still probe with video gallery*: In this experiment, the probe is a still face image and the gallery comprises videos of individuals. The ranked list of a still probe image is compared with the composite ranked list of each video in the gallery. The experiment is divided as: 1) gallery comprises 100 walking videos and 2) gallery comprises 100 activity videos. In both the cases, the probe comprises 1543 still face images pertaining to 100 subjects.

Matching videos with videos: The proposed algorithm is evaluated for matching video-to-video face information where both gallery and probe comprise videos of individuals. The performance of the proposed algorithm on the MBGC v2 video challenge database is evaluated under three different scenarios, 1) walking vs walking (WW), 2) walking vs activity (WA), and 3) activity vs activity (AA). In the MBGC v2 video challenge protocol, verification experiments are specified by two sets: target and query. The protocol requires the algorithm to match each target (gallery) sequence with all the query (probe) sequences. In this experiment, the composite ranked list of a probe video is compared with the composite ranked lists of the gallery videos.

5.3.3 Results and Analysis

The proposed algorithm utilizes the observation that a discriminative video signature can be computed using a dictionary of still face images. Key results and observations from the experiments are summarized below:

- For both still images and videos, a dictionary of non-overlapping individuals is used to generate discriminative signatures represented as ranked lists of images. The results suggest that the representation based on dictionary is very efficient for matching individuals across large intra-personal variations in videos.

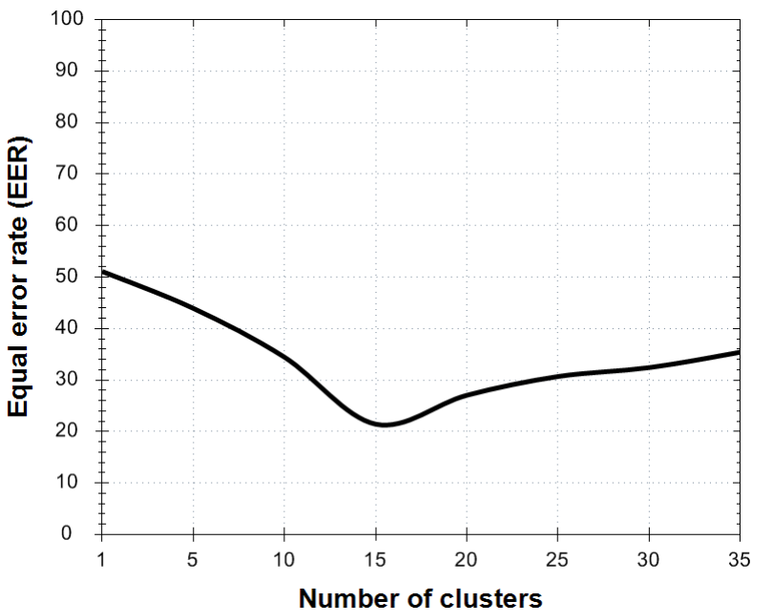

Figure 5.5: Illustrates the variations in equal error rate by varying the number of clusters.

- Figure 5.5 shows that the performance is dependent on the number of clusters. In our experiments, the number of clusters k is varied from 1 to 35. It is observed that all the variations in dictionary images can be broadly grouped into 15 different categories of pose, illumination, and expression. This observation also corroborates with our experiment to empirically determine the number of clusters as shown in Figure 5.5 where $k = 15$ yields the lowest EER. If the number of clusters is less, images are not segregated in the cluster representing the exact viewing conditions. It results in erroneously updating the similarity scores of images based on the reliability of the cluster which increases the error rate. On the other hand, large number of clusters also increases the error rates and the computational cost.

- The proposed algorithm utilizes the taxonomy of facial features to compute the initial ranked lists using computationally efficient level-1 features and further utilizes more discriminative level-2 features to compute the relevance of images in the final composite ranked list. This selection of features for computing the ranked lists and relevance makes the proposed algorithm discriminative and computationally efficient.

5.3.3.1 Results on YouTube database

- The results in Table 5.2 and receiver operating characteristic (ROC) curves in Figure 5.6 demonstrate the performance of the proposed algorithm with benchmark results on the YouTube faces database [6]. The proposed algorithm outperforms existing algorithms and COTS for video-to-video face recognition. The proposed algorithm achieves an average accuracy of 80.7% at EER of 19.4%. The proposed algorithm also achieves a higher area under the curve of 90.5% as compared to other algorithms.

- To evaluate the gain in performance due to clustering, re-ranking, and fusion, the performance of the proposed algorithm is compared when multiple ranked lists are combined using min-max normalization and sum-rule fusion (referred to as MNF). Table 5.2 shows that clustering based re-ranking and fusion reduces the EER by ~9%. This gain can be attributed to the observation that images with similar appearances are clustered together and similarity scores of images are adjusted based on the reliability of the clusters.

- To match two video signatures, a two-sided $nDCG$ measure is used that seamlessly utilizes both level-1 (ranks) and level-2 (relevance) features. The performance gain due to two sided $nDCG$ measure is evaluated by comparing the performance of the

Algorithm	**Verification Accuracy at EER (%)**	**SD (%)**	**AUC (%)**	**EER (%)**
mindst LBP	65.7	1.7	70.7	35.2
mindst FPLBP	65.6	1.8	70.0	35.6
$\|\|U1'U2\|\|$ LBP	65.4	2.0	69.8	36.0
$\|\|U1'U2\|\|$ FPLBP	64.3	1.6	69.4	35.8
MBGS(mean) FPLBP	72.6	2.0	80.1	27.7
MBGS(mean) LBP	76.4	1.8	82.6	25.3
COTS	67.9	2.3	74.1	33.1
MNF	76.4	2.1	81.6	24.3
Schroff *et al.* [198]	77.5	1.6	83.8	23.6
Bhatt *et al.* [189]	78.3	1.7	85.8	21.6
APEM-FUSION [208]	79.1	1.5	86.6	21.4
STFRD+PMML [209]	79.5	2.5	86.6	19.9
VSOF+OSS [210]	79.7	1.8	89.4	20.0
Proposed	**80.7**	**1.4**	**90.5**	**19.4**

Table 5.2: Comparing the proposed algorithm with the benchmark test results and COTS on the YouTube faces database [6].

proposed algorithm when two signatures are matched using the similarity measure used by Schroff *et al.* [198]. Existing approaches only compute the overlap between two lists while discarding other information embedded in the lists, whereas, the results in Table 5.2 show that the two sided $nDCG$ measure reduces the EER by $_{\sim}7\%$.

- Existing approaches that use set-to-set similarities do not consider that multiple frames capture different intra-personal variations. Matching such diverse image sets independently leads to sub-optimal performance. However, the proposed algorithm combines the diverse information from multiple frames to form a composite video signature to match two videos. Figure 5.7 shows some successful and unsuccessful verification examples by the proposed algorithm. Figure 5.8 show the confidence interval for different algorithms proposed for video based face recognition.

- The proposed algorithm has different stages such as computing ranked lists for each video frame, clustering, re-ranking and fusion for combining multiple ranked lists into a discriminative video signature. Finally, two video signatures are matched using two sided $nDCG$ measure. The algorithm takes about 0.06 seconds to compute the ranked list for a single frame, 0.04 seconds to cluster a ranked list, 0.04 seconds for

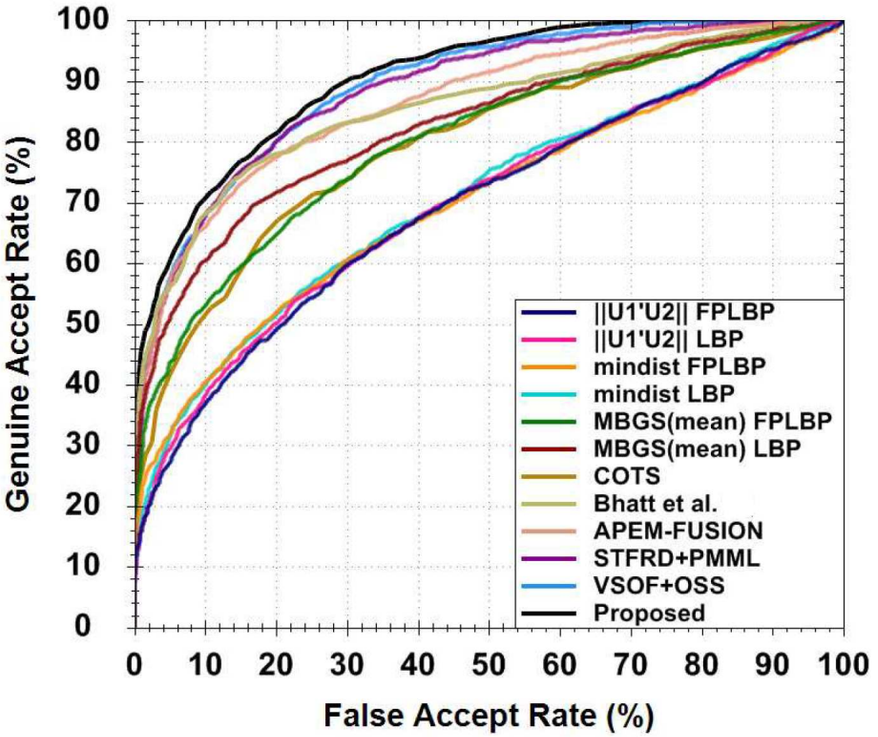

Figure 5.6: ROC curves comparing the performance of the proposed algorithm with benchmark results on the YouTube faces database [6]. (Best viewed in color). The results from the YouTube database website are as of October, 2013.

re-ranking the similarity scores within a ranked list. Further, for computing the signature for a video with 100 frames, fusing 100 ranked lists takes around 1.3 seconds. Therefore, total time to compute a composite ranked list for a video with 100 frames is $100 \times (0.06 + 0.04 + 0.04) + 1.3 = 15.3$ seconds. The time is reported on 2 GHz Intel Duo Core processor with 4 GB RAM under C# programming environment.

5.3.3.2 Results on MBGC v2 database

- Surveillance applications generally require matching an individual in a live-video stream with a watch-list database consisting of still face images. The proposed algorithm can efficiently represent both still face images and videos as ranked lists of still face images from the dictionary. ROC curves in Figure 5.9 show the efficacy of the proposed algorithm for matching both walking and activity videos as probe with still gallery images from the MBGC v2 database. Table 5.3 demonstrates that the proposed algorithm yields at least 1.3% lower equal error rate as compared to

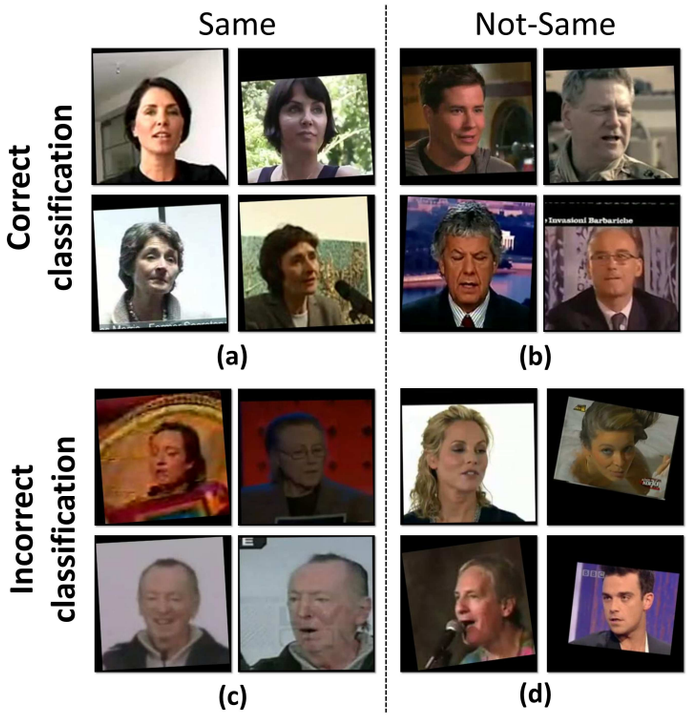

Figure 5.7: Illustrating examples when the proposed algorithm correctly classified (a) 'same', (b) 'not-same' video pairs from the YouTube faces database [6]. Similarly, examples when the proposed algorithm incorrectly classified (c) 'same' and (d) 'not-same' video pairs.

other algorithms for matching video probe with still gallery images.

- Matching a still probe image with video gallery also has a very important law enforcement application when a known individual has be identified at a crime scene using multiple surveillance videos of the crime scene. The results in Figure 5.10 and Table 5.3 demonstrate the efficacy of the proposed algorithm for such scenarios. It yields a lower equal error rate of 17.8% and 20.1% (at least 5.2% lower than other algorithms) for matching still probe images with the gallery consisting of videos of individuals walking or performing some activity from the MBGC v2 database respectively.

- The results in Table 5.4 and Figures 5.11, 5.12, 5.13 show the efficacy of the proposed algorithm for matching unconstrained videos i.e. where the individual is walking or performing some activity. The proposed algorithm outperforms COTS and MNF

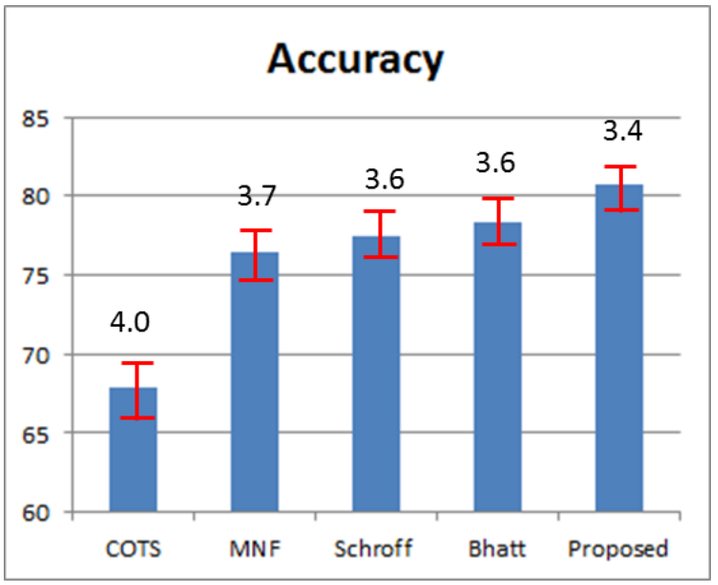

Figure 5.8: Illustrates the confidence interval for different algorithms for video based face recognition on the YouTube faces database.

approach for all the three matching scenarios i.e walking vs walking (WW), walking vs activity (WA), and activity vs activity (AA).

- The proposed algorithm performs better for walking vs walking experiment as compared to the other two scenarios that involve videos of individuals performing some activities. As shown in Figure 5.4, activity videos present a challenging scenario because the quality of facial region is severely deteriorated by the presence of pose, illumination, and expression variations.

- Unlike many existing techniques that are affected by unequal number of frames in two videos, the proposed algorithm mitigates such limitations and can efficiently match two videos regardless of the number of frames in each video. As shown in Table 5.3, the proposed algorithm can also match still face images, analogous to a video with single frame, with videos comprising multiple frames.

5.4 Summary

With advancements in technology, reduction in the cost of sensor (video camera), and several limitations of face recognition from still images in unconstrained scenario, video

Gallery	Probe	Algorithm	Verification Accuracy at EER(%)	SD (%)	AUC (%)	EER (%)
Still images	Walking videos	COTS	73.2	2.1	79.6	27.1
		MNF	78.3	1.7	85.0	22.7
		Proposed	**80.6**	**1.4**	**87.6**	**19.2**
Still images	Activity videos	COTS	68.4	1.9	75.5	31.9
		MNF	76.5	1.8	82.1	24.4
		Proposed	**77.8**	**1.5**	**84.0**	**22.7**
Walking videos	Still images	COTS	70.7	2.2	79.4	29.3
		MNF	77.1	2.0	86.0	23.7
		Proposed	**82.6**	**1.7**	**90.4**	**17.8**
Activity videos	Still images	COTS	69.2	2.0	76.5	31.3
		MNF	74.3	1.7	83.8	25.3
		Proposed	**79.8**	**1.5**	**87.7**	**20.1**

Table 5.3: Comparing the proposed algorithm with COTS and MNF on the MBGC v2 [7] database for matching still face images with videos.

Protocol	Algorithm	Verification Accuracy at EER(%)	AUC (%)	EER (%)
Walking vs walking (WW)	COTS	55.7	59.1	44.3
	MNF	59.9	63.6	40.1
	Proposed	**62.2**	**67.0**	**37.8**
Activity vs walking (AW)	COTS	52.5	53.8	47.5
	MNF	54.4	55.3	45.6
	Proposed	**57.3**	**59.8**	**42.7**
Activity vs activity (AA)	COTS	50.2	50.6	49.8
	MNF	51.5	52.2	48.5
	Proposed	**54.1**	**55.4**	**45.9**

Table 5.4: Comparing the proposed algorithm with COTS and MNF on different protocols of the MBGC v2 video challenge database [7].

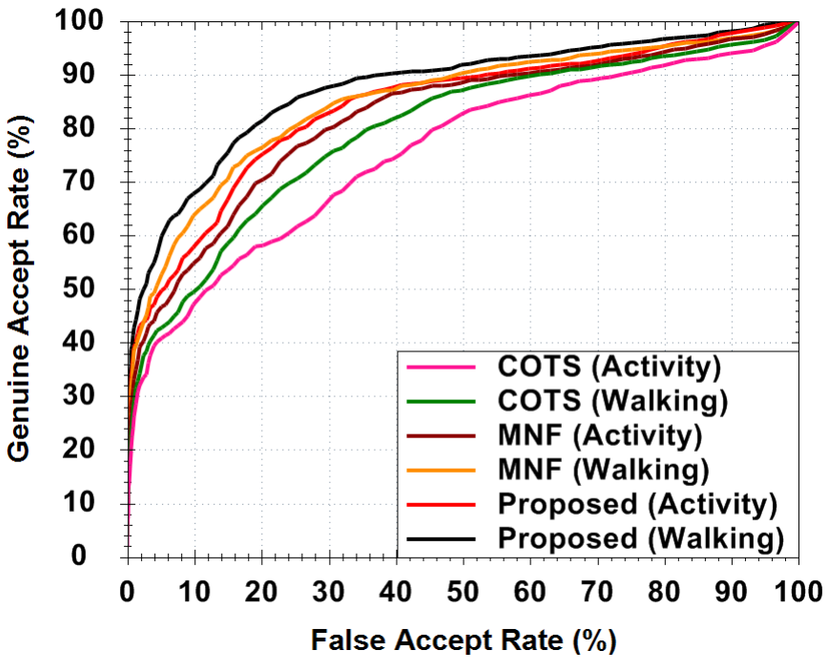

Figure 5.9: ROC curves comparing the performance of the proposed algorithm with COTS and MNF on the MBGC v2 database [7] for matching activity and walking videos with the gallery comprising still face images.

based face recognition has gained significant attention from the research community. Multiple frames in a video depict the temporal and intra-class variations that can be leveraged for efficient face recognition. The proposed video based face recognition algorithm captures the observation that a discriminative video signature can be generated by combining the abundant information available across multiple frames of a video. It assimilates this information as a ranked list of still face images from a large dictionary. It starts with generating a ranked list for every frame in the video using computationally efficient level-1 features. Multiple ranked lists across the frames are then optimized using clustering based re-ranking and finally fused together to generate the video signature. Usefulness (relevance) of images in the video signature is computed using level-2 features. The video signature thus embeds large intra-personal variations across multiple frames which significantly improves the recognition performance. Finally, to match two videos, their video signatures (ordered ranked lists) are compared using a DCG measure that seamlessly utilizes both ranking and relevance of images in the signature. This chapter transforms the problem of video based face recognition into comparing two ordered lists of images. The usability of the proposed algorithm is evaluated under different operating scenarios such as

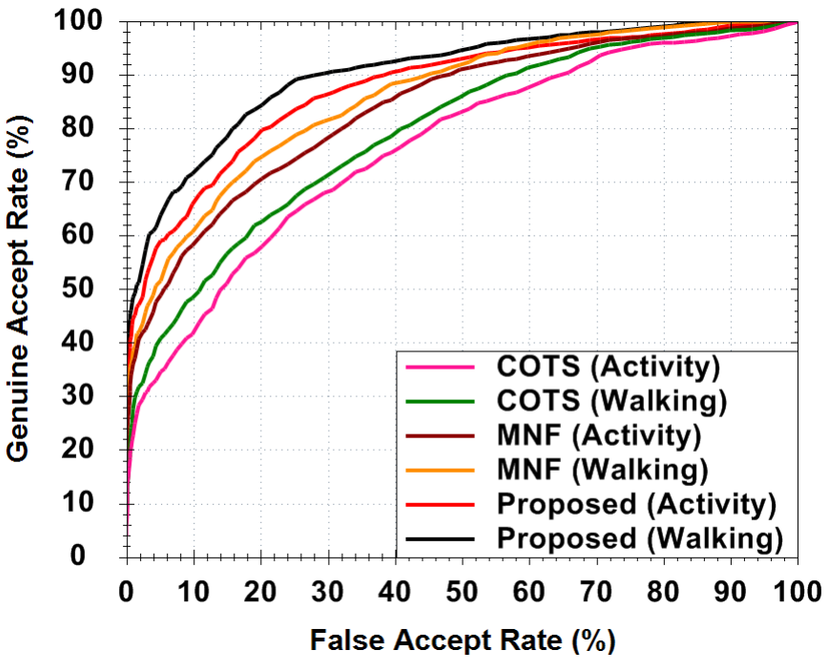

Figure 5.10: ROC curves comparing the performance of the proposed algorithm with COTS and MNF on the MBGC v2 database [7] for matching still face images with gallery comprising activity and walking videos. (Best viewed in color)

matching still images with videos and matching video with videos. Several experiments on unconstrained video databases show that the proposed algorithm consistently outperforms existing algorithms including a commercial face recognition system.

As future research direction, we plan to reduce the computational time of the proposed algorithm. The proposed algorithm utilizes the abundant information in a video to yield better face recognition performance across large variations. Therefore, it requires more computational time as compared to still face recognition algorithms. Multiple frames of a video are processed independently of each other, therefore, processing them in a parallel fashion may enhance the computational efficiency of the proposed approach. Moreover, videos exhibit wide intra-personal variations such as pose, illumination, and expression variations of an individual. Detecting face regions from such unconstrained videos may be challenging and the performance of video based face recognition algorithms is often dependent on the face detection algorithm.

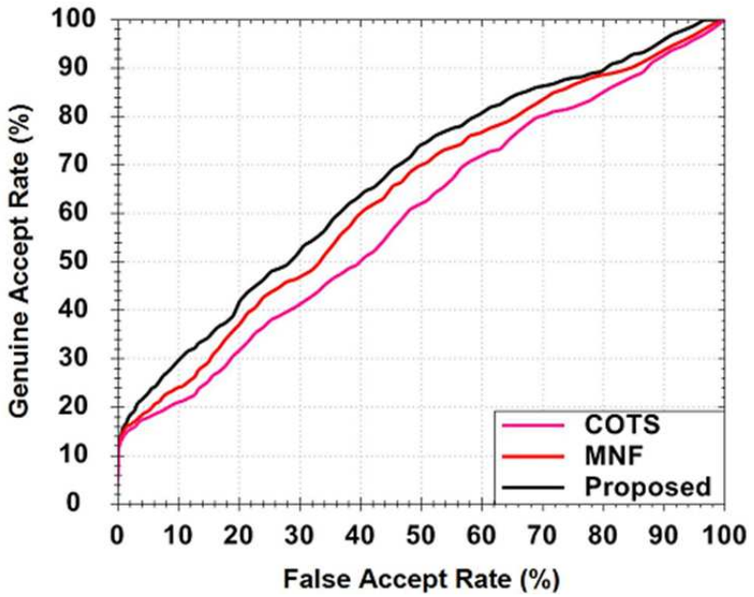

Figure 5.11: ROC curves showing the performance of the proposed algorithm on the MBGC v2 video challenge database [7] for matching walking vs walking (WW).

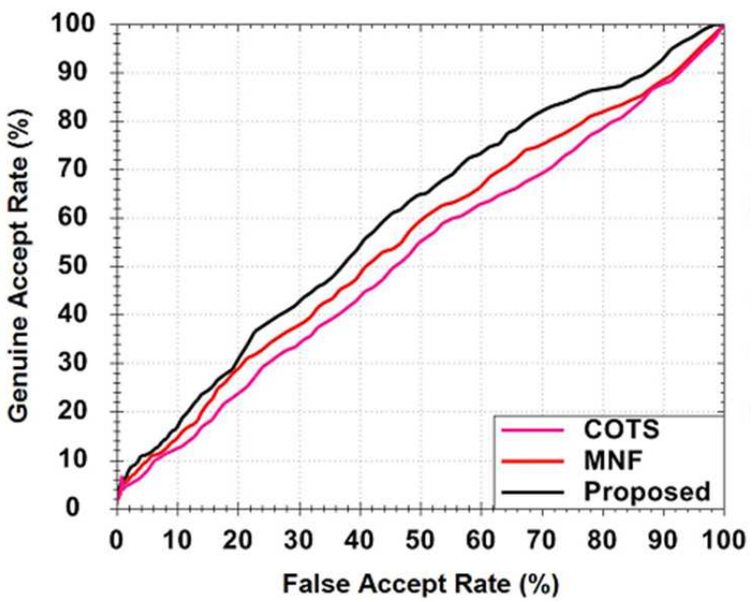

Figure 5.12: ROC curves showing the performance of the proposed algorithm on the MBGC v2 video challenge database [7] for matching walking vs activity (WA).

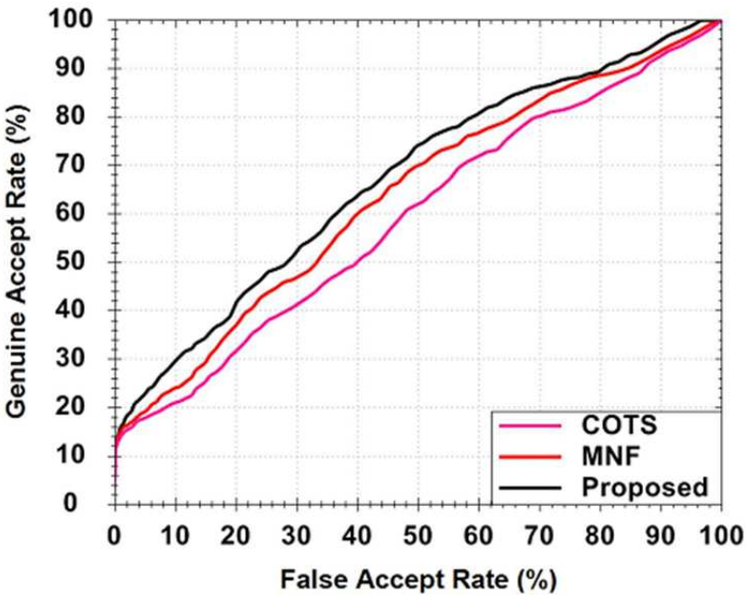

Figure 5.13: ROC curves showing the performance of the proposed algorithm on the MBGC v2 video challenge database [7] for matching activity vs activity (AA) videos (Best viewed in color).

Chapter 6

Conclusions and Future Work

This dissertation formally defines a covariate in face recognition and categorizes several challenges in face recognition as existing and emerging covariates. Figure 6.1 shows the categorization of face recognition techniques based on covariates. Existing covariates of face recognition such as pose, illumination, expression, aging, and disguise have been extensively studied and several algorithms are proposed to mitigate their effect. Apart from existing covariates, the emerging covariates such as matching sketches with digital images, faces altered due to plastic surgery, low resolution face images, and faces from videos are some new research directions in face recognition. The covariates addressed in this dissertation have recently gained attention due to their significance in law enforcement applications. One of the limitations in developing robust solutions for face recognition is the lack of large databases for these emerging covariates. The availability of publicly available large databases will allow better understanding and characterization of these covariates thus leading to better quality solutions. This dissertation presents several algorithms to address these covariates and further instigates multiple research directions.

6.1 Conclusion

In this dissertation, we first developed an automated algorithm for matching forensic sketches with digital face images. The algorithms starts by enhancing the quality of forensic sketches and digital face images to eliminate distortions and noise introduced due to the excessive use of charcoal pencil, paper quality, and scanning (device noise/errors). A multi-scale circular weber's local descriptor (MCWLD) is proposed for encoding discriminative micro patterns from local regions of sketches and digital face images at multiple scales. Further, a memetic algorithm is developed to assign optimal weights to different local regions of face for matching using weighted χ^2 distance. We also evaluated

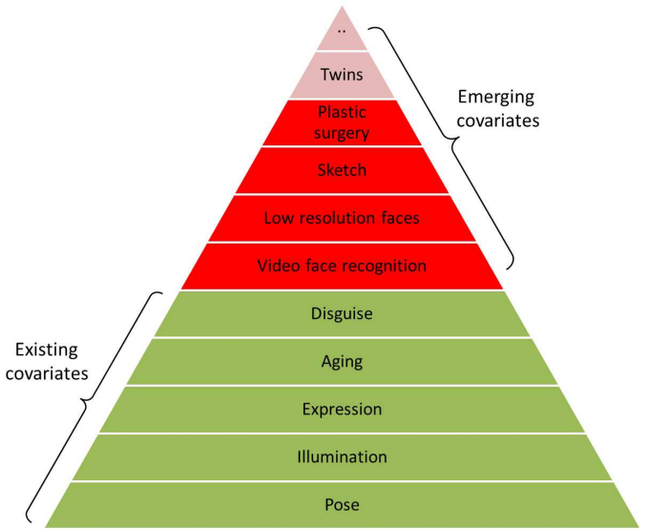

Figure 6.1: Progression in face recognition with respect to different covariates.

human performance for matching sketches with digital face images and found that the information collected from individuals corroborate with our observation that local regions provide discriminating information for efficient face recognition. Comprehensive experimental evaluation on different sketch databases show that the proposed algorithm yields better identification performance compared to existing face recognition algorithms and two commercial face recognition systems. Finally, we prepared a sketch database, namely IIIT-D sketch database [51], that comprises viewed, semi-forensic, and forensic sketches for instigating further research in understanding progression from matching viewed to semi-forensic to forensic sketches.

With widespread popularity and acceptability of plastic surgery procedures, it is imperative for face recognition algorithms to efficiently match pre-and post-surgery images. We developed a multi-objective evolutionary algorithm for matching face images altered due to plastic surgery. The algorithms first generates a set of 40 non-disjoint face granules of varying shapes and sizes. Scale invariant feature transform (SIFT) and extended uniform circular local binary patterns (EUCLBP) features are extracted from different face granules and are selectively combined using evolutionary genetic algorithm. The multi-objective genetic algorithm simultaneously optimizes feature selection and weight assignment for different face granules. The evolutionary selection of feature extractor allows switching between two feature extractors (SIFT and EUCLBP) and helps in encoding discriminatory information for each face granule. We analyzed the effect of different types of plastic surgery procedures (i.e. local and global plastic surgery) and the performance

of individual face granules. Experimental evaluation under different protocols, including large scale matching, on the IIIT-D plastic surgery database [8] show that the proposed algorithm outperforms existing algorithms including a commercial system when matching surgically altered face images.

A very important law enforcement application is performing face recognition from low resolution surveillance quality images. Face recognition algorithms are generally trained for matching high resolution images and the performance is severely compromised when it encounters a situation where a low resolution probe is matched with a high resolution gallery image. We pose the problem of cross-resolution face matching as a transfer learning problem and propose a co-transfer learning framework. To facilitate knowledge transfer with probe instances in the target domain, a co-training algorithm is developed which assigns pseudo labels to the unlabeled probe instances. Cross-pollination of these two paradigms in the proposed framework enhances the performance of cross-resolution face recognition. Experiments are performed on four publicly available databases, namely, CMU Multi-PIE [9], ChokePoint [11], SCface [10], and MBGC v2 [211] databases. The performance evaluation with existing, super-resolution and commercial face recognition algorithms show the efficacy of the proposed co-transfer learning algorithm for cross-resolution face matching.

Videos provide abundant information in terms of multiple frames which capture wide intra-personal variations of an individual. This abundant information can be leveraged for efficient face recognition. This dissertation proposes a video based face recognition algorithm which computes a discriminative video signature as an ordered list of still face images. The algorithm begins by comparing each video frame with all still face images in the dictionary and generates an ordered list in which each image from the dictionary is ranked based on its similarity to the input frame. Multiple ordered lists across different video frames are combined into a composite ranked list using clustering based re-ranking and fusion algorithm. The final composite list constitutes the video signature and minimizes the distance from all the ordered lists corresponding to multiple video frames. To match two videos, their composite video signatures are compared using a normalized discounted cumulative gain measure. The nDCG measure encodes both rank in the ordered list as well as usefulness of images for characterizing the individual in the video. Experimental evaluation on the YouTube faces [6] and the MBGC v2 [211] databases under different video based face recognition scenarios such as matching still face images with videos and matching videos with videos show that the proposed algorithm outperforms existing algorithms including a commercial face recognition system.

6.2 Future Work

This dissertation offers several algorithms for emerging covariates of face recognition; however, it also instigates some future research directions for making face recognition robust and scalable. We conclude this dissertation with some possible future research directions that can be explored for addressing the emerging covariates of face recognition.

- This dissertation presents a classification of different challenges in face recognition as existing and emerging covariates. This classification is based on the maturity of a covariate in terms of how extensively it has been studied in the literature. A possible future research direction could be to have a taxonomy of these covariates beyond simple existing and emerging covariates.

- Matching sketches with digital face images has been one of the most important cues in apprehending criminals, finding missing individuals, and recognizing individuals when the face is reconstructed as a composite sketch post-mortem. It has gained significant attention from the research community and several algorithms have been proposed for matching sketches with digital face images. However, law enforcement agencies are progressively shifting from hand-drawn sketches to composite sketches which are drawn using software tools. These tools facilitates an eyewitness to select the most resembling facial template for each feature based on his/her recollection from the crime scene. Preparing composite sketches require less effort both in terms of cost as well as time as compared to hand-drawn sketches. However, the problem of matching composite sketches with digital face images is not limited to the effects of variations in composite sketches and digital face images. In real world scenarios, it is often required to match composite sketches and digital face images with age variations; e.g., in cases for finding missing individuals and recognizing individuals when the face is reconstructed as a composite sketch after death. Age variations further make this problem arduous as it changes the structural geometry and face texture. Therefore, matching composite sketches with digital face images across age variations is an important research direction. We believe that large databases with composite sketches and digital face images with age variations will lead to better understanding of the problem.

- The allure for plastic surgery is experienced worldwide and is driven by factors such as the availability of advanced technology, affordable cost and the speed with which

these procedures are performed. According to the statistics provided by the American society of aesthetic plastic surgery [101], more and more individuals are expected to undergo facial plastic surgery for cosmetic and medical reasons. Therefore, it is imperative for face recognition algorithms to be robust for matching face images altered due to plastic surgery. This dissertation presents an efficient algorithm for matching pre- and post-surgery images, however, the results and analysis inspire further research in this important area. Face recognition algorithms should be capable of automatically detecting facial regions that have variations possibly due to plastic surgery. Understanding the effects of plastic surgery in thermal-infrared imagery can be one of the possible future research direction. The research in plastic surgery is primarily focussed around a single publicly available database, IIIT-D plastic surgery database [8]. Preparing large scale databases for different types of plastic surgery procedures in visible as well as thermal-infrared imagery will lead to better understanding of the non-linear variations introduced due to plastic surgery.

- The generality of face recognition has lead to several challenging applications such as matching low resolution images from surveillance cameras. Surveillance images serve as the primary evidence in leading the investigation and recognizing the individuals at the end. It is therefore desirable to build a system where surveillance cameras coupled with a face recognition algorithm can be used to automatically identify individuals from a watch-list. The progression in face recognition has made it possible to recognize low resolution surveillance face images against watch-list database [212] to an acceptable level. However, these efforts could not foil any of the anti-social activities. One of the possible future research directions is to develop a real time face recognition algorithm coupled with surveillance system [213] that can upfront raise an alarm by identifying individuals who have committed crime or with the intent to commit crime.

- Video based face recognition has gained significant attention due to limitations of still images in addressing the wide intra-personal variations of face in many real world applications. Unlike still face images, videos provide abundant information that can be leveraged to compensate for these variations and enhance the performance of face recognition. It is our belief that videos have the potential to address face recognition in uncontrolled and unconstrained environments. The results presented in this dissertation encourages further research in video based face recognition i.e. still to video matching and video-to-video matching. Video based face recognition for identifying

an individual from a video against a watch-list requires open-set identification. One of the possible future directions, often required by law enforcement agencies, is video based face recognition in open-set scenarios where the existing research is very limited. Another possible research direction is to combine other modalities such as iris, voice, and gait for more robust identification from videos.

- Real world applications require efficient video based face recognition techniques that can identify individuals from videos captured through surveillance cameras. Currently, surveillance cameras help law enforcement agencies in tracking the activities of individuals or identifying them using manual intervention. However, an efficient low resolution face recognition system coupled with surveillance cameras can significantly speed up the accuracy and speed of this process. Therefore, developing low resolution face recognition algorithms for videos can be one of the interesting future research directions.

Appendix

Appendix A

Dissemination of Research Results

1. H.S. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa, Plastic Surgery and Face Recognition, Encyclopedia of Biometrics, 2nd Edition, 2014.

2. H.S. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa, Recognizing Surgically Altered Face Images with Multi-objective Evolutionary Approach, IEEE Transaction on Information Forensics and Security, vol. 8, no. 1, pp. 89-100, 2013.

3. H.S. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa, Memetic Approach for Matching Sketches with Digital Face Images, IEEE Transaction on Information Forensics and Security vol. 7, no. 5, pp. 1522-1535, 2012.

4. R. Singh, M. Vatsa, H.S. Bhatt, S. Bharadwaj, A. Noore, and S. S. Nooreyezdan, Plastic Surgery: A New Dimension to Face Recognition, IEEE Transaction on Information Forensics and Security, vol. 5, no. 3, pp. 441-448, 2010.

5. H.S. Bhatt, R. Singh, and M. Vatsa, On Combining Multiple Evidences in Videos for Face Recognition using Dictionary of Still Face Images, IEEE Transactions on Information Forensics and Security. (Under Review)

6. T. Chugh, H.S. Bhatt, R. Singh, and M. Vatsa, Age Separated Composite Sketches and Digital Face Images, In Proceedings of International Conference on Biometrics: Theory, Applications and Systems (BTAS), pp. 1-6, 2013.

7. H.S. Bhatt, R. Singh, and M. Vatsa, On Rank Aggregation for Face Recognition from Videos, In Proceedings of International Conference on Image Processing (ICIP), pp. 2993-2997, 2013.

8. H.S. Bhatt, R. Singh, and M. Vatsa, Can Combining Demographics and Biometrics Improve De-duplication Performance?, IEEE Computer Society and IEEE Biometrics Council Workshop on Biometrics (CVPR), pp. 188-193, 2013.

9. H.S. Bhatt, R. Singh, M. Vatsa, and N.K. Ratha, Cross-resolution Face Matching using Co-transfer Learning, In Proceedings of International Conference on Image Processing (ICIP), pp. 1453-1456, 2012.

10. H.S. Bhatt, S. Bharadwaj, R. Singh, M. Vatsa, A. Noore, and A. Ross, On Co-training Online Biometric Classifiers, In Proceedings of International Joint Conference on Biometrics (IJCB), pp. 1-4, 2011.

11. H.S. Bhatt, S. Bharadwaj, M. Vatsa, R. Singh, A. Noore, and A. Ross, Quality Driven Biometric Classifier Selection Framework for Improved Performance, In Proceedings of International Joint Conference on Biometrics (IJCB), pp. 1-7, 2011.

12. S. Bharadwaj, H.S. Bhatt, M. Vatsa, R. Singh, Quality Assessment based Denoising to Improve Face Recognition Performance, In IEEE Computer Society and IEEE Biometrics Council Workshop on Biometrics (CVPR), pp. 140-145, 2011.

13. H.S. Bhatt, S. Bharadwaj, R. Singh, M. Vatsa, Evolutionary Granular Approach for Recognizing faces altered due to Plastic Surgery, In International Conference on Automatic Face and Gesture Recognition (F&G), 2011.

14. H.S. Bhatt, S. Bharadwaj, R. Singh, M. Vatsa, On Matching Sketches with Digital Face Images, In International Conference on Biometrics: Theory, Applications and Systems (BTAS), pp. 1-8, 2010.

15. S. Bharadwaj, H.S. Bhatt, R. Singh, M. Vatsa, Periocular Biometrics: When Iris Recognition Fails, In International Conference on Biometrics: Theory, Applications and Systems (BTAS), pp. 1-6, 2010.

16. S. Bharadwaj, H.S. Bhatt, R. Singh, M. Vatsa, Face Recognition for Newborns: A Preliminary Study, In International Conference on Biometrics: Theory, Applications and Systems (BTAS), pp. 1-6, 2010.

17. M. Vatsa, R. Singh, S. Bharadwaj, H.S. Bhatt and A. Noore, Matching Digital and Scanned Face Images with Age Variation, In International Conference on Biometrics: Theory, Applications and Systems (BTAS), pp. 1-6, 2010.

18. M. Vatsa, R. Singh, S. Bharadwaj, H.S. Bhatt, R. Mashruwala, Analyzing Fingerprints of Indian Population Using Image Quality For Large Scale Application: A UIDAI Case Study, In International Conference on Pattern Recognition (ICPR) - Emerging Techniques and Challenges for Hand-based Biometrics, pp. 1-5, 2010.

19. H.S. Bhatt, S. Bharadwaj, R. Singh, M. Vatsa, Face Recognition and Plastic Surgery: Ethical, Social and Engineering Challenges, In International Conference on Ethics and Policy of Biometrics and International Data Sharing (ICEB), pp. 70-75, 2010.

Appendix B

Error Bounds for the Ensemble

Using the square loss function $l^*(z,y) = (z-y)^2$ and the exponential weighting update function, bounds of an ensemble are given as:

$$\sum_{i=1}^{I} l^*(w_i^S \Pi(C_i^S) + w_i^T \Pi(C_i^T), \Pi(y_i)) \leq 2ln(2) \tag{1}$$
$$+min\{\sum_{i=1}^{I} l^*((\Pi(C_i^S), \Pi(y_i)), \sum_{i=1}^{I} l^*(\Pi(C_i^T), \Pi(y_i))\}$$

The above equation is derived by following the proof in [164]. Using this, the error bounds of an ensemble are derived as follows: The error at the i^{th} step is represented as $|w_i^S \Pi(C_i^S) + w_i^T \Pi(C_i^T) - \Pi(y_i)| \geq \frac{1}{2}$. Therefore, we have

$$\sum_{i=1}^{I} l^*(w_i^S \Pi(C_i^S) + w_i^T \Pi(C_i^T), \Pi(y_i)) \tag{2}$$
$$= \sum_{i=1}^{I} (w_i^S \Pi(C_i^S) + w_i^T \Pi(C_i^T), \Pi(y_i))^2 \geq \frac{1}{4} M$$

Combining Eqs. 1 and 2, we have

$$\frac{1}{4} M \leq min\{\sum C^S, \sum C^T\} + 2ln(2) \tag{3}$$

where $\sum C^S = \sum_{i=1}^{I} l^*(\Pi(C_i^S), \Pi(y_i))$ and $\sum C^T = \sum_{i=1}^{I} l^*(\Pi(C_i^T), \Pi(y_i))$

Bibliography

[1] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikainen, X. Chen, and W. Gao, "WLD: A robust local image descriptor," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1705–1720, 2010. 1, 24, 29, 30, 31, 39, 40, 42

[2] L. Gibson, *Forensic Art Essentials*. Elsevier, 2008. 1, 37, 38

[3] K. Taylor, *Forensic Art and Illustration*. CRC Press, 2001. 1, 37, 38

[4] B. Klare, L. Zhifeng, and A. Jain, "Matching forensic sketches to mug shot photos," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, pp. 639–646, 2011. 2, 18, 23, 24, 25, 39, 40, 42, 43, 44, 46, 48, 51, 52

[5] K. Anderson and P. McOwan, "Robust real-time face tracker for cluttered environments," *Computer Vision and Image Understanding*, vol. 95, pp. 184–200, 2004. 3, 66

[6] L. Wolf, T. Hassner, and I. Maoz, "Face recognition in unconstrained videos with matched background similarity," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2011, pp. 529–534. 6, 8, 13, 14, 129, 130, 141, 143, 146, 147, 148, 149, 159

[7] MBGC-V2, Available at http://www.nist.gov/itl/iad/ig/mbgc.cfm. 6, 8, 14, 95, 102, 105, 111, 113, 115, 129, 130, 141, 144, 151, 152, 153, 154, 155

[8] R. Singh, M. Vatsa, H. Bhatt, S. Bharadwaj, A. Noore, and S. Nooreyezdan, "Plastic surgery: A new dimension to face recognition," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 3, pp. 441 –448, sep. 2010. 7, 14, 18, 59, 60, 61, 63, 72, 73, 81, 82, 159, 161

[9] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," *Image and Vision Computing*, vol. 28, no. 5, pp. 807–813, 2010. 8, 13, 14, 95, 102, 104, 105, 111, 133, 159

[10] M. Grgic, K. Delac, and S. Grgic, "SCface - surveillance cameras face database," *Multimedia Tools and Applications*, vol. 51, no. 3, pp. 863–879, 2011. 8, 14, 95, 102, 104, 111, 112, 115, 159

[11] 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 *Proceedings of International Conference on Computer Vision and Pattern Recognition Workshops*, 2011, pp. 74–81. 8, 14, 95, 102, 104, 111, 112, 115, 159

[12] A. J. O'Toole, P. J. Phillips, F. Jiang, J. Ayyad, N. Penard, and H. Abdi, "Face recognition algorithms surpass humans matching faces over changes in illumination," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 9, pp. 1642–1646, 2007. 9

[13] S. Z. Li and A. K. Jain, *Handbook of Face Recognition, 2nd Edition*. Springer, 2011. 10, 127

[14] M. Yang, D. Kriegman, and N. Ahuja, "Detecting faces in images: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34–58, 2002. 11

[15] H. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 23–38, 1998. 11

[16] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2001, pp. 511–518. 11

[17] C. Zhang and Z. Zhang, "A survey of recent advances in face detection," Tech. Rep. MSR-TR-2010-66, 2010. 11

[18] B. Klare and A. Jain, "On a taxonomy of facial features," in *Proceedings of International Conference on Biometrics: Theory, Applications and Systems*, 2010, pp. 1–8. 11

[19] H. Lamba, A. Sarkar, M. Vatsa, and R. Singh, "Face recognition for look-alikes: A preliminary study," in *Proceedings of International Joint Conference on Biometrics*, 2011, pp. 1–6. 12, 14, 18

[20] B. Klare, A. Paulino, and A. Jain, "Analysis of facial features in identical twins," in *Proceedings of International Joint Conference on Biometrics*, 2011, pp. 1–6. 12, 18

[21] P. J. Phillips, P. J. Flynn, K. W. Bowyer, R. W. V. Bruegge, P. J. Grother, G. W. Quinn, and M. Pruitt, "Distinguishing identical twins by face recognition," in *Proceedings of International Conference on Automatic Face Gesture Recognition and Workshops*, 2011, pp. 185–192. 12, 14, 18

[22] K. Bowyer, "What surprises do identical twins have for identity science?" *IEEE Computer*, vol. 44, no. 7, pp. 100–102, 2011. 12, 18

[23] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, 2003. 12, 16

[24] S. Kong, J. Heo, B. Abidi, J. Paik, and M. Abidi, "Recent advances in visual and infrared face recognition: A review," *Computer Vision and Image Understanding*, vol. 97, no. 1, pp. 103–135, 2005. 12

[25] P. Belhumeur, "Ongoing challenges in face recognition," in *Frontiers of Engineering: Reports on Leading-Edge Engineering*, 2006, pp. 5–14. 12, 13, 16

[26] A. Abate, M. Nappi, D. Riccio, and G. Sabatino, "2D and 3D face recognition: A survey," *Pattern Recognition Letters*, vol. 28, no. 14, pp. 1885–1906, 2007. 13

[27] X. Zhang and Y. Gao, "Face recognition across pose: A review," *Pattern Recognition*, vol. 42, no. 11, pp. 2876–2896, 2009. 13, 16

[28] X. Zou, J. Kittler, and K. Messer, "Illumination invariant face recognition: A survey," in *Proceedings of International Conference on Biometrics: Theory, Applications, and Systems*, 2007, pp. 1–8. 13, 16

[29] Y. Fu, G. Guo, and T. S. Huang, "Age synthesis and estimation via faces: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 1955–1976, 2010. 13

[30] A. Jain, B. Klare, and U. Park, "Face recognition: Some challenges in forensics," in *Proceedings of International Conference on Automatic Face Gesture Recognition and Workshops*, 2011, pp. 726–733. 13

[31] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression database," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1615–1618, 2003. 13, 14, 130

[32] FG-Net, Available at http://www.fgnet.rsunit.com/. 13, 14

[33] K. Ricanek and T. Tesafaye, "MORPH: A longitudinal image database of normal adult age-progression," in *Proceedings of International Conference on Automatic Face and Gesture Recognition*, 2006, pp. 341–345. 13, 14

[34] G. Huang, M. Ramesh, T. Berg, and E. Leonard, "Labeled faces in the wild : A database for studying face recognition in unconstrained environment," University of Massachusetts, Amherst, Tech. Rep. 07-49, 2007. 13, 14

[35] N. Kumar, A. Berg, P. Belhumeur, and S. Nayar, "Attribute and Simile Classifiers for Face Verification," in *Proceedings of International Conference on Computer Vision*, 2009, pp. 365–372. 13, 14

[36] ORL, Available at http://www.cl.cam.ac.uk/research/dtg/attarchive/ facedatabase.html. 14

[37] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, "XM2VTSDB : The extended M2VTS database," in *Proceedings of International Conference on Audio and Video-based Biometric Personal Verification*, 1999, pp. 72–77. 14

[38] P. J. Phillips, H. Moon, S. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000. 14

[39] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997. 14

[40] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, and D. Zhao, "The CAS-PEAL large-scale chinese face database and baseline evaluations," *IEEE Transactions on Systems, Man and Cybernetics-A*, vol. 38, no. 1, pp. 149–161, 2008. 14

[41] A. Martinez and R. Benevento, "The AR face database," *CVC Technical Report #24*, 1998. 14, 36

[42] CASIA-FaceV5, Available at http://biometrics.idealtest.org/. 14

[43] S. Li, R. Chu, S. Liao, and L. Zhang, "Illumination invariant face recognition using near-infrared images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 627–639, 2007. 14

[44] B. Zhang, L. Zhang, D. Zhang, and L. Shen, "Directional binary code with application to PolyU near-infrared face database," *Pattern Recognition Letters*, vol. 31, no. 14, pp. 2337–2344, 2010. 14

[45] M. Lyons, J. Budynek, and S. Akamatsu, "Automatic classification of single facial images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 12, pp. 1357–1362, 1999. 14

[46] P. Lucey, J. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The Extended Cohn-Kanade Dataset: A complete dataset for action unit and emotion-specified expression," in *Proceedings of International Conference on Computer Vision and Pattern Recognition Workshops*, 2010, pp. 94–101. 14

[47] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, "A 3D facial expression database for facial behavior research," in *International Conference on Automatic Face and Gesture Recognition*, 2006, pp. 211–216. 14

[48] X. Liu, T. Chen, and S. M. Thornton, "Eigenspace updating for non-stationary process and its application to face recognition," *Pattern Recognition*, vol. 36, no. 9, pp. 1945–1959, 2003. 14

[49] X. Wang and X. Tang, "Face photo-sketch synthesis and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 11, pp. 1955–1967, 2009. 14, 22, 24, 36

[50] W. Zhang, X. Wang, and X. Tang, "Coupled information-theoretic encoding for face photo-sketch recognition," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2011. 14, 23

[51] H. S. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa, "Memetically optimized MCWLD for matching sketches with digital face images," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 5, pp. 1522–1535, 2012. 14, 18, 23, 24, 158

[52] A. Dantcheva, C. Chen, and A. Ross, "Can facial cosmetics affect the matching accuracy of face recognition systems?" in *Proceedings of International Conference on Biometrics: Theory, Applications and Systems*, 2012, pp. 391–398. 14

[53] C. Chen, A. Dantcheva, and A. Ross, "Automatic facial makeup detection with application in face recognition," in *Proceedings of International Conference on Biometrics*, 2013, pp. 1–8. 14

[54] V. Vijayan, K. Bowyer, and P. Flynn, "3D twins and expression challenge," in *Proceedings of International Conference on Computer Vision Workshops*, 2011, pp. 2100–2105. 14, 18

[55] K. Lee, J. Ho, M. Yang, and D. Kriegman, "Visual tracking and recognition using probabilistic appearance manifolds," *Computer Vision and Image Understanding*, vol. 99, no. 3, pp. 303–331, 2005. 14

[56] K. Lee and D. Kriegman, "Online learning of probabilistic appearance manifolds for video-based recognition and tracking," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2005, pp. 852–859. 14

[57] Z. Huang, S. Shan, H. Zhang, S. Lao, A. Kuerban, and X. Chen, "Benchmarking still-to-video face recognition via partial and local linear discriminant analysis on COX-S2V dataset," in *Proceedings of Asian Conference on Computer Vision*, 2013, pp. 589–600. 14

[58] N. Ramanathan and R. Chellappa, "Face verification across age progression," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3349–3362, 2006. 16

[59] ——, "Modeling age progression in young faces," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2006, pp. 387–394. 16

[60] U. Park, Y. Tong, and A. Jain, "Age-invariant face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 947–954, 2010. 16

[61] N. Ramanathan, R. Chellappa, and A. K. Roy Chowdhury, "Facial similarity across age, disguise, illumination and pose," in *Proceedings of International Conference on Image Processing*, 2004, pp. 1999–2002. 16

[62] R. Singh, M. Vatsa, and A. Noore, "Face recognition with disguise and single gallery images," *Image and Vision Computing*, vol. 27, no. 3, pp. 245–257, 2009. 16

[63] S. Biswas, K. W. Bowyer, and P. J. Flynn, "Multidimensional scaling for matching low-resolution face images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 10, pp. 2019–2030, 2012. 18, 92, 108, 110, 116, 117

[64] H. S. Bhatt, R. Singh, M. Vatsa, and N. K. Ratha, "Matching cross-resolution face images using co-transfer learning," in *Proceedings of International Conference on Image Processing*, 2012, pp. 1453–1456. 18

[65] Z. Lei, S. Liao, A. K. Jain, and S. Z. Li, "Coupled discriminant analysis for heterogeneous face recognition," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 6, pp. 1707–1716, 2012. 18, 94

[66] P. H. Hennings-Yeomans, S. Baker, and V. Bhagavatula, "Simultaneous super-resolution and feature extraction for recognition of low-resolution faces," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8. 18, 92, 93

[67] S. Shekhar, V. M. Patel, and R. Chellappa, "Synthesis-based recognition of low resolution faces," in *Proceedings of International Joint Conference on Biometrics*, 2011, pp. 1–6. 18, 92, 94

[68] S. Biswas, K. W. Bowyer, and P. J. Flynn, "A study of face recognition of identical twins by humans," in *Proceedings of International Workshop on Information Forensics and Security*, 2011, pp. 1–6. 18

[69] X. Tang and X. Wang, "Face photo recognition using sketch," in *Proceedings of International Conference on Image Processing*, 2002, pp. 257–260. 22, 24

[70] ——, "Face sketch synthesis and recognition," in *Proceedings of International Conference on Computer Vision*, 2003, pp. 687–694. 22

[71] Q. Liu, X. Tang, H. Jin, H. Lu, and S. Ma, "A nonlinear approach for face sketch synthesis and recognition," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2005, pp. 1005–1010. 22, 24

[72] L. Yung-hui, M. Savvides, and V. Bhagavatula, "Illumination tolerant face recognition using a novel face from sketch synthesis approach and advanced correlation filters," in *Proceedings of International Conference on Acoustics, Speech and Signal Processing*, vol. 2, 2006. 22

[73] B. Xiao, X. Gao, D. Tao, and X. Li, "A new approach for face recognition by sketches in photos," *Signal Processing*, vol. 89, no. 8, pp. 1576–1588, 2009. 22

[74] W. Zhang, X. Wang, and X. Tang, "Lighting and pose robust face sketch synthesis," in *Proceedings of European Conference on Computer Vision*, 2010, pp. 420–433. 22

[75] A. Sharma and D. W. Jacobs, "Bypassing synthesis: PLS for face recognition with pose, low-resolution and sketch," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2011, pp. 593–600. 22

[76] R. Uhl and N. Lobo, "A framework for recognizing a facial image from a police sketch," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 1996, pp. 586–593. 22

[77] P. Yuen and C. Man, "Human face image searching system using sketches," *IEEE Transactions on Systems, Man and Cybernetics - A*, vol. 37, no. 4, pp. 493–504, 2007. 22

[78] Y. Zhang, C. McCullough, J. Sullins, and C. Ross, "Human and computer evaluations of face sketches with implications for forensic investigations," in *Proceedings of International Conference on Biometrics: Theory, Applications and Systems*, 2008, pp. 1–7. 23

[79] ——, "Hand-drawn face sketch recognition by humans and a PCA-based algorithm for forensic applications," *IEEE Transactions on Systems, Man and Cybernetics - A*, vol. 40, no. 3, pp. 475–485, 2010. 23, 42

[80] H. Nizami, J. P. Adkins-Hill, Y. Zhang, J. Sullins, C. McCullough, S. Canavan, and L. Yin, "A biometric database with rotating head videos and hand-drawn face sketches," in *Proceedings of International Conference on Biometrics: Theory, Applications and Systems*, 2009, pp. 38–43. 23

[81] H. Nejati and T. Sim, "A study on recognizing non-artistic face sketches," in *Proceedings of Workshop on Applications of Computer Vision*, 2011, pp. 240–247. 23

[82] H. Nejati, T. Sim, and E. Martinez-Marroquin, "Do you see what i see? A more realistic eyewitness sketch recognition," in *Proceedings of International Joint Conference on Biometrics*, 2011, pp. 1–8. 23

[83] B. Klare and A. Jain, "Sketch-to-photo matching: A feature-based approach," in *Proceedings of Society of Photo-Optical Instrumentation Engineers Conference Series*, 2010. 23, 24, 29, 39, 40, 42, 44, 46

[84] B. F. Klare and A. K. Jain, "Heterogeneous face recognition using kernel prototype similarities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1410–1422, 2013. 23

[85] Y. Zhang, S. Ellyson, A. Zone, P. Gangam, J. Sullins, C. McCullough, S. Canavan, and L. Yin, "Recognizing face sketches by a large number of human subjects: A perception-based study for facial distinctiveness," in *Proceedings of International Conference on Automatic Face and Gesture Recognition*, 2011. 23, 52

[86] H. S. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa, "On matching sketches with digital face images," in *Proceedings of International Conference on Biometrics: Theory, Applications and Systems*, 2010, pp. 1–7. 23, 24, 29, 36, 39, 40, 42, 44, 46, 62, 67, 76

[87] S. Canavan, X. Zhang, L. Yin, and Y. Zhang, "3D face sketch modeling and assessment for component based face recognition," in *Proceedings of International Joint Conference on Biometrics*, 2011, pp. 1–6. 24

[88] H. K. Galoogahi and T. Sim, "Face sketch recognition by local radon binary pattern: LRBP," in *Proceedings of International Conference on Image Processing*, 2012, pp. 1837–1840. 24

[89] I. Daubechies, *Ten lectures on wavelets*, 1st ed. Society for Industrial and Applied Mathematics, 1992. 25

[90] Z. Rahman, D. Jobson, and G. Woodell, "Multi-scale retinex for color image enhancement," in *Proceedings of International Conference on Image Processing*, 1996, pp. 1003–1006. 26

[91] S. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1532–1546, 2000. 26

[92] T. Ahonen, A.Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006. 27, 33, 70

[93] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004. 29, 62, 68, 76, 102, 103, 110

[94] C. Geng and X. Jiang, "Face recognition using SIFT features," in *Proceedings of International Conference on Image Processing*, 2009, pp. 3313–3316. 29

[95] P. Sinha, B. J. Balas, Y. Ostrovsky, and R. Russell, "Face recognition by humans: 19 results all computer vision researchers should know about," *Proceedings of IEEE*, vol. 94, no. 11, pp. 1948–1962, 2006. 32, 42, 63, 68

[96] N. Krasnogor and J. Smith, "A tutorial for competent memetic algorithms: model, taxonomy, and design issues," *IEEE Transactions on Evolutionary Computation*, vol. 9, no. 5, pp. 474–488, 2005. 32

[97] H. Wang, D. Wang, and S. Yang, "A memetic algorithm with adaptive hill climbing strategy for dynamic optimization problems," *Soft Computing*, vol. 13, pp. 763–780, 2009. 33, 34

[98] F. Vafaee and P. C. Nelson, "A genetic algorithm that incorporates an adaptive mutation based on an evolutionary model," in *Proceedings of International Conference on Machine Learning and Applications*, 2009, pp. 101–107. 35, 70

[99] M. Rocha and J. Neves, "Preventing premature convergence to local optima in genetic algorithms via random offspring generation," in *Proceedings of International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems: Multiple Approaches to Intelligent Systems*, 1999, pp. 127–136. 35, 70

[100] A. O'Toole, P. Phillips, F. Jiang, J. Ayyad, N. Penard, and H. Abdi, "Face recognition algorithms surpass humans matching faces over changes in illumination," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 9, pp. 1642–1646, 2007. 52

[101] "American society for aesthetic plastic surgery 2010 statistics," 2010. [Online]. Available: http://www.surgery.org/media/statistics 57, 84, 161

[102] M. De Marsico, M. Nappi, D. Riccio, and H. Wechsler, "Robust face recognition after plastic surgery using local region analysis," in *Proceedings of International Conference on Image Analysis and Recognition*, 2011, pp. 191–200. 59, 61

[103] G. Aggarwal, S. Biswas, P. J. Flynn, and K. W. Bowyer, "A sparse representation approach to face matching across plastic surgery," in *Proceedings of Workshop on the Applications of Computer Vision*, 2012, pp. 1–7. 59, 61, 80

[104] N. Kose, N. Erdogmus, and J. L. Dugelay, "Block based face recognition approach robust to nose alterations," in *Proceedings of International Conference on Biometrics: Theory, Applications and Systems*, 2012, pp. 121–126. 59

[105] P. J. Phillips, W. T. Scruggs, A. J. O'Toole, P. J. Flynn, K. W. Bowyer, C. L. Schott, and M. Sharpe, "FRVT 2006 and ICE 2006 Large-Scale Experimental Results," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 831–846, 2010. 59

[106] N. Erdogmus, N. Kose, and J. L. Dugelay, "Impact analysis of nose alterations on 2D and 3D face recognition," in *Proceedings of International Workshop on Multimedia Signal Processing*, 2012, pp. 354–359. 59

[107] R. Jillela and A. Ross, "Mitigating effects of plastic surgery: Fusing face and ocular biometrics," in *Proceedings of International Conference on Biometrics: Theory, Applications and Systems*, 2012, pp. 402–411. 59, 61

[108] H. S. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa, "Recognizing surgically altered face images using multiobjective evolutionary algorithm," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 1, pp. 89–100, 2013. 59, 61

[109] J. Liu, X. Yang, T. Xi, L. Gu, and Z. Yu, "A novel method for computer aided plastic surgery prediction," in *Proceedings of International Conference on Biomedical Engineering and Informatics*, 2009, pp. 1–5. 60

[110] S. A. Rabi and P. Aarabi, "Face fusion: An automatic method for virtual plastic surgery," in *Proceedings of International Conference on Information Fusion*, 2006, pp. 1–7. 60

[111] H. S. Bhatt, S. Bharadwaj, R. Singh, M. Vatsa, and A. Noore, "Face recognition and plastic surgery: Social, ethical and engineering challenges," in *Proceedings of International Conference on Ethics and Policy of Biometrics and International Data Sharing*, 2010, pp. 70–75. 60

[112] B. Heisele, P. Ho, J. Wu, and T. Poggio, "Face recognition: component-based versus global approaches," *Computer Vision and Image Understanding*, vol. 91, pp. 6–21, 2003. 62

[113] B. Weyrauch, B. Heisele, J. Huang, and V. Blanz, "Component-based face recognition with 3D morphable models," in *Proceedings of International Conference on Computer Vision and Pattern Recognition Workshop*, 2004, pp. 85–91. 62

[114] F. Li and H. Wechsler, "Robust part-based face recognition using boosting and transduction," in *Proceedings of International Conference on Biometrics: Theory, Applications, and Systems*, 2007, pp. 1–5. 62

[115] B. Gkberk, M. O. Irfanoglu, L. Akarun, and E. Alpaydin, "Learning the best subset of local features for face recognition," *Pattern Recognition*, vol. 40, pp. 1520–1532, 2007. 62, 68

[116] R. Campbell, M. Coleman, W. Michael, B. Jane, J. Philip, S. Wallace, J. Michelotti, and S. Baron-Cohen, "When does the inner-face advantage in familiar face recognition arise and why?" *Visual Cognition*, vol. 6, no. 2, pp. 197–216, 1999. 63, 65

[117] W. Hayward, G. Rhodes, and A. Schwaninger, "An own-race advantage for components as well as configurations in face recognition," *Cognition*, vol. 106, no. 2, pp. 1017–1027, 2008. 63

[118] A. Schwaninger, J. Lobmaier, and S. Collishaw, "Role of featural and configural information in familiar and unfamiliar face recognition," in *Proceedings of International Workshop on Biologically Motivated Computer Vision*, 2002, pp. 245–258. 63

[119] A. Bargiela and W. Pedrycz, *Granular computing: An introduction.* Kluwer Academic Publishers, 2002. 63

[120] T. Lin, Y. Yao, and L. Zadeh, *Data mining, rough sets and granular computing.* Physica-Verlag, 2002. 63

[121] P. Burt and E. Adelson, "A multiresolution spline with application to image mosaics," *ACM Transaction on Graphics*, vol. 2, no. 4, pp. 217–236, 1983. 64

[122] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4–37, 2000. 68, 72, 76

[123] C. Shan, S. Gong, and P. W. McOwan, "Conditional mutual information based boosting for facial expression recognition," in *Proceedings of British Machine Vision Conference*, 2005. 68

[124] E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing Co., Inc., 1989. 69

[125] A. Ross and A. Jain, "Information fusion in biometrics," *Pattern Recognition Letters*, vol. 24, no. 13, pp. 2115–2125, 2003. 71, 80, 85, 108, 110, 117, 118, 138, 141

[126] A. Young, D. Hay, K. McWeeny, B. Flude, and A. Ellis, "Matching familiar and unfamiliar faces on internal and external features," *Perception*, vol. 14, no. 6, pp. 737–746, 1985. 82

[127] U. Park, R. Jillela, A. Ross, and A. Jain, "Periocular biometrics in the visible spectrum," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 1, pp. 96–106, 2011. 85

[128] S. Bharadwaj, H. S. Bhatt, M. Vatsa, and R. Singh, "Periocular biometrics: When iris recognition fails," in *Proceedings of International Conference on Biometrics: Theory, Applications and Systems*, 2010, pp. 1–6. 85

[129] F. Juefei-Xu, K. Luu, M. Savvides, T. Bui, and C. Suen, "Investigating age invariant face recognition based on periocular biometrics," in *Proceedings of International Joint Conference on Biometrics*, 2011, pp. 1–7. 85

[130] "http://www.cbc.ca/news/background/london_bombing/investigation_timeline.html," (last accessed: January, 5, 2013). 89, 120

[131] "http://www.hindustantimes.com/india-news/newdelhi/who-s-keeping-watch/article1-908391.aspx," (last accessed: January, 5, 2013). 89, 120

[132] "http://www.lawisgreek.com/can-surveillance-cameras-preventdeter-terrorist-acts," (last accessed: January, 5, 2013). 89

[133] H. Huang and H. He, "Super-resolution method for face recognition using nonlinear mappings on coherent features," *IEEE Transactions on Neural Networks*, vol. 22, no. 1, pp. 121–130, 2011. 91, 92

[134] K. Jia and S. Gong, "Multi-modal tensor face for simultaneous super-resolution and recognition," in *Proceedings of International Conference on Computer Vision*, 2005, pp. 1683–1690. 92, 93

[135] W. W. W. Zou and P. C. Yuen, "Very low resolution face recognition problem," *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 327–340, 2012. 92, 93

[136] Z. Lei, T. Ahonen, M. Pietikainen, and S. Li, "Local frequency descriptor for low-resolution face recognition," in *Proceedings of International Conference on Automatic Face Gesture Recognition and Workshops*, 2011, pp. 161–166. 92, 94

[137] B. Li, H. Chang, S. Shan, and X. Chen, "Low-resolution face recognition via coupled locality preserving mappings," *IEEE Signal Processing Letters*, vol. 17, no. 1, pp. 20–23, 2010. 92, 93

[138] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1167–1183, 2002. 91

[139] A. Chakrabarti, A. N. Rajagopalan, and R. Chellappa, "Super-resolution of face images using kernel PCA-based prior," *IEEE Transactions on Multimedia*, vol. 9, no. 4, pp. 888–892, 2007. 91

[140] H. Chang, D. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2004, pp. 275–282. 91

[141] B. Li, H. Chang, S. Shan, and X. Chen, "Locality preserving constraints for super-resolution with neighbor embedding," in *Proceedings of International Conference on Image Processing*, 2009, pp. 1189–1192. 91

[142] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861–2873, 2010. 91, 108

[143] C. Liu, H. Shum, and W. Freeman, "Face hallucination: Theory and practice," *International Journal of Computer Vision*, vol. 75, no. 1, pp. 115–134, 2007. 91

[144] S. Biswas, G. Aggarwal, and P. J. Flynn, "Pose-robust recognition of low-resolution face images," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2011, pp. 601–608. 93

[145] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010. 94, 95

[146] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *Proceedings of Conference on Learning Theory*, 1998, pp. 92–100. 94, 99

[147] Y. Zhu, Y. Chen, Z. Lu, S. J. Pan, G. Xue, Y. Yu, and Q. Yang, "Heterogeneous transfer learning for image classification." in *proceedings of AAAI Conference on Artificial Intelligence*, 2011, pp. 1304–1309. 95

[148] A. Quattoni, M. Collins, and T. Darrell, "Transfer learning for image classification with sparse prototype representations," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8. 95

[149] A. Ahmed, K. Yu, W. Xu, Y. Gong, and E. Xing, "Training hierarchical feed-forward visual recognition models using transfer learning from pseudo-tasks," in *Proceedings of European Conference on Computer Vision*, 2008, pp. 69–82. 96

[150] B. Geng, D. Tao, and C. Xu, "Daml: Domain adaptation metric learning," *IEEE Transactions on Image Processing*, vol. 20, no. 10, pp. 2980–2989, 2011. 96

[151] H. Wang, F. Nie, H. Huang, and C. Ding, "Dyadic transfer learning for cross-domain image classification," in *Proceedings of International Conference on Computer Vision*, 2011, pp. 551–556. 96

[152] X. Siyu, S. Ming, and F. Yun, "Kinship verification through transfer learning," in *Proceedings of International Joint Conference on Artificial Intelligence*, 2011, pp. 2539–2544. 96

[153] J. Chen, X. Liu, P. Tu, and A. Aragones, "Person-specific expression recognition with transfer learning," in *Proceedings of International Conference on Image Processing*, 2012, pp. 2621–2624. 96

[154] D. Rim, K. Hassan, and C. J. Pal, "Semi supervised learning for wild faces and video," in *Proceedings of British Machine Vision Conference*, 2011, pp. 1–12. 96

[155] D. Cai, X. He, and J. Han, "Semi-supervised discriminant analysis," in *International Conference on Computer Vision*, 2007, pp. 1–7. 96

[156] R. Gross, L. Sweeney, F. Torre, and S. Baker, "Semi-supervised learning of multi-factor models for face de-identification," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8. 96

[157] Y. Zhang and D. Yeung, "Semi-supervised discriminant analysis using robust path-based similarity," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8. 96

[158] F. Roli and G. Marcialis, "Semi-supervised PCA-based face recognition using self-training," in *Structural, Syntactic, and Statistical Pattern Recognition*, 2006, pp. 560–568. 96

[159] F. Roli, L. Didaci, and G. Marcialis, "Adaptive biometric systems that can improve with use," in *Proceedings of Advances in Biometrics: Sensors, Systems and Algorithms*, 2008, pp. 447–471. 96

[160] X. Zhao, N. Evans, and J. Dugelay, "Semi-supervised face recognition with LDA self-training," in *Proceedings of International Conference on Image Processing*, 2011, pp. 3041–3044. 96

[161] H. S. Bhatt, S. Bharadwaj, R. Singh, M. Vatsa, A. Ross, and A. Noore, "On co-training online biometric classifiers," in *Proceedings of International Joint Conference on Biometrics*, 2011, pp. 1–6. 96, 97, 99, 100

[162] R. Singh, M. Vatsa, A. Ross, and A. Noore, "Biometric classifier update using online learning: A case study in near infrared face verification," *Image and Vision Computing*, vol. 28, no. 7, pp. 1098–1105, 2010. 97

[163] G. Cauwenberghs and T. Poggio, "Incremental and decremental support vector machine learning," in *Proceedings of Advances in Neural Information Processing Systems*, 2000, pp. 409–415. 97

[164] P. Zhao and S. Hoi, "OTL: A framework of online transfer learning," in *Proceedings of International Conference on Machine Learning*, 2010, pp. 1231–1238. 99, 101, 105, 166

[165] T. Ahonen, E. Rahtu, V. Ojansivu, and J. Heikkila, "Recognition of blurred faces using local phase quantization," in *Proceedings of International Conference on Pattern Recognition*, 2008, pp. 1–4. 102, 103, 110

[166] K. I. Kim and Y. Kwon, "Single-image super-resolution using sparse regression and natural image prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 6, pp. 1127–1133, 2010. 108

[167] J. C. Klontz and A. K. Jain, "A case study on unconstrained facila recognition using the boston marathon bombings suspects," Michigan State University, Tech. Rep., 2013. 119, 120

[168] "http://www.reuters.com/article/2013/04/23/us-usa- explosions-boston-injuries-idusbre93m0lw20130423." 120

[169] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys*, vol. 35, pp. 399–458, 2003. 127

[170] Z. Zhang, C. Wang, and Y. Wang, "Video-based face recognition: State of the art," in *Proceedings of Chinese Conference on Biometric Recognition*, 2011, pp. 1–9. 128

[171] N. Poh, C. H. Chan, J. Kittler, S. Marcel, C. McCool, E. A. Rua, J. L. A. Castro, M. Villegas, R. Paredes, V. Struc, N. Pavesic, A. A. Salah, H. Fang, and N. Costen, "An evaluation of video-to-video face verification," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 4, pp. 781–801, 2010. 128

[172] O. Arandjelovic, G. Shakhnarovich, J. Fisher, R. Cipolla, and T. Darrell, "Face recognition with image sets using manifold density divergence," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2005, pp. 581–588. 129, 130

[173] R. Wang, S. Shan, X. Chen, and W. Gao, "Manifold-manifold distance with application to face recognition based on image set," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8. 129, 130

[174] K. C. Lee, J. Ho, M. H. Yang, and D. Kriegman, "Visual tracking and recognition using probabilistic appearance manifolds," *Computer Vision and Image Understanding*, vol. 99, no. 3, pp. 303–331, 2005. 130

[175] R. Gross and J. Shi, "The CMU motion of body (MoBo) database," Tech. Rep. CMU-RI-TR-01-18, 2001. 130

[176] G. Aggarwal, A. K. Roy Chowdhury, and R. Chellappa, "A system identification approach for video-based face recognition," in *Proceedings of International Conference on Pattern Recognition*, 2004, pp. 175–178. 129, 130

[177] K. Fukui and O. Yamaguchi, "The kernel orthogonal mutual subspace method and its application to 3D object recognition," in *Proceedings of Asian Conference on Computer Vision*, 2007, pp. 467–476. 129, 130

[178] M. Nishiyama, M. Yuasa, T. Shibata, T. Wakasugi, T. Kawahara, and O. Yamaguchi, "Recognizing faces of moving people by hierarchical image-set matching," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8. 129, 130

[179] M. T. Harandi, C. Sanderson, S. Shirazi, and B. C. Lovell, "Graph embedding discriminant analysis on grassmannian manifolds for improved image set matching," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2011, pp. 2705–2712. 129, 130

[180] E. Bailly-Baillire, S. Bengio, F. Bimbot, M. Hamouz, J. Kittler, J. Marithoz, J. Matas, K. Messer, V. Popovici, F. Pore, B. Ruiz, and J. Thiran, "The BANCA database and evaluation protocol," in *Proceedings of Conference on Audio- and Video-Based Biometric Person Authentication*, 2003, pp. 625–638. 130

[181] Z. Cui, S. Shan, H. Zhang, S. Lao, and X. Chen, "Image sets alignment for video-based face recognition," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2626–2633. 129, 130

[182] M. Kim, S. Kumar, V. Pavlovic, and H. Rowley, "Face tracking and recognition with visual constraints in real-world videos," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8. 129, 130

[183] K. Lee, J. Ho, M. Yang, and D. Kriegman, "Video-based face recognition using probabilistic appearance manifolds," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2003, pp. 313–320. 129, 130

[184] Y. Hu, A. S. Mian, and R. Owens, "Sparse approximated nearest points for image set classification," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2011, pp. 121–128. 129, 130, 133

[185] S. Zhou, V. Krueger, and R. Chellappa, "Probabilistic recognition of human faces from video," *Computer Vision and Image Understanding*, vol. 91, no. 12, pp. 214–245, 2003. 129, 130

[186] X. Liu and T. Chen, "Video-based face recognition using adaptive hidden Markov models," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2003, pp. 340–345. 129, 130

[187] Y. C. Chen, V. M. Patel, P. J. Phillips, and R. Chellappa, "Dictionary-based face recognition from video," in *Proceedings of European Conference on Computer Vision*, 2012, pp. 766–779. 130, 131, 133

[188] P. J. Phillips, P. J. Flynn, J. R. Beveridge, W. T. Scruggs, A. J. O'Toole, D. Bolme, K. W. Bowyer, B. A. Draper, G. H. Givens, Y. M. Lui, H. Sahibzada, J. A. Scallan, Iii, and S. Weimer, "Overview of the multiple biometrics grand challenge," in *Proceedings of the International Conference on Advances in Biometrics*, 2009, pp. 705–714. 130

[189] H. S. Bhatt, R. Singh, and M. Vatsa, "On rank aggregation for face recognition from videos," in *Proceedings of International Conference on Image Processing*, 2013, pp. 1–5. 130, 131, 143, 147

[190] J. R. Barr, K. W. Bowyer, P. J. Flynn, and S. Biswas, "Face recognition from video : A review," *Journal of Pattern Recognition and Artificial Intelligence*, vol. 26, no. 5, 2012. 129

[191] J. Stallkamp, H. K. Ekenel, and R. Stiefelhagen, "Video-based face recognition on real-world data," in *Proceedings of International Conference on Computer Vision*, 2007, pp. 1–8. 129

[192] G. Shakhnarovich, J. W. Fisher, III, and T. Darrell, "Face recognition from long-term observations," in *Proceedings of European Conference on Computer Vision*, 2002, pp. 851–868. 129

[193] A. Hadid and M. Pietikinen, "Manifold learning for video-to-video face recognition," in *Proceedings of Biometric ID Management and Multimodal Communication*, 2009, pp. 9–16. 129

[194] V. M. Patel, T. Wu, S. Biswas, P. J. Phillips, and R. Chellappa, "Dictionary-based face recognition under variable lighting and pose." *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 3, pp. 954–965, 2012. 129, 133

[195] Y. C. Chen, V. M. Patel, S. Shekhar, R. Chellappa, and P. J. Phillips, "Video-based face recognition via joint sparse representation," in *Proceedings of International Conference and Workshops on Automatic Face and Gesture Recognition*, 2013, pp. 1–8. 131, 138

[196] B. Klare and A. K. Jain, "On a taxonomy of facial features," in *Proceedings of International Conference on Biometrics: Theory Applications and Systems*, 2010, pp. 1–8. 132

[197] K. Järvelin and J. Kekäläinen, "Cumulated gain-based evaluation of IR techniques," *ACM Transactions on Information Systems*, vol. 20, no. 4, pp. 422–446, 2002. 132, 138

[198] F. Schroff, T. Treibitz, D. Kriegman, and S. Belongie, "Pose, illumination and expression invariant pairwise face-similarity measure via doppelganger list comparison," in *Proceedings of International Conference on Computer Vision*, 2011, pp. 2494–2501. 132, 133, 134, 144, 147

[199] Q. Yin, X. Tang, and J. Sun, "An associate-predict model for face recognition," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2011, pp. 497–504. 133

[200] D. A. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, "Incremental learning for robust visual tracking," *International Journal on Computer Vision*, vol. 77, no. 1-3, pp. 125–141, 2008. 134

[201] H. S. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa, "Memetically optimized MCWLD for matching sketches with digital face images," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 5, pp. 1522–1535, 2012. 134

[202] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997. 134

[203] J. Zhang, J. Gao, M. Zhou, and J. Wang, "Improving the effectiveness of information retrieval with clustering and fusion," *Journal of Computational Linguistics and Chinese Language Processing*, vol. 6, no. 1, pp. 109–125, 2001. 135, 137

[204] J. A. Hartigan, *Clustering Algorithms*. Wiley Series in Probability and Mathematical Statistics, 1975. 136

[205] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226–239, 1998. 138

[206] C. D. Manning, P. Raghavan, and H. Schtze, *Introduction to Information Retrieval*. Cambridge University Press, 2008. 138

[207] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recognition*, vol. 38, no. 12, pp. 2270–2285, 2005. 143

[208] H. Li, G. Hua, Z. Lin, J. Brandt, and J. Yang, "Probabilistic elastic matching for pose variant face verification," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3499–3506. 143, 147

[209] Z. Cui, W. Li, D. Xu, S. Shan, and X. Chen, "Fusing robust face region descriptors via multiple metric learning for face recognition in the wild," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3554–3561. 143, 147

[210] H. Mendez-Vazquez, Y. Martinez-Diaz, and Z. Chai, "Volume structured ordinal features with background similarity measure for video face recognition," in *Proceedings of International Conference on Biometrics*, 2013, pp. 1–6. 143, 147

[211] MBGC v2: available at http://www.nist.gov/itl/iad/ig/mbgc.cfm. 159

[212] J. C. Klontz and A. K. Jain, "A case study on unconstrained facial recognition using the boston marathon bombings suspects," Tech. Rep. MSU, 2013. 161

[213] Cognitec Systems GmbH, "Video Surveillance Systems with Face Recognition Technology", Availabe at http://www.security-technologynews.com/article/video-surveillance-systems-with-face-recognition-technology.html. 161