Automatic Target Cueing (ATC)

Task 2 Report – Literature Survey on Facial Recognition

November 19, 2013

Prepared by:
Mohsen Ghazel

MDA Systems Ltd.
13800 Commerce Parkway
Richmond, BC, Canada
V6V 2J3

PWGSC Contract Title:	Automatic target cueing and facial recognition in visible & infrared spectrum for military operations
MDA Project Title:	6484 - ATC
MDA Document Number:	RX-RP-53-5691
Contract No.:	W7701-135590/001/QCL
Project Duration:	August 2 2013 – March 31 2014
Contract Scientific Authority:	Dr. Philips Laou (418) 844-4000 x4218
DRDC-RDDC-2014-C171	

Prepared By:	Mohsen Ghazel	Nov. 20, 2013 Signature and Date
Reviewed By:	Stephen Se	Stephen Se Nov 20, 2013 Signature and Date
Reviewed By:	Rabab Ward	(ON BEHALF OF PROF. WARD) Stephen Se Nov 20, 2013 Signature and Date
Project Manager:	Stephen Se	Stephen Se Nov 20, 2013 Signature and Date

MacDonald, Dettwiler and Associates Ltd
13800 Commerce Parkway
Richmond, BC, Canada
V6V 2J3

CHANGE RECORD

ISSUE	DATE	PAGE(S)	DESCRIPTION
1/0	Nov. 8, 2013	All	First Issue
1/1	Nov. 19, 2013	All	First Issue, First Revision Minor edits to some figures, tables and text

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

ACRONYMS AND ABBREVIATIONS

2D	Two Dimensional
2DLPGNN	2D-log polar Gabor transform and Neural Network
3D	Three Dimensional
3DMM	Three Dimensional Morphable Model
AAM	Active Appearance Model
AdaBoost	Adaptive Boosting
AGES	AGing pattErn Subspace
AI	Artificial Intelligence
AMP	Advanced Multimedia Processing
API	Application Programming Interface
AR	Automatic Recognition
ASHCI	Age Specific Human Computer Interaction
ASM	Active Shape Model
ATC	Automatic Target Cueing
ATE	Assisted Target Engagement
BC	Bi-Cubic
BFM	Basel Face Model
CAS-PEAL	Chinese Academy of Sciences-Pose, Expression, Accessory, and Lighting
CAVIAR	Context Aware Vision using Image-based Active Recognition
CCTV	Closed Circuit Television
CEA	Conformal Embedding Analysis
CFAS	Constant Face Acquisition and Suspect
CIM	Comparison Is Made
CKFE-DB	Cohn-Kanade-Facial-Expression database
CLPM	Coupled Locality Preserving Mappings
CMC	Cumulative Match Characteristic or Cumulative Match Curve
CMS	Cumulative Match Score

CMU	Carnegie Mellon University
COTS	Commercial Off-The-Shelf
CPU	Central Processing Unit
CVC	Computer Vision Center
CVPR	Computer Vision and Pattern Recognition
DCT	Discrete Cosine Transform
DICW	Dynamic Image-to-Class Warping
DND	Department of National Defence (Canada)
DoD	Department of Defense (USA)
DRDC	Defence Research and Development Canada
DTW	Dynamic Time Warping
DWT	Discrete Wavelet Transform
EF	Eigen Face
EHMM	Embedded Hidden Markov Model
EM	Expectation Maximization
EO	Electro-Optical
EQUINOX	EQUINOX Corporation
FACE	Face Analysis Commercial Entities
FaceVACS	Face Video Analysis for Cognitive Systems
FAL	False Alarm Rate
FAR	False Accept Rates
FAVES	Facial Automated Verification Solution
FBI	Federal Bureau of Investigation
FERET	Facial Recognition Technology
FGNET	Face and Gesture Recognition Research Network
FIA	Face In Action
FIMS	FaceR Identity Management System
FLR	False Label Rate
FNR	False Negative Rate

FPGA	Field Programmable Gate Array
FPR	False Positive Rate
FR	Facial Recognition
FRGC	Face Recognition Grand Challenge
FRR	False Recognition Rate
FRS	Face Recognition System
FRT	Face Recognition Technology
FRVT	Face Recognition Vendor Test
FSAR	Future Small Arms Research
GAR	Genuine Accept Rate
GB	Giga Bytes
GF	Geometrical Feature
GHz	Giga Hertz
GMM	Gaussian Mixture Model
GO	Gradient Orientation
GOP	Gradient Orientation Pyramid
GPU	Graphics Processing Unit
GWN	Gabor Wavelet Networks
HCI	Human Computer Interaction
HF	Hallucination Face
HID	Human Identification at a Distance
HM	Hidden Markov
HMM	Hidden Markov Model
HR	High Resolution
HRI	Human Robot Interaction
Hz	Hertz
ICA	Independent Component Analysis
ICBC	Insurance Corporation of British Columbia
ICCV	International Conference on Computer Vision

ICP	Iterative Closest Point
ICVS	International Conference on Computer Vision Systems
IEEE	Institute of Electrical and Electronics Engineers
IGF	Independent Gabor Features
IJCV	International Journal of Computer Vision
IR	Infrared
iRSC	improved Robust Sparse Classification or improved Robust Sparse Classification
JAFFE	Japanese Female Facial Expression
KF	Kernel Face
KLT	Kanada-Lucas-Tomasi
kNN	k-Nearest Neighbor
KPCA	Kernel Principle Component Analysis
LBP	Local Binary Pattern
LC	Lighting Compensation
LDA	Linear Discriminant Analysis
LFA	Local Feature Analysis
LMNN	Large Margin Nearest Neighbor
LPP	Locality Preserving Projection
LR	Low Resolution
LSE	Least Squares Estimation
LSS	Local Self-Similarity
LSVM	Linear Support Vector Machine
LWIR	Long Wave Infra-Red
M2VTS	Multi-Model Verification and Testing System
MBGC	Multiple Biometric Grand Challenge
MDA	MDA Systems Ltd.
MDS	Multidimensional Scalings
MFDA	Multi-Feature Discriminative Analysis
MIT	Massachusetts Institute of Technology

MLBP	Multi-scale Local Binary Patterns
MLE	Maximum Likelihood Estimation
MPEG	Moving Picture Experts Group
NA	Not Available
NBNN	Naïve Bayes Nearest Neighbor
NEC	NEC Technologies Ltd.
NIST	Institute of Standards and Technology
NLPR	National Laboratory of Pattern Recognition
NN	Nearest Neighbor
NVIDIA	NVIDIA Corporation Ltd.
O	Occluded
OLPP	Orthogonal Locality Preserving Projection
OpenCV	Open Source Computer Vision Library
ORL	Olivetti Research Laboratory
PCA	Principal Component Analysis
PCSO	Pinellas County Sheriff's Office
PDM	Point Distribution Model
PICS	Psychological Image Collection at Stirling
PIE	Pose, Illumination and Expression
RAM	Random Access Memory
RBF	Radial Basis Function
RGB	Red Green Blue
RL	Relationship Learning
RMS	Root Mean Squares
RMSE	Root Mean Squared Error
ROC	Receiver Operating Characteristic
ROI	Region of Interest
RSC	Robust Sparse Classification
S2R2	Simultaneous Super Resolution and Recognition

SCface	Surveillance Camera Face
SDK	Software Development Kit
SIFT	Scale Invariant Feature Transform
SR	Super Resolution
SRC	Sparse Representation Classification
STASM	Extended Active Shape Model
SVD	Singular Value Decomposition
SVM	Support Vector Machine
TASC	The Analytic Science Company
U	Un-occluded
UCA	University of California at Davis
UM	Unsharp Masking
UMIST	University of Manchester Institute of Science and Technology
UND	University of Notre Dame
US	United States
USA	United States of America
USC	University of Southern California
UTK-LRHM	University of Tennessee, Knoxville Long Range High Magnification
VLR	Very Low Resolution
YCbCr	Luminance; Chroma: Blue; Chroma: Red

1 INTRODUCTION

1.1 Background

Under the mandate of Future Small Arms Research (FSAR) program, Defence Research and Development Canada (DRDC) is examining existing and future technologies for small arms capabilities with the objective of identifying technologies which could increase shot placement accuracy and reduce engagement time under various situations.

Automatic Target Cueing (ATC) is considered as one of the key enablers in future small arms technology. The goal is to assess the feasibility and capability of Assisted Target Engagement (ATE) in small arms by combining ATC with electronic ignited ammunition and small arm weapons. It is believed that ATE may help improving shot placement and shortening engagement time in some situations.

1.2 Project Objectives

The objectives of this project involve conducting a feasibility study and providing software development support on ATC and Facial Recognition (FR) in visible and infrared spectrum for military operations.

The project consists of seven tasks: the first two tasks are related to literature survey of ATC and FR, the next two tasks are related to the study and evaluation of existing ATC and FR products, while the last three tasks are related to enhancing existing DRDC system and developing new ATC capabilities.

1.3 Task 2 Objectives

The objective of Task 2 is to conduct a review of existing literature on facial recognition methodologies and technologies based on optical imagery, in order to determine the feasibility of performing highly accurate facial recognition in visible spectrum in military operations. These operations will make use of ATC systems, which perform short range facial recognition for personnel identification at standoff distance below 100m. Facial recognition is the critical final step of any ATC system, as discussed in the Task 1 report [R-1]. This task is to conduct a literature survey, as well identify, describe and interpret the key characteristics of facial recognition methodologies and technologies in visible spectrum.

1.4 Scope

This report is to fulfil Task 2 milestone of this contract and contains the following sub-tasks:

- Conduct general literature survey of FR methodologies and in visible spectrum.
- Conduct literature survey of face recognition methodologies, which address some of the challenges encountered in real-world uncontrolled FR applications.
- Review Commercial Off-The-Shelf (COTS) FR software/Software Development Kit (SDK).
- Review some of the performance evaluation studies and real-world applications of FR technologies.
- Review open-source FR databases, used for evaluating FR methodologies and technologies.

2 LITERATURE SURVEY

Face Recognition (FR) is a challenging problem in the field of image analysis and computer vision, and as such has received a great deal of attention over the past decade because of its applications in various domains, including security, surveillance, identification, verification, video indexing, etc.

A Face Recognition System (FRS) is a computer application for automatically identifying or verifying a person from a digital image or a video frame from a video source. A face image or image sequence is given as input to the FRS, which applies proposed methods to match the given input with face images stored in the database. The significant research efforts and investigations of the face recognition problems have resulted in various automated facial recognition systems. Most current face recognition systems work well under controlled conditions, where the input image is acquired under similar conditions as those under which the images in the database were acquired. These image acquisition conditions include illumination, pose, occlusion, facial expression, age, range and resolution [R-2]. Unfortunately, the performance of most of the current state of the art face recognition systems degrades rapidly when they are put to work under conditions where some of these factors are not regulated. Recently, significant efforts have focused on addressing these limitations and developing more robust face recognition systems, which are less sensitive to these variations between the image acquisition properties and conditions of the input and database images.

Facial recognition literature is already quite vast and comprehensive and continues to grow at a fast rate. Conducting a comprehensive literature review of all facial recognition techniques and systems is beyond the scope of this task, given the limited budget and resources. In an effort to achieve a good overview of the facial recognition literature, we shall focus our FR literature review on the following topics:

- FR Databases: Review of available databases for face recognition systems development and testing
- General facial recognition surveys: Review of papers, which highlight the recent advances in FR literature, at a very high level.
- Addressing some of the challenges in facial recognition: Review of papers, which review or propose new methods for addressing some of the challenges in FR related to variations in the acquisition conditions between the input and the database images. As mentioned

above, these variations include changes in illumination, pose, occlusion, facial expression, time frame age, range and resolution.

- Current end-to-end state of the art face real-time recognition systems: Review of papers describing complete real-time FR systems for uncontrolled real-world situations.
- COTS face recognition software/SDK: Review of commercially available COTS FR technologies, software and SDKs.
- Real-world applications of facial recognition technology: Review of examples of real-world case and evaluation studies and real-world application of FR technology.

We begin with a brief discussion of FR databases.

2.1 Face Recognition Databases

Before we present the FR literature review, we shall summarize some of the common FR databases, which are used by most of the reviewed FR techniques for performance assessment and evaluation. Typically, when benchmarking an algorithm, it is recommended to use a standard test data set for researchers to be able to directly compare the results of their different FR approaches. While there are many databases in use currently, the choice of an appropriate database is typically made based on the given task and specifications of the developed FR algorithms. For examples, for FR algorithms, which focus on addressing some of the challenges in FR, such as those related to acquisition conditions including age, expressions, lighting, pose, occlusion, low resolution, etc., then a suitable database with images depicting variations of investigated conditions should be used.

Table A-1 lists many of the FR databases found in the literature, their descriptions and their usability.

2.2 General Face Recognition Literature Survey Papers

In an effort to highlight some of the recent progress and development in facial recognition research, we review one FR literature survey paper, as listed in Table 2-1.

Table 2-1 Selected Paper on FR Literature Survey

#	Paper Title	Authors	Source	Year
P-1	A Survey of Face Recognition Techniques	Rabia Jafri, and Hamid R. Arabnia	Journal of Information Processing Systems, Vol.5, No.2, June 2009	2009

In paper # [P-1], the authors broadly divide face recognition techniques found in the literature into three categories based on the face data acquisition methodology:

1. FR methods that operate on intensity images
2. FR methods that deal with video sequences

3. FR techniques that require other sensory data such as 3D information or infra-red imagery.

An overview of some of the well-known methods in each of these categories is provided and some of the benefits and drawbacks of these schemes are highlighted in this paper.

Next, we summarize some of the discussed FR approaches for each category.

2.2.1 Face Recognition Techniques for Intensity Images

The authors sub-divide facial recognition methods for intensity images into two main categories:

- Feature-based techniques
- Holistic/Global techniques

An overview of some of the well-known methods in these categories is given below.

2.2.1.1 Feature-based FR Methods

As illustrated in Figure 2-1, feature-based approaches first process the input image to identify and extract (and measure) distinctive facial features such as the eyes, mouth, nose, etc., as well as other marks, and then compute the geometric relationships among those facial points, thus reducing the input facial image to a vector of geometric features. Standard statistical pattern recognition techniques are then employed to match faces using these measurements.

The paper discusses different types of feature-based facial recognition techniques, depending on the types of used facial features and how they are matched. These techniques are summarized next.

2.2.1.1.1 Techniques using Vectors of Geometric Features

Early feature-based facial recognition techniques employ simple image processing methods to extract an *N-dimensional* vector of *geometric* facial features and parameters. These parameters include the ratios of distances, areas and angles (to compensate for the varying size of the pictures) - and used a simple Euclidean distance measure for matching. The paper reports on many variations of these techniques, which mainly differ in the number of types of selected features.

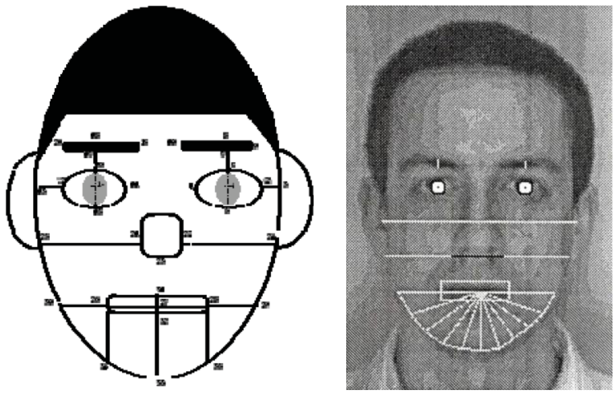


Figure 2-1 Extracted Facial Features in FR Techniques Based on Vectors of Geometric Features (Source: [P-1])

The reported detection rates of these types of techniques range from 75% for fully automated systems to 95%, for semi-automated systems where the key facial features were manually extracted. The authors conclude that, for these types of techniques, current algorithms for automatic feature extraction generally do not provide a high degree of accuracy and require considerable computational capacity.

2.2.1.1.2 Techniques using Elastic Face Bunch Matching

Another group of feature-based techniques discussed in this paper are based on the elastic bunch graph matching method, where a graph for an individual face is generated as follows: a set of fiducial points on the face are chosen. As illustrated in Figure 2-2, each fiducial point is a node of a full connected graph, and is labeled with the Gabor filters' responses applied to a window around the fiducial point. Each arch is labeled with the distance between the correspondent fiducial points. A representative set of such graphs is combined into a stack-like structure, called a *face bunch graph*. Once the system has a face bunch graph, graphs for new face images can then be generated automatically by Elastic Bunch Graph Matching method documented in the literature. Recognition of a new face image is performed by comparing its image graph to those of all the known face images and picking the one with the highest similarity value.

It is reported that this method was among the best performing ones in a reported recent facial recognition evaluation study, achieving a recognition rate of around 98% in some cases. However, it does suffer from the serious drawback of requiring the graph placement for the first 70 faces to be done manually before the elastic graph matching becomes adequately dependable. The paper reports on recent efforts, which automate the graph placement and fiducial points selection while achieving comparable performance as the semi-automated original approaches.

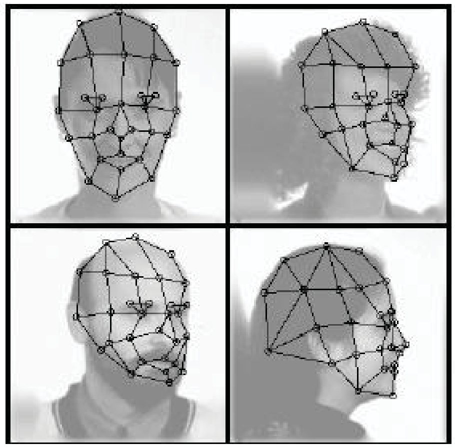


Figure 2-2 Extracted Facial Features in FR Techniques Based on Elastic Face Bunch Matching (Source: [P-1])

2.2.1.1.3 Techniques using Facial Profile Features

The paper also reports on considerable FR research efforts devoted to recognizing faces from their profiles. The approach is based on selecting an N-dimensional feature vector to describe the face profiles and employing a Euclidean distance measure to match them. As illustrated in Figure 2-3, feature extraction becomes a somewhat simpler one-dimensional problem in this case. Several methods related to this approach are reports and they mainly differ in the number and type of selected features.

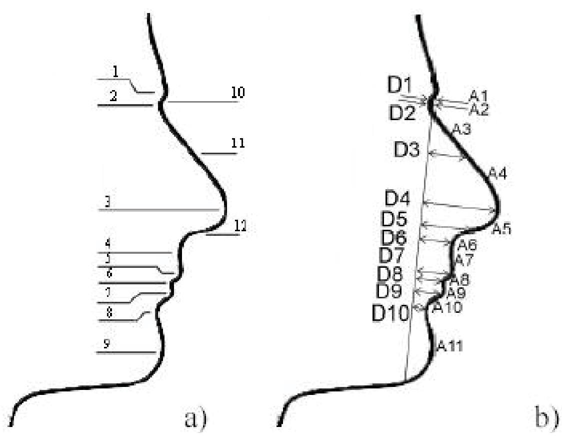


Figure 2-3 Extracted Facial Features in FR Techniques Based on Facial Profile Matching (Source: [P-1])

Recognition rates ranging between 90% and 96% are reported for various facial profile based methods are reported. However, these rates are reported using relatively small data sets.

2.2.1.1.3.1 Advantages and Disadvantages

This paper reports that the main advantage offered by the featured-based techniques is that since the extraction of the feature points precedes the analysis done for matching the image to that of a known individual, such methods are relatively robust to pose variations in the input image. Other benefits of these schemes include the compactness of representation of the face images and high speed matching.

The major disadvantage of these approaches is the difficulty of automatic feature detection and the fact that the implementer of any of these techniques has to make arbitrary decisions about which features are important.

2.2.1.2 Global FR Methods

The paper reviews holistic approaches, which attempt to identify faces using global representations, i.e., descriptions based on the entire image rather than on local features of the face. The schemes are further subdivided into two groups: statistical and Artificial Intelligence (AI) approaches. An overview of some of the methods in these categories is presented next.

2.2.1.2.1 Statistical Approaches

The reviewed statistical approaches are based on the Principal Components Analysis (PCA) algorithm, which attempts to economically represent face images. As illustrated in Figure 2-4, a particular face can be efficiently represented along the Eigenfaces coordinate space, and that any face can be approximately reconstructed by using just a small collection of Eigenfaces and the corresponding projections (‘coefficients’) along each Eigenface. Projections along Eigenfaces could be used as classification features to recognize faces. They employed this reasoning to develop a face recognition system that builds Eigenfaces, which correspond to the eigenvectors associated with the dominant eigenvalues of the known face (patterns) covariance matrix, and then recognizes particular faces by comparing their projections along the Eigenfaces to those of the face images of the known individuals. The Eigenfaces define a feature space that drastically reduces the dimensionality of the original space, and face identification is carried out in this reduced space. The paper reported that, PCA-based methods yielded recognition rates ranging from 64% to 96% under lighting, orientation and scale variation.

Figure 2-4 (a) An Example Training Set, and (b) Seven of the Eigenfaces Calculated from the Training Images (Source: [P-1])

The paper cites works which indicated that PCA appears to work well when a single image of each individual is available. However, when multiple images per person are present, it was shown that by choosing the projection which maximizes total scatter, PCA retains unwanted variations due to lighting and facial expression. As such, the variations between the images of the same face due to illumination and lighting direction are almost always larger than image variations due to a change in face identity. The Fisher's Linear Discriminant Analysis (LDA), which maximizes the ratio of the between-class scatter and the within-class scatter and is thus proposed as a better alternative for classification than PCA. As illustrated in Figure 2-5, the sample of Eigenfaces generated by the PCA shows the tendency of the principal components to capture major variations in the training set such as lighting direction, while the corresponding sample of Fisherfaces, generated by the LDA shows the ability of Fisherfaces to discount those factors unrelated to classification. It is reported that the LDA is better at simultaneously handling variations in lighting and expression than the PCA and yielded recognition result of about 90% accuracy. However, it was noted that, when the training set is small, the PCA is less sensitive to the use of different training sets than the LDA.

Figure 2-5 Top: Eigenfaces (PCA), Bottom: Fisherfaces (LDA) (Source: [P-1])

The paper also briefly cites numerous variations on and extensions to the standard Eigenfaces (PCA) and the Fisherfaces (LDA) approaches, which have been developed since their introduction. Some of these variations are applied in the non-spatial domain of the images, including Fourier and Wavelet domains.

2.2.1.2.2 Artificial Intelligence (AI)

The paper cites several AI-based facial recognition approaches utilize tools such as neural networks and machine learning techniques to recognize faces. The paper cites several such methods, with reported recognition rates varying from 60% to 97.9%, using different datasets. It was also observed that some AI-based approaches use Support Vector Machine (SVM) and Hidden Markov (HM) models for pattern classification for the purpose of facial recognition.

2.2.1.2.3 Multiple Classifier Systems

The paper cites several recent works which propose hybrid facial recognition systems, composed of distinct facial recognition algorithms. The idea is, since the performance of an individual classifier is more sensitive to some factors and relatively invariant to others, individual classifiers can be combined in order to integrate their complementary information and thereby create a system that is more robust than any individual classifier to variables that complicate the recognition task. Figure 2-6 illustrates a hybrid facial recognition system composed of the PCA and LDA-based facial recognition components.

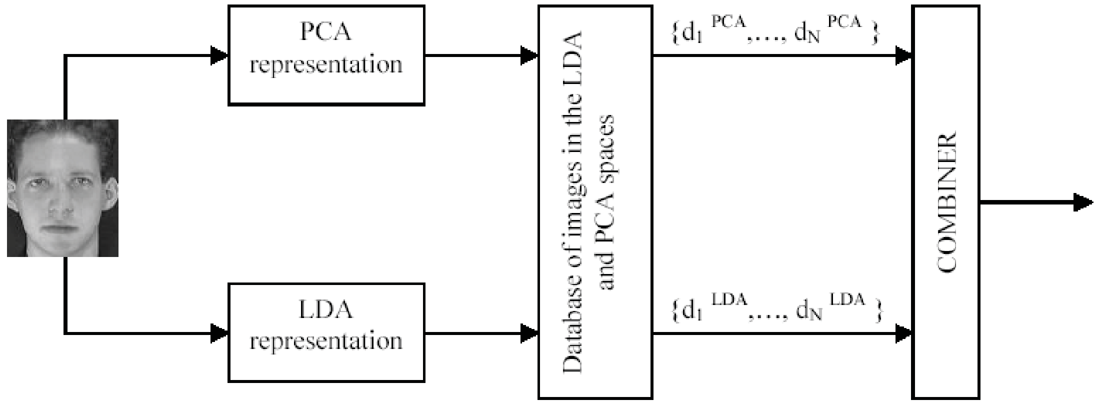


Figure 2-6 Overview of a Hybrid PAC/LDA Facial Recognition Methodology (Source: [P-1])

2.2.1.2.4 Advantages and Disadvantages

The paper presents a brief overall assessment of the holistic facial recognition approaches. The main advantage of these approaches is that they do not destroy any of the information in the images by concentrating on only limited regions or points of interest. However, this same property is also their greatest drawback, since most of these approaches start out with the basic assumption that all the pixels in the image are equally important. Consequently, these techniques are not only computationally expensive but require a high degree of correlation between the test and training images, and do not perform effectively under large variations in pose, scale and illumination, etc. Nevertheless, several of these algorithms have been modified and/or enhanced to compensate for such variations, and dimensionality reduction techniques have been exploited. As a result of which these approaches appear to produce better recognition results than the feature-based ones in general.

2.2.2 Face Recognition Techniques for Video Sequences

Since one of the major applications of face recognition is surveillance for security purposes, which involves real-time recognition of faces from an image sequence captured by a video camera, a significant amount of research has been directed towards this area in recent years.

A video-based face recognition system typically consists of three modules:

1. Face detection
2. Face tracking
3. Face recognition.

Most of these systems choose a few good frames and then apply one of the recognition techniques for intensity images to those frames in order to identify the individual. This paper briefly describes a few of these approaches.

One of the reported recognition system which uses skin color modeling to detect the face, then utilizes Gabor Wavelet Networks (GWN) to detect prominent facial landmarks (i.e., the eyes, nose, and mouth) and to track those features. For each individual frame, Eigenfeatures are then extracted and a feature selection algorithm is applied over the combination of all the Eigenfeatures, and the best ones are selected to form the feature space. Two classifiers are then applied to identify the individual in the frame and, finally, a super-classifier based on a voting scheme performs the final classification for the entire video sequence. Figure 2-7 illustrates an overview of this system. Good recognition results (97.7% accuracy) have been reported, using this FR system.

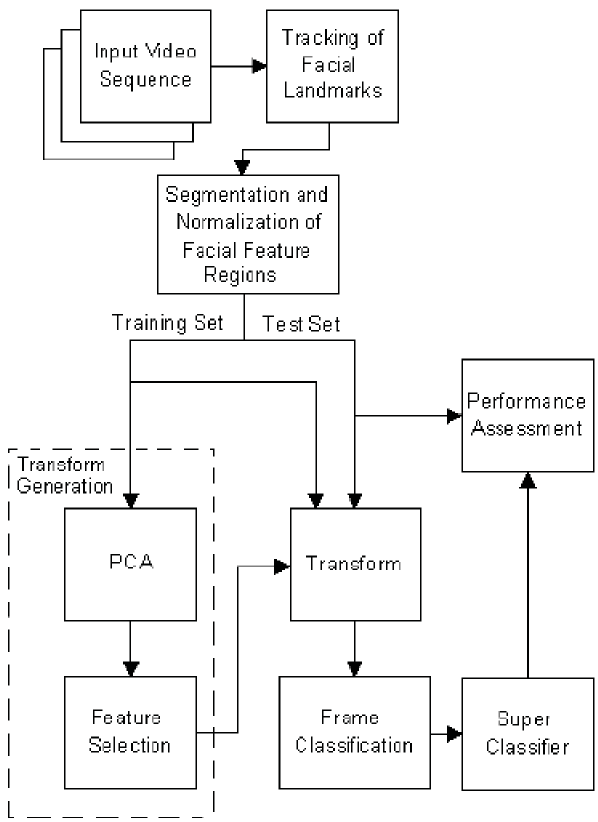


Figure 2-7 Overview of the Facial Recognition System Based on Skin Color Modeling (Source: [P-1])

The paper also reports that several data fusion-based FR techniques have attempted to incorporate information from other modalities in order to recognize facial images acquired from video clips. For instance, one approach makes use of stereo information has reported a recognition accuracy of 90%. Another approach exploits both audio and video cues as well as 3D information about the head to achieve a 100% accuracy rate.

2.2.2.1.1 Advantages and Disadvantages

The paper presents a brief overall assessment of the video-based facial recognition approaches. It is concluded that dynamic face recognition from video schemes appear to be at a disadvantage relative to their static counterparts in general, since they are usually hampered by one or more of the following: low quality images (though image quality may be enhanced by exploiting super-resolution techniques); cluttered backgrounds (which complicate face detection); the presence of more than one face in the picture; and a large amount of data to process. Furthermore, the face image may be much smaller than the size required by most systems employed by the recognition modules.

However, dynamic schemes do have the following advantages over static techniques: the enormous abundance of data empowers the system to choose the frame with the best possible image and discard less satisfactory ones. Video provides temporal continuity, so classification information from several frames can be combined to improve recognition performance. Moreover, video allows the tracking of face images such that variations in facial expressions and poses can be compensated for, resulting in improved recognition. Dynamic schemes also have an edge over static ones when it comes to detecting the face in a scene, since these schemes can use motion to segment a moving person’s face.

2.2.3 Face Recognition Techniques for other Sensors Data

Although the bulk of the research on face recognition has been focused on identifying individuals from 2D intensity images, in recent years some attention has nevertheless been directed towards exploiting other sensing modalities, such as 3D or range data and infra-red imagery, for this purpose.

Just like it did for image and video-based facial recognition literature, the paper presents a high-level description of various face recognition approaches using multi-sensor data. However, for the purpose of the ATC project, the input data is limited to still imagery and video visible spectrum data. Thus, the facial recognition approaches using other sensor data, highlighted in this paper, will not be summarized here.

2.3 Literature Addressing Challenges in Face Recognition

Over the last decade, academic computer vision researchers and commercial product developers have improved the performance of automated face recognition algorithms on a variety of challenging face recognition tasks. Today, there are many state-of-the-art FR systems that perform sufficiently well, at least in controlled environment applications. However, the need for robust and reliable FR systems is more urgent in real-world security situations and uncontrolled environments. Thus, the real challenge and objective in face detection and recognition technologies is the ability to handle all these practical and uncontrolled real-world applications where subjects are non-cooperative and the acquisition phase is unconstrained. There are numerous factors that cause the appearance of the face to vary. These sources of variation in the facial appearance can be categorized into two groups: *intrinsic* factors and *extrinsic* ones:

- *Intrinsic factors*: These factors are due purely to the physical nature of the face and are independent of the observer. These factors can be further divided into two classes: intrapersonal and interpersonal. Intrapersonal factors are responsible for varying the facial appearance of the same person, some examples include:
  - Age
  - Facial expression
  - Facial paraphernalia (facial hair, glasses, cosmetics, etc.).

  Interpersonal factors, however, are responsible for the differences in the facial appearance of different people, some examples being ethnicity and gender.
- *Extrinsic factors*: These factors cause the appearance of the face to alter via the interaction of light with the face and the observer. These factors include:
  - Pose
  - Illumination
  - Scale and imaging parameters (e.g., resolution, focus, imaging, noise, etc.).

Following are the common problems and challenges that a face recognition system can have while detecting and recognizing faces:

Next, we shall review some of the recent works, reported in the literature, which were devoted to overcoming some of these limitations and challenges. For each considered FR challenge, we review four research papers, which are categorized as follows:

- The first paper presents a high-level literature survey of the various proposed techniques, which address the challenge under consideration.
- The remaining three papers focus on proposed specific techniques and methodologies addressing the considered challenge.

2.3.1 Pose and Illumination Variations

Pose variations in an image is a key matter of concern in face recognition in uncontrolled environment, where the subject is usually un-cooperative. As illustrated in Figure 2-8, the pose of a face changes with viewing angle of the observer and rotation in the head position. These changes in the posture strike a serious problem for the identification of the input image. A FR system can tolerate cases with small rotation angles, but it becomes a challenging problem when the rotation angle is higher and the available imagery in the database may only have the frontal view of the face. In such cases, the frontal view face image may differ significantly in pose with the input image, and the FR system may results in in false identification or no recognition results.

Figure 2-8 Variations in Pose (Source: [R-2])

Also, in uncontrolled environment, illumination or light variations between the input and database images poses another key challenge to FR systems. Illumination changes can vary the overall magnitude of light intensity reflected back from an object, as well as the pattern of shading and shadows visible in an image. The variation in illumination can drastically change the appearance of the face as shown in Figure 2-9. Indeed, varying the illumination can result in larger image differences than varying either the identity or the viewpoint of a face. In some cases, the same individual imaged with the same camera and seen with nearly the same facial expression and pose may appear dramatically different with changes in the lighting conditions. Furthermore, the difference between two images of the same person taken under varying illumination can be greater than the difference between the images of two different persons under same illumination. The problem of face recognition over changes in illumination is widely recognized to be difficult for humans and for algorithms. The difficulties posed by variable illumination conditions, therefore, remain a significant challenge for automatic face recognition systems.

In the FR literature, pose and illumination variations are usually treated together because changes in pose often lead to variation in illumination.

Figure 2-9 Variations in Illumination (Source: [R-2])

Table 2-2 lists the papers, which address the illumination and pose variation challenges, reviewed in this section. As mentioned above, the first paper ([P-2]) presents a literature survey of the various proposed techniques, which address the pose and illumination challenges, while the other three papers focus on specific reported methodologies addressing these challenges.

Table 2-2 Selected Papers on Addressing Pose and Illumination Variations

#	Paper Title	Authors	Source	Year	Type
P-2	Illumination and Pose Invariant Face Recognition: A Technical Review	Kavita. R. Singh, Mukesh. A. Zaveri, Mukesh. M. Raghuwanshi	International Journal of Computer Information Systems and Industrial Management Applications (IJCISIM) http://www.mirlabs.org/ijcisim ISSN: 2150-7988 Vol.2 (2010),pp.029-038	2010	Survey
P-3	Pose Invariant Face Recognition Under Arbitrary Illumination Based on 3D Face Reconstruction	Xiujuan Chai, Laiyun Qing, Shiguang Shan, Xilin Chen, and Wen Gao1	T. Kanade, A. Jain, and N.K. Ratha (Eds.): AVBPA 2005, LNCS 3546, pp. 956.965, 2005. © Springer-Verlag Berlin Heidelberg 2005	2005	Proposed new technique(s)
P-4	Pose and Illumination Invariant Face Recognition in Video	Yilei Xu, Amit Roy-Chowdhury, Keyur Patel	IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR '07, pp. 1-7.	2007	Proposed new technique(s)
P-5	A 3D Face Model for Pose and Illumination Invariant Face Recognition	Pascal Paysan, Reinhard Knothe, Brian Amberg	Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance, 2009. AVSS '09.	2009	Proposed new technique(s)

2.3.1.1 Paper # [P-2]

In this paper, the authors review different techniques proposed to address the illumination and pose-related challenges encountered in conducting facial recognition in uncontrolled environment and review some of the classifiers that have been successfully used for face recognition in general.

First, the authors categorize the predominant face recognition approaches, which have been proposed over the past decade into four main categories:

1. Holistic method, which uses whole face region
2. Model based methods which employ shape and texture of the face, along with 3D depth information
3. Template based face recognition, where face templates are extracted and used for recognition
4. Techniques using Neural Networks.

While the various types of techniques developed so far have solved many challenges encountered in developing automated facial recognition systems, many other challenges remain unresolved. In particular, this paper provides an overview of some of the leading FR approaches which are widely used in FR systems to overcome the illumination and pose problems.

Before addressing the pose and illumination-related challenges, the paper presents an overview of a typical facial recognition system, as discussed next.

2.3.1.1.1 FR System Overview

The paper presents an overview of a typical COTS FR system, as depicted in Figure 2-10. The steps of this system can be summarized as follows:

1. First an image of the face is acquired. This acquisition can be accomplished by digitally scanning an existing photograph or by using an electro-optical camera to acquire a live picture of a subject.
2. Secondly, face detection and tracking algorithms are employed to detect the location of any faces in the acquired image. This task is difficult, and often generalized patterns of what a face “looks like” (two eyes and a mouth set in an oval shape) are employed to pick out the faces.
3. Feature extraction being the third step is important towards the classification task. Different vendors use different methods to extract the identifying features of a face.
4. The fourth step is to compare the features generated in step three with those in a database of known faces.
5. Finally, just as the human perception system uses both local features and the whole face region to recognize a face, a machine recognition system should use both.

The paper then reviews some of the challenges faced by FR systems, as discussed next.

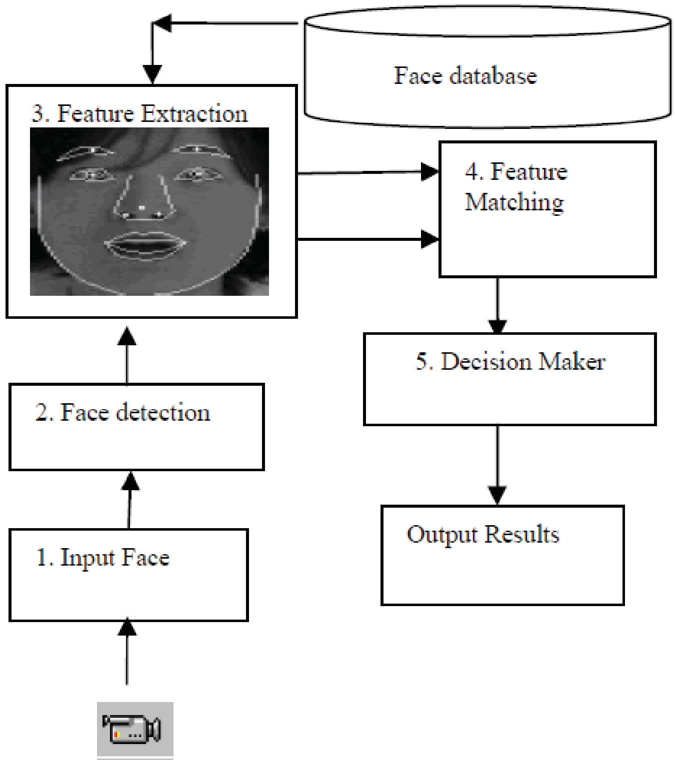


Figure 2-10 Overview of a Face Recognition System (Source: [P-2])

2.3.1.1.2 Challenges in FR Systems Related to Pose and Illumination

Figure 2-11, illustrates some illumination and expressions, illumination and pose, and lastly illumination and expression and pose variations. The paper cites FR systems studies that have revealed that most facial recognition techniques were successful on large face databases recorded in well-controlled environments. However, under uncontrolled environments their performance gets deteriorated mainly due to variations in illumination and head rotations. Such variations have proven to be one of the biggest problems of face recognition systems.

Figure 2-11 Some of the Challenges in Face Recognition Systems (Source: [P-2])

The paper then focuses on the pose and illumination challenges and reviews some of the recent literature related to addressing these challenges in facial recognition, starting with addressing the illumination variations.

2.3.1.1.3 Approaches to Illumination Variations

As illustrated in Figure 2-12, the illumination problem in facial recognition arises due to uneven lightning on faces. In the past few years, many approaches to cope up with illumination variations have been proposed. The paper broadly divides these approaches towards the illumination problem according to the following four categories:

- Transformation of images with variable illumination to a canonical representation
- Extracting illumination invariant features
- Modeling of illumination variation
- Utilization of some 3D face models whose facial shapes and albedos are obtained in advance.

Next, we summarize some of the highlighted techniques related to each of these categories.

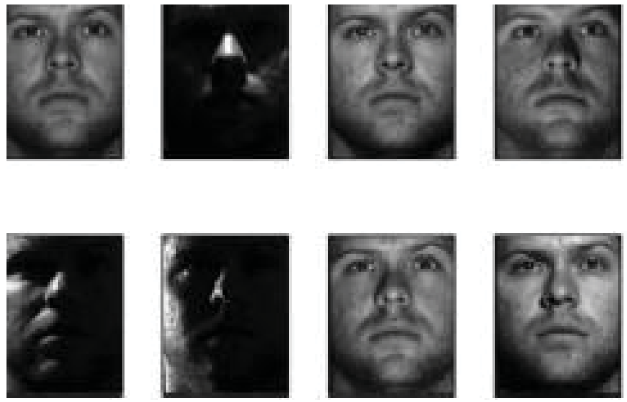


Figure 2-12 Illustration of the Illumination Variations Problem (Source: [P-2])

2.3.1.1.3.1 Transforming Images to a Canonical Representations

These approaches are based on applying PCA algorithm and discarding the first few principal components, which achieved better performance for images under different lighting conditions. The underlying assumption of these methods is that first principal components mainly capture variations due to lighting. One of the drawbacks is that some important discarded components can influence the recognition under normal lighting conditions. The paper reports that some of these techniques have demonstrated a reduction of 50–75% in recognition error rates, with a recognition rate of 98% of subjects in a database.

2.3.1.1.3.2 Extracting Illumination Invariant Features

These approaches are based on extracting only those features that are not affected by variations in lighting conditions. Some of such image representations include gradient faces, 2D Gabor Filter, DCT coefficients, and Local Binary Pattern (LBP) features. The paper highlights a few feature-based methods in each of these image representations.

2.3.1.1.3.3 Modeling of Illumination Variation

This approach is appearance-based method, where a small number of training images are used to synthesize novel images under changes in lighting and viewpoint. A convex cone, termed as illumination cone is formed from the set of images of an object in fixed pose but under all possible illumination conditions. This illumination cone can be well approximated by a low-dimensional linear subspace. Under variable lighting, the set of images is characterized by a family of illumination cones parameterized by the pose. The illumination cones for non-frontal poses can be constructed by applying an image warp on the extreme rays defining the frontal cone. To construct the illumination cone, the shape and albedo of each face is reconstructed.

2.3.1.1.3.4 Using 3D Face Models

A 3D morphable model is used to generate 3D face models from three input images from each person in the training database. Thus the 3D models are rendered under varying pose and illumination conditions to build a large set of synthetic images. Figure 2-13 illustrates an example of the triplet of training images used and two synthetic images created by rendering the newly generated 3D face model, generated from the training images.

Figure 2-13 Generation of a 3D Model (Source: [P-2])

In all the recent approaches, a 3D model of a face is utilized to transform the input image into the same pose as the stored prototypical faces, and then direct template matching is used to recognize faces. In other approaches, an Active Appearance Model of a generic face is deformed to fit to the input image, and the control parameters are used as a feature vector for classification.

2.3.1.1.4 Approaches to Pose Variations

The pose problem illustrated in Figure 2-14 where the same face appears differently due to changes in viewing condition. Post-invariance recognition capability is crucial to a face recognition system because in general it is difficult, if not possible, to control the imaging direction when acquiring images of human faces.

The paper divides approaches dealing with pose variations into three classes:

- Multiple images based approaches where multiple images per person are available during training and recognition.
- Hybrid approaches in which multiple training images are available during training but only one database image per person is available during recognition.
- Single image/shape approaches which require no training.

The paper presents and overview of some of the well-known methods in each of these categories. Next, we summarize some of the discussed general approaches for each category.

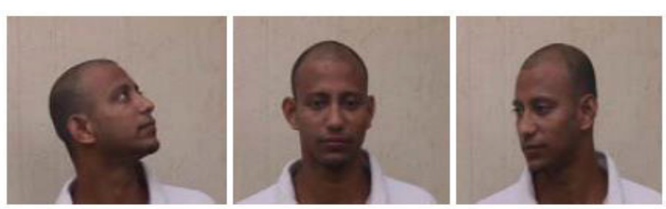


Figure 2-14 Illustration of the Pose Variations Problem (Source: [P-2])

The paper presents and overview of some of the well-known methods in each of these categories. Next, we summarize some of the discussed general approaches for each category.

2.3.1.1.4.1 Multi-image Based Approaches

In multi-image based approaches, the template based correlation matching scheme was earliest one. In this approach, pose estimation and face recognition are coupled in an iterative loop which makes the computational cost high. A new approach to handling illumination and pose variations which is motivated by the two-dimensional view-based face recognition methods having multiple Eigenspaces was also proposed. The geometrical relations between the facial feature points were used to assign facial images to the proper pose class, achieving a rate of 99.5%. The proposed method has several advantages. Since the method is based on 2D images and does not need to estimate the 3D shape, it is computationally much more efficient than the other methods based on 3D models.

2.3.1.1.4.2 Hybrid Approaches

The paper reviews three representative methods. The first one is the linear class based method, the second one is the graph matching based method and the third is the view-based-Eigenface approach. The image synthesis method is based on the assumption of linear 2D object classes and the extension of linearity to images which are 2D projections of the 2D objects. A correspondence between images of the input object and a reference object is established. And then the correspondence field for the input image is linearly decomposed into the correspondence fields for the examples. Compared to the parallel deformation scheme this method reduces the need to compute the correspondence between images of different poses.

2.3.1.1.4.3 Low-level feature based Approaches

The third class of approaches includes *low-level feature based methods*, invariant feature based methods, and the 2D model based method. The paper cites many references on invariant features in the computer vision literature. The main drawback of these methods is cited to be their complexity and computational cost. Other cited methods use a face re-rotating approach based on linear shape prediction and image warp based on the strategy of generating virtual views from one single face image, that is, synthesize novel views of the given face. Other cited approaches attempt to exploit the similarities of a face image against a set of faces from a training set at the same view to establish pose-invariant representations of a person in different poses.

2.3.1.1.5 Classifiers

In addition to outlining some of the recent approaches in the literature, which have been developed in order to address some of the pose and illumination-related challenges in conducting automated facial recognition, this paper also outlines some of the classification approaches implemented during the image or feature matching phase. Technically speaking, face recognition is actually a pattern matching classification task. That is, the given facial image is transformed into features, after which a classifier trained on example faces decides whether that particular facial image is present in the database or not. The paper highlights some of the classifiers that have been used successfully in face recognition domain. These classifiers include Euclidian distance, nearest neighbor classifier, neural network, Bayesian classifier, Hidden Markov Model (HMM), Support Vector Machine (SVM), rough neural network, Radial Basis Function (RBF) and Adaboost (Adaptive Boosting) classification techniques.

2.3.1.1.6 Concluding Remarks

The paper concludes by remarking that although, face recognition has been claimed to be an almost solved problem; recognition under uncontrolled conditions remains a field of research, as there still many challenges and unresolved issues. This review paper mainly focuses on the research efforts dedicated to addressing illumination and pose problems. Most of the reviewed strategies claim satisfactory recognition rates only when tested on standard databases or some part of them. It is also noted that one particular database, Carnegie Mellon University- Pose, Illumination and Expression (CMU-PIE) as described in Table A-1, has been widely used for pose and illumination problem. Similarly, for the reviewed classifiers, the reported recognition rates depend on the database used and the number of subjects on which classification task has been performed.

2.3.1.2 Paper # [P-3]

As mentioned earlier, the review of this paper, which proposes a new pose and illumination FR approach, is organized in subsections summarizing its abstract, the motivation of the undertaken work, the proposed method, sample experimental results and the key contributions of the paper.

2.3.1.2.1 Abstract

In this paper, a novel method to handle both pose and lighting condition simultaneously is proposed, which calibrates the pose and lighting condition to a pre-set reference condition through an illumination invariant 3D face reconstruction. First, some located facial landmarks and a priori statistical deformable 3D model are used to recover an elaborate 3D shape. Based on the recovered 3D shape, the texture image calibrated to a standard illumination is generated by spherical harmonics ratio image and finally the illumination independent 3D face is reconstructed completely. The proposed method combines the strength of statistical deformable model to describe the shape information and the compact representations of the illumination in spherical frequency space, and handle both the pose and illumination variation simultaneously. This algorithm can be used to synthesize virtual views of a given face image and enhance the performance of face recognition. The experimental results on CMU PIE database, as described in Table A-1, show that this method can significantly improve the accuracy of the existed face recognition method when pose and illumination are inconsistent between gallery and probe sets.

2.3.1.2.2 Motivation

Pose and illumination changes from picture to picture are two main barriers toward full automatic face recognition. The appearance of faces may looks quite different when pose or illumination change, and this issues an imperfect task for face recognition when only the 2D appearance-based method is applied. Although some 2D-based methods are proposed to tackle pose or illumination variation problem, the authors believe that a 3D-based method is a better approach to address both pose and illumination blending problem since the pose and illumination variations are both related to the 3D face structure.

2.3.1.2.3 Method

The framework of the proposed pose and illumination calibration for face recognition is given in Figure 2-15. First, the irises are located by a region growing searching algorithm and the rude pose class is defined for labeling the sparse feature points in the given facial image. Then 3D shape is reconstructed based on a 2D geometry driven statistical deformable model. Recurring to the recovered 3D shape of the specific person, the illumination independent "texture image" is obtained by relighting the face region extracted from the given image with spherical harmonic ratio image strategy. The pose and lighting calibrated image are used as the input of face recognition and get the identity result. The proposed algorithm can be regarded as a "*pre-processing*" step of any face recognition system.

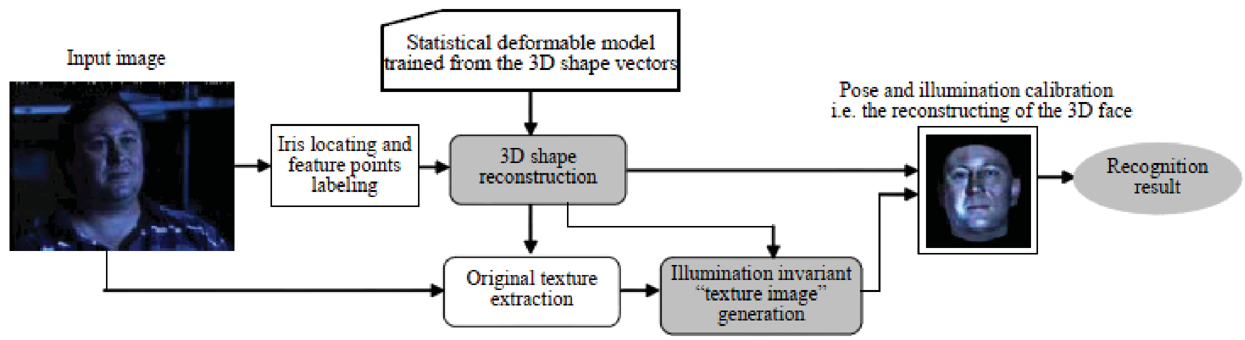


Figure 2-15 Framework of Pose and Illumination Calibration for Face Recognition (Source: [P-3])

Note that the paper does not explore developing new FR algorithms. Its main focus is to pre-process the input data in order to reduce or eliminate the effects of the pose or illumination variations. The pre-processed data is then fed into one of the many FR systems found in the literature.

2.3.1.2.4 Experimental Results

The paper evaluates the performance of the proposed approach for conducting pose and illumination invariant face recognition, using input data acquired with different illumination and pose parameters.

2.3.1.2.4.1 Database

The evaluation is based on subset of the CMU PIE database, as described in Table A-1, which provides the facial images under well-controlled poses and lightings.

2.3.1.2.4.2 FR Algorithm(s)

As mentioned above, the proposed approach is a pre-processing of the input data in order to reduce or eliminate the effects of the pose and illumination variations. The pre-processed data is then fed into one of the many FR systems found in the literature. In this paper, the Gabor PCA plus LDA facial recognition algorithms are applied on the images, after they have been pre-processed using the proposed method.

2.3.1.2.4.3 Sample Results

Some of the experimental results for the pose and illumination variations are presented here.

Pose Variation Experiment

First, the experiment on face recognition across pose only is carried out on 4 pose subsets of CMU PIE database, which are pose set 05 (turn right 22.5 degree), pose set 29 (turn left 22.5 degree), pose set 37 (turn right 45 degree) and 11 (turn left 45 degree) respectively, and the gallery images are from the pose set 27, which are all frontal images.

Some pose normalization results based on the 3D face reconstruction are presented in Figure 2-16 to give a visualize evaluation in the 4 pose sets. The first row is the original masked images. The second row is the corresponding pose normalized images, and right to which are the gallery images in Pose 27 to be references.

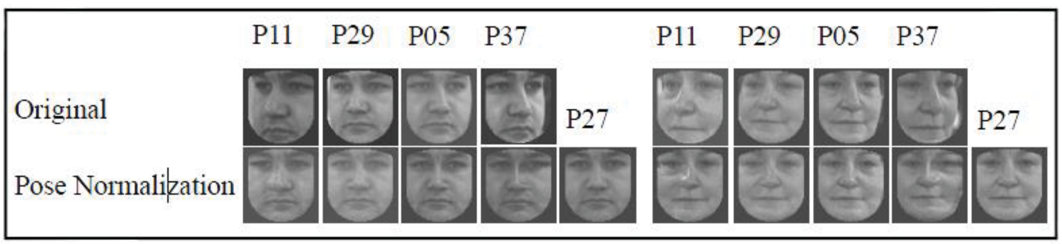


Figure 2-16 Pose Normalized Images (Source: [P-3])

The recognition results are listed in Figure 2-17, which shows the good performance of the pose normalization based on the proposed 3D face reconstruction. The recognition match scores for the 4 pose sets are improved significantly compared with the original recognitions, and the rank-1 recognition rate reaches to 94.85% on average, after pose normalization. Throughout this report, *Rank-1* recognition rate is the ratio between number of faces correctly recognized and the number of probe faces. In general, *Rank-n* means that the *n* highest scored faces instead of one face are selected.

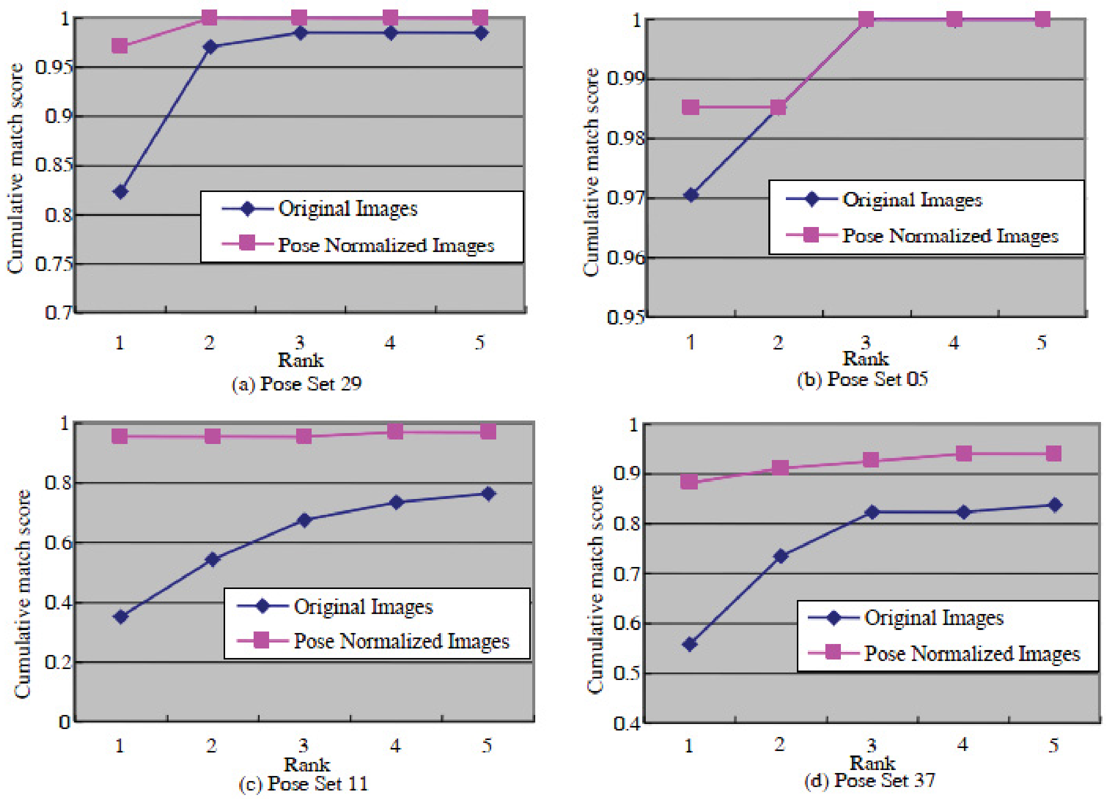


Figure 2-17 Recognition Results on the Original and the Pose Normalized Images in the 4 Different Pose Sets (Source: [P-3])

Illumination Variation Experiment

An experiment is conducted on 2856 images from the 2 pose subsets, 05 and 29, each subset including 21 different kinds of illuminations and the flash numbers are 02-21. The frontal pose set 27 under flash “11” is taken as the gallery, and the other probe images are all aligned to the frontal pose and the standard light as flash number “11”. The experimental results of face recognition with correlation matching strategy across pose and lighting are listed in Table 2-3, and they illustrate the recognition results on 2 pose subsets under 21 different lighting conditions in CMU PIE database with the correlation matching strategy used for recognition.

Table 2-3 Recognition Results on 2 Pose Subsets under 21 Different Lighting Conditions (Source: [P-3])

F\C	Pose 05 (original)	Pose 05 (calibrated)	Increase	Pose 29 (original)	Pose 29 (calibrated)	Increase
02	0.044	0.206	0.162	0.015	0.235	0.230
03	0.059	0.412	0.353	0.029	0.324	0.295
04	0.103	0.735	0.632	0.059	0.612	0.553
05	0.397	0.897	0.500	0.103	0.882	0.779
06	0.735	0.882	0.147	0.162	0.926	0.764
07	0.676	0.912	0.236	0.118	0.912	0.794
08	0.544	0.897	0.353	0.588	0.956	0.368
09	0.235	0.897	0.662	0.676	0.985	0.309
10	0.324	0.912	0.588	0.088	0.838	0.750
11	0.676	0.912	0.236	0.838	0.971	0.133
12	0.309	0.926	0.617	0.765	0.941	0.176
13	0.074	0.868	0.794	0.221	0.882	0.661
14	0.088	0.897	0.809	0.235	0.912	0.677
15	0.029	0.750	0.721	0.059	0.750	0.691
16	0.029	0.368	0.339	0.044	0.471	0.427
17	0.015	0.221	0.206	0.029	0.279	0.250
18	0.250	0.838	0.588	0.074	0.750	0.676
19	0.647	0.912	0.265	0.118	0.926	0.808
20	0.662	0.912	0.250	0.838	0.971	0.133
21	0.265	0.926	0.661	0.706	0.941	0.235
22	0.074	0.838	0.764	0.118	0.824	0.706
Average	0.296	0.768	0.472	0.280	0.776	0.496

2.3.1.2.5 Key Contributions

In this paper, a novel illumination independent 3D face reconstruction is proposed to recognize facial images across pose and illumination. The key features of the proposed method are as follows:

- The 3D shape is recovered from single non-frontal facial image based on a statistical deformable model regressed through 2D geometry formed by some facial landmarks.
- Recurring to the reconstructed 3D shape, the illumination independent facial “texture image” is achieved with spherical harmonic ratio image.
- The experimental results show that the pose and illumination calibrating strategy largely improves the performance of the general face recognition for the probe images under uncontrolled pose and lighting.
- A key feature of the proposed method is that it is not a standalone FR approach, but rather it can be regarded as a “*pre-processing”* step of any face recognition system.

2.3.1.3 Paper # [P-4]

2.3.1.3.1 Abstract

The paper presents a framework for face recognition from video sequences that is robust to large changes in facial pose and lighting conditions. The proposed method is based on a recently obtained theoretical result that can integrate the effects of motion, lighting and shape in generating an image using a perspective camera. This result can be used to estimate the pose and illumination conditions for each frame of the probe sequence. Then, using a 3D face model, images corresponding to the pose and illumination conditions estimated in the probe sequences are synthesized. Similarity between the synthesized images and the probe video is computed by integrating over the entire sequence. The method can handle situations where the pose and lighting conditions in the training and testing data are completely disjoint.

2.3.1.3.2 Motivation

Many that video-based face recognition systems hold promise in certain applications where motion can be used as a cue for face segmentation and tracking, and the presence of more spatio-temporally coherent data can increase recognition performance. However, these systems have their own challenges such as tracking the face over time, 3D modeling, changes of pose and lighting over the sequence duration, and developing efficient measures for integrating information over the entire sequence.

In this paper, the spatio-temporally coherent data is exploited further in order to develop a novel framework for video-based face recognition that is based on learning joint illumination and motion models from video, synthesizing novel views based on the learned parameters, and using a metric that can compare two time sequences while being robust to outliers.

2.3.1.3.3 Method

The proposed method for estimating the illumination and 3D motion parameters is based on reported recent studies, which have shown that for a fixed Lambertian object, the set of reflectance images under distant lighting without cast shadows can be approximated by a linear combination of nine basis images, defined using spherical harmonics. Several reported papers have shown the suitability of this model for faces.

The authors expand on these previous contributions to show that each new frame of a video sequence spans a bilinear space of six motion and approximately nine illumination variables, assuming that the motion between two consecutive frames is small enough. They also show that, when the illumination changes gradually, it can be shown that the bilinear space becomes a combination of two linear subspaces, defined by the motion and illumination variables. This result is valid for both slow and sudden changes in illumination and also accounts for attached shadows. It can handle large changes of scale of the object. Another nice feature of the proposed formulation is that it depends upon the image intensities, not the image gradients; this will make the algorithms proposed for the inverse problems described later less sensitive to image noise. The theory is applicable to color images by treating each color channel separately.

In this work, it is assumed that, the gallery is represented by a 3D model of the face. The model can be built from a single image or video sequence, using cited techniques from found in the literature, or can be obtained directly from 3D sensors. For the purpose of face recognition, given a probe sequence, the motion and illumination conditions are estimated using the proposed approach. The 3D model is registered to the first frame of the probe sequence manually by choosing five points on the face. Given the motion and illumination estimates, images are then rendered from the 3D models in the gallery. The rendered images can then be compared with the images in the probe sequence. Given the rendered images from the 3D models in the gallery and the probe images, a robust metric for comparing these two sequences is then applied. This metric computes the distance between the frames in the probe and each synthesized sequence and chooses the identity as the individual with the smallest distance in the gallery.

2.3.1.3.4 Experimental Results

The paper evaluates the performance of the proposed algorithm through pose and illumination invariant face recognition.

2.3.1.3.4.1 Database

One of the challenges in video-based face recognition is the lack of a good dataset, unlike in image-based approaches, which usually use the CMU PIE dataset. The authors collected a new set, which has large, simultaneous pose, illumination and expression variations. It is similar to the PIE dataset, but with video sequences instead of discrete poses and uses pre-existing natural indoor and outdoor lighting that changes randomly. An example of some of the images in the video database is shown in Figure 2-18.

Figure 2-18 Sample Frames from the Collected Video Sequences Database (Source: [P-4])

2.3.1.3.4.2 Sample Results

A 3D model of each face was from constructed from the gallery sequence, using a cited method from the literature. As can be seen from Figure 2-18, the pose and illumination varies randomly in the video. From the portion of each sequence designated as probe, three experiments were designed by choosing different parts of it, as described below:

- Exp. A: A video sequence was used as the probe with the average pose of the face in the video being about 15 degrees from frontal.
- Exp. B: A video sequence was used as the probe with the average pose of the face in the video being about 30 degrees from frontal.
- Exp. C: A video sequence was used as the probe with the average pose of the face in the video being about 45 degrees from frontal.

Each probe sequence has about 20 frames around the average pose. The variation of pose in each sequence was less than 10 degrees.

The results on tracking and synthesis on three of the probes are shown in Figure 2-19, which illustrates the Cumulative Match Characteristic or Cumulative Match Curve (CMC) for the three experiments with distance measure equals to unity. The CMC is used as a measure of 1: m identification system performance. It judges the ranking capabilities of an identification system. It measures how well an identification system ranks the identities in the enrolled database with respect to “unknown” probe image. The proposed algorithm gives relatively high performance – about 95% for Exp. B and 93% for Exp. C, which deals with video sequences with an average pose far away from frontal. In Exp. A, where pose is 15 degrees away from frontal, recognition rate reaches 100% on videos with large and arbitrary variations illumination changes.

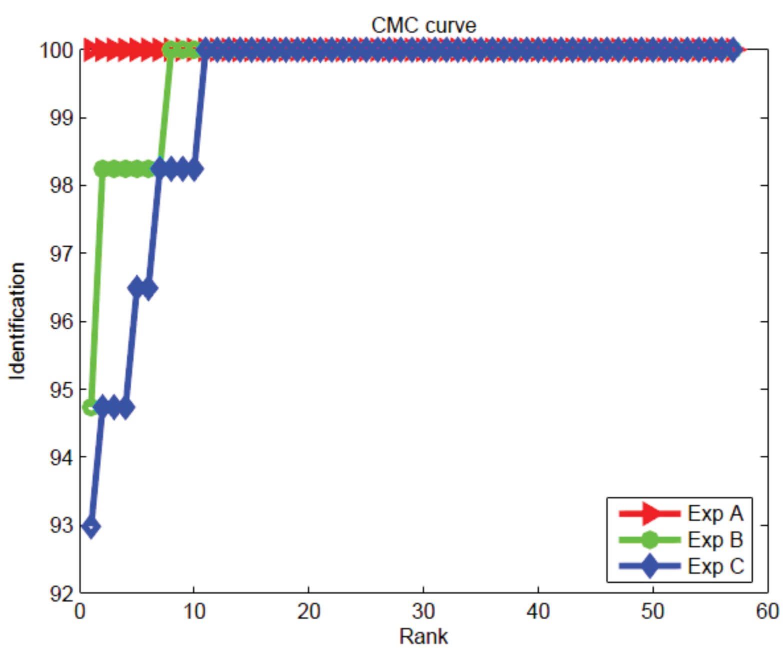


Figure 2-19 CMC Curve for Described Video-Based FR Experiments A to C (Source: [P-4])

2.3.1.3.5 Key Contributions

This paper proposes a method for video-based face recognition that relies upon a novel theoretical framework for integrating illumination and motion models for describing the appearance of a video sequence. The key features of the proposed method are as follows:

- The recognition algorithm relies on synthesis of video sequences under the conditions of the probe.
- The effectiveness of the method was demonstrated on video databases with large and arbitrary variations in pose and illumination.
- The method assumes that the gallery is already represented by a 3D model of the face. The model can be built from a single image or video sequence, using cited techniques from found in the literature, or can be obtained directly from 3D sensors.

2.3.1.4 Paper # [P-5]

2.3.1.4.1 Abstract

This paper publishes a generative 3D shape and texture model, the Basel Face Model (BFM), and demonstrates its application to several face recognition task. The proposed model improves on previous models by offering higher shape and texture accuracy due to a better scanning device and less correspondence artifacts due to an improved registration algorithm. The same 3D face model can be fit to 2D or 3D images acquired under different situations and with different sensors using an analysis by synthesis method. The resulting model parameters separate pose, lighting, imaging and identity parameters, which facilitates invariant face recognition across sensors and data sets by comparing only the identity parameters. The developed registered face model is made available as open source in order to facilitate research in generative models.

2.3.1.4.2 Motivation

Automatic face recognition from a single image is still difficult for non-frontal views and complex illumination conditions. To achieve pose and light invariance, 3D information of the object is useful. For this reason, 3D Morphable Models (3DMM) have been introduced. However, the widespread use of 3DMMs has been held back by their difficult construction process, which requires a precise and fast 3D scanner, the scanning of several hundreds of individuals and the computation of dense correspondence between the scans. Numerous face recognition articles acknowledge the fact that 3DMM based face image analysis constitutes the state of the art, but note that the main obstacle resides in the complications of their construction. Hence, there is a demand from the face image analysis community for a publicly available 3D Morphable Face Model. The aim of this paper is to fill this gap and describe 3D Morphable Model (3DMM) of the human face that is made publicly available.

2.3.1.4.3 Method

This paper proposes 3DMM of the human face known as the Basel Face Model (BFM). The construction of a 3DMM requires a training set with a large variety in face shapes and appearance. The training data should be a representative sample of the target population. To make the raw data usable it needs to be brought in correspondence by applying a registration algorithm. After registration the faces are parameterized as triangular meshes with m = 53490 vertices and shared topology. Facial recognition is then achieved by matching these vectors of the input image to the vectors corresponding to the database images and selecting the images with minimum distance between these vectors.

2.3.1.4.4 Experimental Results

The paper evaluates the performance of the proposed algorithm using input data with pose and illumination variations.

2.3.1.4.4.1 Database

The generated BFM standard training set for face recognition algorithms is used together with test sets such as Facial Recognition Technology (FERET), CMU-PIE and UND, as described in Table A-1. This allows for a fair, data independent comparison of face identification algorithms.

2.3.1.4.4.2 Sample Results

To demonstrate the quality of the presented model it is assessed using the CMU-PIE/FERET datasets. None of the individuals in the test sets is part of the training data for the 3DMM approach. The test sets cover a large ethnic variety.

The subset of the FERET data set consists of 194 individuals across 9 poses at constant lighting condition except the frontal view taken under a different illumination condition.

The subset of the CMU-PIE data set consists of 68 individuals (28 wearing glasses) at 3 poses (frontal, side and profile) under illumination from 21 different directions and ambient light only.

To perform the identification, the BFM is fitted to the images of the test sets. In the fitting process, three error terms based on landmarks, the contour and the shading are optimized. To extend the flexibility, four facial regions (eyes, nose, mouth, etc.) are fitted separately and combined later by blending them together. The obtained shape and albedo model parameters for the global fitting α_0 and β_0 and for the facial segments $\alpha_{1,\ldots,}\alpha_4$ and $\beta_{1,\ \ldots,}\beta_4$ represent the identity in the model space. These parameters are stacked together into one identity vector:

$$c = (\alpha_0, \beta_0, \ldots, \alpha_4, \beta_4).$$

The similarity of two scans is measured by the angle between their identity vectors.

Figure 2-20 illustrates an example of fitting result for CMU-PIE with BFM Face Model: the original image, the fitting result rendered into the image and right the resulting 3D model, respectively from left to right.

Figure 2-20 Exemplary Fitting Result from CMU-PIE Database with BFM Face Model (Source: [P-5])

Table 2-4 and Table 2-5 list the percentages of correct *rank-1* identification obtained on the CMU-PIE and the FERET subset, respectively. As expected, the best results are obtained for frontal views.

Table 2-4 Rank 1 Identification Results Obtained on a CMU-PIE Subset (Source: [P-5])

Gallery / Probe	front	side	profile	mean
front	98.9 %	96.1 %	75.7 %	90.2 %
side	96.9 %	99.9 %	87.8 %	94.9 %
profile	79.0 %	89.0 %	98.3 %	88.8 %
mean	91.6 %	95.0 %	87.3 %	91.3 %

Table 2-5 Rank 1 Identification Results Obtained on a FERET Subset (Source: [P-5])

Gallery / Probe	Pose Φ	Identification rate
bb	38.9°	97.4 %
bc	27.4°	99.5 %
bd	18.9°	100.0 %
be	11.2°	Gallery
ba	1.1°	99.0 %
bf	-7.1°	99.5 %
bg	-16.3°	97.9 %
bh	-26.5°	94.8 %
bi	-37.9°	83.0 %
bk	0.1°	90.7 %
mean		95.8 %

2.3.1.4.5 Key Contributions

In this paper, 3DMM of faces was developed and made publicly available. The key features of the model include:

- The model addresses the lack of universal training data for face recognition. Due to its 3D structure it can be used indirectly to generate images with any kind of pose and light variation or directly for 2D and 3D face recognition.
- Using these standardized training and test sets makes it possible for researchers to focus on the comparison of algorithms independent of the data.
- One of the key contributions of this paper is that the developed 3DMM is made available as open source in order to facilitate research in generative models.

2.3.2 Occlusion and Disguise

Occlusion is one of the important challenges of the face recognition as shown in Figure 2-21. This is due to presence of various occluding objects such as glasses, beard, moustache etc. on the face and when an image is captured from a surveillance camera; the face lacks some parts. In real world applications also it is very common situation to acquire imagery of persons talking on the phone or wearing hats, glasses, scarves, etc. Occlusion and disguise pose major challenges to facial recognition systems. Most of the commercial face recognition engines reject an input image when the eyes cannot be detected.

Figure 2-21 Examples of Partial Facial Occlusion (Source: [R-2])

Table 2-6 lists the specific facial recognition papers, which address the occlusion and disguise variation challenges, reviewed in this section.

Table 2-6 Selected Papers on Addressing the Occlusion and Disguise Challenges

#	Paper Title	Authors	Source	Year	Type
P-6	Recognizing Face Images with Disguise Variations	Richa Singh, Mayank Vatsa and Afzel Noore	Recent Advances in Face Recognition, Book edited by: Kresimir Delac, Mislav Grgic and Marian Stewart Bartlett, ISBN 978-953-7619-34-3, pp. 236, December 2008, I-Tech, Vienna, Austria	2008	Survey

#	Paper Title	Authors	Source	Year	Type
P-7	An improved Robust Sparse Coding for Face Recognition with Disguise	Dexing Zhong, Peihong Zhu, Jiuqiang Han, Shengbin Li	International Journal of Advanced Robotic Systems, 2012, Vol. 9, 126:2012, pp. 1 – 7.	2012	Proposed new technique(s)
P-8	Face Recognition with Occlusion Using Dynamic Image-to-Class Warping (DICW)	Xingjie Wei, Chang-Tsun Li and Yongjian Hu	10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 2013, pp. 1 – 6.	2013	Proposed new technique(s)
P-9	Face Recognition with Occlusion in Both Reference and Query Images	Xingjie Wei, Chang-Tsun Li and Yongjian Hu	International Workshop on Biometrics and Forensics (IWBF), 2013, pp. 1-4.	2013	Proposed new technique(s)

2.3.2.1 Paper # [P-6]

In this paper, the authors review different techniques proposed to address the partial occlusion challenges encountered in conducting facial recognition in uncontrolled environments and review some of the classifiers that have been successfully used for face recognition in general.

The paper begins by summarizing the challenges in automatic face recognition in uncontrolled environments, which were classified into six categories: illumination, image quality, expression, pose, aging, and disguise. Among these challenges, the paper focuses on recognition of faces with disguise as a major challenge and has only been recently addressed by few researchers. As shown in Figure 2-22, the inter-personal and intra-personal characteristics can be modeled using disguise accessories to alter the appearance of an individual, to impersonate another person, or to hide one’s identity. The challenges due to disguise cause changes in visual perception, alter actual data, make pertinent facial information disappear, mask features to varying degrees, or introduce extraneous artifacts in the face image. Existing face recognition algorithms may not be able to provide the desired level of security for such cases.

Figure 2-22 Face disguise: Use of Makeup Tools and Accessories to Alter Facial Features and Appearance of the Same Individual (Source: [P-6])

2.3.2.1.1 Types of disguises

The paper classifies the possible variations of disguise into the following eight categories depending on their effect on facial appearance and features:

1. **Minimal variations**: Two face images captured at different time instances can have minimal variations in appearance and features. In such cases, face recognition algorithms usually yield correct results.
2. **Variations in hair style**: Hair style can be changed to alter the appearance of a face image or hide facial features. Figure 2-23 shows an example of facial variations of an individual due to changes in hair style.

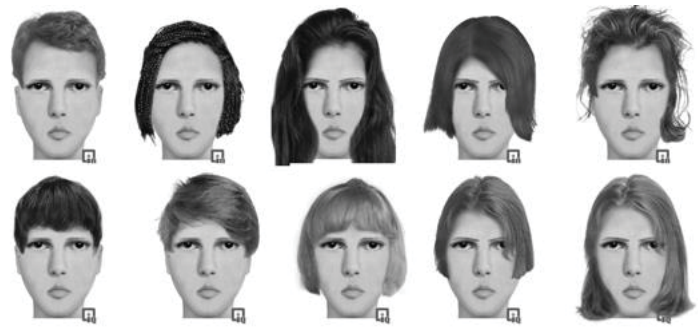


Figure 2-23 Face Images with Variation in Hair Style (Source: [P-6])

3. **Variations due to beard and moustache**: Facial hair such as beard and moustache can alter facial appearance and features in the lower half of the face, specifically near mouth and chin regions. Figure 2-24 shows an example where face images with and without beard and moustache show different appearance.

Figure 2-24 Face Images with Variation in Beard and Moustache (Source: [P-6])

4. **Variations due to glasses**: Glasses, especially sun-glasses are one of the easiest ways to alter facial appearance. In general, glasses affect upper facial region by hiding the facial features (e.g. eyes and eyebrows). As shown in Figure 2-25, structural differences in glasses and opacity of lens can also change the appearance of an individual.

Figure 2-25 Face Images with Variation in Eye Glasses (Source: [P-6])

5. **Variations due to cap and hat**: In general, use of cap and hat hides hairs and some part of the forehead, which are not used by the recognition algorithms. However, as shown in Figure 2-26, some specific types of cap and hat (e.g., monkey cap) can hide pertinent facial features, thereby affecting the performance of face recognition algorithms.

Figure 2-26 Face images with Variation in Cap and Hat (Source: [P-6])

6. **Variations due to lips, eyebrows and nose**: Makeup tools can be used to alter the shape and size of lips, eyebrows and nose, as illustrated in Figure 2-27. These key local features, if altered, can affect the performance of feature based face recognition algorithms.

Figure 2-27 Face Images with Variation in Lips, Eyebrow and Nose Characteristics (Source: [P-6])

7. **Variations due to aging and wrinkles**: Aging can be natural (due to age progression) and artificial (using makeup tools). In both the cases, aging and wrinkles can severely affect the performance of face recognition algorithms. An example of facial variations due to aging and wrinkles is shown in Figure 2-28.

Figure 2-28 Face Images with Aging and Wrinkle Variations (Source: [P-6])

8. **Multiple variations**: A combination of the above mentioned variations can be used to disguise and defraud law enforcement. Figure 2-29 shows an example in which multiple variations are used to alter the appearance and features of an individual.

Figure 2-29 Face Images of an Individual with Multiple Disguise Variations (Source: [P-6])

2.3.2.1.2 Evaluated FR Algorithms

In general, face recognition algorithms can be broadly classified into three classes (1) appearance based algorithms, (2) feature based algorithms, and (3) texture based algorithms. To compare the performance of these algorithms; eight algorithms are selected and are briefly explained below.

1. **Appearance based algorithms**: Three appearance based algorithms are used in experiments that are specifically tailored for recognizing individuals with altered appearances. These algorithms are: (1) PCA algorithm with Mahalanobis distance, (2) Half-face based algorithm, and (3) Eigen-eyes based algorithm.
2. **Feature based algorithms**: Two feature based algorithms are selected namely, Geometrical Feature (GF) and Local Feature Analysis (LFA). Geometrical feature based recognition algorithm uses mixture distances of the facial features for matching. This algorithm works on the distance between geometrical features. Facial features such as nose, mouth, eyes, and ears are extracted and their shape information is computed. For matching two images, this shape information is matched using Euclidean distance measure. Hence, the algorithm depends on the correspondence between the facial features and works only in cases when this information is preserved. If the facial features are occluded using accessories such as glasses, beard, moustache, and scarf, then the performance decreases. Local feature analysis (is one of the most widely used face recognition algorithms which can accommodate some changes in facial expression. LFA refers to a class of algorithms that extract a set of geometrical metrics and distances from facial images and use these features as the basis for representation and comparison. The recognition performance is dependent on a relatively constant environment and quality of the image.
3. **Texture based algorithms**: In the third class of face recognition algorithms, i.e., texture based, three algorithms are selected namely Independent Gabor Features (IGF), Local Binary Pattern (LBP), and 2D-log polar Gabor transform and Neural Network (2DLPGNN). Independent Gabor features-extract Gabor features from the face image and then reduces the dimensionality using PCA. The independent Gabor features are obtained from the reduced dimensionality feature vector by applying Independent Component Analysis. These independent Gabor features are classified using Bayes classifier and then matched using the Manhattan distance measure. Local binary pattern based face recognition algorithm extracts textural feature from the face images. In this algorithm, a face image is divided into several regions and weighted LBP features are extracted to generate a feature vector. Matching of two LBP feature vectors is performed using weighted Chi square distance measure based algorithm.

2.3.2.1.3 Face Databases

The experiments are divided into two parts. In the first part, the performance of face recognition algorithms is evaluated on a heterogeneous face database that contains variations due to pose, expression and illumination. This experiment is performed as the baseline experiment. The second experiment is performed to evaluate the effect of disguises on face recognition algorithms. For these experiments, two face databases were created: (1) heterogeneous face database and (2) face disguise database.

1. **Heterogeneous face database**: For evaluating the performance on a large database with challenging intra-class variations, images from multiple face databases were combined to create a heterogeneous database of 882 subjects. Table 2-7 lists the databases used and the number of subjects selected from the individual databases.

Table 2-7 Composition of the Heterogeneous Face Databases (Source: [P-6])

Face database	Number of subjects
CMU-AMP	13
CMU-PIE	65
Equinox	90
AR	120
FERET	300
Faces in the Wild	294
Total	882

2. **Face disguise database**: This database was prepared by the authors. It contains real and synthetic face images from 125 subjects. For every subject, disguise variations are collected based on the eight classes of disguise variations described above. The database contains real face images of 25 individuals with 15-25 different disguise variations of each individual. An example of the synthetic face database is shown in Figure 2-30. The complete face disguise database is used to broadly evaluate the performance of the proposed algorithm for disguised images.

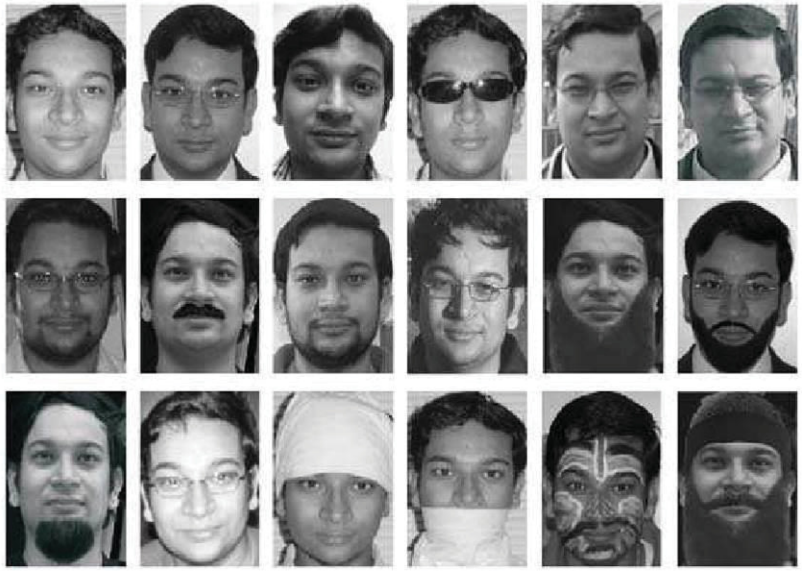


Figure 2-30 Sample Images from the Real Face Database of the Same Individual (Source: [P-6])

2.3.2.1.4 Performance Evaluation

The experiments are divided into two parts:

1. evaluation using the heterogeneous face database, and
2. evaluation using the disguise database.

The Experimental protocol adopted in this paper is as follows: For both experiments, the images are partitioned into two sets:

1. the training dataset is used to train the individual face recognition algorithms and
2. the gallery-probe dataset (the test set) is used to evaluate the performance of the recognition algorithms.

The training set comprises of randomly selected three images of each subject and the remaining images are used as the test data to evaluate the algorithms. This train-test partitioning is repeated 20 times (cross validation) and the Receiver Operating Characteristics (ROC) curves are generated by computing the Genuine Accept Rates (GAR) over these trials at different False Accept Rates (FAR). Furthermore, verification accuracies are reported at 0.01% FAR.

2.3.2.1.4.1 Evaluation using Heterogeneous Face Database

This experiment is conducted to evaluate the effects of three covariates namely pose, expression, and illumination on the performance of face recognition. Figure 2-31 shows the ROC plot and Table 2-8 illustrates the verification accuracies at 0.01% FAR.

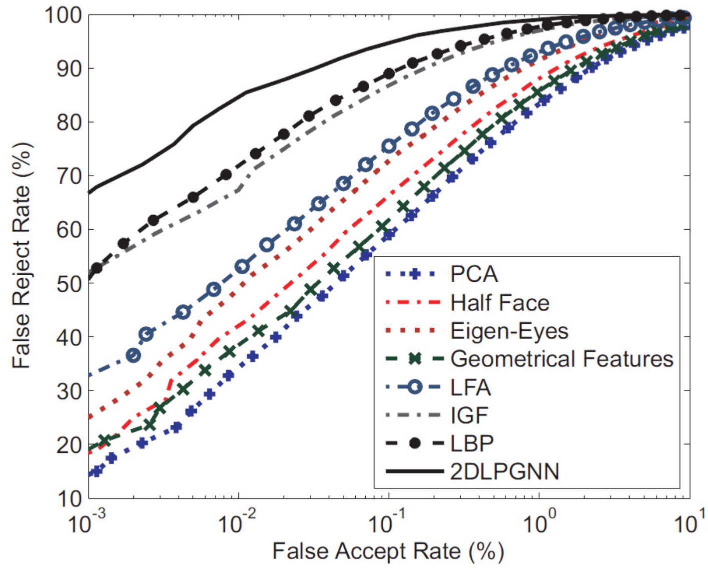


Figure 2-31 ROC to Evaluate the Performance of FR Algorithms on the Heterogeneous Face Database (Source: [P-6])

Table 2-8 Verification Performance of Appearance, Feature and Texture Based FR Algorithms for Different Covariates (Source: [P-6])

Verification accuracy at 0.01% FAR								
Covariates	Appearance based algorithms			Feature based algorithms		Texture based algorithms		
	PCA	Half-face	Eigen-eyes	GF	LFA	IGF	LBP	2D-LPGNN
Pose	31.9	36.6	29.7	35.4	50.1	60.7	73.2	75.3
Expression	35.5	42.1	78.6	38.2	52.3	69.8	71.4	87.6
Illumination	35.3	45.2	45.8	41.7	53.5	68.6	70.9	86.5
Overall	34.4	41.8	49.3	38.9	52.7	67.3	72.1	84.2

2.3.2.1.4.2 Evaluation using Disguise Face Database

This experiment is conducted to evaluate the performance of face recognition algorithms for each disguise using the disguise database. Figure 2-32 shows the ROC plot and Table 2-9 illustrates the verification accuracies at 0.01% FAR.

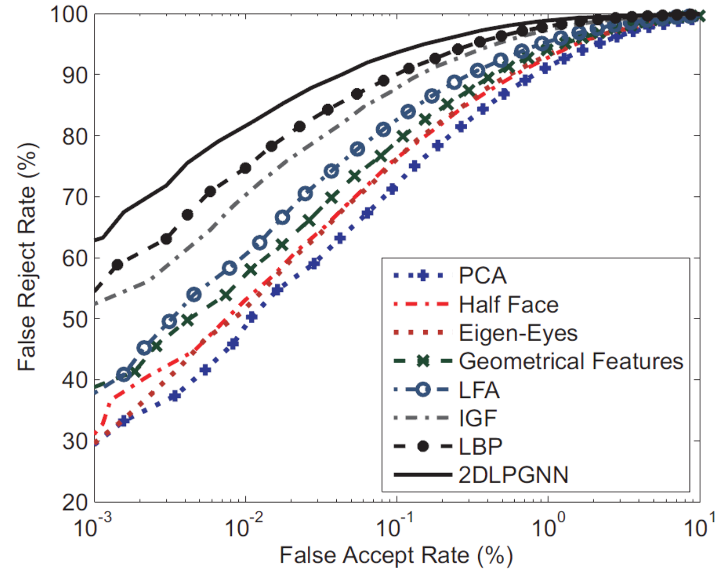


Figure 2-32 ROC to Evaluate the Performance of FR Algorithms on Face Disguise Database (Source: [P-6])

Table 2-9 Verification Performance of Appearance, Feature and Texture Based FR Algorithms for Different Disguise Variations (Source: [P-6])

Verification accuracy at 0.01% FAR								
Variations	Appearance based algorithms			Feature based algorithms		Texture based algorithms		
	PCA	Half-face	Eigen-eyes	GF	LFA	IGF	LBP	2DLPGNN
Minimal variations	60.3	59.5	63.1	61.3	63.7	74.2	85.5	96.9
Hair	56.8	61.4	57.2	61.9	63.1	73.8	85.1	96.4
Beard and moustache	32.2	34.1	60.5	53.2	54.5	58.7	61.0	77.3
Glasses	41.4	44.6	6.9	52.4	53.8	57.6	62.5	81.9
Cap and hat	55.7	56.8	50.4	58.9	61.4	71.0	80.4	86.3
Lips, nose, and eyebrow	56.3	59.2	47.7	49.1	56.3	70.9	78.6	89.2
Aging and wrinkles	49.6	53.9	41.6	51.8	54.9	55.1	70.3	80.8
Multiple variations	14.7	16.4	30.3	31.6	32.0	32.8	50.3	65.6
Overall	48.2	52.9	51.1	58.0	60.4	70.1	74.7	82.0

These comprehensive experimental results show that most tested methods yield relatively low verification rates, which indicates that further research is needed to address high degree of disguise variations.

2.3.2.1.5 Summary and Concluding Remarks

Currently, many security applications use human observers to recognize the face of an individual. In some applications, face recognition systems are used in conjunction with limited human intervention. For autonomous operation, it is highly desirable that the face recognition systems be able to provide high reliability and accuracy under multifarious conditions, including disguise. However, most of the algorithms are not robust to high security applications such as border crossing and terrorist watch list, when an individual attempts to defraud law enforcement by altering his or her physical appearance with disguises. This chapter emphasizes this important aspect of face recognition. It describes different types of disguise variations and experimentally analyzes their effect on face recognition algorithms. The performance of appearance based algorithms, feature based algorithms, and texture based algorithms are compared using the heterogeneous face database and the disguise face database. Experimental results suggest that high degree of disguise variations is more challenging to address compared to variations in pose, expression and illumination. Furthermore, it also suggests that a careful and thorough investigation is required to develop a robust face recognition algorithm that can fulfill the operational needs of real world applications.

2.3.2.2 Paper # [P-7]

2.3.2.2.1 Abstract

This paper proposes an improvement of the efficiency of the implementation of the algorithm and increase the robustness of sparse representation-based classification (RSC) based methods dealing with real face disguise. In particular, an improved robust sparse coding (iRSC) algorithm, where in each step of iteration, the dictionary, which consists of all training samples, is reduced by eliminating the objects with larger coding residuals. The reduced dictionary is used to obtain the convergence result of the Maximum Likelihood Estimation (MLE) solution of the sparse coding problem. By eliminating the interference of the objects with larger coding residual errors, iRSC is fast convergence and more efficient. Presented experiments in Automatic Recognition (AR) face database show that iRSC achieves better performance than Sparse Representation Classification (SRC) and RSC when dealing with real face disguise.

2.3.2.2.2 Motivation

Robust vision-based face recognition is one of most challenging tasks for robots. Recently the sparse representation-based classification (SRC) has been proposed to solve the problem. All training samples without disguise are used to compose an over-complete dictionary, and the testing sample with disguise is represented by the dictionary with a sparse coding coefficients plus an error. The coding residuals between the sample and each class of training samples are measured and the minimum of them is the identified class to which the sample belongs. The Robust Sparse Coding (RSC) seeks for the MLE solution of the sparse coding problem, so it is more robust to disguise. However, the iterative algorithm to solve RSC problem is computationally expensive and highly time consuming. This paper proposes an improved Robust Sparse Coding (iRSC) algorithm for practical application conditions, which achieves better performance than RSC algorithm.

2.3.2.2.3 Method

In sparse representation based face recognition algorithms, given a face image sample from a certain face database, its storage format is $M\times N$ color or gray image. The face image is stretched into a *d-dimensional* face vector **x** $(d=M\times N)$. Then face recognition algorithms can directly applied in the *d-dimensional* face space. Sparse Representation-based Classification (SRC) and Robust Sparse Coding (RSC) are two such sparse representation based face recognition algorithms.

In both SRC and RSC, all training samples are involved to compose the over-complete dictionary, each query sample is represented as a sparse linear combination over the dictionary. The sparsity constraint on the coding coefficients and the iteratively solving algorithm make the computational cost of RSC very high. Since the dictionary is over-complete for sparse representation, for example AR database has 100 classes and even more classes in practical systems, the dictionary can be reduced in the iterative steps to calculate the weight matrix W. Actually, the objects with larger coding residual errors have less contribution to representing the query sample, and the coding coefficients associated with those objects are usually equal or close to zero under the sparsity constraint. Therefore those objects could be omitted from the dictionary without loss of the over-complete property. The iRSC algorithm is proposed according to the above principle. At the beginning, more irrelevant objects can be omitted to reduce the total computing cost. When the over-complete dictionary is small enough, all remained objects should be reserved to keep the condition for well sparse representation.

The paper defines a retention factor R of the dictionary for the step t as follow:

$$R_t = \begin{cases} 0.1t + 0.5, t \le 5 \\ 1, t > 5 \end{cases} \qquad (10)$$

After the step t, only $R_t\times 100\%$ of the dictionary with minor coding residual errors can be reserved for the next step. Moreover, the retention factor R can be set to a *fixed ratio* or a *median ratio*. Based on this observation, the paper outlines the improved iRSC algorithm.

2.3.2.2.4 Experimental Results

The paper evaluates the performance of the proposed algorithm and compares its performance to the SRC and RSC algorithms, using data with occlusion and disguise variation.

2.3.2.2.4.1 Database

The evaluation experiments are conducted on the AR face database, using samples with sunglasses or scarf.

2.3.2.2.4.2 FR Algorithm(s)

The paper proposes an improvement on benchmarks methods using sparse representation for face recognition, which address the occlusion problem, namely the SRC and RSC methods. These methods are standalone FR approaches, which address the occlusion challenge.

2.3.2.2.4.3 Sample Results

The performance of the proposed iRSC method is assessed and compared with SRC and RSC which are benchmark methods using sparse representation for face recognition, dealing with occlusion and disguise.

The face recognition results by SRC, RSC and iRSC are listed in Table 2-10. Although the dictionary is reduced in iRSC, it still can achieve competitive recognition rates with RSC dealing with both sunglasses and scarf disguises. SRC did not get good performance with scarf (only 38% accuracy) in which about 40% face region are covered. The reason is that SRC cannot handle the case with large occlusion more than around 30%.

Table 2-10 Recognition Rates by Competing Methods on the AR Database with Disguise Occlusion (Source: [P-7])

Algorithms	Sunglasses	Scarves
SRC	87%	38%
RSC	100%	99%
iRSC (Eq. (10))	98%	99%
iRSC (fixed ratio 0.8)	100%	98%

In the experiments, the programming environment is Matlab 7.0a. The computer used is of 3.10 GHz Intel(R) Core(TM) i5-2400 CPU and with 4.00 GB RAM. Average runtimes by the above three methods are listed in Table 2-11. As a result of the size reduction of dictionary, the average runtime of iRSC is much shorter than both SRC and RSC.

Table 2-11 Average Runtimes by Competing Methods on the AR Database with Disguise Occlusion (Source: [P-7])

Algorithms	Sunglasses	Scarves
SRC	17.86 s	20.09 s
RSC	28.32 s	23.35 s
iRSC (Eq. (10))	4.43 s	4.03 s
iRSC (fixed ratio 0.8)	5.85 s	4.80 s

2.3.2.2.5 Key Contributions

This paper presented an iRSC algorithm for robust face recognition with real disguise. The key features of the proposed method include:

- The advantages of the original RSC algorithm, which include its robustness to various types of outliers and large region of occlusion, have been well preserved.
- By the size reduction of dictionary in each iterative step of iRSC, its computational complexity is reduced significantly. Its average runtime is only about 16% of RSC. In this

process, the over-complete property of dictionary is not affected, and therefore, iRSC can still achieve competitive recognition rates with RSC.

- The experimental results on AR face database demonstrated that iRSC has better comprehensive performance than SRC and RSC.
- With high recognition rate but low computational cost, iRSC is assessed to be a good candidate for practical robotic systems to fulfill robust face recognition tasks.

2.3.2.3 Paper # [P-8]

2.3.2.3.1 Abstract

This paper proposes a novel approach Dynamic Image-to-Class Warping (DICW) to deal with partially occluded face recognition. An image is partitioned into sub-patches, which are then concatenated in the raster scan order to form a sequence. A face consists of forehead, eyes, nose, mouth and chin in a natural order and this order does not change despite occlusion or small rotation. Thus, in this work, a face is represented by the aforementioned sequence which contains the order of facial features. Taking the *order information* into account, DICW computes the distance between a query face and an enrolled person by finding the optimal alignment between the query sequence and all sequences of that person along both *time* dimension and *within-class* dimension. Extensive experiments on public face databases with various types of occlusion have verified the effectiveness of the proposed method. In addition, the proposed method, which considers the inherent structure of the face, performs with greater consistency than current methods when the number of enrolled images per person is limited. The proposed method does not require any training process and has a low computational cost, which makes it applicable for real-world FR applications.

2.3.2.3.2 Motivation

In recent years, FR under unconstrained conditions has attracted wide research interests. FR systems in unconstrained conditions need to deal with variations in illumination, pose, expression, etc. This work focuses on the recognition of partially occluded faces, which often occurs in real-world environments but has not attracted much attention yet. Most of the existing works that simply treat occluded FR as a signal recovery problem or just employ the framework for general object classification, do not consider the inherent structure of the face. In this paper, a novel approach that takes the *order information* of facial features into account when dealing with partially occluded faces.

2.3.2.3.3 Method

A patch-matching method DICW is proposed to deal with occlusion. As such, an image is partitioned into sub-patches, which are then concatenated in the raster scan order to form a sequence. In this way, a face is represented by a patch sequence which contains *order information* of facial features, as illustrated in Figure 2-33. DICW calculates the *Image-to-Class* distance between a query face and an enrolled person by finding the optimal alignment between the query sequence and all the enrolled sequences of that person. Then the probe image is classified as the class with the shortest distance.

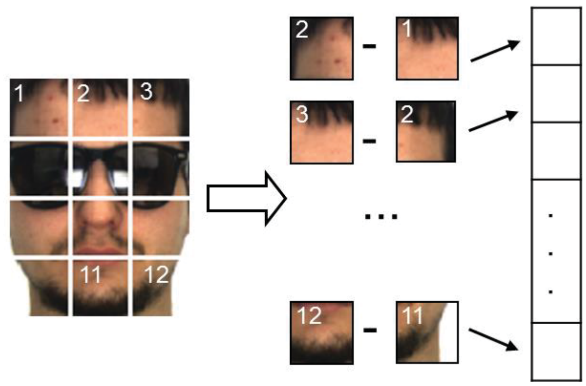


Figure 2-33 Image Representation using the DICW Method (Source: [P-8])

Figure 2-34 illustrates alignment by Dynamic Time Warping (DTW) (left) and DICW (right). The arrows indicate the alignment for each patch. The dashed line marks the optimal warping path between an image and a gallery set.

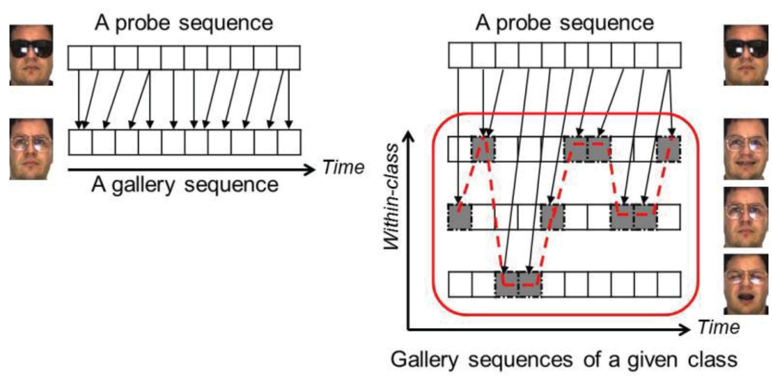


Figure 2-34 Illustration of Alignment by DTW (Left) and DICW (Right) (Source: [P-8])

Figure 2-35 shows an illustration of the *Image-to-Class* warping. In (a), the occluded face belongs to class 2 but is misclassified when using the *Image-to-Image* distance (the nearest neighbor is from class 1). Although the *Image-to-Image* distance is large to each individual gallery image of class 2, the *Image-to-Class* distance is small and leads to the correct classification. As shown in (b), in each step, each patch in the probe image is matched to the most similar patch from one gallery image. The *Image-to-Class* distance is calculated mainly based on these most similar patch pairs (matched patches are indicated by the same color).

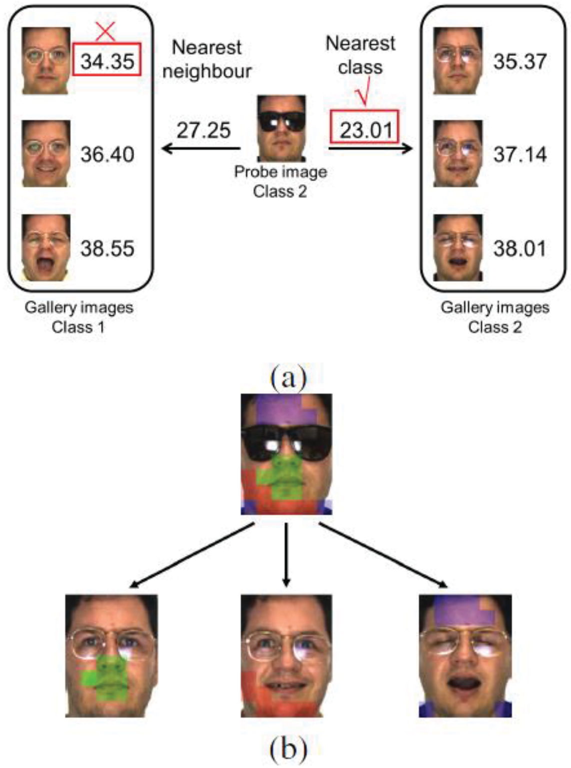


Figure 2-35 (a) Comparison of the Image-to-Image Distance and the Image-to-Class Distance (b) The Distance between a Probe Image and a Class (Source: [P-8])

The reported advantages of DICW are as follows:

- It can directly deal with occluded images without occlusion detection
- It does not need any training process.
- It works well with limited gallery images per person.
- It is appropriate for real FR applications.

2.3.2.3.4 Experimental Results

The paper presents a quantitative comparison of the proposed method DICW with four representative methods in the literature: the linear SVM using Eigenface as features (LSVM), the reconstruction based Sparse Representation (SRC), the patch-matching based Naive Bayes Nearest Neighbor (NBNN) and the baseline Nearest Neighbor (NN).

2.3.2.3.4.1 Database

A set of large-scale experiments are conducted using two public databases: the FRGC2.0 (Face Recognition Grand Challenge) database and the AR database and one outdoor environment database.

2.3.2.3.4.2 FR Algorithm(s)

The paper proposes a new standalone occlusion robust FR approach.

2.3.2.3.4.3 Sample Results

Figure 2-36 illustrates this comparative assessment using the FRGC database with synthetic occlusion. One hundred subjects were chosen and 8 images for each subject. For each subject, the number of images is set as follows: K = 1, 2, 3 and 4 un-occluded images as gallery sets respectively and the other 4 images with synthetic occlusion as the probe set.

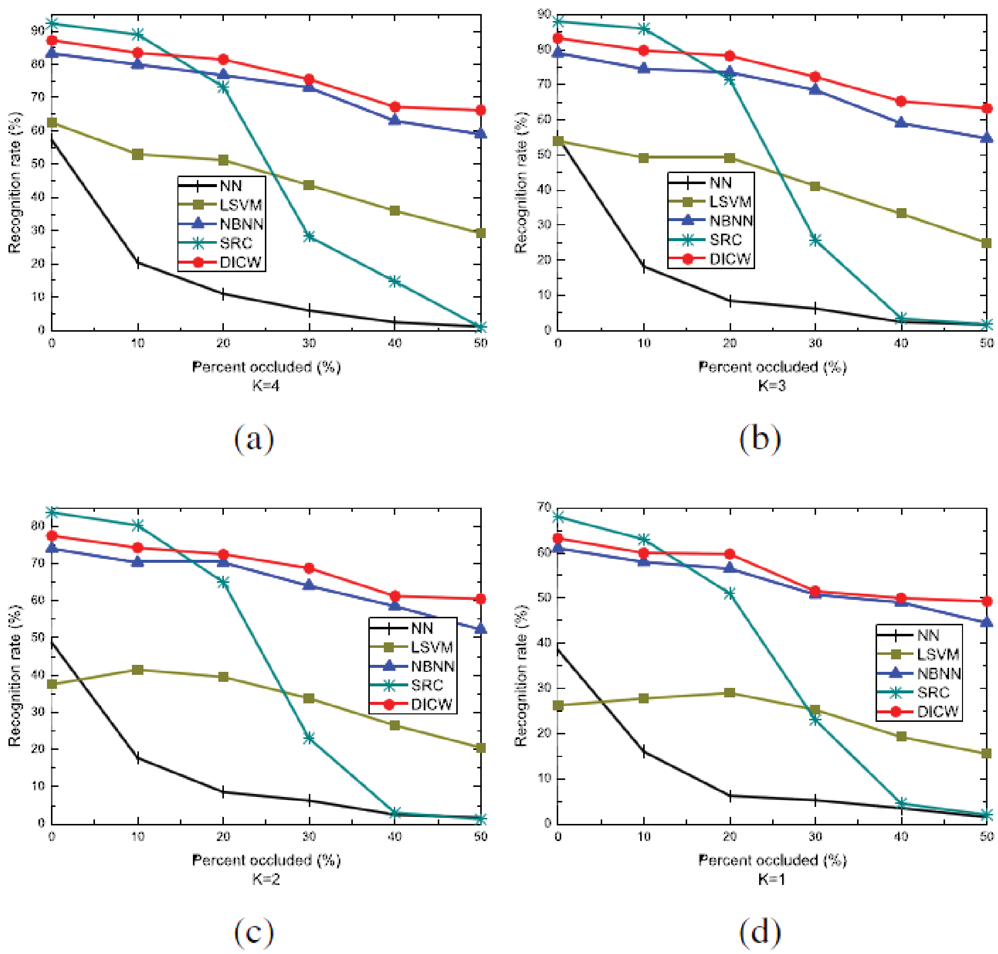


Figure 2-36 Recognition Rates on the FRGC Database with Different Number of Gallery Images per Person (Source: [P-8])

Figure 2-37 illustrates the quantitative assessment using the AR database with real disguise. A subset of the AR database (50 men and 50 women) containing varying illumination conditions, expressions and occlusion is used. The un-occluded frontal view images with various expressions are used as the gallery images (8 images per person). For each person, K = 1, 2, 4, 6 and 8 images are selected from those images as gallery sets respectively.

Notice that compared with other methods, the proposed DICW method achieves comparable or better recognition rates among these methods and about 15 times faster than the work, as claimed in the paper.

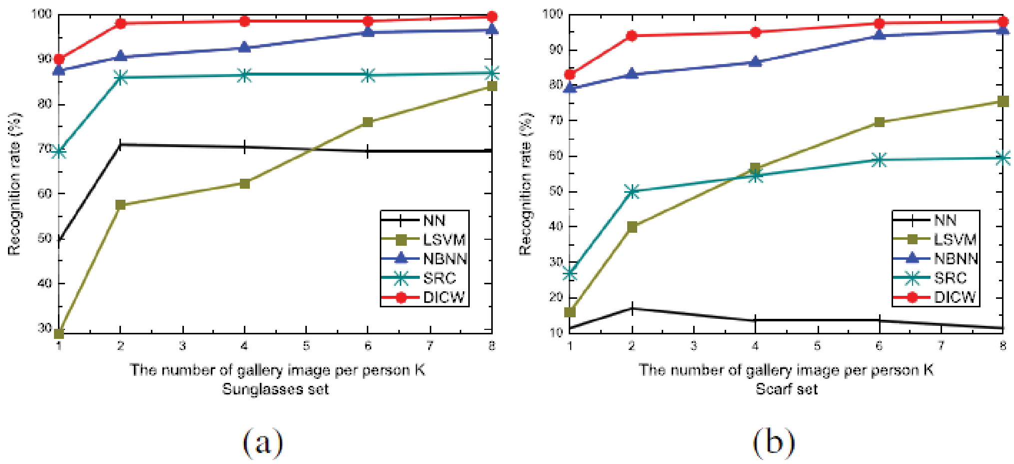


Figure 2-37 Recognition Rates on the AR Database with (a) Sunglasses Occlusion (b) and Scarf Occlusion (Source: [P-8])

2.3.2.3.5 Key Contributions

This paper proposed a novel approach, DICW, to the recognition of occluded faces. The key features of the proposed method include:

- Extensive experiments in three face databases show that DICW achieves much better performance than four representative methods in the literature.
- In the extreme case where only single gallery image is available for each person, DICW still performs consistently.
- The DICW can be applied directly to raw data without performing occlusion detection in advance and does not require a training process.
- All of these make the proposed approach more applicable in real-world scenarios.

2.3.2.4 Paper # [P-9]

2.3.2.4.1 Abstract

This paper explores three occlusion cases that a realistic FR system should take into account. A novel non-parametric classification method to handle the occlusion related problems is presented. This proposed method represents a face image as a sub-patch sequence which maintains the inherent structure information of the face. Matching is based on the *Image-to-Class* distance from a query sequence to all reference sequences of an enrolled class. Experimental results on public databases verify the effectiveness and robustness of the proposed method.

2.3.2.4.2 Motivation

Face Recognition (FR) systems in real environment need to deal with uncontrolled variations in face images such as occlusions and disguise. Most of the current FR algorithms do not consider the fact that occlusions may exist in both reference and query images. However, in the real-world environment, occlusions may occur in both reference and query data, especially in surveillance scenarios. Another example is that, for Human Computer Interaction (HCI) applications which are totally open to users, reference faces may be occluded in the enrolment stage. Most current FR approaches found in the literature are easily affected by occlusions in the training data. This paper proposes a novel non-parametric classification method to handle problems related to occlusion in both reference and query data.

2.3.2.4.3 Method

The method applied in this paper is similar as that proposed in Paper # [P-8], reviewed above. The proposed method tests all possible matching correspondence between patches and selects the combination with minimal overall cost for the whole sequence. So the occluded patches which cause large distance error can be ignored. Moreover, when occlusions exist in the gallery set, the proposed model is capable of exploiting the visible information from different gallery images. Thus, the proposed method, which uses the *Image-to-Class* distance and considers the inherent structure of the face, is able to deal with occlusions in both reference and query images.

2.3.2.4.4 Experimental Results

Table 2-12 summarizes three occlusion cases, which a FR system may encounter in the real world. The latter two cases would also occur in real environment but have not yet received much attention. Most of the current methods rely on a clean gallery/training set and only consider the first case. In this section, we present experimental results related to addressing these cases using the proposed method.

Table 2-12 Three Typical Occlusion Cases in the Real World (Source: [P-9])

	Gallery	Probe	Scenarios
Uvs.O:	Unoccluded	Occluded	Access control, boarder check
Ovs.U:	Occluded	Unoccluded	Suspect detection, shoplifter recognition
Ovs.O:	Occluded	Occluded	

The paper presents a quantitative comparison of the proposed method DICW with four representative methods in the literature: The reconstruction based SRC which achieves good performance for FR recently, the original DTW which considers the order information as the proposed method, the NBNN which also uses the *Image-to-Class* distance as the proposed method, and the baseline NN.

2.3.2.4.4.1 Database

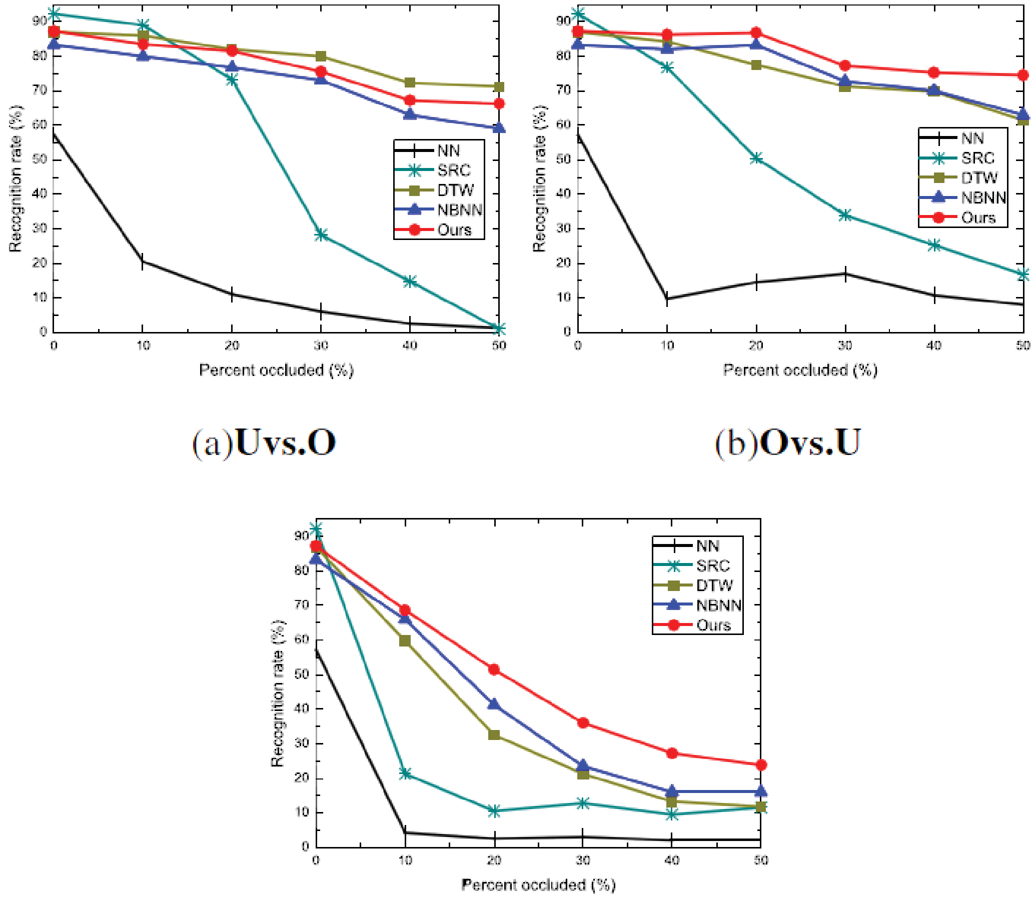
A set of large-scale experiments are conducted using two public databases: the FRGC2.0 (Face Recognition Grand Challenge) database and the AR database and one additional realistic database.

2.3.2.4.4.2 FR Algorithm(s)

The paper proposes a new standalone pose occlusion robust FR approach.

2.3.2.4.4.3 Sample Results

Figure 2-38 shows the recognition results in the three cases with different levels of occlusions. The performance of SRC is slightly better than others when the occlusion level is less or equal to 10%, however, drops sharply as the occlusion increases. As discussed before, reconstruction based methods like SRC are easily affected by occlusions in the gallery data. In the *U vs. O* case, DTW and the proposed method which consider the inherent structure of the face perform better than NBNN. In the following two cases, the proposed *Image-to-Class* method outperforms others, especially when both the gallery and probe images are occluded (c). On the whole, the proposed method performs consistently better than all other tested methods, in all three occlusion cases.



(a)**Uvs.O** (b)**Ovs.U**

(c)**Ovs.O**

Figure 2-38 Recognition Rates on the FRGC Database (Source: [P-9])

The recognition results are shown in Table 2-13. The performance of SRC in the *O vs. O* case is poor since the task is very challenging. The sunglasses cover about 30% of the face and the scarves covers nearly 50%. Unlike DTW and NBNN whose recognition rates fluctuate according to different occlusions (from sunglasses to scarf), the proposed method behaves consistently and outperforms other methods in most cases.

Table 2-13 Recognition Rates (%) on the AR Database (Source: [P-9])

		NN	SRC	DTW	NBNN	Ours
Uvs.O: Gallery-unoccluded						
Probe	sunglasses	69.5	87.0	99.0	96.5	**99.5**
	scarf	11.5	59.5	96.5	95.5	**98.0**
Ovs.U: Gallery-occluded						
Probe	unoccluded	42.6	85.7	87.7	94.4	**94.6**
Ovs.O: Gallery-occluded						
Probe	sunglasses	5.5	18.0	55.0	49.0	**56.0**
	scarf	5.5	10.0	**61.5**	52.5	55.5

2.3.2.4.5 Key Contributions

This paper presents a novel non-parametric classification method for occluded face recognition. The key features of the method include:

- Compared with the current methods, the proposed method is able to deal with occlusions which exist in both gallery and probe sets.
- When the gallery images are contaminated, the *Image-to-Class* distance leads to better performance than the *Image-to-Image* distance.
- Experimental results demonstrate that the proposed method outperforms the state-of-the-art methods for occluded FR.

2.3.3 Facial Expression

Face is one of the most important human's biometrics which, due to its unique characteristics, plays a major role in conveying human identity and emotion. As illustrated in Figure 2-39, human emotions and moods are often expressed with different facial expressions. These differences in facial expressions change the appearance of the face and it becomes difficult for a Face Recognition System to match the accurate face stored in the database.

Figure 2-39 Variations in Facial Expression (Source: [R-2])

Table 2-14 lists the specific facial recognition papers, which address the facial expression variations challenges in FR systems, reviewed in this section.

Table 2-14 Selected Papers on Addressing Facial Expression Variations in FR Systems.

#	Paper Title	Authors	Source	Year	Type
P-10	Analysis of Face Recognition Under Varying Facial Expression: A Survey	Marryam Murtaza, Muhammad Sharif, Mudassar Raza, and Jamal Hussain Shah	The International Arab Journal of Information Technology, Vol. 10, No. 4, July 2013	2013	Survey
P-11	Exploring Facial Expression Effects in 3D Face Recognition Using Partial ICP	Yueming Wang, Gang Pan, Zhaohui Wu, and Yigang Wang	P.J. Narayanan et al. (Eds.): ACCV 2006, LNCS 3851, pp. 581–590, 2006. Springer-Verlag Berlin Heidelberg 2006	2006	Proposed new technique(s)
P-12	Expression-invariant Facial Identification	Pohsiang Tsai, Tich Phuoc Tran, Longbing Cao	Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics San Antonio, TX, USA - October 2009	2009	Proposed new technique(s)
P-13	Model Based Face Recognition Across Facial Expressions	Zahid Riaz, Christoph Mayer, Matthias Wimmer, and Bernd Radig	Biometric ID Management and Multimodal Communication Lecture Notes in Computer Science Volume 5707, 2009, pp. 122-129.	2009	Proposed new technique(s)

2.3.3.1 Paper # [P-10]

In this paper, the authors review different techniques proposed to address the variation in facial expression challenges encountered in conducting facial recognition in uncontrolled environment. It is a survey-based timeline view that performs an analysis on different techniques to handle facial expressions in order to recognize faces. Figure 2-40 illustrates the general frame work for automatic facial expression presented in this paper:

- Primarily, face images are first acquired and normalized in order to eliminate the complications like pose and illumination factor during face analysis.
- It has been proven that feature extraction is a great milestone which uses various techniques to characterize facial features like motion, model, and muscles-based approaches.
- Finally these features are classified and trained in different subspaces and then used for recognition.

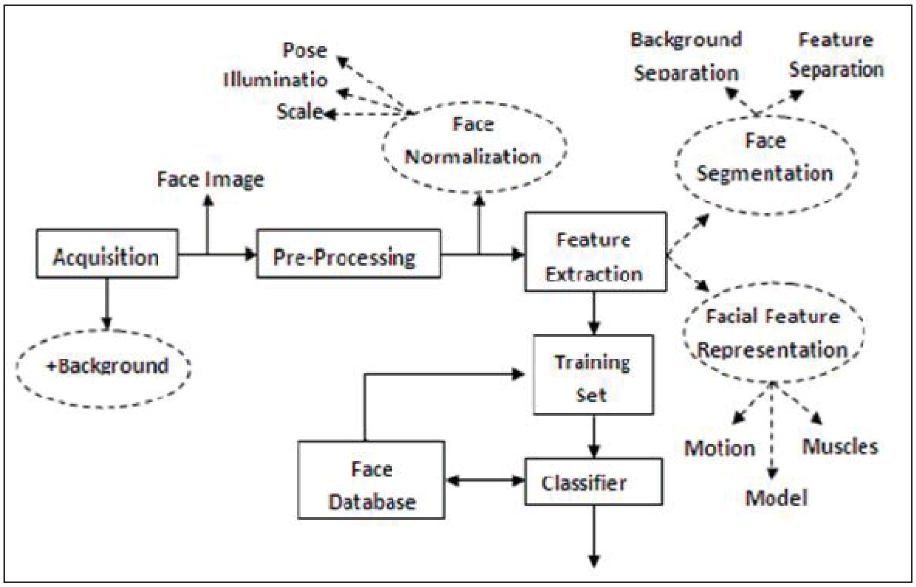


Figure 2-40 Presented Framework for Automatic Facial Expressions (Source: [P-10])

2.3.3.1.1 Pre-Processing Steps

The pre-processing steps are applied to locate faces from chaotic backgrounds and normalize images so that it can easily be processed. For pre-processing images different researchers used different algorithms. Pre-processing is not the necessary steps. Image acquisition is also a part of pre-processing. Ideally, faces are acquired by locating faces from the chaotic backgrounds. Accurately measuring the position of faces plays a vital role in order to extract facial features. Active Shape Model (ASM) is a feature orientation method that is used to extort transient and intransient facial features. Automatic facial expression is a complex task because the outer appearance of a person may vary by the changed mood of a person and thus subsequently affect the facial expression. These expressions may vary with age, ethnicity, gender and occluding objects like facial hair, cosmetic products, glasses and hair etc. Additionally, pose, lighting conditions, and expression variations also affects face recognition rate. Although, face normalization is not mandatory, it is used to overcome the harmful effects of illumination and pose variations because facial expression recognition depends on the angle, distance and illumination invariant conditions against each face.

2.3.3.1.2 Feature Extraction

Physically face is classified as many local and global features which may change with the change of facial muscles and skin tissues. This alteration causes a serious dilemma in automatic face recognition that downgrades the performance of recognition rate. Human faces can be classified based on the following features:

- Intransient facial features permanently lie on the face, however may be deformed with the change of facial expressions.
- Transient facial features are like creases i.e. wrinkles bulges etc., which affect the skin texture but do not notify the type of emotion.

Transient and Intransient facial features can be categorically divided as to whether they rely on certain actions of muscles or warping of faces and facial features respectively as shown in Figure 2-41.

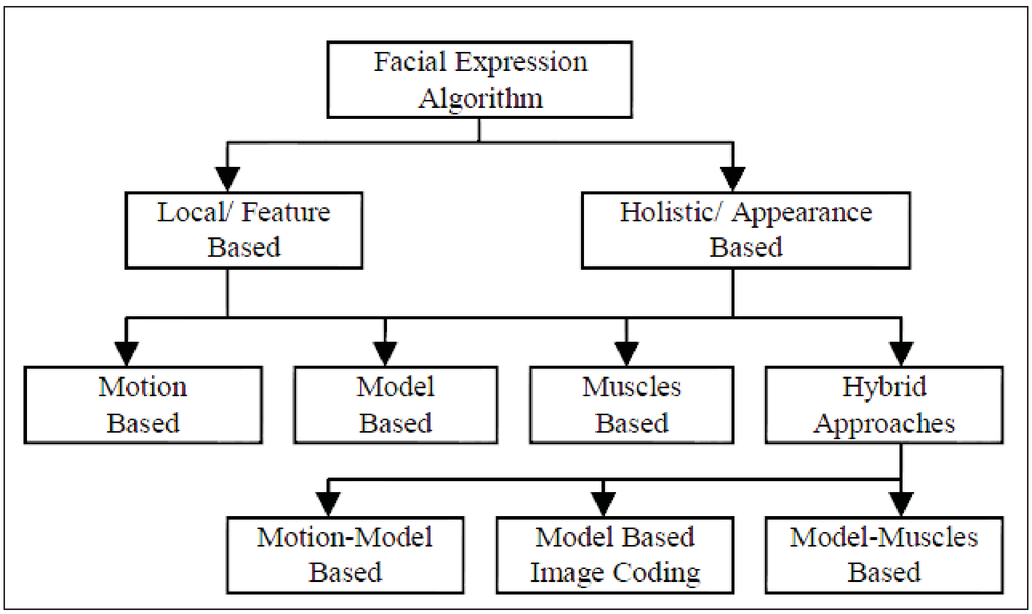


Figure 2-41 Facial Features Extraction and Representation (Source: [P-10])

2.3.3.1.2.1 Feature vs. Appearance Based Approaches

Basically, face recognition under varying facial expression algorithms is categorized as local or feature based approaches and holistic or appearance based approaches that process the whole face for extracting facial features and hence give the detailed information. But sometimes, all the information becomes irrelevant because all facial features are not changed by the appearance of single emotion as, for example, the degree of smile doesn't affect all features but influences only the appearance of cheeks and lips, etc. So, in contrast to holistic based approaches local based approaches provide a way to process only the affected facial features

Feature and appearance based approaches are further categorized as motion, model, muscles-based and hybrid approaches which provide further distinctions of motion-model based, motion based image coding and model-muscles based approaches. There are also hybrid approaches, which combine two or more of these classes of algorithms.

The paper presents a comprehensive comparative assessment of several cited facial expression algorithms in each of these classes. The reader is referred to the paper for more details on the illustrated assessment.

2.3.3.1.3 Classification

The final phase of automatic facial expression recognition classifies the transient and in-transient facial features in accordance with the desired result. Selecting a low dimensional feature subspace from thousands of features is a key phenomenon for optimal classification. The main ambition to use subspace classifiers is to convert high dimensional input data into low dimensional feature subspace. Subspace classifiers selectively represent the features that minimize the processing area. Feature extraction plays a vital role to reduce the computational cost and progress the classification results because selecting a low dimensional feature subspace from bundle of features is very crucial for optimal classification. Wrong features selection degrades the performance of face recognition; even though superlative classifier may be used. There are bunch of linear and non-linear classifier's that offers categorization between correlated and uncorrelated variables. The two basic linear classification techniques are principal component analysis PCA, and Linear discriminant analysis LDA. Others classifiers are Independent Component analysis Independent Component Analysis (ICA), Support vector machine SVM, Singular Value Decomposition (SVD). The paper references several other less commonly used classifiers.

2.3.3.1.4 Concluding Remarks

This paper provides a snapshot and comparative assessment of different facial expression algorithms, which can be applied for the purpose of enhancing facial recognition systems and achieving better recognition results.

2.3.3.2 Paper # [P-11]

2.3.3.2.1 Abstract

This paper investigates facial expression effects in face recognition from 3D shape using partial Iterative Closest Point (ICP) algorithm. The partial ICP method could implicitly and dynamically extract the rigid parts of facial surface by selecting a part of nearest points pairs to calculate dissimilarity measure during registration of facial surfaces. The method is expected to be able to get much better performance than other methods in 3D face recognition under expression variation for its dynamic extraction of rigid parts of facial surface at the same time of matching. An effective method for coarse alignment of facial shape, which is fully automatic, is also presented. Experiments on 3D face database of 360 models with 40 subjects, 9 scans with four different kinds of expression for each subject, show partial ICP is very promising, compared with PCA baseline.

2.3.3.2.2 Motivation

Although the 2D face recognition system has good performance under constrained conditions, since the 2D image essentially is a projection of the 3D human face, it is still challenged by changes in illumination, pose and expression. Utilizing 3D information can improve the system performance due to its explicit representation of facial surface. However, facial expression is still a big challenge even using 3D data in face recognition because in fact facial surface is a non-rigid object. Thus, just extracting and matching the same relatively rigid parts for all facial surfaces partially solve the expression problem but is often not be perfect a perfect solution.

2.3.3.2.3 Method

The paper proposes an ICP based, which yields the partial ICP for 3D facial shape recognition which can implicitly extract variant rigid regions of the face according to deformation extent under different expression during matching. The method does a proper selection of nearest points pairs to calculate Root Mean Squares (RMS) when using ICP to match two surfaces. When applied the method to three expressions, smile, surprise and sad in the conducted experiments, 96.88% rank-one matching rate is obtained.

2.3.3.2.4 Experimental Results

The paper evaluates the performance of the proposed expression-invariant FR approach and its performance is compared to the PCA-based 3D face recognition, which was implemented as a baseline algorithm.

2.3.3.2.4.1 Database

Experiments use the 3D facial expression database ZJU-3DFED, collected by the authors.

2.3.3.2.4.2 FR Algorithm(s)

This is a standalone expression-invariant FR system.

2.3.3.2.4.3 Sample Results

PCA-based method is implemented in the conducted experiments for comparison. After models are trimmed, PCA-based method can easy be applied to 3D face recognition by projecting the trimmed models to x-y plane. The first 40 eigenvectors are used when test PCA-based method which hold 96.46% energy. The performances between PCA-based method and partial ICP method on all five probe sets are compared. The results are shown in Figure 2-42.

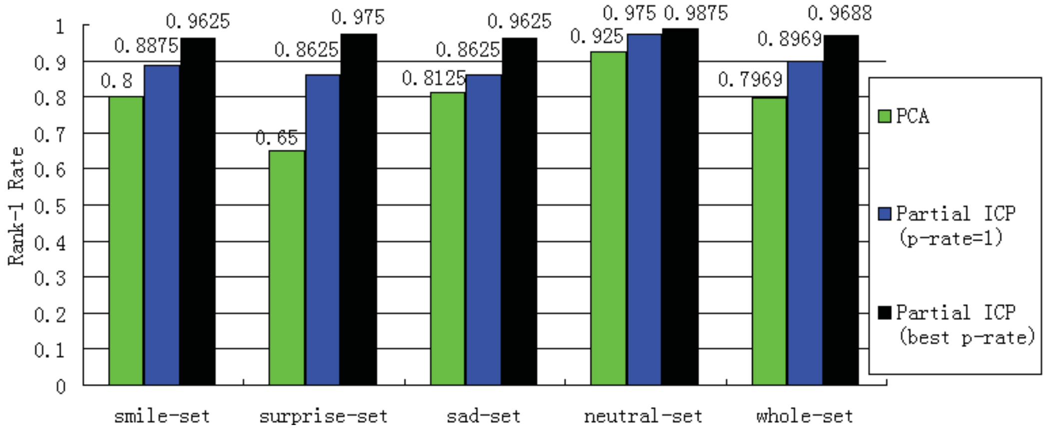


Figure 2-42 Rank-1 Rates: PCA vs. Partial ICP (Source: [P-11])

2.3.3.2.5 Key Contributions

This paper proposed a method, partial ICP method, which is capable of dynamically extracting rigid parts of facial surface. The key features of the proposed method include:

- The extraction is completely dependent on the deformation extent of facial surface and extracted areas are varied between different expressions.
- Several experiments based on partial ICP for 3D face recognition are conducted using database with 360 models. A rank-1 rate 96.88% demonstrates the high performance of the proposed method in 3D face recognition with different expressions.
- The experimental results also show that the proposed method significantly outperforms PCA-based method.

2.3.3.3 Paper # [P-12]

2.3.3.3.1 Abstract

This paper proposes a new approach, which transforms facial expressions to neutral-face like images. Hence, this enables image retrieval systems to robustly identify a person's face for which the learning and testing face images differ in facial expression.

2.3.3.3.2 Motivation

Facial identification has been recognized as most simple and non-intrusive technology that can be applied in many places. However, there are still many unsolved facial identification problems due to different intra-personal variations. In particular, when images of the databases appear at different facial expressions, most currently available facial recognition approaches encounter the expression-invariant problem in which neutral faces are difficult to be recognized.

2.3.3.3.3 Method

In facial recognition, facial expressions are typically considered as the abnormal behaviors in comparison with expressionless faces as the normal behaviors. The paper proposes a boosting framework that first converts different expression changes (abnormal behavior) into their corresponding pseudo neutral facial images in the preprocessing stage and then adopts the AdaBoost.M1 algorithm to do the classification.

The conversion method in the preprocessing stage is motivated by the idea of image segmentation. In image segmentation the aim is to find Region of Interests (ROIs) in any image; that is, to segment any ROI (foreground) from an image common background. Accordingly, neutral faces are treated as mean faces (background) and facial expression as variance image (foreground) that deviates from its mean face (neutral face). Therefore, the aim is to restore the subject's neutral face when one displays his/her deviated face (facial expression).

Once any facial expression images have been converted into their corresponding pseudo neutral facial images, the AdaBoost.M1 algorithm is applied to do the recognition. AdaBoost.M1 is a proposed extension of the AdaBoost algorithm.

2.3.3.3.4 Experimental Results

The paper evaluates the performance of the proposed expression-invariant FR approach and compares its performance to standard FR approaches, including: EigenExpression and FisherExpression SimNeu using PCA and SimNeu using LDA methods.

2.3.3.3.4.1 Database

The proposed approach was applied on ten subjects of the Japanese Female Facial Expression (JAFFE) database. The neutral facial images were used for training, while the others for identification.

2.3.3.3.4.2 FR Algorithm(s)

Once any facial expression images have been converted into their corresponding pseudo neutral facial images, the AdaBoost.M1 algorithm is used to do the recognition. In practice, one may be able to use any other FR algorithm after the facial expression conversion. Thus, the method can be considered as a pre-processing step to convert facial expression images into their corresponding pseudo neutral facial images. The pre-processed data can then be fed into one of the many FR systems.

2.3.3.3.4.3 Sample Results

Figure 2-43 shows the conversion results of the preprocessing method. This method converts all the different facial expressions into their corresponding pseudo facial images. The second row for each subject shows the pseudo facial images. These converted images are then used in the recognition stage. The geometric-based (Euclidean distance-based facial fiducial points) facial features of facial neutral and pseudo neutral images will be used in the recognition stage.

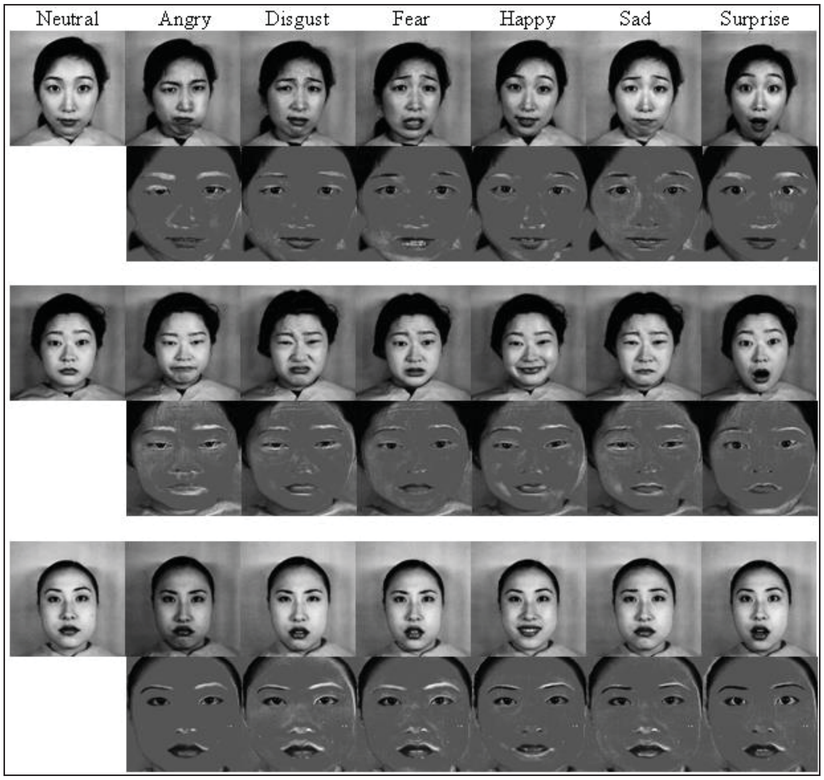


Figure 2-43 Pseudo Facial Images of Different Facial Expressions after the Preprocessing Stage (Source: [P-12])

Figure 2-44 shows the recognition performances of different approaches. EigenExpression and FisherExpression showed the low recognition rates before converting to their pseudo facial images. Their performances were 33% and 40% (when number of features is 9), respectively. SimNeu using PCA and SimNeu using LDA showed increased recognition rates after using pseudo facial images as inputs. Their performances were 41% and 57% (when the number of features is 9), respectively. Although the SimNeu using PCA had slightly increased performance, the SimNeu using LDA had approximately 20% increased performance. The findings showed that their pseudo facial images had become closer to their corresponding neutral faces. However, due to the limitations of the standard algorithms, the recognition performances of pseudo facial images only showed slight improvement. Another reason could be because the human facial structures were not very different. The Euclidean distance-based approach using facial fiducial points could only carry precise individual differences to a limited extent. Therefore, an advanced method should be employed to be able to distinguish the differences among human faces.

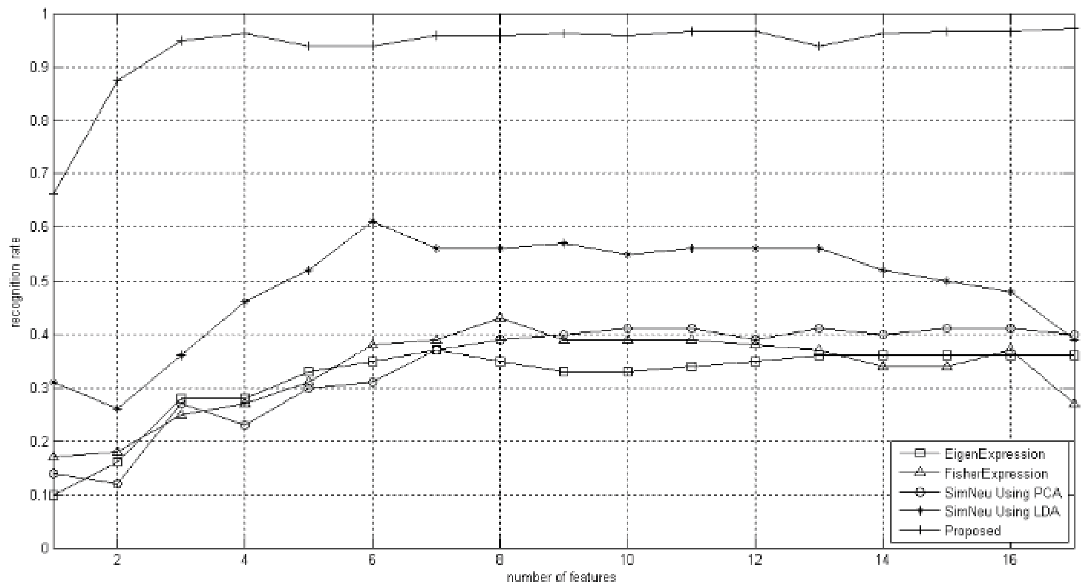


Figure 2-44 Recognition Performances of Different Approaches (Source: [P-12])

2.3.3.3.5 Key Contributions

This paper explores the issue of expression-invariant facial recognition by using an image conversion method to transform facial expressions to pseudo neutral images in the pre-processing stage and a boosting algorithm to do the classification. Some of the key features of the proposed method include:

- The facial expression images are first into their corresponding pseudo neutral facial images.
- The experimental results showed that the proposed framework provided a boosting recognition capability for discerning among human faces on the pseudo neutral images.
- Facial expressions in the testing set were converted into their corresponding pseudo neutral images before testing; making them closer to their corresponding neutral faces
- The AdaBoost.M1 algorithm was capable of dealing with hard-to-classify patterns.
- The experimental showed that the proposed method had an increased performance rate. It is possible that the pseudo facial images of the converted facial expressions had become closer to the corresponding neutral faces.

2.3.3.4 Paper # [P-13]

2.3.3.4.1 Abstract

This paper describes a novel idea of face recognition across facial expression variations using model based approach. The approach follows in 1) modeling an Active Appearance Model

(AAM) for the face image, 2) using optical flow based temporal features for facial expression variations estimation, 3) and finally applying binary decision trees as a classifier for facial identification. The novelty lies not only in generation of appearance models which is obtained by fitting ASM to the face image using objective but also using a feature vector which is the combination of shape, texture and temporal parameters that is robust against facial expression variations. Experiments have been performed on Cohn-Kanade facial expression database using 61 subjects of the database with image sequences consisting of more than 4000 images. This achieved successful recognition rate up to 91.17% using decision tree as classifier in the presence of six different facial expressions.

2.3.3.4.2 Motivation

Over the past three decades of face recognition technology, there exist many commercially available systems to identify human faces. However face recognition is still an outstanding challenge against different kinds of variations like facial expressions, poses, non-uniform light illuminations and occlusions. Higher level applications like facial expression recognition and face tracking still remain outstanding along with person identity. This gives rise to an idea for generating a framework suitable for solving these issues together.

2.3.3.4.3 Method

The paper proposes a hierarchal utilization of shape model fitting, texture mapping and estimating optical flow-based parameters for feature vector extraction. The feature vector consists of the shape, texture and temporal variations, sufficient for considering local variations in shapes. All the subjects in the database are labeled for identification. A fitted face model, on the training images is then used for defining the reference shape in the conducted experiments. This reference shape is calculated by finding the mean shape of the all shapes in the database.

Figure 2-45 illustrates a detailed process flow of the prosed FR model-based technique, which splits the challenge of image interpretation into computationally independent modules.

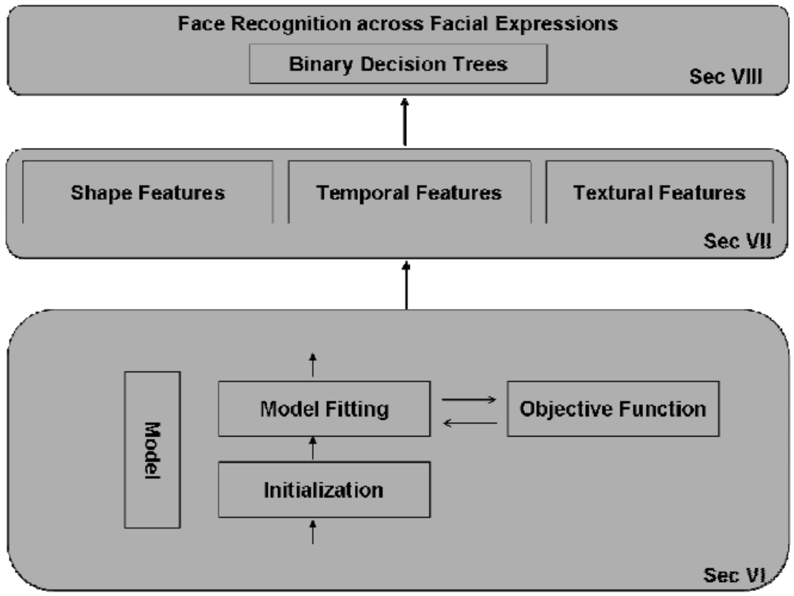


Figure 2-45 Framework of the Proposed Model-based FR Method (Source: [P-13])

An explicit 2D face model is used to develop a baseline for feature extraction. A Point Distribution Model (PDM) is used as an active shape model. Face is localized in the image. An objective function is learned for fitting this model to the faces. After fitting the model to the example face image, texture information is extracted from the example image on a reference shape which is the mean shape of all the shapes of database. Image texture is extracted using planar subdivisions of the reference and the example shapes. Texture warping between the subdivisions is performed using affine transformation. The resulting image texture is normalized to remove the global lighting effects. This image texture is now normalized both in the sense of shape and varying illuminations effects, making the image robust for shape and illumination. PCA is used to obtain the texture and shape parameters of the example image. This approach is similar to extracting AAM parameters. In addition to AAM parameters, temporal features of the facial changes are also calculated. Local motion of the feature points is observed using optical flow. Reduced descriptors are used by trading off between accuracy and runtime performance. A binary decision tree is used as classifier for person identification.

2.3.3.4.4 Experimental Results

The paper evaluates the performance of the proposed expression-invariant FR approach.

2.3.3.4.4.1 Database

The proposed framework was tested on images from The Cohn-Kanade-Facial-Expression database (CKFE-DB).

2.3.3.4.4.2 FR Algorithm(s)

This is a standalone expression-invariant FR system.

2.3.3.4.4.3 Sample Results

Experiments have been performed on Cohn-Kanade facial database for human faces and Figure 2-46 and Figure 2-47 show true positive and true negative for the database respectively. Since this database consists of six standard facial expressions, it is useful to perform person identity experiment against these expressions. For experimental purposes, image sequences of 61 persons have been used which consists of overall 4060 images. A binary decision tree is trained as classifier in 22.99 sec. A sample of 1381 images is used for testing the recognition results and successfully recognized 1259 images. The recognition rate achieved was 91.17% in the presence of facial expressions.

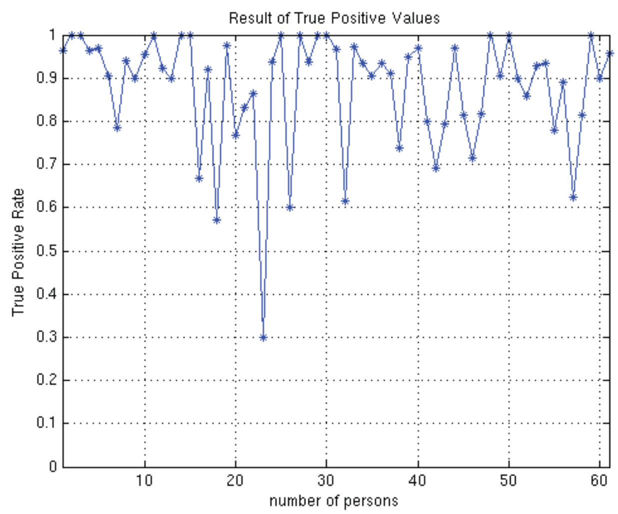


Figure 2-46 True Positive for 61 Persons in the Experiments (Source: [P-13])

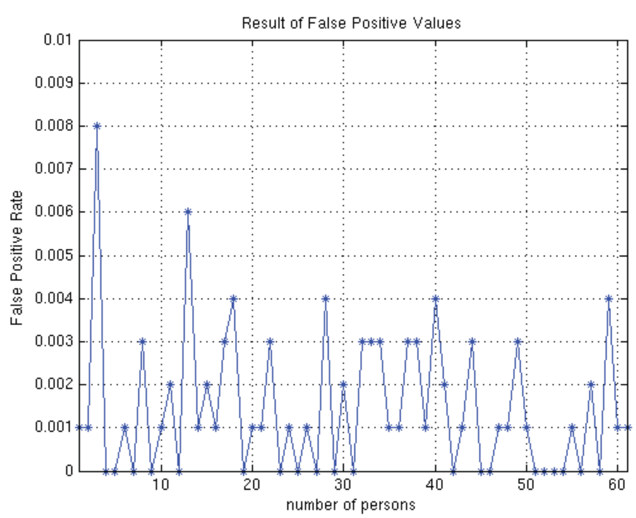


Figure 2-47 False Positive for 61 Persons in Experiments (Source: [P-13])

2.3.3.4.5 Key Contributions

This paper introduced an idea to develop a feature vector which consists of three types of facial variations and is robust against the expressional changes in the human faces in real environments. Some of the key features of the proposed method include:

- Since the training set consists of the facial expressions information of a person, it can recognize the person even under various expressions.
- A binary decision tree is efficient to train and classify. However the benchmarked database consists of frontal faces only.
- This technique is capable of working in real time environment and can be applied to Human Robot Interaction (HRI) very efficiently.
- It can keep the person identity information even under the presence of facial expressions which could originate under human machine interaction scenarios.

2.3.4 Aging Variation

Aging related changes on the face appear in a number of different ways, including: wrinkles and speckles, weight loss and gain, and change in shape of face primitives (e.g., sagged eyes, cheeks, or mouth). The human face is not a unique, rigid object. Everything changes with time, so with the increasing age the appearance of a person also changes which affect the face recognition system as shown in figure. All these aging related variations degrade face recognition performance.

Figure 2-48 Aging Variations (Source: [R-1])

Table 2-15 lists the specific facial recognition papers, which address the aging variations challenges, reviewed in this section.

Table 2-15 Selected Papers on Addressing the Aging Variations Challenges

#	Paper Title	Authors	Source	Year	Type
P-14	Survey: Techniques for Facial Aging Problems in Face Recognition	Deepali T. Biradar, Anil Z. Chhangani, K.T.Patil	2nd International Conference on Emerging Trends in Engineering & Technology, April12, 13, 2013 College of Engineering, Teerthanker Mahaveer University	2013	Survey
P-15	Age Invariant Face Recognition Using Graph Matching	Gayathri Mahalingam, and Chandra Kambhamettu	Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on Biometrics Compendium, IEEE, pp. 1 – 7, 2010	2010	Proposed new technique(s)
P-16	A Discriminative Model for Age Invariant Face Recognition	Zhifeng Li, Unsang Park, and Anil K. Jain	IEEE Transactions on Information Forensics and Security, (Volume:6 , Issue: 3), Biometrics Compendium, IEEE, pp. 1028 – 1037, 2011	2011	Proposed new technique(s)
P-17	Wavelet Based Age Invariant Face Recognition using Gradient Orientation	Saeid. Fazli, and Leila. Ali Heidarloo	International Conference on Advances in Computer and Electrical Engineering (ICACEE'2012) Nov. 17-18, 2012 Manila (Philippines).	2012	Proposed new technique(s)

2.3.4.1 Paper # [P-14]

In this paper, the authors review different techniques proposed to address the aging variation challenges encountered in conducting facial recognition in uncontrolled environment. In particular, the paper presents an overview of some of the recent methods for age estimation, age modeling and age verification across age progression to overcome the aging problem is given.

The paper cites the growing interest in achieving age invariant face recognition to meet the requirements of several applications in law enforcement and forensic investigation, missing individual, Multimedia and human computer interaction (HCI), and security. An Age Specific Human Computer Interaction (ASHCI) system controls access to young kids to specified internet pages. By securing a vending machine, with ASHCI system can refuse to sell alcohol or cigarettes to the underage people. In image and video retrieval, Users could retrieve their photographs or videos by specifying a required age range in image and video retrieval.

As illustrated in Figure 2-49, facial aging affects both the shape and texture of a face. This aging process also appears in different ways in different age groups. Compared with other facial variations, aging effects has three unique characteristics:

1. The aging progress is uncontrollable: It cannot be advanced or delayed and it is slow and irreversible.

2. Personalized aging patterns: Every person has a different aging process. The aging pattern of each person depends on his/her genes as well as many external factors, such as health, lifestyle, weather conditions, etc.
3. The aging patterns depend on time: The face status at a particular age will affect all older faces, but will not affect those younger ones.

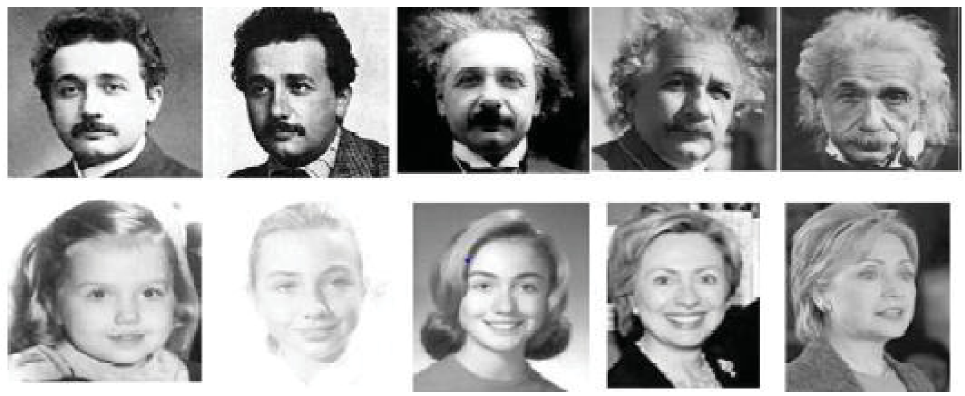


Figure 2-49 Face Aging of Two Famous Individuals at Different Ages (Source: [P-14])

Physiologically, it has been shown that during the early growth and development of the face, from birth to adulthood, the greatest change is the craniofacial growth (shape change). The human face undergoes more significant texture changes at the early age of a child and minor shape changes in older age groups. Therefore, an age correction scheme needs to be able to compensate for both types of aging processes.

The paper outlines an approach for compensating for age variation in face recognition systems. This approach has the following the three main steps:

1. Age estimation
2. Modeling and simulating the aging process
3. Face verification across age progression.

Each of these steps is discussed in more details next.

2.3.4.1.1 Age Estimation

According to the paper, there are three main approaches for image-based age estimation, as discussed next.

2.3.4.1.1.1 Anthropometric Modeling

This approach is based on the craniofacial development theory and facial skin wrinkle analysis. Methods in this approach are suitable for coarse age estimation or modeling ages just for young. Since the human head shape does not change too much in its adult period. In a cited work, wrinkles can be computed from face images to separate young adults from seniors. The wrinkles were computed in several regions, such as on the forehead, next to the eyes, and near the cheek bones. The presence of wrinkles in a region is based on the detection of curves in that region,

2.3.4.1.1.2 Aging Pattern Subspace

An aging pattern is defined as a sequence of personal face images, coming from the same person, sorted in the temporal order. A sequence of an individual's aging face images might be used all together to model the aging process, which is called Aging pattErn Subspace (AGES). When the face images of all ages are available for an individual, the corresponding aging pattern is called a complete aging pattern, otherwise, it is called an incomplete aging pattern as in Figure 2-50, where the missing parts in the aging pattern vector are marked by "m." The AGES method can synthesize the missing ages by using an EM-like iterative learning algorithm. The age of a test face is determined by the projection in the subspace that can best reconstruct the face image.

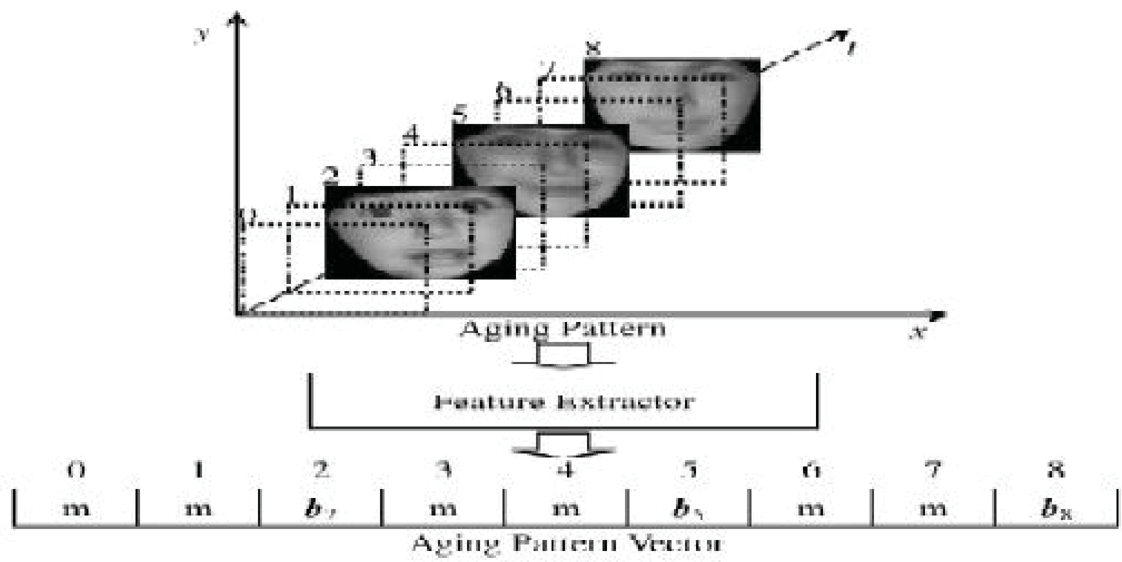


Figure 2-50 Vectorization of the Aging Pattern (Source: [P-14])

2.3.4.1.1.3 Age Regression

For the age regression approach, facial features are extracted by AAM, which incorporates the shape and appearance information together. An input face image is then represented by a set of fitted model parameters. The regression coefficients are finally estimated according to a known regression function. Reviewed works investigated three formulations for the aging function: linear, quadratic, and cubic respectively. The paper highlighted some of the limitation to these Least-Squared Estimated (LSE) based methods, including lack of robustness to outliers, the small sample problem, etc. The paper also cites various techniques to overcome these limitations, including the use of refined age estimation technique by age manifold analysis, Locality Preserving Projections (LPP), Orthogonal Locality Preserving Projections (OLPP), CEA, and LBP. A flow diagram of one of the cited manifold-based methods is illustrated in Figure 2-51.

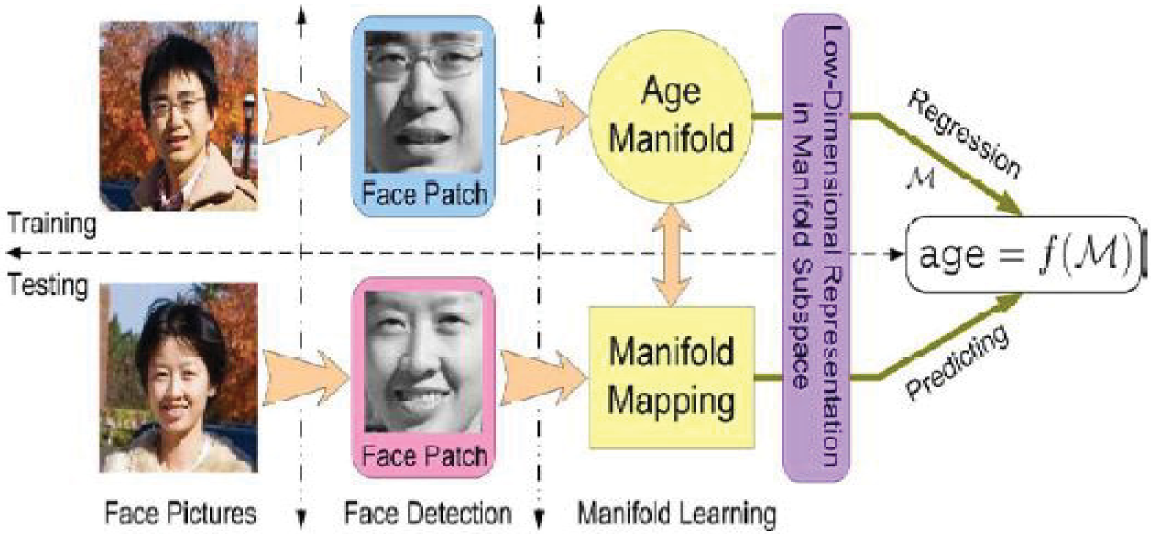


Figure 2-51 Framework of Age Estimation via Face Image Analysis (Source: [P-14])

2.3.4.1.2 Modeling/Simulating the Aging Process

For age synthesis (also called age progression) is often implemented by first building a generic face model. Modeling facial appearances across different ages amounts to building computational models for facial aging that account for a large number of factors.

From a computer vision perspective, the aging problem involves the following factors:

1. **Shape vs. Texture**: There is change in facial shape in childhood, while there are textural variations in the form of wrinkles and other skin artifacts take during later stages of adulthood. Hence, facial aging can be described as a problem of characterizing facial shape and facial texture as functions of time. To deal with the temporal nature of the induced variation, models having same characteristics should be built.
2. **Feature selection:** For developing facial growth model or facial aging model requires identification of the appropriate form of data that will give all required description of the event. The data could be individual specific or population-specific. Fiducial features (2D or 3D) extracted from age-separated faces, 2D facial imagery or 3D facial scans extracted from individuals across different ages.
3. **Factors**: Apart from biological factors such as bone growth, loss in elasticity of facial muscles, facial fat etc., numerous other factors have been shown to influence facial aging effects such as one's ethnicity, gender, dietary habits and climatic conditions. Also facial appearances get altered with increase in age due to factors such as changes in hair-style. Hence, the proposed model for facial aging should take into consideration all these factors.

Several recent efforts for modeling and simulating the aging process have been cited in the paper. These approaches include the techniques which take both shape and color information into consideration, the use of wavelet transform to that captures the age information lost in the blending process, a new framework called Merging Face (M-Face) for appearance-based photorealistic facial modeling, a two-step approach for modeling aging in adults, which comprised of a shape and texture variation model and a method to simulate the adult aging effects on face images by means of super resolution in tensor space and active appearance models.

2.3.4.1.3 Face Verification across Age Progression

The different methods that were proposed to perform face verification across ages can be classified:

- Generative approaches: apply a computational model to simulate the aging and then apply subsequent recognition algorithms.
- Non-generative approaches: concentrate on deriving "age-invariant signatures" from faces.

When comparing two photos, these methods either transform one photo to have the same age as the other, or transform both to reduce the aging effects.

Generative approaches have some limitations, including:

- Construction of face models is difficult and sometimes they do not represent the aging process very well, especially when the training sample size is limited.
- Strong parametric assumptions are needed, which are often unrealistic in real-world face recognition scenarios.
- For constructing the aging model, additional information in the form of the true ages of the training faces and the locations of landmark points on each face image are needed.
- Constraint on the training set is that the images should be captured under controlled conditions (e.g., frontal pose, normal illumination, neutral expression).

In order to overcome these problems, non-generative approaches have been proposed for the aging problem. One cited such approach derives an age-invariant signature from faces and use the same to perform face verification across age progression. Another such approach proposed a face operator derived based on the image gradient orientations derived from multiple resolutions and used Support Vector Machines to perform face verification across ages. A third cited non-generative approach uses location drift of facial features across ages. The idea is to look for forward coherencies in facial feature drifts across ages. It is observed that the coherency of some selected facial features is larger on two different images of the same person with different ages. As illustrated in Figure 2-52, the drifts across images of same individuals appear coherent *(top two rows);* while they are somewhat incoherent *(third row)* when the images belong to different individuals.

Figure 2-52 Drifts in Facial Features for Age Separated Face Images (Source: [P-14])

2.3.4.1.4 Aging Databases

The paper briefly describes there databases, which include significant sets for aging individuals, among many available face databases. These include the MORPH, FG-NET and the FERET databases.

2.3.4.1.5 Concluding Remarks

The paper concludes by emphasizing the importance of compensating for the human aging variation, when using facial recognition biometrics and other facial processing applications. Unlike other variations, such as facial hair, glasses, etc., aging is a natural process in human life. Age estimation is the first key step to accounting for age variations, then the modeling the aging process before age-invariant facial recognition can be conducted.

2.3.4.1 Paper # [P-15]

2.3.4.1.1 Abstract

This paper presents a graph based face representation for efficient age invariant face recognition. The graph contains information on the appearance and geometry of facial feature points. An age model is learned for each individual and a graph space is built using the set of feature descriptors extracted from each face image. A two-stage method for matching is developed, where the first stage involves a Maximum a Posteriori solution based on PCA factorization to efficiently prune the search space and select very few candidate model sets. A simple deterministic algorithm which exploits the topology of the graphs is used for matching in the second stage. The experimental results on the FGnet database show that the proposed method is robust to age variations and provides better performance than existing techniques.

2.3.4.1.2 Motivation

In recent years, face recognition across aging has gained attention from computer vision researchers and has been looked into using age estimation and aging models. Many researchers have studied how age differences affect the face recognition performance, using different approaches, including non-generative approaches, manifold learning schemes, transformation algorithms. This paper present a graph based feature representation of the face images, and builds a probabilistic aging model for each individual using Gaussian Mixture Model (GMM) which incorporates both the shape and texture information. The proposed method differs significantly from other approaches in the literature.

2.3.4.1.3 Method

The proposed method is based on graph based feature representation of the face images, and building a probabilistic aging model for each individual using GMM which incorporates both the shape and texture information. A simple graph construction algorithm is presented which uses the feature points of an image as vertices, and their corresponding feature descriptors as labels. Matching is performed in two stages. In the first stage, a Maximum a posteriori solution is computed using the aging model of the individuals to effectively reduce the search space and identify potential individuals for the second stage. In the second stage, a simple deterministic graph matching algorithm that exploits the spatial similarity between the graphs is proposed.

2.3.4.1.4 Experimental Results

The paper evaluates the performance of the proposed age-invariant FR approach and compares its performance to a 3D aging model technique proposed in the literature.

2.3.4.1.4.1 Database

The proposed framework was tested the FG-NET database.

2.3.4.1.4.2 FR Algorithm(s)

This is a standalone FR system.

2.3.4.1.4.3 Sample Results

Figure 2-53 illustrates a comparative assessment of the proposed and other comparable age-invariant FR methods. The proposed Graph-based GMM method appears to consistently perform better than the other methods.

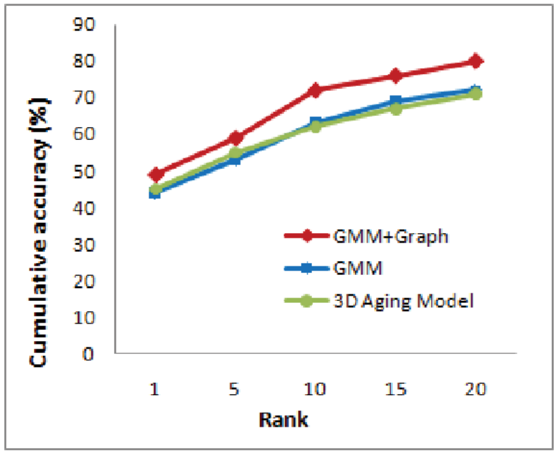


Figure 2-53 CMC Curve for Age in Range [18, 69] for FG-NET Database (Source: [P-15])

2.3.4.1.5 Key Contributions

In this paper, a graph based image representation is proposed and an aging model constructed using GMM for each individual to model their age variations mainly in shape and texture. Some of the key features of the proposed approach include:

- A modified Local Feature Analysis that uses Fisher score to extract the feature points has been used effectively to extract feature points.
- Uniform LBP operator is applied to each feature point to compute a feature descriptor for each feature point, and is used in the graph representation.
- A two stage approach for recognition has been proposed in which a simple deterministic algorithm that exploits the topology of the graphs is proposed for efficient graph matching between the probe image and the gallery image.
- The experimental results indicate that the combination of aging model and the graph representation perform well in age invariant face recognition.
- Thus, an effective representation of the spatial relationship between the feature points of an image can improve the performance of a face recognition system across age progression.

2.3.4.2 Paper # [P-16]

2.3.4.2.1 Abstract

In this paper, a discriminative model is proposed to address face matching in the presence of age variation. In this framework, each face is first represented by designing a densely sampled local feature description scheme, in which Scale Invariant Feature Transform (SIFT) and Multi-scale Local Binary Patterns (MLBP) serve as the local descriptors. By densely sampling the two kinds of local descriptors from the entire facial image, sufficient discriminatory information, including the distribution of the edge direction in the face image (that is expected to be age invariant) can be extracted for further analysis. Since both SIFT-based local features and MLBP-based local features span a high-dimensional feature space, to avoid the over-fitting problem, an algorithm, called Multi-Feature Discriminative Analysis (MFDA), is designed to process these two local feature spaces in a unified framework. The MFDA is an extension and improvement of the LDA using multiple features combined with two different random sampling methods in feature and sample space. By random sampling the training set as well as the feature space, multiple LDA-based classifiers are constructed and then combined to generate a robust decision via a fusion rule.

2.3.4.2.2 Motivation

Aging variation poses a serious problem to automatic face recognition systems. Most of the face recognition studies that have addressed the aging problem are focused on age estimation or aging simulation. Designing an appropriate feature representation and an effective matching framework for age invariant face recognition remains an open problem. Published approaches to age invariant face recognition are limited. Most of the available algorithms dealing with facial aging problem are focused on age estimation and aging simulation. The approach to match two face images of the same person acquired at different ages proposed in this paper differs significantly from the previously published approaches.

2.3.4.2.3 Method

The proposed discriminative model consists of two components: densely sampled local feature description and MFDA, as described next.

2.3.4.2.3.1 Densely Sampled Local Feature Description

Compared to the global appearance features, local features have been shown to be more effective in representing face images at diverse scales and orientations and robust to geometric distortions and illumination variations. Hence, the local image descriptor-based technique for face representation is adopted.

First, the whole face image is divided into a set of overlapping patches and then the selected local image descriptors are applied to each patch. The extracted features from these patches are concatenated together to form a feature vector with large dimensionality for further analysis. This process is illustrated in Figure 2-54.

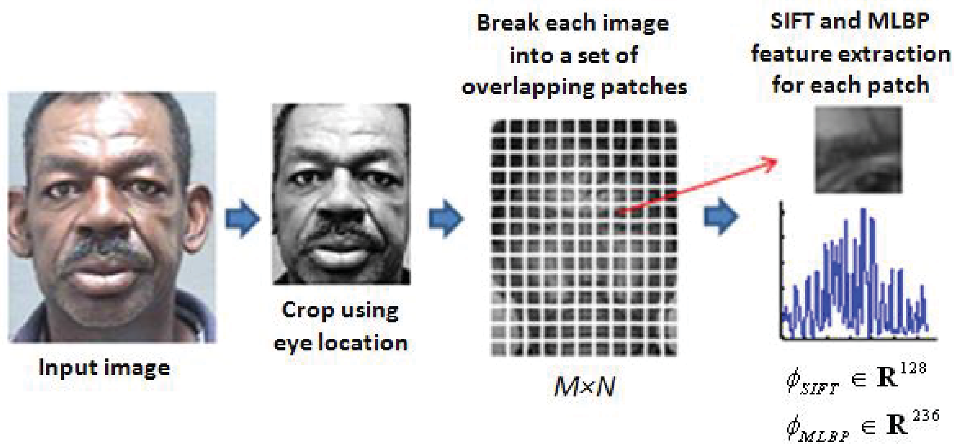


Figure 2-54 Illustration of Local Features Representation of a Face Image (Source: [P-16])

2.3.4.2.3.2 Multi-Feature Discriminant Analysis (MFDA)

The MFDA is proposed specifically for handling multiple feature sets with large dimensionality and with different scales and measurements. The output of the previous step (densely sampled local feature description), there are two kinds of local features (SIFT and MLBP), each with two different feature sets corresponding to two different patch sizes. In order to effectively handle these large numbers of features for enhanced performance, two problems need to be addressed:

1. different incompatibility in scale and measurement and
2. over-fitting problem.

The MFDA algorithm solves the traditional dimensionality reduction problem. In MFDA, different kinds of features are broken into slices and then scaled by PCA normalization, and the over-fitting problem is solved by the random sampling.

2.3.4.2.4 Experimental Results

The paper evaluates the performance of the proposed age-invariant FR approaches and compares its performance to several other methods found in the literature.

2.3.4.2.4.1 Database

The proposed framework was tested the FG-NET as well as the MORPH album 2, which is considered to be the largest face aging dataset available in the public domain.

2.3.4.2.4.2 FR Algorithm(s)

This is a standalone age-invariant FR system.

2.3.4.2.4.3 Sample Results

Table 2-16 illustrates a comparative assessment of the proposed age-invariant FR methods. The proposed discriminative model method appears to consistently perform better than the other tested methods.

Table 2-16 Comparison of Age Invariant Face Recognition Methods (Source: [P-16])

	Approach	Database (# subjects, # images) in probe and gallery	Rank-1 recognition accuracy reported (%)
Lanitis et al. (2002) [14]	Build an aging function in terms of PCA coefficients of shape and texture	Private database (12,85)	68.5%
Ramanathan et al. (2006) [10]	Shape growth modeling up to age 18	Private database (109,109)	15.0%
Wang et al. (2006) [11]	Build an aging function in terms of PCA coefficients of shape and texture	Private database (NA,2000)	63.0%
Geng et al. (2007) [4]	Learn aging pattern on concatenated PCA coefficients of shape and texture across a series of ages	Public domain FG-NET (10,10)	38.1%
Park et al. (2010) [18]	Learn aging pattern based on PCA coefficients in separate 3D shape and texture spaces from the given 2D database	Public domain FG-NET (82,82)	37.4%
		Public domain MORPH Album 1 (612,612)	66.4%
		Public domain MORPH Album 2 (10000,20000)	79.8%
Proposed discriminative model	Use discriminative analysis method with densely sampled local descriptors	Public domain FGNET (82,82)	47.50%
		Public domain MORPH Album 2 (10000,20000)	83.9%

In order to verify the generality of the proposed discriminative model, additional experiments on the Face and Gesture Recognition Research Network (FGNET) database to compare the discriminative-based approaches with the Face Video Analysis for Cognitive Systems (FaceVACS), one of the best state-of-the-art face recognition systems, as well as the generative model approach, as illustrated in Figure 2-55. Both the generative and discriminative approaches outperform one of FaceVACS. However, the discriminative approach offers more significant improvement, yielding rank-1 accuracy of 83.9% compared to ~79% rank-1 accuracy of both generative model and FaceVACS.

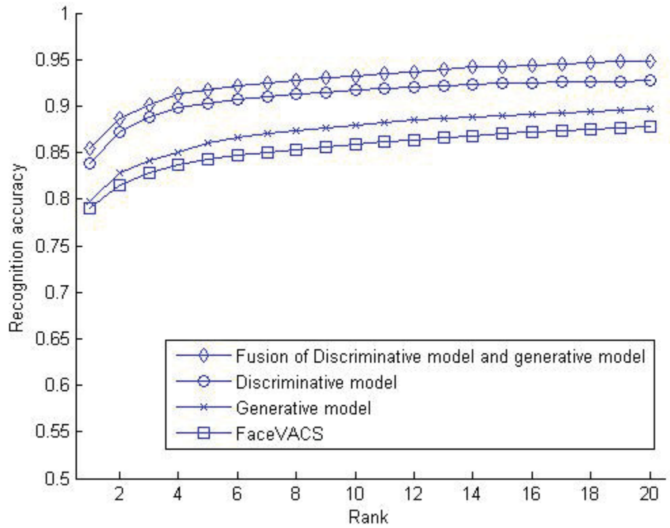


Figure 2-55 Cumulative Matching Characteristic (CMC) Curves of Different Aging Models (Source: [P-16])

2.3.4.2.5 Key Contributions

This paper proposes a discriminative model for age invariant face recognition. The key features of the proposed method are as follows:

- The proposed approach addresses the face aging problem in a more direct way without relying on a generative aging model. This obviates the need of a training set of subjects that differ only in their age with minimal variations in illumination and pose, which is often a requirement to build a generative aging model
- Each face is represented with a patch-based local feature representation scheme. In order to overcome the large feature dimensionality problem, a multi-feature discriminant analysis (MFDA) method is adopted to refine the feature space for enhanced recognition performance.
- Experimental results on two public domain databases (MORPH and FGNET) show the effectiveness of the proposed method.
- The performance of the proposed method surpasses that of FaceVACS, a commercial state-of-the-art face recognition engine.

2.3.4.3 Paper # [P-17]

2.3.4.3.1 Abstract

This paper proposes two methods: discrete wavelet transforms and combination discrete wavelet transforms with gradient orientation for feature representation, followed by a principal component analysis to reduce the features dimensions, then recognize by Euclidean distance. Experimental results are presented on the FGNET database. The proposed methods results compared with gradient orientation pyramid. Experimental results show that the proposed method is efficient and it can achieves better recognition performance.

2.3.4.3.2 Motivation

Face recognition has many applications such as biometric, image retrieval, surveillance. Face recognition across age is challenging problem which that has not been widely studied compared to other facial variations due to pose, lighting, and expression. And it is practical to verify passport photo and identifying missing children. Most of works on facial aging focus on modeling the aging process, age estimation and age simulation. Existing face recognition methods can be mainly classified into two categories: generative and non-generative. In this paper, two methods proposed for feature representation: discrete wavelet transform and combination discrete wavelet transform with gradient orientation for feature extraction and PCA applied for dimension reduction of features.

2.3.4.3.3 Method

The overall flowchart of the proposed face recognition system is shown in Figure 2-56, which consists of three steps:

1. Pre-processing step that in this stage, the non-face parts is removed from face images.
2. The next step is feature extraction that methods used to overcome the effects of aging. A Discrete Wavelet Transform (DWT) and DWT+GO for feature extraction have been proposed.
3. Finally, the feature vector dimensions are reduced by the PCA and recognition is performed by the Euclidean distance between features.

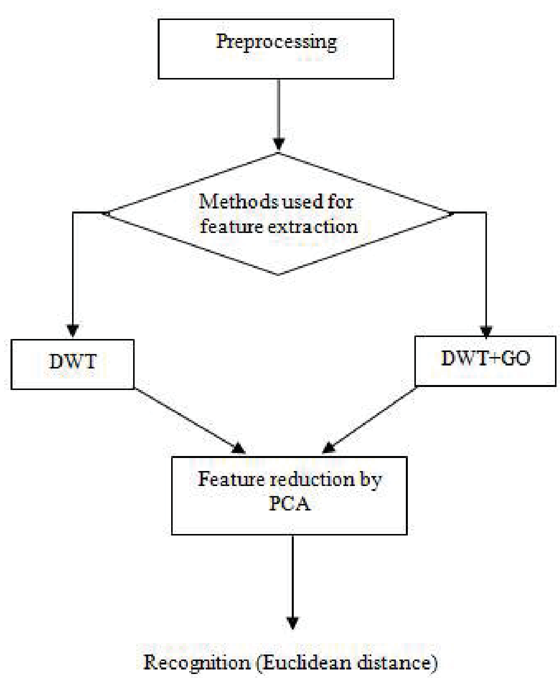


Figure 2-56 Framework of the Proposed Model-Based FR Method (Source: [P-17])

The paper explored two methods: DWT+PCA and DWT+GO+PCA for age invariant face recognition.

2.3.4.3.4 Experimental Results

The paper evaluates the performance of the proposed age-invariant FR approaches. In particular, the following approaches were compared:

- DWT+GO+PCA (proposed method)
- DWT+PCA (proposed method): In this approach, a two-level wavelet decomposition is applied on each image, low frequency sub-band used as feature vector and PCA applied on feature vector.
- GOP+PCA: This approach combined gradient orientation pyramid and PCA.

2.3.4.3.4.1 Database

The proposed framework was tested on different age groups from FGNET database images.

2.3.4.3.4.2 FR Algorithm(s)

This is a standalone age-invariant FR system.

2.3.4.3.4.3 Sample Results

Table 2-17 illustrates a comparative assessment of the proposed age-invariant FR methods. The proposed DWT+GO+PCA appears to consistently perform better than the other two methods.

Table 2-17 Assessment of the Proposed Age-Invariant FR Methods (Source: [P-17])

Database	Age range	Method	Accuracy
1	0-2	GOP+PCA	44.12%
		WT+PCA	8.8%
		WT+GO+PCA	*47.05%*
2	3-5	GOP+PCA	45.45%
		WT+PCA	12.12%
		WT+GO+PCA	*51.51%*
3	6-8	GOP+PCA	50%
		WT+PCA	22.73%
		WT+GO+PCA	*59.09%*
4	9-11	GOP+PCA	41.66%
		WT+PCA	8.33%
		WT+GO+PCA	*41.66%*

The recognition performance of low frequency subbands at the different wavelet decomposition level was also evaluated. The DWT+GO+PCA method was applied on 3 set images, using an increasing number of wavelet decomposition levels (from 1 to 3 levels). The results are shown in Table 2-18. The results show that recognition accuracy increases with the number of DWT level of devecomposition.

Table 2-18 Assessment of the Recognition Rate Dependence of the Number of Levels of the DWT for the Proposed DWT+GO+PCA Method (Source: [P-17])

Approximation Coefficients	First Level	First and Second Level	First, Second and Third Level
Recognition Accuracy	54.54%	59.09%	59.09%

2.3.4.3.5 Key Contributions

This paper explores the problem of age invariant face recognition. Some of the key features of the proposed method include:

- Combination discrete wavelet transform of gradient orientation are proposed for representation feature, followed by a PCA to reduce the dimensions of the extracted features.

- The proposed methods performance is evaluated on FGnet database. This approach compared with discrete wavelet transform with PCA (DWT+PCA) method and (GOP+PCA) method.
- The experiment shows that DWT+GO+PCA method gives best performance for age invariant face recognition in compare with two other methods.

2.3.5 Low Resolution

As illustrated in Figure 2-57, low resolution problem occurs in a face recognition system when resolution of the face image to be recognized is lower than 16×16. This problem happens in many surveillance applications, such as small scale standalone camera applications in supermarkets and banks, Closed Circuit Television (CCTV) in public streets, etc. where images taken from a surveillance camera generally consists of very small face area and cannot provide enough resolution of face for recognition. As the person face is not close to the camera, the face region will be smaller than 16x16. Such a low resolution face image consists of very limited information as most of the details are lost. This can drop down the recognition rate drastically.

Figure 2-57 Typical Frame from a Surveillance Video (CAVIAR database) (Source: [R-2])

Table 2-19 lists the specific facial recognition papers, which address the low image resolution challenges, reviewed in this section.

Table 2-19 Selected Papers on Addressing the Low Resolution Challenges

#	Paper Title	Authors	Source	Year	Type
P-18	Low-resolution face recognition: a review	Zhifei Wang, Zhenjiang Miao, Q.M. Jonathan Wu Yanli Wan, Zhen Tang	The Visual Computer, Springer-Verlag Berlin Heidelberg 2013 http://link.springer.com/article/10.1007/s00371-013-0861-x?no-access=true	2013	Survey

#	Paper Title	Authors	Source	Year	Type
P-19	Improving long range and high magnification face recognition: Database acquisition, evaluation, and enhancement	Yi Yao a, Besma R. Abidi a,*, Nathan D. Kalka b, Natalia A. Schmid b, Mongi A. Abidi	Computer Vision and Image Understanding 111(2):111-125 (2008)	2008	Proposed new technique(s)
P-20	Very Low Resolution Face Recognition Problem	Wilman, W.W. Zou and Pong C. Yuen	Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on Biometrics Compendium, Washing DC, 27-29 Sept. 2010.	2010	Proposed new technique(s)
P-21	Pose-robust recognition of low-resolution face images	Soma Biswas, Gaurav Aggarwal and Patrick J. Flynn	2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 601 – 608.	2011	Proposed new technique(s)

2.3.5.1 Paper # [P-18]

This paper provides a comprehensive survey on these methods and discusses many related issues. First, it gives an overview on Low Resolution Face Recognition (LR FR), including concept description, system architecture, and method categorization. Second, many representative methods are broadly reviewed and discussed. They are classified into two different categories, super-resolution for LR FR and resolution-robust feature representation for LR FR. Their strategies and advantages/disadvantages are elaborated.

2.3.5.1.1 Challenges in Low Resolution Face Recognition (LR FR)

Low Resolution FR aims to recognize faces from small size or poor quality images with varying pose, illumination, expression, etc. In particular, the challenges associated with LR FR can be attributed to the following factors:

- **Misalignment:** Inaccurate alignment will severely affect the performance of High Resolution (HR) as well as LR face recognition systems and it is difficult to perform automatic alignment on LR images.
- **Noise affection:** As a generalized issue, LR problem causes a lot of chain effects for face recognition. The degradation in resolution together with pose, illumination, and expression variations adds complexity to the recognition. In other words, these variations produce much more noises for LR FR.
- **Lack of effective features:** LR leads to the loss of large amounts of information. Most effective features used in HR FR such as Gabor and LBP may fail in LR case, especially very LR case, e.g., 6 × 6. Novel features insensitive to resolution are essential for LR FR.
- **Dimensional mismatch:** Different resolutions between gallery images and probe ones in LR FR systems cause dimensional mismatch in traditional subspace learning methods.

2.3.5.1.2 Low Resolution Imagery

The addressed low resolution FR problem involves automatically recognizing people by their face images in LR. A LR face image means that a face size is smaller than 32 × 24 pixels (with an eye-to-eye distance about 10 pixels), typically taken by surveillance cameras (maximum resolution 320 × 240) without subject's cooperation and normally contains noises and motion blurs. In an LR image, exact delineation of facial features is not so trivial both for humans and machines.

Generally, LR images discussed here roughly fall into three cases as follows:

- **Small size**: For which the probe images have the insufficient number of pixels. According to the reports of FRVT2000 on resolution experiment, the metric used to quantify resolution is eye-to-eye distance in pixels. Nevertheless, the distance is not usually adopted in most LR FR systems but replaced by face size. In the conventional methodology, small size is sufficient for face recognition. However, when the size of face captured from surveillance camera is smaller than 32 × 24 pixels, or the size of face down-sampled from HR static image is smaller than 16 × 16 and even 6 × 6 pixels, most conventional methods will be of no effect.
- **Poor quality**: Which is from the fact that probe images are provided in the resized form with blur (e.g., out of focus, interlace, and motion blur), variation of illumination and loss of details. Therefore, the underlying resolutions of the images are comparatively low, which means that even if the size of face is 200 × 200 pixels, there is no guarantee that the face is in HR. The levels of "focus" and "illumination" are taken as image quality measures. However, they just consider LR problem from vision perspective rather than recognition purpose.
- **Small size & Poor quality**: This is combination of the two weaknesses.

Examples of LR images depicting some of the above weaknesses are demonstrated in Figure 2-58.

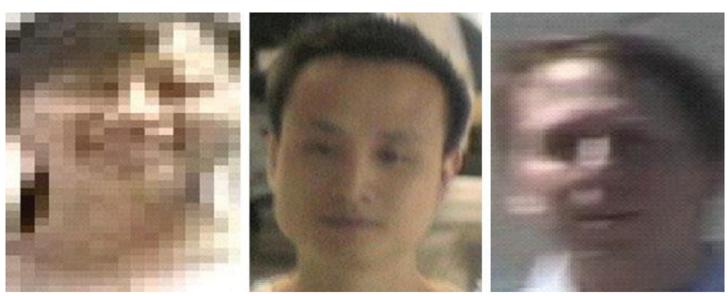


Figure 2-58 Some Examples of LR Face Images (Source: [P-18])

2.3.5.1.3 Low Resolution FR System Architecture

Similar to conventional HR FR system, LR FR system also includes three main parts. That is LR face detection or tracking, LR feature extraction and LR feature classification, as illustrated in Figure 2-59. In general, the latter two stages are collectively referred to as LR FR, which is the focus in this review.

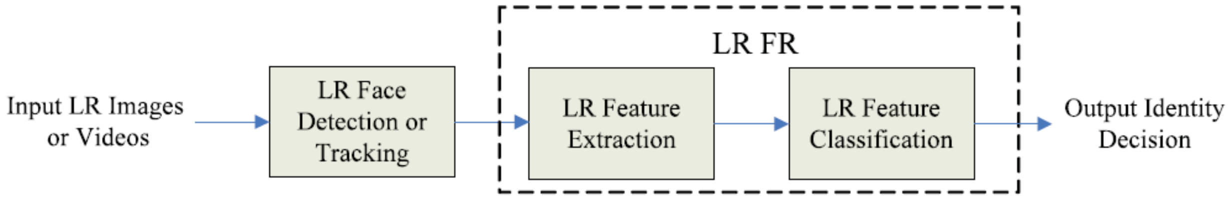


Figure 2-59 System Architecture of LR FR (Source: [P-18])

LR FR initially extracts resolution-robust features, and performs classification by matching the features to obtain the identity decision. The steps are similar to HR FR system from general framework. However, as opposed to HR FR, LR FR needs to consider the particular problem of dimensional mismatch. In practical face recognition applications, it is reasonable to assume that all gallery images are in HR. From classification perspective, LR will obviously cause the mismatch problem between gallery/probe pairs, as illustrated in Table 2-20. To deal with the problem, three general ways can be considered as follows:

1. **Up-scaling (or interpolation)**: This uses interpolation methods, such as cubic interpolation, is conventionally adopted in most of the subspace-based face recognition methods. For LR images, it does not introduce any new information but potentially brings noises or artifacts. Therefore, the process of up-scaling can be feasible under high-resolution or middle resolution, but may drop in performance confronting with much lower resolution. Thus, it is generally not a good way for solving LR problem. For further refined solution, *super-resolution or hallucination* can be employed to estimate HR faces from LR ones. However, it usually requires a lot of images, which belongs to the same scene with precise alignment, and it also needs large computation cost.

2. **Unified feature space**: This method is used to project HR gallery images and LR probe ones into a common space. This idea seems to be direct and reasonable for solving LR problem. However, it is difficult to find the optimal inter-resolution space. And the two bidirectional transformations from both HR to IR and LR to IR may bring much more noises.

3. **Down-scaling:** This seems to be the feasible solution for the mismatch problem. Unfortunately, it reduces the amount of available information, especially the high-frequency information mainly for recognition. However, downscaling on both training/gallery and test/probe may improve the performance under very LR case such as 7×6.

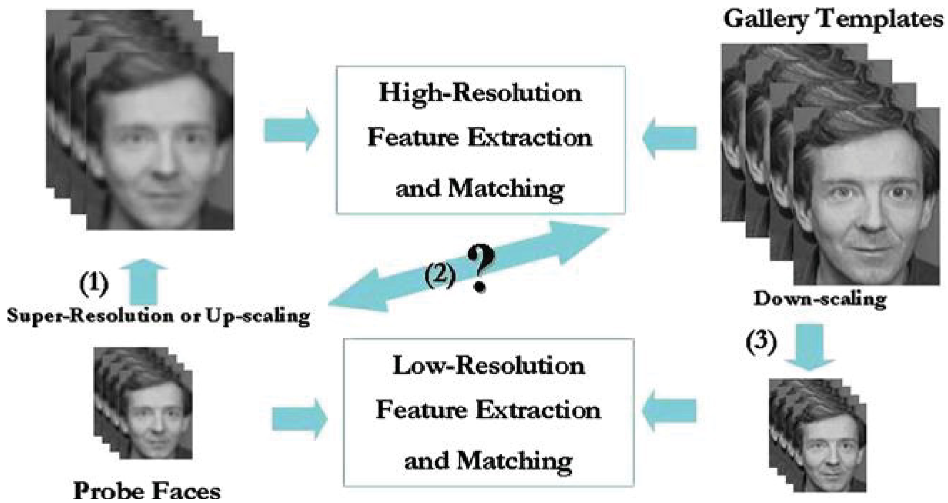


Figure 2-60 Three General Approaches for LR FR (Source: [P-18])

The paper generally classifies LR FR methods into two categories:

1. Indirect method: Super Resolution (SR) is initially used to synthesize the higher-resolution images from the LR ones, and then traditional HR FR methods could be applied for recognition. The landmark works in this category are face hallucination and simultaneous SR and recognition (S_2R_2). Two criteria are considered for SR applications: visual quality and recognition discriminability. However, most of these methods aimed to improve face appearance but failed to optimize face images from recognition perspective. Recently, a few attempts were made to achieve these two criteria under very LR case.
2. Direct method: resolution-robust feature representation is the process of directly extracting the discriminative information from LR images. The landmark works are color feature and Coupled Locality Preserving Mappings (CLPM). These methods can be separated into two groups further. One is feature-based method in which the resolution-robust features, such as texture, and subspace information, are used to represent faces. However, some features used in traditional HR FR methods are sensitive to resolution. The other is structure-based method, e.g., Multidimensional Scalings (MDS) in which the relationships between LR and HR are explored in resolution mismatch problem.

Table 2-20 Categorization of LR FR Methods (Source: [P-18])

Categorization		Representatives
Indirect method: Super-resolution for LR FR	Vision-oriented SR	Face Hallucination [52, 57];
		Two-step Statistical Approach [53, 58];
		Eigentransformation [36];
		Extended Morphable Face Model [59].
	Recognition-oriented SR	Simultaneous SR and Recognition (S^2R^2) [16, 60–62];
		Multi-Modal Tensor SR (M^2TSR) [12];
		Discriminative SR (DSR) [3, 63];
		Support Vector Data Description (SVDD) [64, 65].
Direct method: Resolution-robust feature representation for LR FR	Feature-based method	Color-based Feature [17, 66–68];
		Texture-based Feature: Local Frequency Descriptor (LFD) [18, 69];
		Kernel Class-dependence Feature Analysis (KCFA) [56].
	Structure-based method	Eigenspace Estimation (EE) [15];
		Coupled Locality Preserving Mappings (CLPMs) [19, 70];
		Multi-Dimensional Scaling (MDS) [20, 71, 72];
		Coupled Kernel Embedding (CKE) [21].

2.3.5.1.4 Super-Resolution for Low Resolution FR

Many researchers want to build face recognition systems with LR images obtained by web cameras or close-circuit television. However, the overall performance of LR FR needs great improvement. Compared with the development of resolution-robust face recognition methods, super-resolution (SR), or hallucination methods have gained much more attentions, due to many problems that degrade the quality of face images in LR case. In this section, some typical SR methods specifically satisfying the requirement for face recognition will be reviewed. In the last decade, most of the conventional SR methods called *vision-oriented SR* were taken as the indirect way, which is reconstruction followed by recognition. Recently, some researchers focused on simultaneous SR and recognition, and SR mainly for recognition, called as *recognition-oriented SR* obtaining promising results for LR classification.

2.3.5.1.4.1 Vision-Oriented Super-Resolution for LR FR

The simplest way to increase resolution is direct interpolation of input images with methods such as nearest neighbor, bilinear, and bi-cubic. However, its performance is usually poor since no new information is added into the process. In contrast to the interpolation, SR increases resolutions of images or video frames using the relationships among several images. Generally, SR can be divided into two classes: *reconstruction-based method* (from input images alone), and *learning-based method* (from other images).

The reconstruction-based method reconstructs HR images based on sampling theory by simulating the image formation process. However, the method has some fatal shortcomings. It was shown that the method inherits limitations when the magnification factor increases and the question: “Do fundamental limits exist for the reconstruction-based SR?” was explored. An explicit bound of the magnification factor based on perturbation theory analysis was derived and more recent cited efforts tried to solve the problem and improve the magnification factor from about two to almost four by using multilayer perceptron.

The learning-based method, also known as face hallucination, which is the focus in this review, is always used to enhance resolutions of face images compared with the reconstruction-based method. In essence, the learning-based SR is used to learn the relationships between LR and HR corresponding to different face images in a training set, and then use these learnt relationships to predict fine details for LR probe images (stored by image pixels, image patches, or coefficients of alternative representations). Establishing a good learning model to obtain the prior knowledge is the key to the learning-based method. At present, the commonly used learning models include the PCA model, image pyramid model, Markov model, etc.

2.3.5.1.4.2 Recognition-Oriented Super-Resolution for LR FR

Recognition-oriented SR is not to use SR before recognition in the conventional way. It embeds the elements of SR methods into face recognition. Specifically, it fuses the models of the image formation process and the prior information, together with feature extraction and classification to design methods for recognition. Compared with vision-oriented SR, recognition-oriented SR maybe more suitable for LR FR due to the following two observations. Firstly, it simultaneously performs *SR and feature extraction* with the direct goal of recognition. Secondly, it performs *feature SR* with the aim of reconstructing not only the low-frequency content (structure information) but also the high-frequency content (discriminative information) for recognition.

2.3.5.1.5 Resolution-Robust Feature Representation for LR FR

In contrast to super-resolution for LR FR, researches on resolution-robust feature representation in LR problem started around year 2008. The difficulties of finding the effective features in LR case render face recognition more complicated. Some typical features in HR case such as texture, shape, and color may fail in the LR case. Therefore, the only feasible option is to explore the potential of these features for LR FR. In addition, LR results in a dimensional mismatch problem under the subspace framework. Building inter-resolution space may provide a promising direction for solving this problem. Resolution-robust feature representation method is classified into two groups: *feature-based method* and *structure-based method.*

2.3.5.1.5.1 Feature-Based Method

This method identifies an LR face directly using the features extracted from probe images in resized forms. However, all the existing resolution-robust features are improved from the features used in HR FR, such as the improved color space and the improved local binary pattern descriptor. The feature-based method can be further classified into two categories, that is, the *global feature-based method* and *local feature-based method*.

2.3.5.1.5.2 Structure-Based Method

Compared with the feature-based method concerning resolution-robust features, the structure-based method focuses on constructing the relationships between LR and HR feature space for facilitating direct comparison of LR probe images with HR gallery ones from a classification perspective. The method aims to build the holistic framework for LR matching by especially solving the particular problem in LR FR, namely dimensional mismatch. Here, three kinds of structure-based methods are introduced. *Coupled mappings* aim to find the structure relationships. *Resolution estimation* determines the kinds of structures chosen for building LR FR system. Finally, the *sparse representation based method* is adopted for representing LR probe images using HR training images from a structure perspective.

2.3.5.1.6 SR vs. Resolution-Robust Feature Representation: A Comparison

A summary of the SR and resolution-robust feature representation approaches for LR FR is as follows:

- Most of the vision-oriented SR methods focus on obtaining a good visual reconstruction rather than a higher recognition rate; however, the essence of the recognition-oriented SR methods is to satisfy the need of recognition with LR images.
- Both vision-oriented SR and recognition-oriented SR are sensitive to different variations such as pose, and require lots of training samples of the same scene.
- In general, although SR methods require large computation costs, they have the potential advantages in very LR cases like 6×6.
- In general, resolution-robust feature representation methods are mainly applicable in relatively LR cases rather than very LR cases like 6 ×6.
- The feature-based methods can be used for multiple resolutions from HR to LR, but they need online training. However, the structure-based methods are more suitable for offline training, but they are mainly used for a single resolution application with the balance between efficiency and speed.

Table 2-21 illustrates a comparison between super-resolution and resolution-robust feature representation.

Table 2-21 Comparison Between Super-Resolution and Resolution-Robust Feature Representation (Source: [P-18])

Comparison issues	Super-resolution	Resolution-robust feature representation
Main purpose	For visual quality	For recognition discriminability
Resolution tolerance	Very LR	Relatively LR
Performing mode	Indirect way	Direct way
Computational complexity	High	Low
Unconstrained variations	Sensitive	Very sensitive
Requirements of training samples	Many	Few

2.3.5.1.7 Evaluations of Low Resolution FR methods

In order to have a clear idea on various LR FR methods, it is important to evaluate them based on certain evaluation criteria with some standard LR face databases. Unfortunately, such a requirement is seldom satisfied in practice due to the lack of general criteria and databases originally developed for LR FR. At present, LR FR methods are just evaluated based on HR FR criteria and databases.

The paper summarizes the performances of LR FR methods tested on several databases, using various LR FR methods. We limit our illustration to the performances of LR FR methods tested on FERET database, as summarized in Table 2-22.

Table 2-22 Experiments and Performances of LR FR on FERET Database (Source: [P-18])

No. of subjects	Probe resolution	Gallery/Probe	Primary variation	Method	Accuracy	Baseline
1196	12×12	1196/1195	Frontal, Expression	CLPMs [19]	90.1 %	61.8 % (PCA)
1196	33×30	1196/1195	Frontal, Blur	LFD [18]	86 %	48 % (LBP)
865	6×6	N/A	Frontal, Expression	S^2R^2 [16]	62 %	60 % (LDA)
295	14×9	1/4	Pose, Illumination	M^2TSR [12]	74.6 %	51.4 % (Tensor)
1196	12×12	1196/1195	Frontal, Expression	NMCF [92]	84.4 %	36.9 % (PCA)
140	15×15	1/4	Frontal, Blur, Color	RQCr [17]	68 %	54 % (R)

2.3.5.1.8 Existing Problems and Future Trends

Compared with other subareas such as pose-invariant face recognition and illumination-invariant face recognition, LR FR is a new sub-area in face recognition. Thus, its first priority is to guarantee efficiency and accuracy rather than real time and low computational complexity. Although researchers have exerted efforts on improving LR FR methods, some specific problems still exist in real applications, especially in surveillance scenarios. Based on the four main parts of low resolution FR systems, including preprocessing, facial representation, feature extraction, and feature classification, four challenges and corresponding four future directions are identified. These challenges are as follows:

1. **Automatic alignment**: Alignment is one of the most important preprocessing issues in face recognition, especially in LR FR. In most LR FR methods such as S_2R_2 and CLPMs, face images are manually aligned and cropped with the positions of eyes and mouth. However, the way of manual alignment is very difficult in practical LR FR applications such as video surveillance. Furthermore, state-of-the-art face detectors such as AdaBoost remain poor in detecting LR face images.
2. **Insensitivity to multiple variations**: The performance of face recognition highly depends on unconstrained variation, e.g., pose, illumination, expression. Many methods have been proposed for reducing the effects of the changes, such as the Gabor filter function and the Discrete Cosine Transform (DCT) on an edge map. However, most of the successful techniques could not be efficiently applied to LR data. For the LR case, the variations are the largest noises to some extent. Most of the existing LR FR methods assumed the constrained cases such as frontal pose, good illumination, and neutral expression, while few methods focused on facial representation against different variations.
3. **Resolution-robust feature extraction**: Most of the effective features used in HR FR such as texture and color may fail in LR case. Thus, it is difficult to find resolution-robust features for LR FR, especially under facial and environmental variations such as pose, illumination, and expression.

4. **Discriminative nonlinear coupled mapping**: Different resolutions between HR gallery images and LR probe images cause a dimensional mismatch problem in the traditional classification framework. In fact, this problem is the most essential problem in LR FR compared with HR FR. Other problems such as misalignment, noise affection, and lack of effective features commonly exist in face recognition system no matter HR or LR case.

2.3.5.1.9 Concluding Remarks

This paper devotes to providing a comprehensive survey on LR FR. After giving an overview of the LR FR concept and introducing some structural categories of the existing methods, many representative methods have been reviewed and evaluated in detail. Discussions on major challenges as well as future research directions toward complete LR FR are provided.

Although significant progress has been made in the last decade, a robust Low resolution FR system should be effective under the following variations:

- Multiple resolutions for imagery acquired at a distance
- Noise and motion blur
- Orientation, pose and partial occlusion
- Illumination, age and facial expression variations.

From the discussions above, the conclusion can be drawn that LR is a challenging and interesting subarea in face recognition. The existing methods in LR FR fail to form a unified theoretical framework, and there are no standard databases and criteria to evaluate the performances of the LR FR methods. What are the effective features representing LR faces for recognition? How to improve super-resolution processing to satisfy both vision and recognition purpose? Is there a unified feature space where LR faces are separable? These questions will spur researchers to create more effective methods. Answers to these questions may lead to a clearer understanding of LR FR even general LR object recognition, which are similar to the findings of nonlinear mappings in pose-invariant face recognition and linear subspaces in illumination-invariant face recognition.

2.3.5.2 Paper # [P-19]

2.3.5.2.1 Abstract

This paper describes a face video database, University of Tennessee, Knoxville Long Range High Magnification (UTK-LRHM), acquired from long distances and with high magnifications. Both indoor and outdoor sequences are collected under uncontrolled surveillance conditions. It is the first database to provide face images from long distances (indoor: 10–16 m and outdoor: 50–300 m). The corresponding system magnifications range from 3× to 20× for indoor and up to 284× for outdoor. This database has applications in experimentations with human identification and authentication in long range surveillance and wide area monitoring. Deteriorations unique to long range and high magnification face images are investigated in terms of face recognition rates based on the UTK-LRHM database. Magnification blur is shown to be a major degradation source, the effect of which is quantified using a novel blur

assessment measure and alleviated via adaptive de-blurring algorithms. A comprehensive processing algorithm, including frame selection, enhancement, and super-resolution is introduced for long range and high magnification face images with a large variety of resolutions. Experimental results using face images of the UTK-LRHM database demonstrate a significant improvement in recognition rates after assessment and enhancement of degradations.

2.3.5.2.2 Motivation

Substantial developments have been made in face recognition research over the last two decades and multiple face databases have been collected, as illustrated in Table A-1. In these databases, face images or videos, typically visible Red Green Blue (RGB) or monochromatic, are recorded under different resolutions, illuminations, poses, expressions, and occlusions. The data is mostly collected from close distances and with low and constant camera zoom and is only well suited for close range applications, such as identity verification at access points. The rapidly increasing need for long range surveillance and wide area monitoring calls for a passage from close-up distances to long distances and accordingly from low and constant camera zoom to high and adjustable zoom. This enables the detection and verification of humans from long distances and provides critical early threat assessment and situational awareness. The work and database described in this paper serve this purpose and provide the research community with a standard testing foundation for long range face related research.

2.3.5.2.3 Method

The paper described the collected long-range database, with imagery data acquired at various distances and magnification levels. The paper then explores the effects of magnification blur on face recognition rates and the resulting degradation is quantified using adaptive sharpness measures. An image enhancement algorithm is then proposed to enhance long-range imagery data degraded by high magnificence.

2.3.5.2.3.1 Database

The paper first describes the establishes long range and high magnification face video database, UTK-LRHM database , which lays the foundation for long range face related research work. The database includes indoor and outdoor imagery and contains frontal view face images and videos collected with various system magnifications (10×– 284×), observation distances (10–300 m), indoor (office ceiling light and side light) and outdoor (sunny and cloudy) illuminations, still/moving subjects, and constant/varying camera zooms. Small expression and pose variations are also included in the video sequences of the used database. Table 2-23 and Table 2-24 illustrate some of the specifications of the indoor and outdoor databases, respectively. Also, additional descriptions of collected indoor and outdoor databases are provided in Table A-1.

Table 2-23 Indoor Sequence Specifications (Source: [P-19])

Indoor sequence specifications							
Still images							
Magnification (×)	1	10	12	14	16	18	20
Distance (m)	1	9.5	10.4	11.9	13.4	14.6	15.9
Inter-ocular distance (pixel)	60	57	57	58	60	60	60

Conditions	Magnification (×)	Distance (m)
Various video sequence conditions		
1. Constant distance & varying system magnification	10 → 20	13.4 and 15.9
2. Varying distance & constant system magnification	10 and 15	9.5 → 15.9
3. Varying distance & varying system magnification	10 → 20	9.5 → 15.9
Varying illumination, constant distance & system magnification	20	15.9

Table 2-24 Outdoor Sequence Specifications (Source: [P-19])

Outdoor sequence specifications						
Magnification (×)	66	109	153	197	241	284
Distance (m)	50	100	150	200	250	300
Inter-ocular distance (pixel)	79	76	79	76	78	78

This paper then makes use of the collected databases in order to explore the influence of magnification blur on face recognition rates is studied and the resulting degradation is quantified using adaptive sharpness measures.

2.3.5.2.3.2 Face Image Enhancement

High magnification images suffer from both increased image blur and high noise levels. In general, deblurring algorithms are sensitive to image noise unless an appropriate probabilistic model is used. Denoising algorithms, on the other hand, smooth out valuable image details. The resulting images are either low on details or too noisy. Since most face recognition engines are sensitive to both types of degradations, a good balance needs to be found to achieve an optimal recognition rate. A multi-scale processing algorithm based on wavelet transforms is proposed in this paper to restore and enhance data with high magnificence.

The proposed face image enhancement algorithm is outlined in the paper and its framework of the proposed algorithm is illustrated in Figure 2-61.

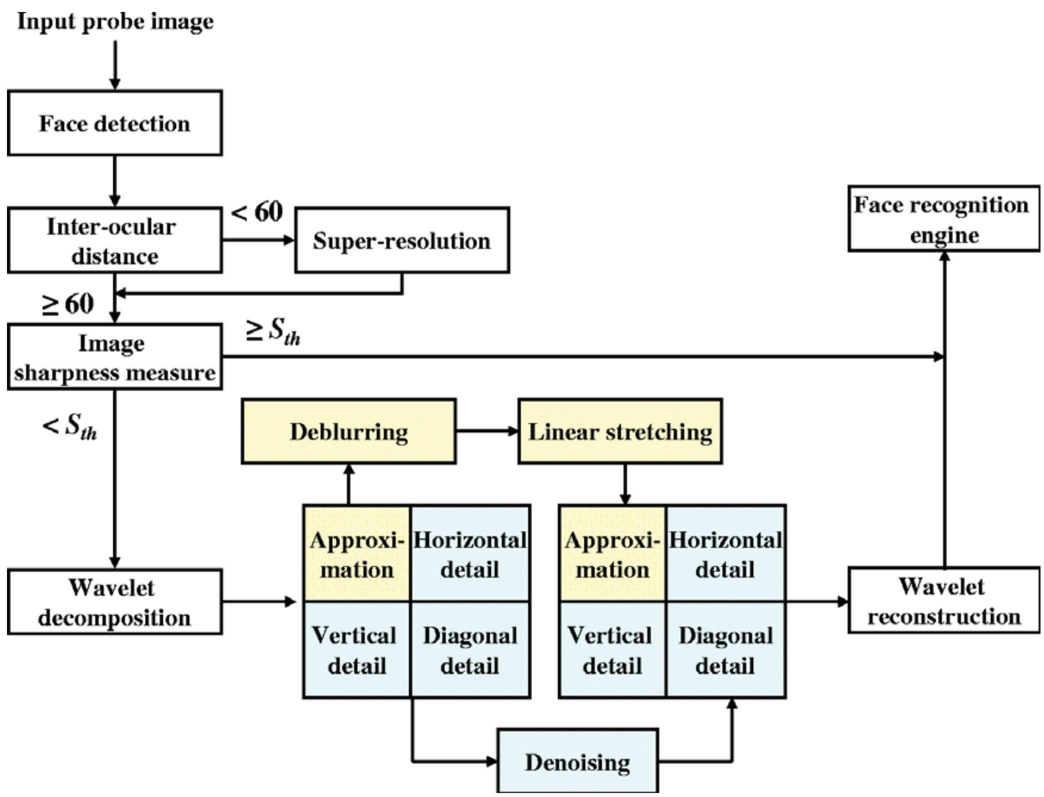


Figure 2-61 Framework of the Enhancement Algorithm for Long Range and High Magnification Face Images (Source: [P-19])

2.3.5.2.4 Experimental Results

The paper evaluates the performance of the proposed enhanced SR approach and compares its performance to other newly developed SR in the FR literature. .

2.3.5.2.4.1 Database

The paper uses the developed face video database, UTK-LRHM, acquired from long distances, indoor and outdoor, and with high magnifications.

2.3.5.2.4.2 FR Algorithm(s)

The proposed SR method is pre-processing step to enhance LR input images and generate enhanced SR versions of the input images. The output data are fed into two commercial recognition engines, FaceIt and VeriLook, in order to conduct FR processing. FaceIt employs LFA while VeriLook utilizes Gabor wavelets. Face recognition engines based on different approaches are selected to validate the generality of the assessment and conclusions.

2.3.5.2.4.3 Sample Results

The paper studied the effects of magnification blur on face recognition rates. The CMC curves with respect to various system magnifications are illustrated in Figure 2-62 and Table 2-25. Note how image deterioration from limited fine facial details causes the recognition rate to drop gradually as the system magnification increases.

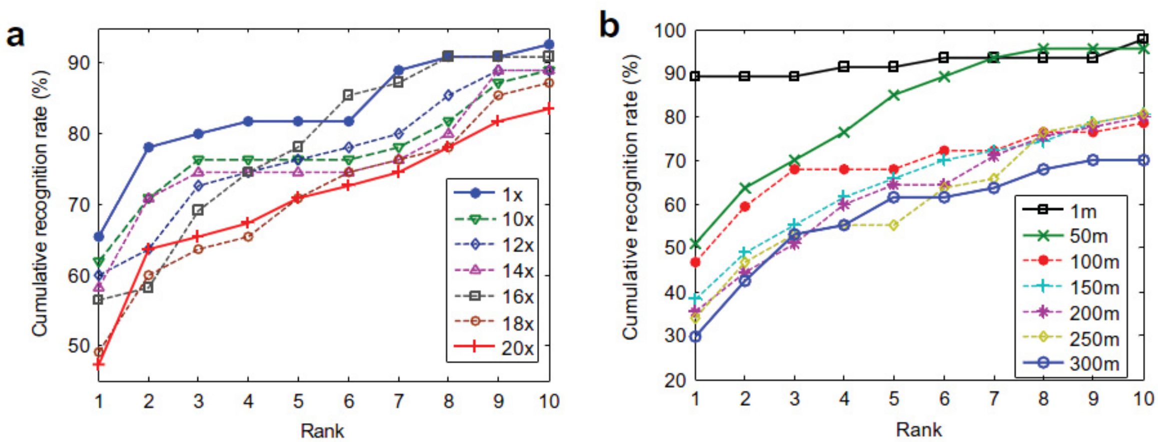


Figure 2-62 CMC Comparison across Probes with Different System Magnifications and Observation Distances: (a) Indoor and (b) Outdoor sessions (FaceIt FR Software) (Source: [P-19])

Table 2-25 Comparison of Rank-One Recognition Rates Across Probes with Different System Magnifications and Observation Distances (Source: [P-19])

Indoor session							
System magnification (×)	**1**	**10**	12	14	16	18	**20**
FaceIt® (%)	**65.5**	**61.8**	60.0	58.2	56.4	49.1	**47.3**
VeriLook® (%)	**63.6**	**41.8**	40.0	38.2	34.6	32.7	**32.7**
Outdoor session							
Observation distance (m)	**1**	**50**	100	150	200	250	**300**
FaceIt® (%)	**80.9**	**51.1**	46.8	38.3	35.6	34.0	**29.8**
VeriLook® (%)	**76.1**	**58.7**	41.3	34.8	30.4	21.7	**10.9**

This paper also studies and compares the performance of SR algorithms when applied to long range and high magnification face images. Image sets with different resolutions, measured by the inter-ocular distance in pixels, are selected from the UTK-LRHM database. Different probe sets are obtained by enhancing the same image set via various enhancement methods, including UM, de-convolution (Deconv), and the proposed wavelet based method. In addition, two probe sets, the unprocessed face images and the 1× reference face images, are also included and their performances serve as references for comparison. The same experiments are repeated for image sets at different magnifications and observation distances. The output data are fed into, FaceIt, a commercial recognition engine and the resulting face recognition rates tabulated and compared, as illustrated in Table 2-26. The proposed wavelet based algorithms are able to achieve the most improvements with an increase of 16.3% in the rank-one recognition rate for the 20× 16 m 60 p data set, yielding a performance comparable to the 1× reference. It is seen from these results that with the proper processing, the degradation in face recognition rate caused by magnification blur can be successfully compensated for.

Table 2-26 Comparison of Rank-One Recognition Rates Across Probes Processed by Different Enhancement Algorithms (Source: [P-19])

	10 × 9 m 60 p	20 × 16 m 60 p
Indoor		
Original	**61.8**	**47.3**
Eigen-face [19]	55.2	48.2
UM	54.5	50.9
Deconv	54.5	56.4
Wavelet + Deconv	52.7	56.4
Wavelet + Lasso SMS (no contrast stretching)	67.2	59.7
Wavelet + UM	65.5	65.5
Wavelet + UM SMS	63.6	65.5
Wavelet + Lasso	63.6	**63.6**
Wavelet + Lasso SMS	**69.1**	**63.6**
1× reference	**65.5**	
	109 × 100 m 80 p	284 × 300 m 80 p
Outdoor		
Original	**46.8**	**29.8**
UM	53.2	40.4
Deconv	50.4	34.0
Wavelet + UM SMS	59.0	36.9
Wavelet + Lasso SMS	**61.1**	**36.9**
1× reference	**89.4**	

Results are obtained using FaceIt®.

2.3.5.2.5 Key Contributions

The key contributions of this paper are as follows:

- A long range and high magnification face video database is established, which lays the foundation for long range face related research work. The observation distances and system magnifications reach up to 300 m and 284×, respectively.
- The influence of magnification blur on face recognition rates is studied and the resulting degradation is quantified using adaptive sharpness measures
- A wavelet based de-convolution algorithm is proposed to reduce computational complexity and preserve image edges especially facial features. Significant improvements in face recognition rates are also achieved.

2.3.5.3 Paper # [P-20]

2.3.5.3.1 Abstract

This paper addresses the Very Low Resolution (VLR) problem in face recognition in which the resolution of face image to be recognized is lower than 16x16. The VLR problem happens in many surveillance camera-based applications and existing face recognition algorithms are not able to give satisfactory performance on VLR face image. While face SR methods can be employed to enhance the resolution of the images, the existing learning-based face SR methods do not perform well on such a very low resolution face image. To overcome this problem, this paper models the SR problem under VLR case as a regression problem with two constraints. First, a new data constraint is designed to perform the error measurement on high resolution image space which provides more detailed and discriminative information. Second, a discriminative constraint is proposed and incorporated in the training stage so that the reconstructed HR image has higher discriminability. CMU-PIE, FRGC and Surveillance Camera face (SCface) databases are selected for experiments. Experimental results show that the proposed method outperforms the existing methods, in terms of image quality and recognition accuracy.

2.3.5.3.2 Motivation

While face recognition research has been studying for more than three decades and many promising practical face recognition systems have been developed, it is assumed the face region is large enough and contains sufficient information for recognition. Cited empirical studies showed that minimum face image resolution between 32×32 and 64×64 is required for existing algorithms. However, a wide range of applications cannot provide enough resolution of face for recognition, such as small-scale stand-alone camera applications in banks and supermarkets, large-scale multiple networked CCTV in law enforcement applications in public streets, etc. As shown in Figure 2-63, when the person is not close to the camera, the face region will be smaller than 16x16 pixels. Recognition of such a very low-resolution face image is called VLR face recognition problem.

The paper cites several SR algorithms have been proposed to increase the resolution of the face images so that the reconstructed higher resolution image is used for recognition. However, it is argued that most of the proposed SR methods require data constraints, which may not work well under the VLR problem. The paper proposes a SR problem, which overcomes some of these limitations and is shown to out-perform the existing methods, in terms of image quality and recognition accuracy.

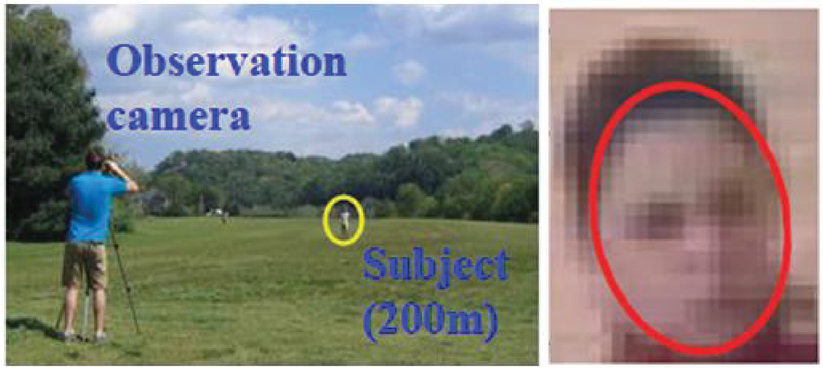


Figure 2-63 Low Resolution Due to Long Range Acquisition (Source: [P-20])

2.3.5.3.3 Method

As illustrated in Figure 2-64, super-resolution problem is modeled by learning the relationship, R, from the given training data, and the reconstructed HR image is recovered by applying R on the testing image. The relationship R, namely mapping pattern, is learnt by a regression model which minimizing the proposed new data constraint and discriminative constraint. Unlike existing data constraint verifying the reconstructed HR image in LR image space, the new data constraint perform this in HR image space which help the proposed algorithm better make use of the information from HR training images. Discriminative constraint is designed for recognition purpose, so that the proposed SR algorithm can use the class label information to boost the performance. The proposed SR method is referred to as the Relationship Learning (RL) based super-resolution method.

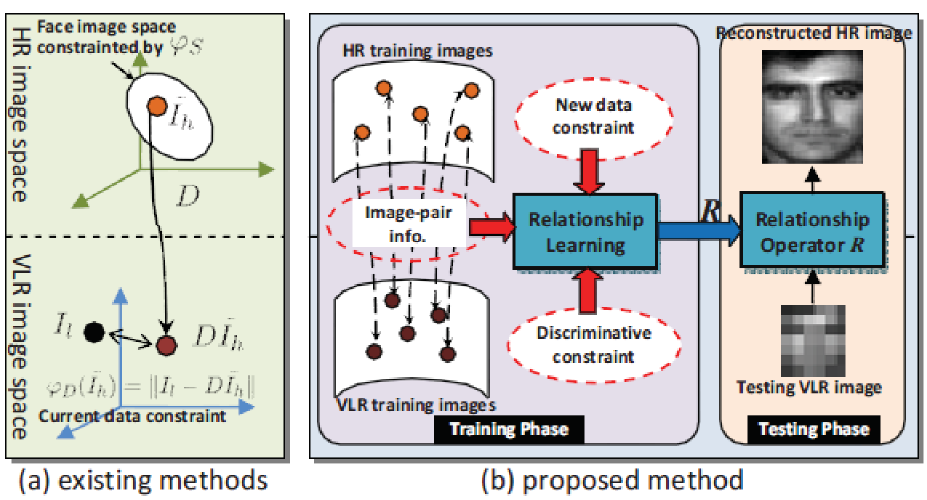


Figure 2-64 Overview of Existing Methods vs. the Proposed Method (Source: [P-20])

2.3.5.3.4 Experimental Results

The paper evaluates the performance of the proposed SR method by comparing its performance to three existing face SR methods, namely Hallucination Face (HF) method, Eigen-transformation based Eigen Face (EF) SR method and KPCA-based Kernel Face (KF) SR method. The SR images generated using each of the SR methods are fed into selected FR systems to compare their performances.

2.3.5.3.4.1 Database

Three databases, two public available face databases (CMU PIE and FRGC 2.0) and one surveillance camera face database, SCface are used for experiments in this paper.

2.3.5.3.4.2 FR Algorithm(s)

The proposed SR method is pre-processing step to enhance LR images and generate SR versions of the input images. The pre-processed SR data can then be fed into one of the many FR systems.

The paper uses three popular face recognition algorithms, namely Eigenfaces (linear method), KPCA (kernel method) and SVM (discriminative method) for the experiments.

2.3.5.3.4.3 Sample Results

Figure 2-65 shows some of the reconstructed images using proposed RL method and existing methods. The columns of the figure are explained as follows:

a) Input VLR images (7 x 6),
b) SR results by Bi-cubic interpolation,
c) Hallucination Face method (HF),
d) Eigen-transformation based Face SR method (EF),
e) KPCA-based Face SR method (KF),
f) The proposed method (RL),
g) Original HR images.

The resolution of reconstructed HR images is 56 × 48 pixels. The told images are from the CMU PIE database, while the bottom images are from the FRGC database.

It can be seen that both bi-cubic interpolation (BC) method and KF method give a relatively blur image and high frequency details cannot be recovered. Both HF method and EF method could recover some high frequent details. However, HF method generates some artifacts which degrade the human visual quality. The visual quality of the reconstructed images from EF method is good. However, when comparing with the original HR image, these HR images do not look like the original HR image. It can be seen that the proposed method gives a good visual quality image which also look like the original one.

Figure 2-65 Comparative Qualitative Assessment for the Various SR Methods (Source: [P-20])

Table 2-27 illustrates the comparison results of rank-1ne FR rates, using different SR and FR methods. The experimental results show that:

- There is a significant drop of recognition accuracy (as high as 30%) for VLR image, comparing with the original HR image, for all recognition engines on both CMU-PIE and FRGC databases.
- The proposed method outperforms existing SR methods. It implies that the reconstructed HR image using the proposed method has high discriminability for recognition purpose.
- The recognition accuracy of the proposed method on real surveillance video also have considerable improvement, even though such database is very challenging.

Table 2-27 Comparison Results of Rank-One Recognition Rates Across Different SR and FR Methods (Source: [P-20])

Database	face recognition algorithm	VLR image	original HR image	HF [1]	EF [18]	KF [2]	Proposed DSR
CMU PIE	Eigenface	61.3	86.7	80.8	71.3	71.7	**83.9**
	Kernel PCA	59.5	90.4	86.0	75.9	77.8	**89.5**
	SVM	87.1	93.9	84.7	86.2	89.6	**90.8**
FRGC 2.0	Eigenface	36.0	57.1	40.5	37.9	31.1	**49.0**
	Kernel PCA	34.4	55.8	39.6	37.5	26.4	**47.9**
	SVM	50.4	70.9	49.2	45.0	50.4	**55.5**
SCface	Eigenface	14.9	/	14.3	14.7	17.3	**19.6**
	Kernel PCA	16.3	/	14.3	14.5	17.3	**20.0**
	SVM	15.5	/	13.5	13.1	17.0	**20.2**

2.3.5.3.5 Key Contributions

The very low resolution face recognition problem is defined and discussed in this paper. The key contributions are as follows:

- To solve the problem, a new learning-based face super-resolution framework, a new data constraint and a learning-based discriminative super-resolution algorithm are developed and reported.
- Comprehensive experiments on three publicly available face databases are performed. Experimental results show that the proposed algorithm outperforms existing algorithms at VLR, in terms of both image quality and recognition accuracy.

2.3.5.4 Paper # [P-21]

2.3.5.4.1 Abstract

This paper proposes a novel approach for matching surveillance quality facial images to high resolution images in frontal pose which are often available during enrollment. The proposed approach uses Multidimensional Scaling to simultaneously transform the features from the poor quality probe images and the high quality gallery images in such a manner that the distances between them approximate the distances had the probe images been captured in the same conditions as the gallery images. Thorough evaluation on the Multi-PIE dataset and comparisons with state-of-the-art super-resolution and classifier based approaches are performed to illustrate the usefulness of the proposed approach. Experiments on real surveillance images further signify the applicability of the framework.

2.3.5.4.2 Motivation

Face images captured by surveillance cameras usually have poor resolution in addition to uncontrolled poses and illumination conditions which adversely affect performance of face matching algorithms. Traditionally, research in the area of face recognition has been concentrated on recognizing faces across changes in illumination and pose, but there is a growing interest in handling poor resolution facial images. The difference in resolution in addition to pose and illumination variations adds to the complexity of the task and limited attention has been given to addressing all these variations jointly. Most of the existing work that addresses the problem of matching faces across changes in pose and illumination cannot be applied when the gallery and probe images are of different resolutions. The commonly used approach for matching a LR probe image with a HR gallery is to use super-resolution to construct a higher resolution image from the probe image and then perform matching. But the primary goal of super-resolution approaches is to obtain a good visual reconstruction and they are usually not designed from a recognition perspective. To this end, there have been a few recent efforts that address recognition and super-resolution simultaneously. But most of them assume that the probe and gallery images are in the same pose making them not directly applicable for more general scenarios. This paper builds on the success of these initial efforts and proposes an approach to match LR probe images taken under uncontrolled pose and illumination conditions with HR gallery images in frontal pose.

2.3.5.4.3 Method

The proposed method is based on developing a suitable transformation applied on the selected SIFT features, features used for matching, and make them less sensitive to external significant differences in the external imaging factors, such as pose, illumination and resolution, as compared to the enrolled gallery images, and hence resulting in higher recognition rates.

The paper first explores the limitation of the SIFT-based FR approach.

2.3.5.4.3.1 SIFT-Based Approach

As part of developing the methodology, the paper first investigates the robustness of the SIFT features across significant variations of the illumination, pose and resolution imaging factors. The paper cites recent works based on using SIFT features for matching facial images. Local feature descriptors describe a pixel in an image through its local neighborhood content, and recent studies have shown the effectiveness of local features for the task of face recognition in unconstrained environments with variations in pose and illumination. Additionally, unlike most holistic face representations, these local descriptors allow for direct comparison of images across resolution with suitable scale changes while computing the descriptors. Though these features are known to be robust to changes in pose and scale, they have never been used to match LR face images with a HR gallery with considerable pose and illumination difference.

To analyze the robustness of SIFT features across significant variations of these external imaging factors, a recognition experiment is conducted on the Multi-PIE dataset with HR frontal gallery and LR probe images in different poses. The SIFT descriptors at fiducial locations as the features for performing recognition. Suitable scale changes in SIFT descriptor computation are made to make the comparison across resolution feasible. Figure 2-66 illustrates the fiducial locations used for extracting SIFT descriptors for representing the input faces (top) and the recognition accuracy on the Multi-PIE dataset for LR probe images with different poses as shown in the top row, when compared with HR frontal gallery (bottom). In this experiment, the LR probe images are of size 20x18 while the HR gallery images are of size 60x55 (scale factor of three). The solid red line represents the recognition accuracy when the probe images are of the same resolution and pose as the gallery images.

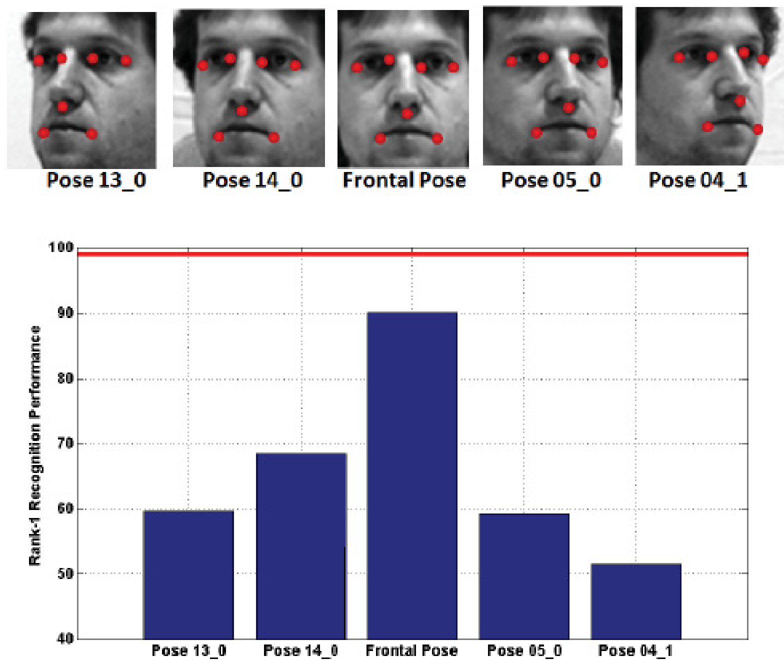


Figure 2-66 Performance of the SIFT-Based Method When Probe and Gallery Images Have Different Pose Resolution (Source: [P-21])

The recognition accuracy for probe images with decreasing resolutions is shown in Figure 2-67. For this experiment, the pose of the probe images is fixed at *05_0* (as labeled in Multi-PIE data). From these results, it was concluded that even though SIFT descriptors are fairly robust to modest variations in pose and resolution, large variations in these factors between the probe and the gallery images result in significant decrease in recognition performance.

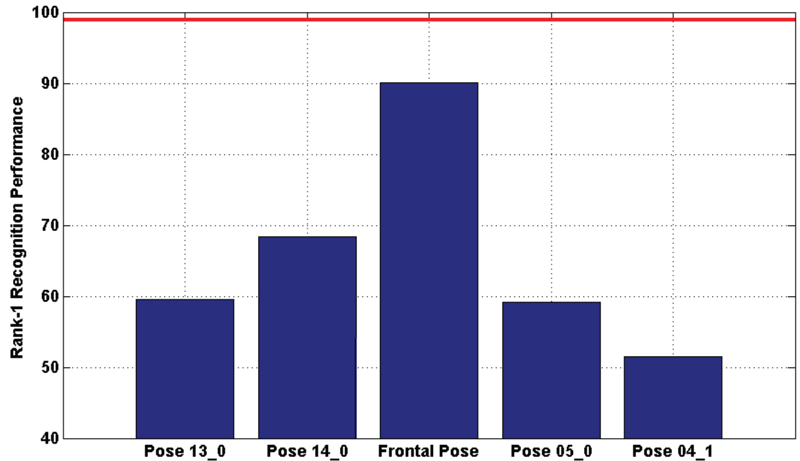


Figure 2-67 Recognition Accuracy for Non-frontal Probe Images with Decreasing Resolutions when Compared Against Frontal HR Gallery Image (Source: [P-21])

2.3.5.4.3.2 Proposed Method

To overcome the above mentioned limitations of the SIFT-based approaches, the paper proposed a transformation applied on the SIFT features in order to improve the recognition performance of matching probe images with significant differences in the external imaging factors, such as pose, illumination and resolution, as compared to the enrolled gallery images. In particular, the paper propose a MDS-based approach to transform the features from LR non-frontal probe images and the HR frontal gallery images to a common space in such a manner that the distances between them approximate the distances had the probe images been of the same resolution and pose as the gallery images, as illustrated in Figure 2-68. The method uses SIFT-based descriptors at fiducial locations on the face image as the input feature. The desired transformation is learned from the training images using iterative majorization algorithm. To perform facial recognition or verification, the SIFT descriptors of the gallery and probe facial images are transformed using the learned transformation, followed by computation of Euclidean distances between the transformed features.

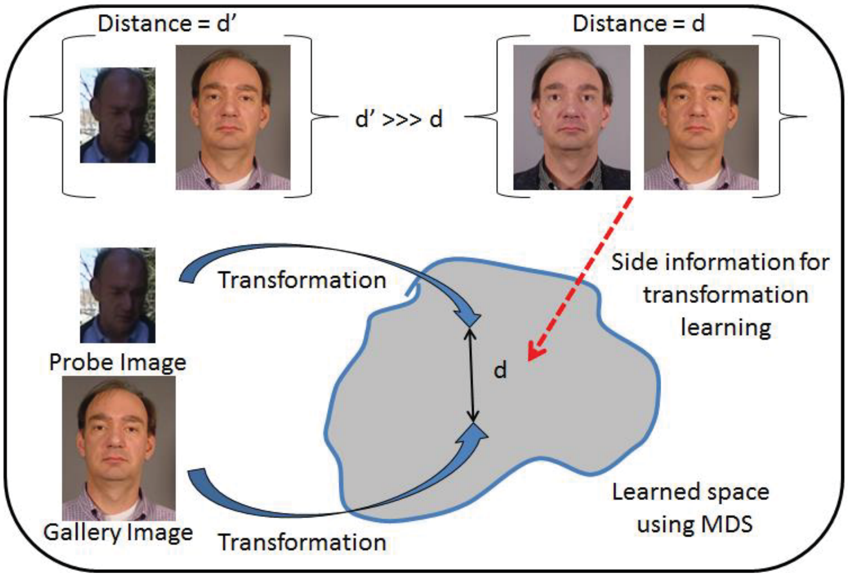


Figure 2-68 Overview of the Proposed Approach (Source: [P-21])

2.3.5.4.4 Experimental Results

Extensive experimental evaluation on the Multi-PIE dataset is performed to evaluate the proposed approach. Comparisons with state-of-the-art super-resolution and classifier-based approaches are performed to illustrate the usefulness of the proposed approach. Experiments on real surveillance images further signify the applicability of the framework for matching LR images in uncontrolled pose and illumination conditions.

2.3.5.4.4.1 Database

Extensive experimental evaluations using the Multi-PIE database are presented in this paper.

2.3.5.4.4.2 FR Algorithm(s)

The proposed algorithm is a standalone FR feature-algorithm. Although, feature based algorithms, such as PCA, can be applied on the generated features before for matching.

2.3.5.4.4.3 Sample Results

Extensive experimental evaluation on the Multi-PIE dataset is performed to evaluate the proposed approach. Comparisons with state-of-the-art super-resolution and classifier-based approaches are performed to illustrate the usefulness of the proposed approach. Experiments on real surveillance images further signify the applicability of the framework for matching LR images in uncontrolled pose and illumination conditions. Some of these results are illustrated next.

A. Recognition Across Resolution, Pose and Illumination

First a recognition experiment with HR frontal images as gallery and LR images in different poses and illuminations as the probe set is performed. Table 2-28 shows the rank-1 recognition performance averaged over the probe images under all illumination conditions. The recognition performance using SIFT+PCA directly without the proposed learning is also given as baseline for comparison The proposed approach significantly improves the recognition accuracy as compared to directly using the SIFT+PCA features.

Table 2-28 Rank-1 Recognition Percentages for Different Gallery Illumination (Source: [P-21])

Gallery Illum. Probe Pose	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
13_0 (SIFT)	61	54	58	60	60	60	57	53	54	57	59	58	58	56	67	68	64	64	63	61
13_0 (Ours)	**76**	**67**	**76**	**81**	**81**	**76**	**75**	**70**	**72**	**76**	**78**	**76**	**76**	**72**	**85**	**81**	**76**	**77**	**79**	**76**
14_0 (SIFT)	71	64	68	68	70	71	68	62	63	68	68	66	64	58	76	76	73	72	71	71
14_0 (Ours)	**83**	**73**	**82**	**88**	**88**	**87**	**85**	**82**	**82**	**87**	**89**	**84**	**81**	**74**	**91**	**91**	**87**	**89**	**89**	**83**
05_0 (SIFT)	63	48	51	55	52	53	55	55	58	65	62	62	61	55	60	60	62	73	71	62
05_0 (Ours)	**81**	**75**	**81**	**84**	**82**	**80**	**81**	**79**	**80**	**83**	**82**	**83**	**79**	**70**	**85**	**84**	**83**	**85**	**86**	**80**
04_1 (SIFT)	51	51	51	50	47	46	46	45	51	57	53	54	52	46	51	52	54	61	61	50
04_1 (Ours)	**72**	**67**	**75**	**77**	**71**	**71**	**69**	**68**	**71**	**75**	**78**	**76**	**70**	**62**	**77**	**73**	**71**	**79**	**80**	**71**

B. Comparison with Super-Resolution Approach

For matching LR images, the most commonly used approach is to first obtain a HR image using SR techniques which are then used for matching. The performance of the proposed approach is compared with a referenced state-of-the art SR technique, where the different patches of a HR image are assumed to have a sparse representation with respect to an over-complete dictionary of prototype signal atoms.

Figure 2-69 shows the recognition performance obtained. Though super-resolution improves upon the baseline performance, it can be seen that the proposed approach performs considerably better than the SR approach. The performance improvement is even more when the features computed from the SR images are used, implying that SR algorithms can be used along with the proposed approach for further performance improvement.

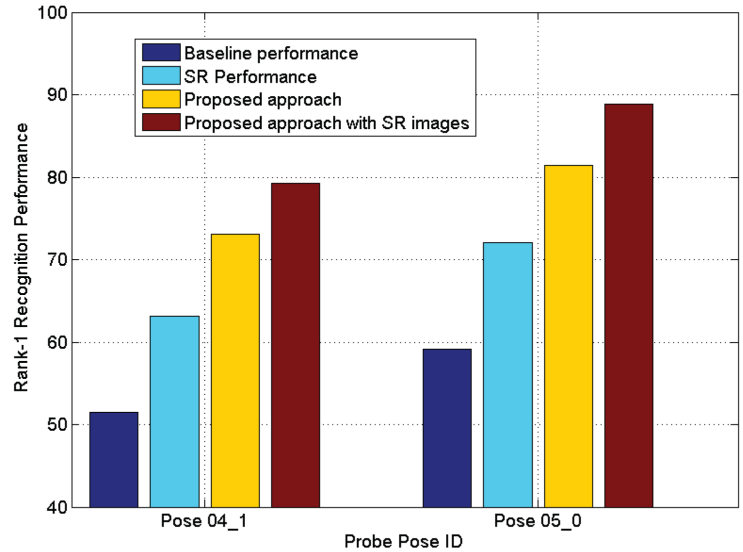


Figure 2-69 Comparison with SR Approach for Two Different Probe Poses (Source: [P-21])

C. Comparison with Classifier-Based Approach

Recently, metric learning approaches like Large Margin Nearest Neighbor (LMNN), cited in the paper, on local features like SIFT have been used successfully for recognizing faces in unconstrained environments. LMNN is used to learn a Mahalanobis distance metric for k-Nearest Neighbor (kNN) classification by semi-definite programming.

Figure 2-70 compares the performance of the proposed approach with LMNN for HR frontal gallery and LR probes in two different poses. We see that for large variations of scale and pose between the gallery and probe images, the proposed approach performs considerably better than this state-of-the art learning approach.

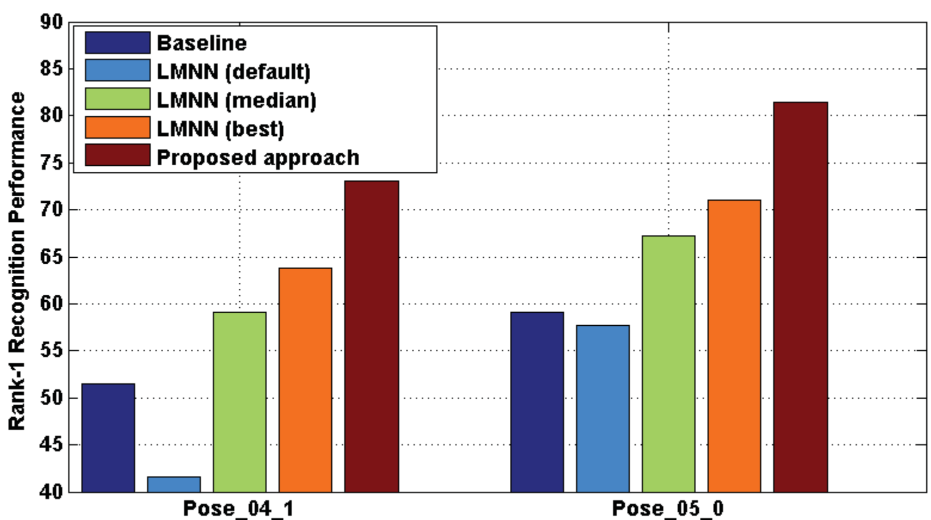


Figure 2-70 Comparison with LMNN for HR Gallery and LR probe Images for two Different Poses (Source: [P-21])

D. Comparison with Classifier-Based Approach

In all the experiments so far, the scale factor between the HR gallery and LR probe images is fixed at 3. Here, the performance of the proposed approach is analyzed for varying resolutions of the probe images. Figure 2-71 shows example gallery and probe images. The bottom row of the figure shows the recognition performance using the SIFT+PCA features and the proposed approach. The proposed approach is successful in significantly improving the recognition performance for a wide range of resolutions.

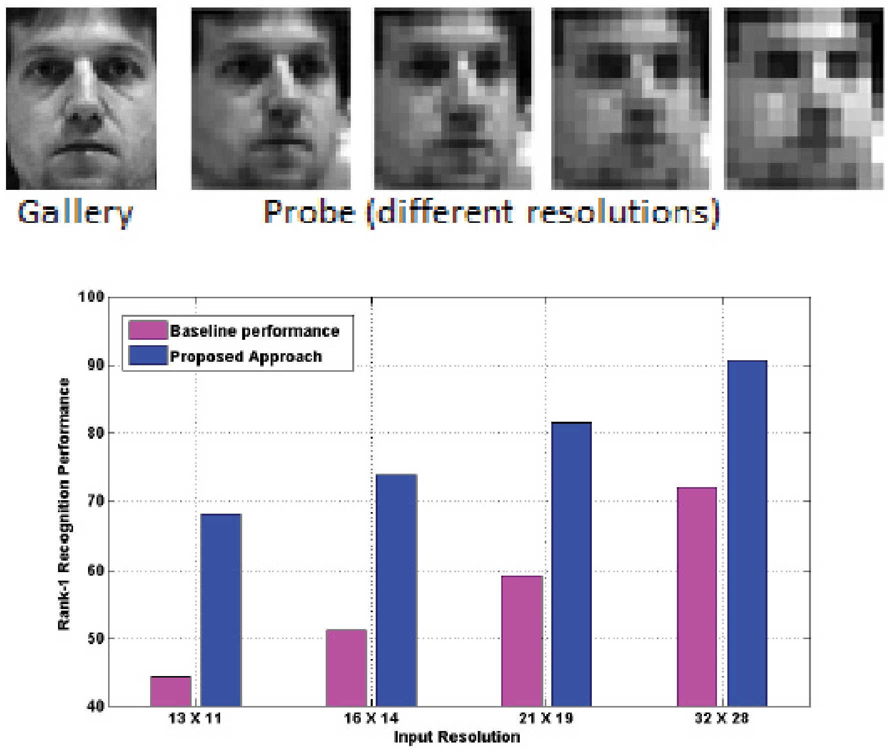


Figure 2-71 (Top) Gallery and Probe Images at Different Resolutions. (Bottom) Recognition Performance of the Baseline and the Proposed Approach for Different Probe Resolutions (Source: [P-21])

E. Evaluation on Surveillance Quality Images

The usefulness of the proposed approach is also testing on surveillance quality LR non-frontal face images. Here the probe images are obtained from the Multiple Biometric Grand Challenge (MBGC) video challenge data, cited in the paper.

Figure 2-72 shows sample gallery and probe images of the same subjects. Images of 25 subjects are used for training while the rest are used for testing. There is no identity overlap between the training and test images.

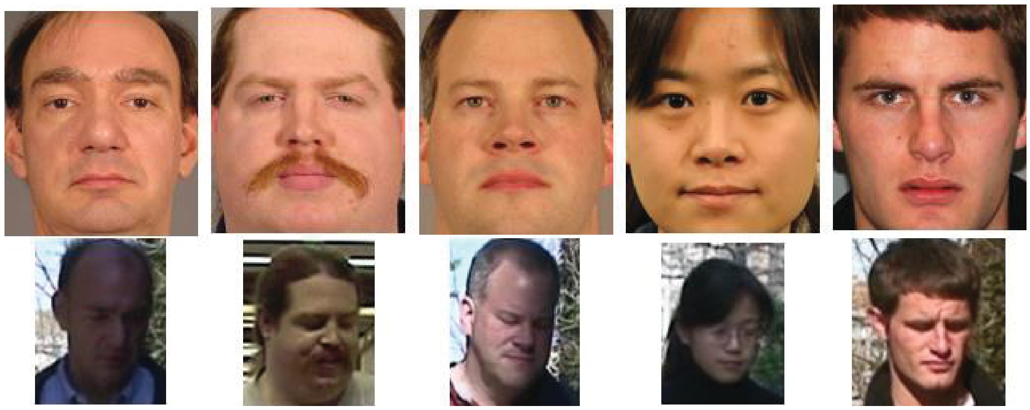


Figure 2-72 Top: Example Gallery Images; Bottom: Example Probe Images of the Same Subjects as the Gallery Images (Source: [P-21])

The experiment is repeated ten times with different random sampling of these subsets. The Receiver Operating Characteristics (ROC) obtained using SIFT+PCA features directly and using the proposed approach on these features is shown in Figure 2-73. Error bars indicate the range of the ROCs obtained in different runs of the experiment. We see that the proposed approach is able to significantly improve the results as compared to the baseline on this challenging dataset.

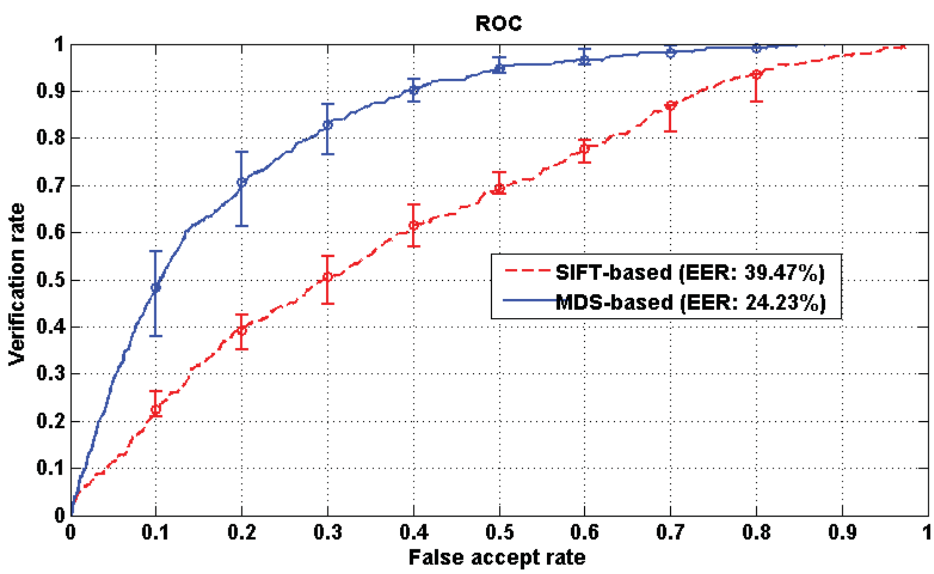


Figure 2-73 ROC using the SIFT+PCA Features Directly and using the Proposed Approach (Source: [P-21])

Figure 2-74 shows top matches obtained for three probe images. For each probe image, the top row shows the five most similar images in the gallery using the baseline SIFT+PCA algorithm and the bottom row shows the results using the proposed approach. The images surrounded by red boxes denote correct match, i.e. these gallery images are of the same subject as the corresponding probe images. The retrieval results using the proposed approach are much better than the results using the SIFT+PCA features directly.

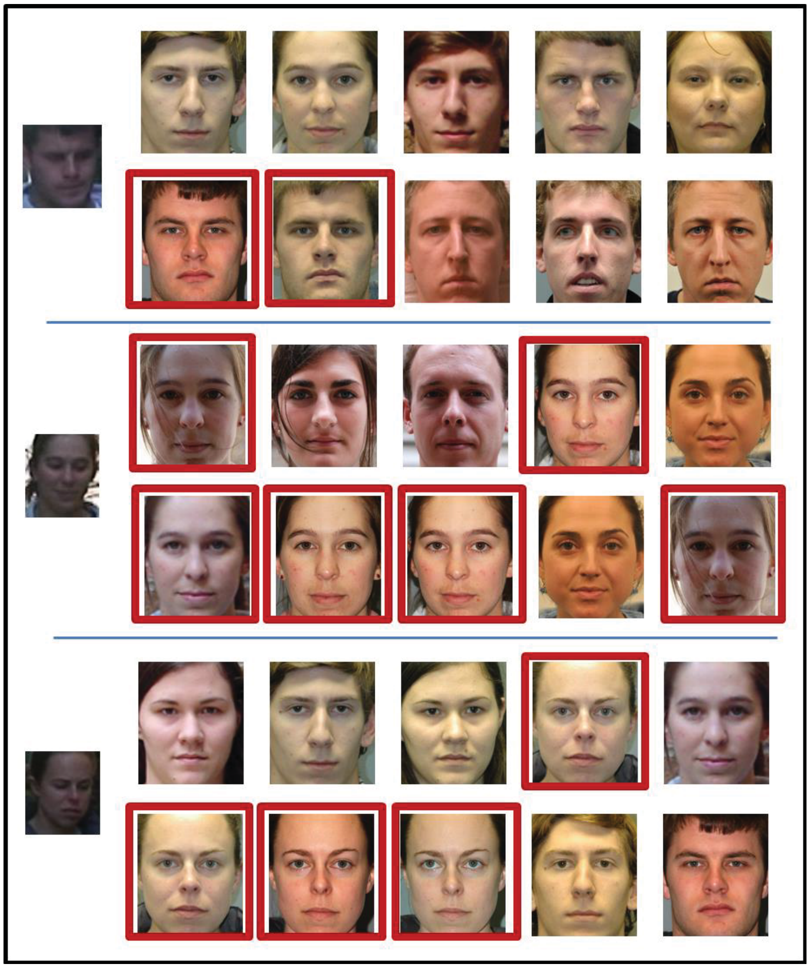


Figure 2-74 Retrieval Results for Three Probe Images (first column) (Source: [P-21])

2.3.5.4.5 Key Contributions

This proposed a novel MDS-based approach for matching LR facial images captured from surveillance cameras with considerable variations in pose and illumination to HR gallery images in frontal pose. The key features of the method are as follows:

- First, the features are simultaneously transformed from the probe and the gallery images such that the distances between them approximate the distances had the probe image been taken in the same conditions as the gallery.
- Extensive evaluation on the Multi-PIE data and surveillance quality images show the usefulness of the proposed approach.

Note that, as we discuss in Section 3, the above reviewed challenges and limitations to any real-world face recognition systems are especially applicable to the FR sub-system of the ATC system.

2.4 End-to-End Real-Time Face Recognition Systems

In this section, we review FR literature papers related to developing end-to-end, real-time FR systems applicable in real-world uncontrolled environments, using still and moving camera video imagery. We begin by reviewing such systems for fixed-camera video imagery.

2.4.1 FRS from Fixed-Camera Video

As illustrated in Table 2-29, we review literature papers proposing end-to-end FR systems for fixed-camera systems.

Table 2-29 Selected Papers on End-to-End FR Systems from Fixed-Camera Videos

#	Paper Title	Authors	Source	Year	Type
P-22	Real-Time Face Detection and Recognition for Video Surveillance Applications	Zhen Lei, Chao Wang, Qinghai Wang, Yanyan Huang	2009 WRI World Congress on Computer Science and Information Engineering, Vol. 5, pp. 168 – 172.	2009	Proposed End-to-End FR System
P-23	Facial Recognition in Uncontrolled Conditions for Information Security	Qinghan Xiao, Xue-Dong Yang	EURASIP Journal on Advances in Signal Processing 2010, Volume 2010: Article ID 345743, 9 pages.	2010	Proposed End-to-End FR System)
P-24	Robust Face Recognition for Uncontrolled Pose and Illumination Changes	Maria De Marsico, *Member, IEEE*, Michele Nappi, Daniel Riccio, and Harry Wechsler, *Fellow, IEEE*	IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, VOL. 43, NO. 1, pp. 149 – 163, JANUARY 2013	2013	Proposed End-to-End FR System

2.4.1.1 Paper # [P-22]

2.4.1.1.1 Abstract

This paper proposes a real-time human face detection and recognition system from video sequences, which is applicable in real-world surveillance applications. Such a system is a challenging task due to the variances in background, facial expression and illumination. The face detection approach is based on modest AdaBoost algorithm and can achieve fast, accurate face detection that is robust to changes in illumination and background. The detection stage provides good results maintaining a low computational cost. The recognition stage is based on an improved independent components Analysis approach which has been modified to cope with the video surveillance application. In the recognition stage, the Hausdorff distance is used as a similarity measure between a general face model and possible instances of the object within the image. After the integration of the two stages, several improvements are proposed which increase the face detection and recognition rate and the overall performance of the system. The experimental results demonstrate the significant performance improvement using the proposed approach over others. It can be seen that the proposed method is very efficient and has significant value in application.

2.4.1.1.2 Motivation

Face detection and recognition have been regarded as a challenging problem in the field of computer vision, due to the large intra-class variations caused by the changes in facial appearance, lighting, and expression in real-world applications. Face detection and recognition have been subject of much interest in the last years. They have many applications in a variety of fields such as identification for law enforcement, personal identification and security system access. The objective of this work is the integration and optimization of an automatic face detection and recognition system for video surveillance applications.

2.4.1.1.3 Method

The paper proposes two novel methods for improving the face detection and recognition stages of the full FR system, as follows.

2.4.1.1.3.1 Face Detection Stage

A novel classifier, which is built using modest AdaBoost algorithm to select a small number of extended Haar-like features from a very large set of potential features, and the criteria for choosing the reduced set of features are discussed. Faces are divided into several categories according to their poses, and for each of these categories a form of weak classifiers in look-up-table type is designed. This face detection approach can achieve fast, accurate face detection that is robust to changes in illumination and background.

2.4.1.1.3.2 Face Recognition Stage

At the face recognition stage, another novel method for fast face recognition using an improved ICA and Hausdorff distance. The Hausdorff distance is used as a similarity measure between a general face model and possible instances of the object within the image. Compared with conventional approaches, the method for face recognition is shown to be more effective and capable of handling more complicated variations.

2.4.1.1.4 Experimental Results

An experimental evaluation of the proposed FR system is performed. The performance of the proposed real-time FR system is compared to two other comparable systems, referred to as the: Rowley and Viola methods. The experimental results demonstrate the significant performance improvement using the proposed approach over the two benchmark methods.

2.4.1.1.4.1 Database

The experimental results and performance evaluation of the proposed method were conducted on the CMU+MIT database.

2.4.1.1.4.2 FR Algorithm(s)

The proposed algorithm is a standalone real-time FR system

2.4.1.1.4.3 Sample Results

Figure 2-75 shows the output of the proposed face detector on some test images from the CMU+MIT frontal, profile and rotated test set.

Figure 2-75 Sample Output of the Proposed Face Detection (Source: [P-22])

Table 2-30 lists the detection rate for various numbers of false detections for the proposed method as well as other published detectors. Obviously, the proposed method achieved a better detection rate than other two methods on the CMU+MIT test set.

Table 2-30 Detection Results Comparison (Source: [P-22])

Method \ False	10	65	78	167
Our method	91.7%	93.6%	94.5%	95.2%
Rowley	83.2%	/	/	90.1%
Viola	76.1%	92.0%	92.1%	93.9%

2.4.1.1.5 Key Contributions

This paper proposed an extended Haar-like feature based system for face detection using modest AdaBoost algorithm. The key contributions of this paper are as follows:

- The Hausdorff distance to improve the ICA algorithm performance for face recognition.
- The system is based on modest AdaBoost and an improved ICA proposed in previous work.
- The system is intended for video surveillance applications and is able to tell whether or not a specific person is present in a surveillance area.
- Both, the face detection and the face recognition method for video surveillance applications can be applied to other object detection and recognition tasks in computer vision.

2.4.1.2 Paper # [P-23]

2.4.1.2.1 Abstract

With the increasing use of computers nowadays, information security is becoming an important issue for private companies and government organizations. Various security technologies have been developed, such as authentication, authorization, and auditing. However, once a user logs on, it is assumed that the system would be controlled by the same person. To address this flaw, the paper proposes a demonstration system that uses facial recognition technology to periodically verify the identity of the user. If the authenticated user's face disappears, the system automatically performs a log-off or screen-lock operation. This paper presents an improvement on the previous system and further efforts in developing image preprocessing algorithms and dealing with angled facial images. The objective is to improve the accuracy of facial recognition under uncontrolled conditions. To compare the results with others, the frontal pose subset of the FERET database was used for the test. The experiments showed that the proposed algorithms provided promising results.

2.4.1.2.2 Motivation

As networks become larger, more complex, and more distributed, information security has become more critical than it has ever been in the past. Many efforts have been made aiming to accurately authenticate and authorize trusted individuals, and audit their activities. Once a user is successfully logged in, the prevailing technique assumes that the system is controlled by the same person. Focusing on this security challenge, an enhanced authentication method that uses video-based facial recognition technology to monitor the user during the entire session in which the person is using the system is developed.

2.4.1.2.3 Method

Figure 2-76 illustrates the system diagram, and the module functionalities are briefly summarized, as follows:

1. Input: The system captures video images and processes either 24-bit or 32-bit color images. The captured images are displayed on the computer screen viaDirectX in real time.
2. Face Detection: A face detection algorithm subsequently examines the captured images. The locations of potential human faces are recorded. Later, when a video image is finally rendered to the monitor, a red rectangle encloses each potential face.
3. Face Segmentation: The detected face is then segmented from the frame image.
4. Face Matching: An alignment operation is first applied on the segmented faces. Each face image is first converted to a feature vector. This vector is projected onto Eigenfaces through inner product calculations, using the PCA algorithm. Each face produces a weight vector. The Euclidean distance between two weight vectors is used to measure the similarity between the two faces. This distance is then mapped to a normalized matching score.
5. Relighting: This module provides a histogram-based intensity mapping function to normalize the intensity distribution of the segmented face image
6. Facial Database: It is assumed that data from up to eight users may be saved in the database. Each user is required to take at least one picture in the user's normal working environment within the normal sitting space and under the normal lighting conditions.
7. Output: The demonstration system has three main outputs.
    - Live video: the detected face in the scene is surrounded by a red rectangle.
    - Matching results: the segmented face image from the current test scene is displayed, along with up to five candidate faces from the database in descending priority order.
    - Performance Data: several performance data are displayed in real time, such as the overall frame rate, the face detection time, the face recognition time, and the best matching score.

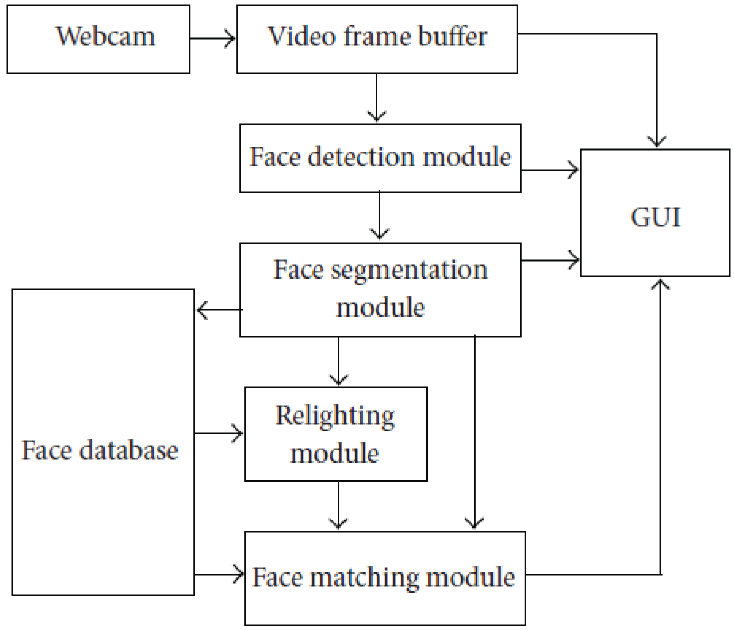


Figure 2-76 Overall System Architecture (Source: [P-23])

A study on image preprocessing algorithms has been carried out to improve the accuracy performance. It focused on the areas that affect the accuracy of facial recognition, which include geometric correction, face alignment, masking and photometric normalization. Also, because there were more side-view face images than front view images in the captured video stream, a study has been conducted to explore the possibility of using multi-angle face images to increase the recognition rate.

2.4.1.2.4 Experimental Results

The paper presents experimental results to assess the performance of the developed system.

2.4.1.2.4.1 Database

Three sets of experiments were conducted on the Comparison Is Made (CIM) and FERET face databases.

2.4.1.2.4.2 FR Algorithm(s)

The proposed algorithm is a standalone real-time FR system

2.4.1.2.4.3 Sample Results

As illustrated in Figure 2-77, the FR system accuracy is evaluated by comparing the FAR to the False Recognition Rate (FRR) for different choices of the threshold *T*, resulting in ROC curve. As a common rule, the threshold value is adjusted based on the classification confidence values to evaluate the trade-off between FAR and FRR. The mean classification rate obtained on the FERET database was 92.5% (an EER of 7.5%), demonstrating that the presented algorithm is scalable to relatively large databases. The paper reports that, this result is only 1.5% lower than the best published classification rate (94%) in the literature for the same database.

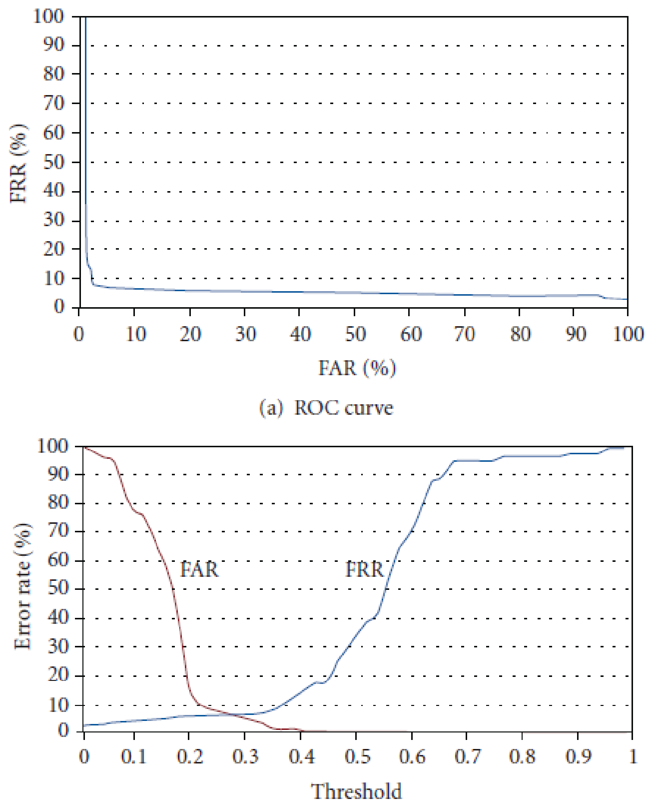


Figure 2-77 ROC of the Proposed System (Source: [P-23])

2.4.1.2.5 Key Contributions

This paper develops an enhanced authentication method that uses video-based facial recognition technology to monitor the user during the entire session in which the person is using the system. The key contributions of the proposed system are:

- It can automatically lock the screen or log out the user when the authenticated user's face disappears from the vicinity of the computer system for an adjustable time interval.
- New image preprocessing algorithms for dealing with angled facial images are proposed, resulting in improved facial recognition under uncontrolled conditions.

2.4.1.3 Paper # [P-24]

2.4.1.3.1 Abstract

This paper proposes a novel framework for real-world face recognition in uncontrolled settings named Face Analysis for Commercial Entities (FACE). Its robustness comes from normalization ("correction") strategies to address pose and illumination variations. In addition, two separate image quality indices quantitatively assess pose and illumination changes for each biometric query, before submitting it to the classifier. Samples with poor quality are possibly discarded or undergo a manual classification or, when possible, trigger a new capture. After such filter, template similarity for matching purposes is measured using a localized version of the image correlation index. Finally, FACE adopts reliability indices, which estimate the "acceptability" of the final identification decision made by the classifier. Experimental results show that the accuracy of FACE (in terms of recognition rate) compares favorably, and in some cases by significant margins, against popular face recognition methods. In particular, FACE is compared against SVM, incremental SVM, principal component analysis, incremental LDA, ICA, and hierarchical multi-scale local binary pattern, using different facial databases. The results obtained by FACE witness a significant increase in accuracy when compared with the results produced by the other algorithms considered.

2.4.1.3.2 Motivation

Real-world face recognition in unconstrained scenarios is still a major challenge for biometrics. The reasons are manifold. Among them is the fact that gallery (stored) enrolled face images are usually captured in controlled settings, using a predefined arrangement of subjects and capture devices, whereas the probes (test images) are captured in quite different settings. The test images are usually acquired under looser restrictions, which significantly increase intra-class variability, so that pose and illumination, as well as expression and occlusions, may become disturbing factors. The goal for the developed Face Analysis for Commercial Entities (FACE), is to develop a novel framework for automatic authentication of biometric face images and address the existing challenges and to advance biometric identity management for real-world commercial applications.

2.4.1.3.3 Method

Figure 2-78 illustrates the architecture of the developed FACE system. The robustness of the FACE system comes from normalization ("correction") strategies to address pose and illumination variations. In addition, two separate image quality indices quantitatively assess pose and illumination changes for each biometric query, before submitting it to the classifier. Samples with poor quality are possibly discarded or undergo a manual classification or, when possible, trigger a new capture. After such filter, template similarity for matching purposes is measured using a localized version of the image correlation index. Finally, FACE adopts reliability indices, which estimate the "acceptability" of the final identification decision made by the classifier.

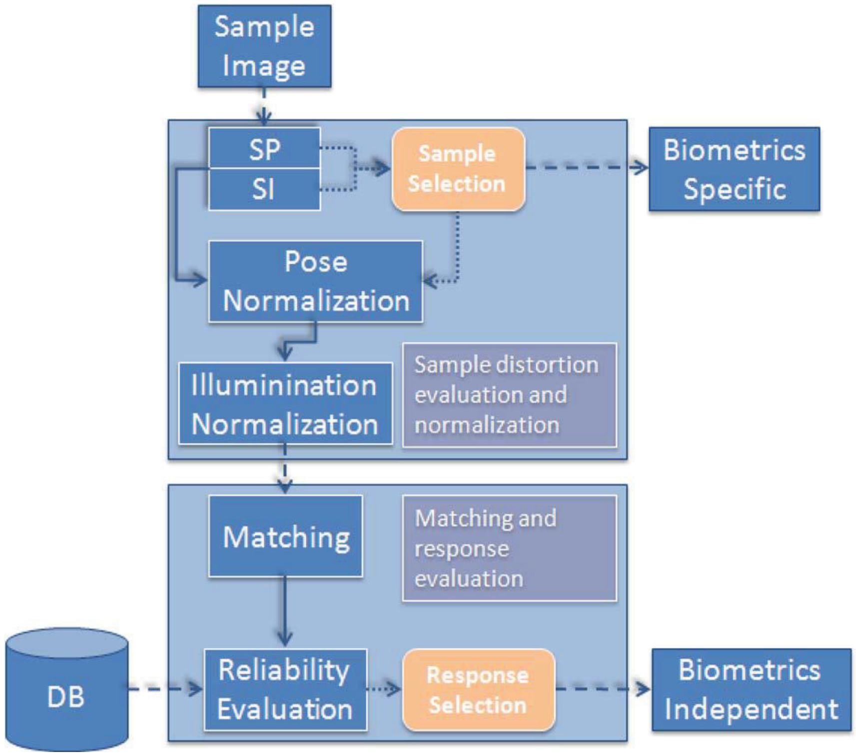


Figure 2-78 Proposed FACE Architecture (Source: [P-24])

2.4.1.3.4 Experimental Results

To assess and evaluate its performance, the developed FACE FR system is compared against SVM, incremental SVM, principal component analysis, incremental LDA, ICA, and hierarchical multi-scale local binary pattern.

2.4.1.3.4.1 Database

Testing exploits data from different data sets: CelebrityDB, Labeled Faces in the Wild, SCface, and FERET.

2.4.1.3.4.2 FR Algorithm(s)

The proposed algorithm is a standalone real-time FR system

2.4.1.3.4.3 Sample Results

Table 2-31 illustrates a comparison of the performances of the local correlation matching component of FACE, with the other techniques, using various data sets.

Table 2-31 Performance Comparison using Local Correlation (Source: [P-24])

DB	Method						
	FACE	SVM	ISVM	PCA	ILDA	ICA	HMLBP
FERET *fa*	93% (only LCM)	75%	78%	74%	79%	73%	84%
CDB	82% (only LCM)	44%	43%	27%	38%	31%	66%
CDB (3 img)	60% (only LCM)	38%	42%	28%	45%	30%	51%
LFW	54% (only LCM)	44%	42%	35%	46%	35%	47%
SCface	81% (only LCM)	70%	69%	50%	62%	55%	73%

Table 2-32 illustrates a comparison of the performances of FACE including the normalization component with other techniques, using various data sets.

Table 2-32 Performance Comparison using Normalization Component (Source: [P-24])

DB	Method						
	FACE	SVM	ISVM	PCA	ILDA	ICA	HMLBP
FERET *fa*	93% (only LCM)	75%	78%	74%	79%	73%	84%
CDB	82% (only LCM)	44%	43%	27%	38%	31%	66%
CDB (3 img)	60% (only LCM)	38%	42%	28%	45%	30%	51%
LFW	54% (only LCM)	44%	42%	35%	46%	35%	47%
SCface	81% (only LCM)	70%	69%	50%	62%	55%	73%

Figure 2-79 shows the FACE accuracy using Cumulative Match Score (CMS). core and its dependence on STASM results, using images from subsets L1 and R1 of SCface are normalized according to the points returned by Extended Active Shape Model (STASM) algorithm (circles) and when the coordinates of the centers of the eyes, nose, and mouth are substituted with the manually located ones (squares). FACE, uses the STASM algorithm to locate the face and its characteristic points (“landmarks”).

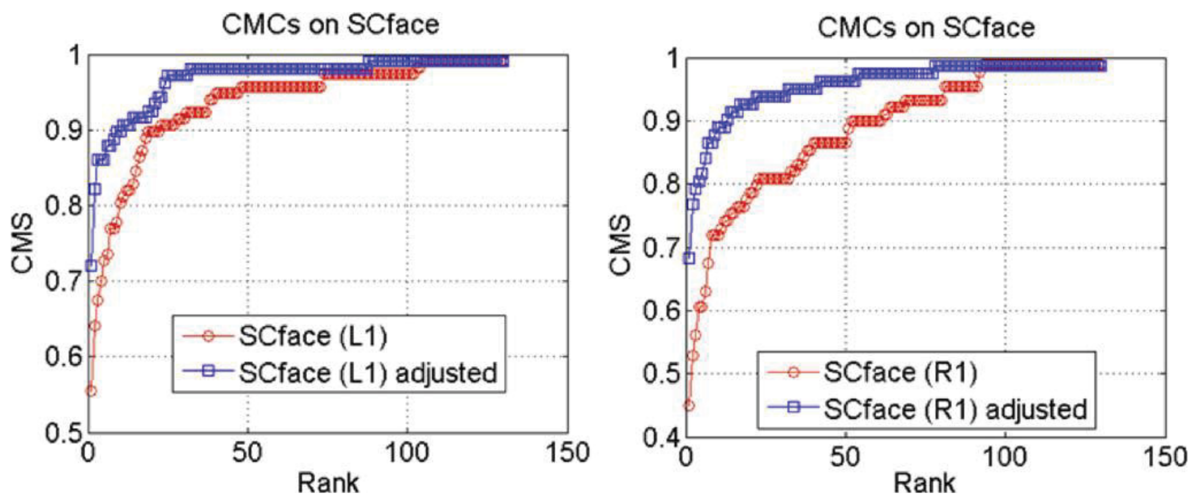


Figure 2-79 CMC of FACE System (Source: [P-24])

These experimental results demonstrate a significant increase in FR recognition accuracy when compared with the results produced by the other algorithms considered.

2.4.1.3.5 Key Contributions

This paper has described FACE, a new framework for face analysis including classification. The key contributions of the developed system are as follows:

- FACE improves accuracy performance compared to state-of-the-art methods, for uncontrolled settings when the image acquisition conditions are not optimal. This is typical of applications such as photo tagging over social networks like Facebook or cataloguing of celebrities' images in a magazine editorial office.
- FACE has access to multiple gallery instances for each subject and does not require expensive training to learn the face space, using instead straightforward correlation of local regions after proper pose and illumination normalization.
- FACE also has access to pose and illumination image quality indices, respectively, which can be used to *a priori* discard images whose quality is not sufficient to guarantee an accurate recognition response.
- Experimental results show that FACE outperforms competing methods, with a significant increment in accuracy versus the next ranked methods. The improvement depends on the complexity of the data set at hand but is always worth of consideration.

2.4.2 FRS from Moving-Camera Video

As illustrated in Table 2-19, we review one paper proposing end-to-end FR system, using moving camera videos, and assessing its performance.

Table 2-33 Selected Paper on End-to-End FR Systems from Moving-Camera Videos

#	Paper Title	Authors	Source	Year	Type
P-25	Video-based Face Recognition and Tracking from a Robot Companion	T.Germa, F.Lerasle, T.Simon	International Journal of Pattern Recognition and Artificial Intelligence Vol. 23, No. 3 (2009) 591–616 © World Scientific Publishing Company	2009	Proposed End-to-End FR System

2.4.2.1 Paper # [P-26]

2.4.2.1.1 Abstract

This paper deals with video-based face recognition and tracking from a camera mounted on a mobile robot companion. All persons must be logically identified before being authorized to interact with the robot while continuous tracking is compulsory in order to estimate the person's approximate position. A first contribution relates to experiments of still-image-based face recognition methods in order to check which image projection and classifier associations give the highest performance of the face database acquired from the robot. The proposed approach, based on PCA and SVMs improved by genetic algorithm optimization of the free parameters, is found to outperform conventional appearance-based holistic classifiers (Eigenface and Fisherface) which are used as benchmarks. Relative performances are analyzed by means of Receiver Operator Characteristics which systematically provide optimized classifier free-parameter settings.

2.4.2.1.2 Motivation

The development of autonomous robots acting as human companions is a motivating challenge and a considerable number of mature robotic systems have been implemented which claim to be companions, servants or assistants in private homes. This is of particular interest for elderly and disabled people given that Europe is to experience significant ageing over the next two decades. Automatic visual person recognition is therefore crucial to this process. Person recognition based on video is preferable to using still images as motion helps in recognition. This paper explores fusing multiple visual cues, face and clothing appearance, within the well-known particle filtering formalism.

2.4.2.1.3 Method

The main focus in this article is on the design of visual functions in order to recognize individuals, verify their presence and track them in the robot's vicinity. These functionalities include robust and efficient basic navigation and object recognition abilities. Figure 2-80 illustrates the robot companion called Jido.

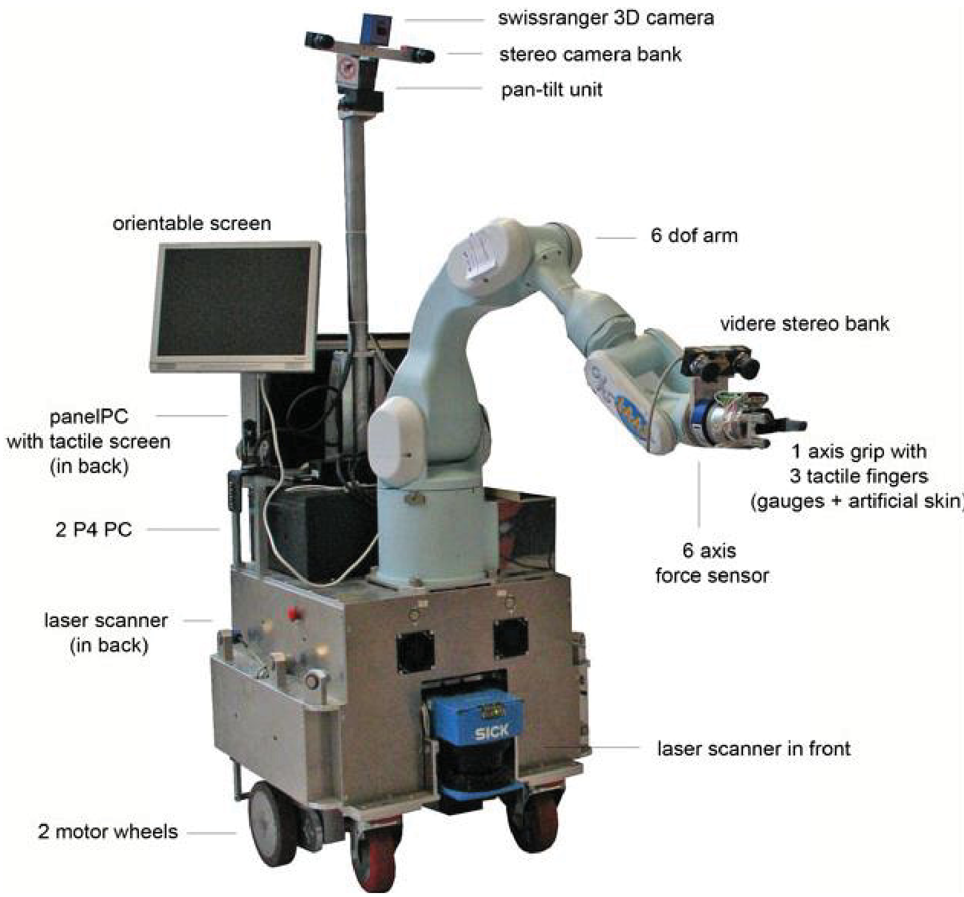


Figure 2-80 The Jido Robot Companion (Source: [P-26])

Figure 2-81 illustrates a typical scenario involving peer-to-peer and human/robot interaction. The left and right columns show the current human/robot situation as well as the video stream from the on-board camera, respectively. In this scenario, the challenge is to recognize a given person in the video stream despite temporary occlusions by other persons, 3D rotations and out-field sight of the targeted person.

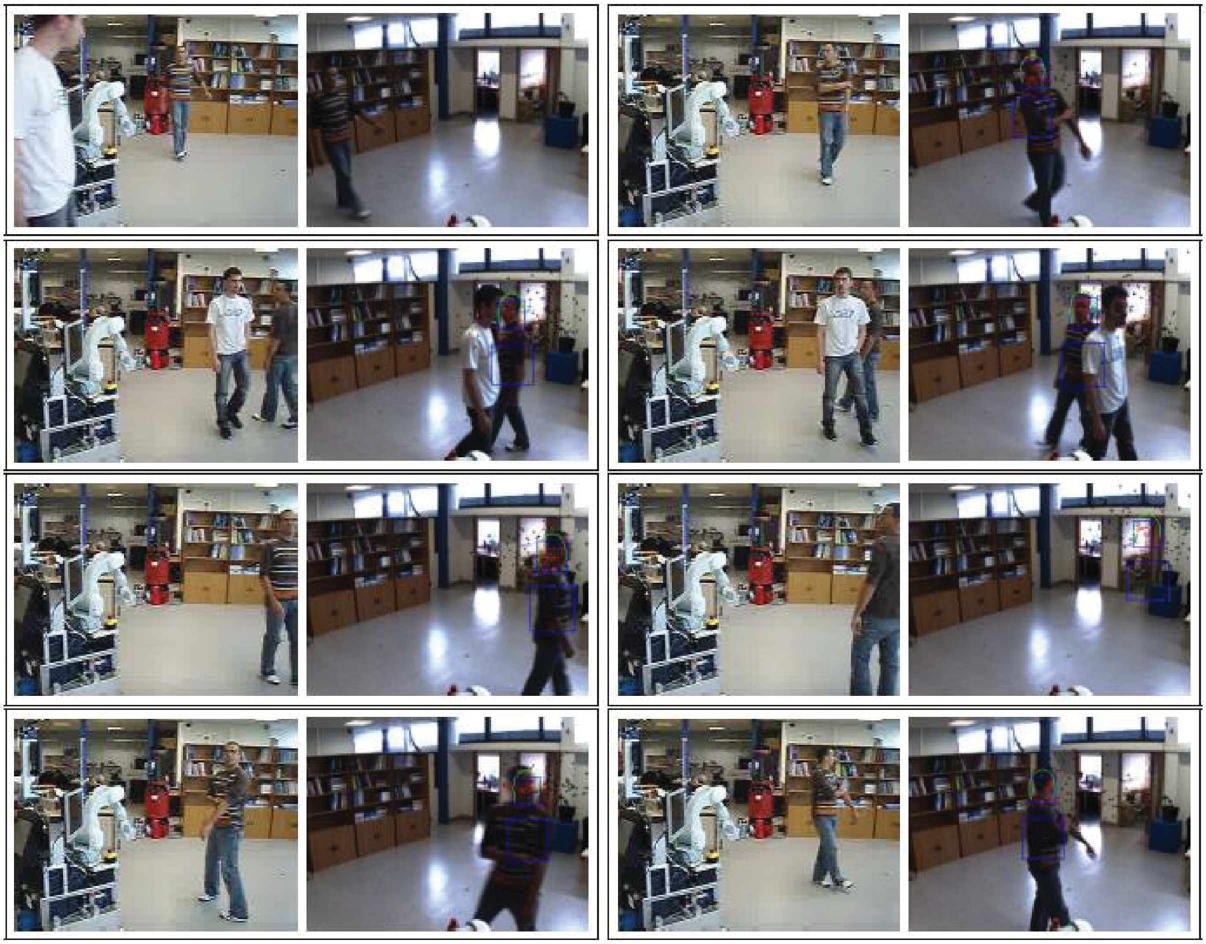


Figure 2-81 From Top-Left to Bottom-Right: Progress of a Peer-to-Peer H/R Interaction session. The Rectangle Represents the Template for the Targeted Person (Source: [P-26])

Given this framework, the proposed FR system must be capable of handling:

- Poor video quality and low image resolution which is computationally faster
- Heavier lighting changes
- Larger pose variations in the face images i.e. 2D (image plane) but also 3D rotations
- Occlusion or background clutter.

These requirements have led to interest in the design of a system that can fuse other cues in addition to face appearance and recognize these faces from video sequences instead of still images. This requires solving tracking (estimation of the targeted person image location) with automatic re-initialization capabilities, apart from the recognition task. However, the robot does not deal solely with still images. By considering subsequent frames and, as a result, spatiotemporal relationships, it is possible to make the FR problem more tractable.

2.4.2.1.4 Experimental Results

The above tracker has been prototyped on a 1. 8GHz Pentium Dual Core using Linux and the OpenCV library. Both quantitative and qualitative off-line evaluations on sequences are reported below. This database of two different sequences (800 images) acquired from Jido mobile robot in a wide range of realistic conditions allows us to:

1. determine the optimal parameter values of the tracker,
2. identify its strengths and weaknesses, and in particular characterize its robustness to environmental artifacts: clutter, occlusion or out-field of sight, lighting changes.

Several filter runs per sequence are performed and analyzed.

2.4.2.1.4.1 Database

This database consists of two different sequences (800 images) acquired from Jido mobile robot in a wide range of realistic conditions.

2.4.2.1.4.2 FR Algorithm(s)

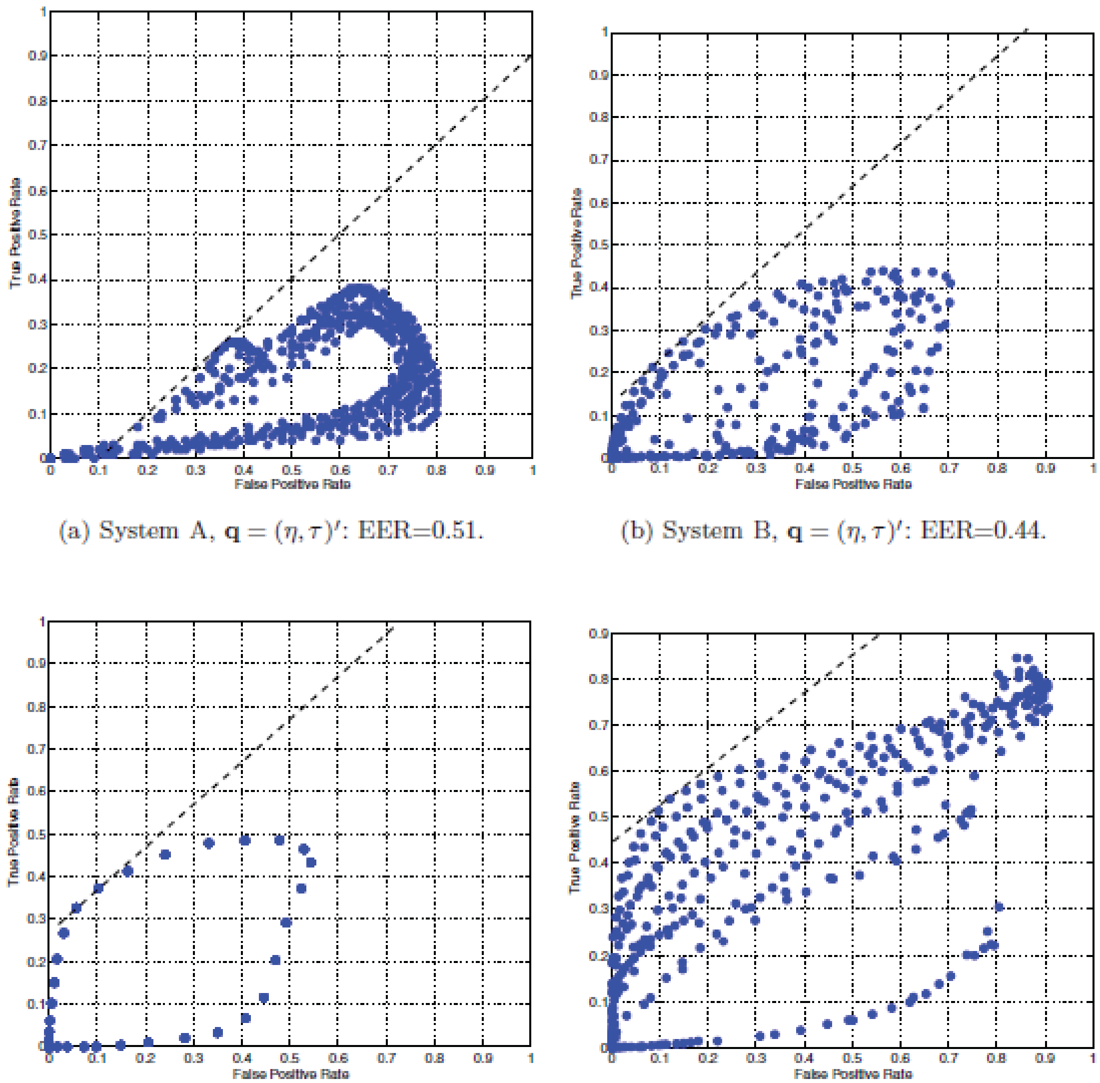
The proposed algorithm is a standalone real-time FR system

2.4.2.1.4.3 Sample Results

The final goal is to classify facial regions *F*, segmented from the input image, into either one class C_t out of the set $\{C_l, l = 1, 2, \ldots, M\}$ of *M* subject faces using training algorithms. The following four developed FR systems are evaluated:

1. FSS+EN system — Face-Specific Subspace and error nor
2. GPCA+MD system — global PCA and Mahanalobis distance
3. LDA+MD system — Fisherfaces and Mahanalobis distance
4. GPCA+SVM system — global PCA and SVM.

Each of these classifiers depends also on a set of free parameters **q**. The performances of these classifiers are analyzed by means of ROCs when varying the free-parameter vector **q** subject to optimization for each classifier and the results, as illustrated in Figure 2-82.



(a) System A, $\mathbf{q} = (\eta, \tau)'$: EER=0.51.

(b) System B, $\mathbf{q} = (\eta, \tau)'$: EER=0.44.

(c) System C, $\mathbf{q} = (\eta, \tau)'$: EER=0.37.

(d) System D, $\mathbf{q} = (\eta, C, \gamma, \tau)'$: EER=0.29.

Figure 2-82 ROC Points for Each Classifier and the Associated Iso-Cost Line for EER (Source: [P-26])

Figure 2-83 illustrated the quantitative performance evaluations, which have also been carried out on the sequence database. Since the main concern of tracking is the accuracy of the tracker results, location as well as face label, the tracking performance is compared quantitatively by defining the False Position Rate (FPR) and the False Label Rate (FLR). If the tracker locks onto none of the observed person, this is considered as a position failure while a tracker lock onto the non-desired person is considered as a label failure.

Figure 2-83 (a) presents the performance considering or not the FR in the tracking loop whereas Figure 2-83 (b) considers the FR performance with or without tracking. The proposed advanced tracker is shown to outperform the conventional tracker (without FR) with much lower false position and label rates for slight additional time consumption.

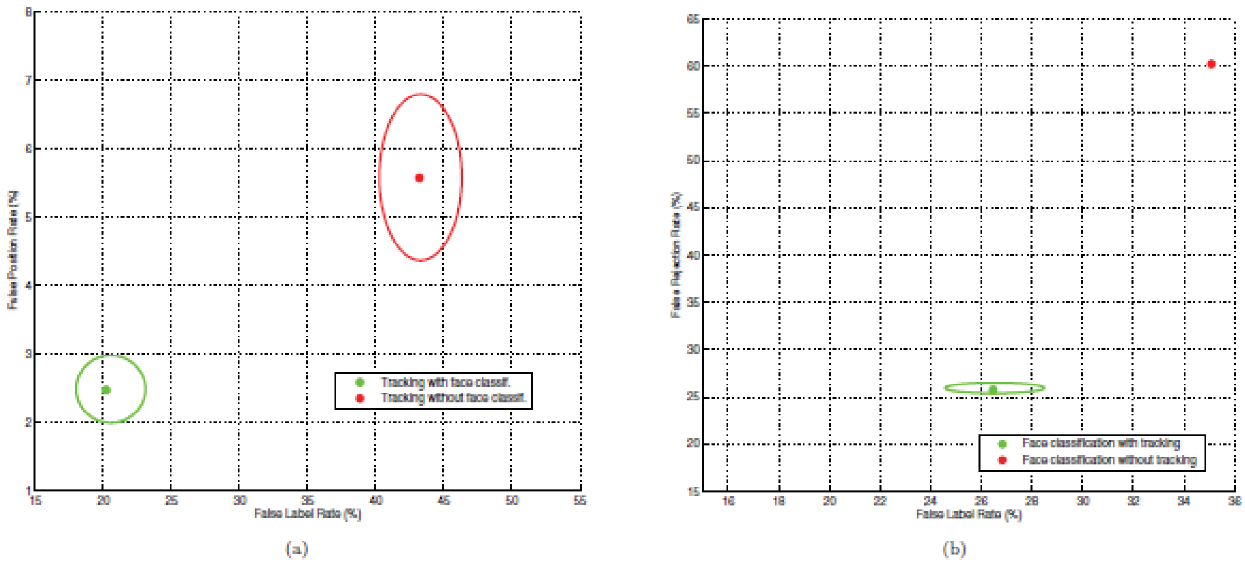


Figure 2-83 (a) Face Tracker Performance for the whole Sequence Database, (b) Face Classification Performance for the Database Image Subset Involving Detected Frontal Faces (Source: [P-26])

2.4.2.1.5 Key Contributions

This paper presented the development of a set of visual functions dedicated to Human/Robot interaction in a household framework used for face recognition. The key contributions of the proposed system are as follows:

- First, a non-dominated sorting genetic algorithm is proposed to find the optimal free-parameters of a SVM-based face classifier in an optimized fashion.
- The main contribution is the design of a video-based face recognition process integrated through a particle filtering framework combining both intermittent features (face and skin blob detection, face recognition) and multiple persistent visual cues (shape and color) in a principled, robust and probabilistically motivated way.
- Off-line evaluations on sequences acquired from the robot show that the overall system enjoys the valuable capabilities:
    1. remaining locked on to the targeted person in populated and continuously changing environments,
    2. recovering this person automatically after full occlusion or temporary disappearance from the field of view.
- Eigenface subspace and SVM makes it possible to improve the face recognition process while the multi-cue fusion in the tracking loop is proven to be more robust than any of the individual cues.
    - Clothing color and also face classification probabilities increase tracker reliability in presence of several persons in the vicinity of the robot.

- Finally, this advanced tracker was integrated into a mobile robot companion called Jido. The visual-based tracker was then successfully tested in Jido's long-term operations in natural settings. The paper claims that the developed robotics system has superior scalable human perception capabilities, as compared to most available mature robotic systems.

Next, we review some of the state-of-art commercial FR technology.

2.5 Commercial Face Recognition Technology

2.5.1 COTS Face Recognition Software/SDK

This Section reviews COTS software or SDK applicable to FR systems. COTS software refers to software packages that can be used out of the box to perform automated facial recognition, while SDK refers to libraries that offer functionalities for FR systems software development. The information is based on publicly available information on vendor websites, which typically do not provide the pricing information.

For the FSAR application, since the rifle-mounted video camera would be moving, we are mainly interested in COTS FR software that can process mobile cameras video imagery, typically with lower resolution and different types of degradation and uncontrolled acquisition conditions. However, to the best of our knowledge, no such commercial FR software packages are found in the literature.

Table 2-34 summarizes a number of COTS FR software for fixed camera. Even though these COTS are mainly designed to handle fixed-camera video data, their suitability for moving camera video imagery should be explored. It is expected that some pre-processing steps will have to be performed on the input mobile-camera video imagery in order to stabilize the video and enhance the image resolutions at the very least. An assessment of the suitability of the listed COTS FR software packages to handle low-resolution imagery is provide in the table, as this is a key requirement for the envisioned FSAR system.

There are a number of SDKs that address some of the FR processing steps. Such functionalities can be integrated with FR system development to avoid re-inventing the wheels and may be applicable, during the development of the FSAR system. Table 2-35 summarizes a number of these SDKs, which are suitable for FR system development and implementation.

Table 2-34 Comparison of COTS FR Software

Company	COTS Software	Web-page	Key Capabilities	Input Data	Low Resolution	Real-time Processing	Reported Identification Rate
FaceFirst	FaceFirst	[R-9]	• Provides fast and accurate automatic FR using a robust information retrieval system • Delivers reliable performance (low false-accept/reject and false-match/non-match rates) • Is easy to use, with a plug-and-play interface that requires no training • Uses many safeguards to secure stored data against theft, loss or reproduction • Applies 3D algorithms to enhance identification	photographs or video from still camera	Limited: The system requires reasonably sharp black-and-white or color photos showing the full face. Because the system must determine the position of both eyes, profile shots are not acceptable.	Yes	About 95% as measured in the U.S. government-sponsored 2006 Face Recognition Vendor Test (FRVT).
IntelliVision	Intelligent Video Analytics	[R-10]	• Recognition of faces and objects • Detection of motion, intrusion, object left, loitering • Video search and summary	Photographs or video from still camera	Yes	Yes	NA

Company	COTS Software	Web-page	Key Capabilities	Input Data	Low Resolution	Real-time Processing	Reported Identification Rate
Avalon Biometrics – A Gemalto Company	FAVES - Facial Automated Verification Solution	[R-11]	• Compatible with images in existing industry standard • Detects more than 400 characteristics/features from each face. • String-based face encoding, which makes matching is extremely fast Can be easily networked, allowing users to connect to many image databases under appropriate controls • Available also in a 64-bit version. • Has text filters, e.g. on gender, which can be used to narrow down the number of potential matches • Accesses existing images in read-only mode so data integrity is assured,	photographs or video from still camera	NA	Yes	NA
Ex-Sight.Com LTD	CFAS - Constant Face Acquisition and Suspect Detection System	[R-12]	• Built-in access to multiple biometric databases • Control center functionalities • Intrusion detecting capabilities • Multi-source enrollment • Automatic face search • Backward face search • Offline image processing • Watch list alerts • Batch faces enrollment • Use of various biometric data (fingerprint, iris, retina, etc.)	Fixed-camera streaming video source, video files, or still images.	Probably not: Has a minimal input Image Size of 640×480 pixels	Yes: Face detection time = 0.07 sec. Matching speed = 100,000 faces/sec. (As reported)	NA

Company	COTS Software	Web-page	Key Capabilities	Input Data	Low Resolution	Real-time Processing	Reported Identification Rate
Cognitec	FaceVACS-VideoScan (Face recognition technology for anonymous people analytics and real-time video screening)	[R-13]	• Tracks multiple faces simultaneously • Compares faces against image galleries in real time • Displays real-time signals when user-defined events occur • Displays and exports statistics about people flow, visitor demographics, and client behavior • Sends signals and statistics to mobile devices • Supports interactive and batch enrollment from still image, still image camera and live video stream via manual and automatic mode.	Fixed-camera streaming video source, video files, or still images.	NA	Yes	NA
herta Security	BioSurveillance	[R-14]	• Real-time multiple face recognition • Analyzes multiple cameras simultaneously • Fully automatic and non-invasive technology • Works properly on changes of facial expression, illumination, beard, eyeglasses, scarfs and caps • Subjects can be enrolled from one or more photos or video • Integration Application Programming Interface (API) available	Fixed-camera streaming video source, video files, or still images.	No Required video capture resolution: HD Resolution at detection: Faces larger than 20x20 pixels Resolution at recognition: Faces larger than 70x70 pixels (100x100 pixels recommended)	Yes Recognition of unlimited number of faces, up to 20 simultaneously without delay.	NA

Company	COTS Software	Web-page	Key Capabilities	Input Data	Low Resolution	Real-time Processing	Reported Identification Rate
SAFRAN Morpho	MorphoFace Investigate – A Facial Recognition System for Intelligence and Investigation Applications	[R-15]	• ISO 19794-5 compliant face image acquisition • Automatic face finding and tracking in photos or videos • Enhancement of poor quality images and non-frontal shots • Use images other than facial photos to aid the operator in the identification process • Navigate within the database to capitalize on information from previous searches • Produce reports on offenders, cases and facial comparisons • Traceability of all image enhancements for submission to courts • Manage access control and user rights	Fixed-camera streaming video source, video files, or still images.	Possibly: It is reported that this software has enhancement of poor quality images and non-frontal shots capabilities	Yes Reported: Fast and accurate facial search, based on photos and videos	NA
Animetrics	FIMS - FaceR Identity Management System (Cloud-powered Facial Recognition)	[R-17]	• Facial recognition for smartphones and web connected digital devices • Real-time in the field ID processing • Real-time watch list search • Advanced face biometrics for uncontrolled facial imagery • Highly scalable • Facial database fault tolerance	Fixed-camera streaming video source, video files, or still images.	No: Saliency requirements: minimum 64 pixels between eye centers	Yes (Cloud-powered)	NA

Company	COTS Software	Web-page	Key Capabilities	Input Data	Low Resolution	Real-time Processing	Reported Identification Rate
iOmniscient	iQ Face	[R-16]	• Face detection: provides the customers the unique ability to detect a face in a crowd. This feature optimizes the capture of the face for recording, matching and recognition. • Face recording: enables the system to record the faces that have been detected with date and time stamp. This feature establishes the presence of persons in the crowd or at the site. • Face matching: enables the system to present to the operator possible matches from faces stored in a database. This also allows the customer to manually and visually match the detected face to the presented images. • Face recognition: automates the process of matching to find the perfect match or the closest match.	Fixed-camera streaming video source, video files, or still images.	Somewhat: Operates with low resolution cameras, but needs at least 22 pixels between the eyes	Yes	NA

Table 2-35 Comparison of SDK Applicable to FR Systems Development

Company	COTS Software	Web-page	Key Capabilities	Supported Platforms	Languages / Programming Environment	Price
OpenCV	OpenCV	[R-18]	• Open source computer vision library for real-time computer vision, with more than 2500 optimized algorithms including state-of-the-art computer vision and machine learning algorithms • Detect and recognize faces, identify objects, classify human actions, track moving objects, etc.	• Windows • Linux/Mac • Android • iOS • NVIDIA GPU	• C/C++ • Wrappers for C#, Python, Java	Free
2D3	Tungsten Media Toolkit	[R-19]	• Provide comprehensive modules to develop real-time digital media solutions and computer vision applications • Moving target detection • Super resolution • Stabilization • Object tracking • Image enhancement	• Windows • Linux • Solaris • Android	• GCC • Visual C++ • Visual Basic • Visual C# • Java and NetBeans • Sun Studio	Depends on modules • Super resolution$1,495 • Stabilization $695 • Object tracking $995 • 30 days trial available

Company	COTS Software	Web-page	Key Capabilities	Supported Platforms	Languages / Programming Environment	Price
Cognitec	FaceVACS-SDK (Development kit brings FaceVACS technology to integrators)	[R-20]	• Powerful face localization on images and video streams • State of the art face tracking on video streams • Industry leading face recognition algorithms (sample quality determination, enrollment, verification, identification) • Substantial and accurate portrait characteristics determination (such as pose deviation, exposure determination, eye glasses detection, eye closed detection, uniform lighting detection, unnatural color check, gender determination, image and face geometry determination) • ISO 19794-5 Full Frontal Image type checks and formatting as required for ePassports	• Windows XP • Professional Windows 2003 and 2008 server Windows Vista, Windows 7 • Linux x86 • Android SDK	• state-of-the-art face recognition API (C++) • BioAPI 2.0 Verification Engine BSP (C - API) • complete Microsoft .NET programming environment integration • Java API • detailed documentation including API reference and user guide • documented examples illustrating the main use cases and providing hints how to create customized implementations • suite of tools to perform biometric evaluations on data: e.g. generation of identification match lists and similarity matrix data	NA

Company	COTS Software	Web-page	Key Capabilities	Supported Platforms	Languages / Programming Environment	Price
Kee-square – Intelligent Sensing for Safety and Security	Morpheus FR SDK	[R-21]	• Face Detection: finds the position and the size of each one of the visible faces in the image • Face Normalization: marks points of morphological interest throughout the face area (eyes, mouth, and eyebrows) in order to properly rotate and scale the image. This operation moves all such points to pre-defined locations. • Feature Extraction: selects relevant features from the normalized image in order to maximize the robustness against environmental disturbance and noise, non-optimal pose, non-neutral facial expressions and variable illumination conditions. • Biometric Template Creation: a trained statistical engine processes the extracted feature vector template. The feature vector is reduced to a smaller one that optimally describes the user's identity. • Biometric Template Matching: the face recognition and the face verification processes are based on measurements of differences between biometric templates. The normalized template distance is the similarity value of the compared identities.	• MS Windows XP, Vista, 2003 • Linus	• C, C++, C#, Visual Basic API for MS Visual Studio 6 or adobe	NA

2.5.2 FERET: US Government-Sponsored FR Technology Program

The FERET program was sponsored by The US Department of Defense (DoD) Counterdrug Technology Development Program Office [R-6]. The goal of the FERET program was to develop automatic face recognition capabilities that could be employed to assist security, intelligence, and law enforcement personnel in the performance of their duties. The FERET database was collected to support the sponsored research and the FERET evaluations. The FERET evaluations were performed to measure progress in algorithm development and identify future research directions.

The program consisted of three major elements:

1. **Sponsoring face recognition research**: A number of institutions have been participating in the FERET FR algorithm development, including MIT, Rutgers University, The Analytic Science Company (TASC), University of Illinois, and University of Southern California.
2. **Collecting the FERET database**: A standard database of face imagery was essential to the success of the FERET program, both to supply standard imagery to the algorithm developers and to supply a sufficient number of images to allow testing of these algorithms. The FERET program set out to establish a large database of facial images that was gathered independently from the algorithm developers. As described in Table A-1, The FERET database is an extensive FR database containing face imagery data, acquired under controlled and uncontrolled conditions, suitable for testing and evaluating the performance of wide range of FR algorithms and techniques. .
3. **Performing the FERET evaluations**: The FERET database made it possible for researchers to develop algorithms on a common database and to report results in the literature using this database. Results reported in the literature did not provide a direct comparison among algorithms because each researcher reported results using different assumptions, scoring methods, and images. The independently administered FERET evaluations allowed for a direct quantitative assessment of the relative strengths and weaknesses of different approaches. Several FERET evaluations have taken place since the start of the FERET program. These FR evaluations are known as the FRVTs, which provide independent government evaluations of commercially available and prototype face recognition technologies. These evaluations are designed to provide U.S. Government and law enforcement agencies with information to assist them in determining where and how facial recognition technology can best be deployed. In addition, FRVT results help identify future research directions for the face recognition community. The last FRVT evaluation was conducted in 2012.

The FERET program is still on-going and thanks to its efforts, new state-of-the- art FR algorithms are being developed and tested on a regular basis. Also, the FERET database continues to expand and grow to include more practical FR test data.

2.5.3 Real-World FR Applications

In this section, we highlight very few of the numerous facial recognition real-world applications.

2.5.3.1 Insurance Corporation of British Columbia (ICBC)

ICBC has been using facial recognition software developed by SAFRAN Morpho [R-15] to check for identify theft, fraud and other illegal activities [R-8]. In 2010, the technology, which enables ICBC to compare a cardholder's image with their existing image on file and with an entire database of millions of images, played a vital role in a number of convictions. The technology works by analyzing facial characteristics that do not change, such as the size and location of cheekbones and the distance between the eyes. Two years after it was introduced, ICBC's use of facial recognition technology has had a dramatic impact on helping to protect their customers from identity theft and fraud [R-8].

2.5.3.2 Passport Canada

The Passport Office has increased scrutiny of all applications for Canadian travel documents since the events of September 11th, 2001. As one of the measures to increase the scrutiny of applications, the Facial Recognition Proof of Concept Project was initiated [R-5]. The Project was initiated to investigate whether FR technology could further improve the security of Canadian travel documents [R-5]. The objective of the FR Proof of Concept Project was to determine whether a Facial Recognition System was feasible, affordable and whether it effectively verified a travel document applicant’s photograph or its digital rendering against those in a query database. The FR processing is accomplished by using FR software to take biometric measurements from a digital image then converting the measurements into an alphanumeric photo biometric identifier using FR algorithms. The derived image-specific biometric identifier is then matched against the identifiers derived from all images in the database, in attempt to match the input photo and identify its subject. The name of the specific COTS face recognition software used by Passport Canada is not made public in [R-5].

2.5.3.3 Identification of Riots Suspects

Face recognition software has also been used to identify offending rioting protestors, such as the Stanley cup riots in Vancouver [R-3] and the G20 riots in Toronto [R-4] with some degree of success. As illustrated in Figure 2-84, the input images of the suspects taken during the Toronto G20 riots illustrate the various challenges of FR software, including occlusion, disguise, illumination, pose, etc. It is reported that, some of the pictured suspects were quickly identified by the Toronto police, using a FR software. The software may not have identified the suspects uniquely, but it succeeded to narrow the set of potential matches significantly. It is reported that, the used facial recognition software can be used to narrow a search. Even if only a person’s eyes are visible, the software can significantly narrow the possibilities by listing, perhaps, 500 possible matches out of a database of 100,000. A policeman then has to go through 500 photographs instead of 100,000 [R-4]. The name of the used COTS FR software was not made public in [R-4].

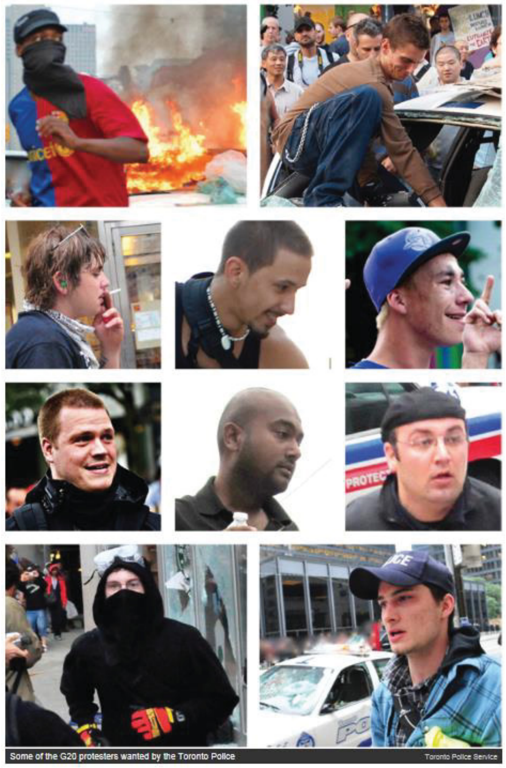


Figure 2-84 Some of the G20 Protesters Wanted by the Toronto Police (Source: [R-4])

2.5.3.4 Case Study: Identification of the Boston Marathon Bombing Suspects

As illustrated in Table 2-40, we review a research paper examining the use of FR systems to identify the Boston marathon bombing suspects.

Table 2-36 Selected Paper on the Use of FR Technology to Identify the Boston Marathon Bombing Suspects

#	Paper Title	Authors	Source	Year	Type
P-26	A Case Study on Unconstrained Facial Recognition Using the Boston Marathon Bombings Suspects	Joshua C. Klontz, Anil K. Jain,	Technical Report MSU-CSE-13-4, 2013, Open access.	2013	Case study

2.5.3.4.1 Abstract

The investigation surrounding the Boston Marathon bombings was a missed opportunity for automated facial recognition to assist law enforcement in identifying suspects. This paper simulates the identification scenario presented by the investigation using two state-of-the-art commercial face recognition systems, and assesses the maturity of face recognition technology in matching low quality face images of uncooperative subjects. The presented experimental results show one instance where a commercial face matcher returns a rank-one hit for suspect Dzhokhar Tsarnaev against a one million mug-shots background database. Though issues surrounding pose, occlusion, and resolution continue to confound matchers, there have been significant advances made in face recognition technology to assist law enforcement agencies in their investigations.

2.5.3.4.2 Motivation

Shortly after the bombing, more than 1,000 law enforcement officers across many agencies began canvassing sources, reviewing government and public databases, and conducting interviews with eyewitnesses. Businesses were asked to review and preserve surveillance video and police received a "huge amount of video evidence" from the public. Figure 2-85 Facial images and videos released by the Federal Bureau of Investigation (FBI) of the two suspects in the Boston Marathon bombings. Suspect 1, Tamerlan Tsarnaev, is wearing a black hat. Suspect 2, Dzhokhar Tsarnaev, is wearing a white hat. The public was asked to help identify these two individuals.

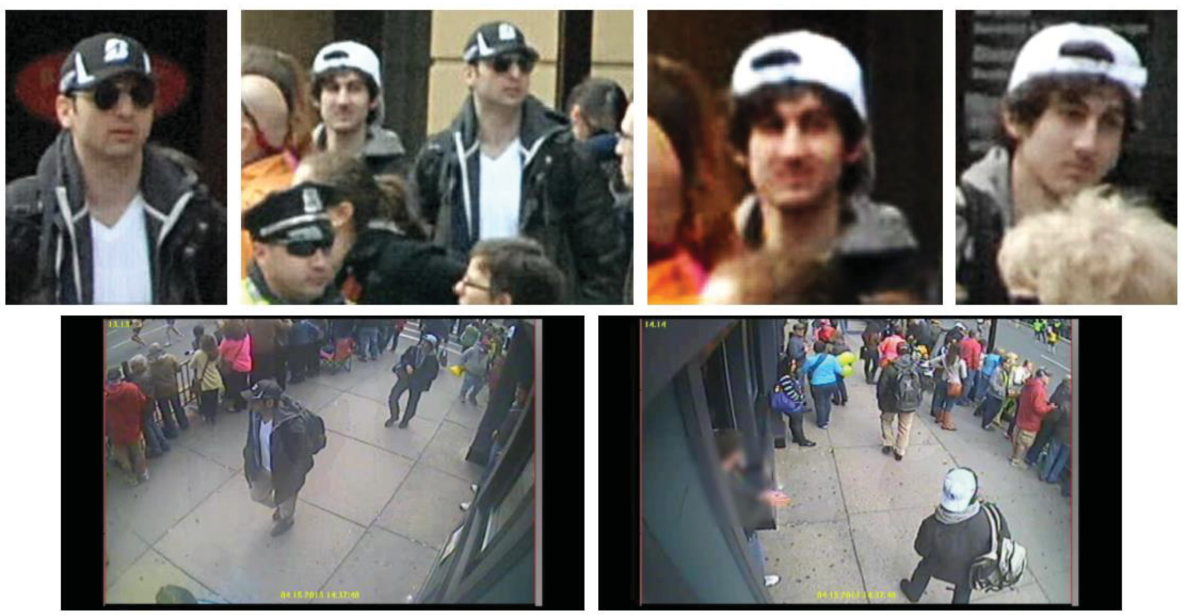


Figure 2-85 Facial images and videos released by the FBI of the two suspects in the Boston Marathon bombings (Source: [P-26])

The investigation of the Boston Marathon bombings has been widely viewed by the media as a failure for automated facial recognition. The technology came up empty even though both Tsarnaevs' photos exist in official government databases: Dzhokhar had a Massachusetts driver's license; the two brothers had legally immigrated to the United States; and Tamerlan had been the subject of an FBI investigation. This paper presents a case study in unconstrained facial recognition, using public domain images of the two suspects in the Boston Marathon bombings.

2.5.3.4.3 Objective

This paper presents a case study in unconstrained facial recognition, using public domain images of the two suspects in the Boston Marathon bombings. Suspects' photographs are matched against a background set of mug-shots with two state-of-the-art commercial face recognition systems. Results are used to gauge the maturity of available technology in unconstrained facial recognition scenarios.

2.5.3.4.4 Experimental Results

The paper simulates the automated facial recognition scenario presented by the Boston Marathon bombings using two state-of-the-art commercial face recognition systems, and images published by law enforcement and news agencies.

2.5.3.4.4.1 Tested COTS Face Recognition Software

The two commercial face recognition systems used in this study were NEC NeoFace and PittPatt.

- NeoFace was chosen based on its top performance in the National Institute of Standards and Technology Multiple Biometrics Evaluation 2010 test.
- PittPatt was selected due to its prevalent use within the law enforcement community and superior performance on non-frontal facial images. In general, matchers were run with their most permissive settings in order to enroll the unconstrained query images, though no other parameter tuning was conducted.

2.5.3.4.4.2 Database

The gallery dataset consists of over 1.6 million law enforcement booking images from the Pinellas County Sheriff's Office (PCSO). Additional photos provided by the public, such as high-school yearbook photo of one of the suspects, were also added to the database. These are older photos of the suspects and they were acquired before the day of the attacks.

Figure 2-86 illustrates face images *1x, 1y* and *1z* are the three gallery images of Suspect 1. Face images *2x, 2y* and *2z* are the three gallery images of Suspect 2.

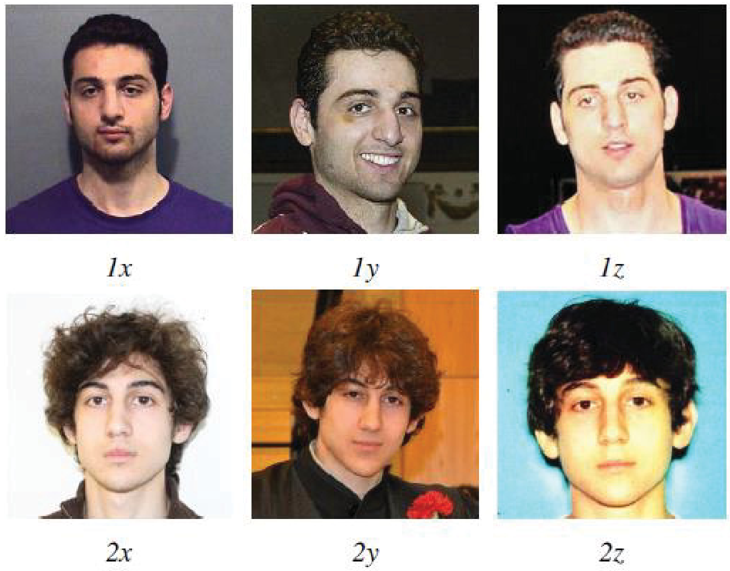


Figure 2-86 Selected Gallery Images of the Two Suspects Selected from the Database (Source: [P-26])

2.5.3.4.4.3 Probe Images

Figure 2-87 shows the five probe images considered in the conducted experiments, cropped from photographs in Figure 2-85. Face images *1x, 1y* and *1z* are the three gallery images of Suspect 1. Face images *2x, 2y* and *2z* are the three gallery images of Suspect 2. No preprocessing was performed prior to enrollment, though probes *2a* and *2b* appear to originate from the same image, suggesting *2b* may have been modified before it was published. Given the difficulty of automatic face detection, quality estimation, tracking, and activity recognition in uncontrolled environments, it is assumed that these face images were extracted manually by law enforcement officials.

Figure 2-87 Selected Probe Images of the two Suspects from Media Released by the FBI (Source: [P-26])

2.5.3.4.4.4 Sample Results

The paper reports experimental results for three separate experiments measuring ranked retrieval rate were conducted to assess the performance of the face matchers in different configurations.

1. Blind Search

In the blind search, each probe is compared against all gallery images without utilizing the demographic information (e.g., gender, ethnicity and age) associated with gallery faces. Table 2-37 shows the retrieval rankings for each probe. Each row contains the ranks at which the true mated gallery images were returned for a given probe. Bold numbers indicate the lowest rank true mate returned for each probe. PittPatt automatic eye detection failed for probes 1a and 1b, and these images could not be enrolled as its SDK does not allow for manual eye localization. NeoFace outperforms PittPatt on all probe images in the conducted experiments.

Table 2-37 Blind (exhaustive) Search Rankings (Source: [P-26])

NeoFace 3.1	1x	1y	1z
1a	116,342	**12,446**	87,501
1b	471,165	438,207	**236,343**
	2x	2y	2z
2a	**213**	308	3,353
2b	7,460	**260**	34,013
2c	1,869	**1**	12,622
PittPatt 5.2.2	2x	2y	2z
2a	14,965	5,556	7,470
2b	997,871	9,002	5,779
2c	139	636	39,943

Figure 2-88 shows the top three returns of each probe for NeoFace FR software. The sunglasses worn by the older brother, Tamerlan Tsarnaev appear to have significantly degraded his top matches.

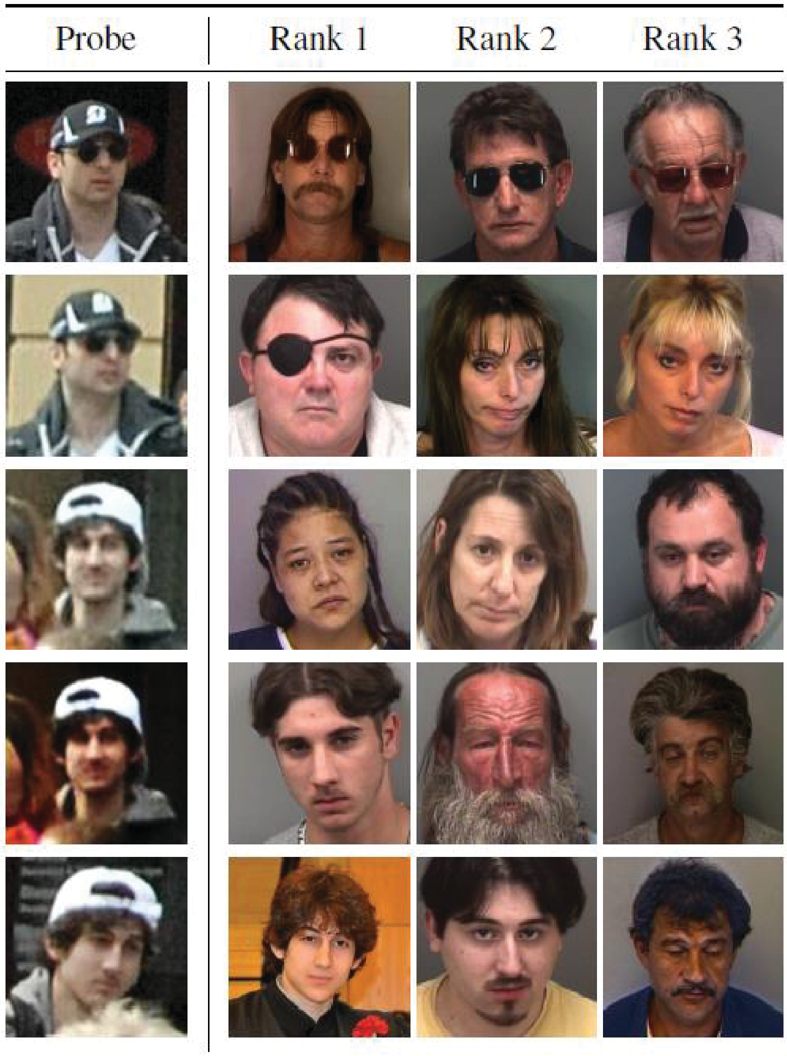


Figure 2-88 Top Three Retrievals in a Blind Search with NeoFace (Source: [P-26])

2. Filtered Search

In the filtered search, each probe is only compared against gallery images with similar demographic data. For Suspect 1 (white, male, 20 to 30 years old) and Suspect 2 (white, male, 15 to 25 years old), filtering reduced the size of the PCSO background gallery from one million to 174,718 and 131,462 images, respectively. Table 2-38 shows the gallery retrieval rankings for each probe for NeoFace and PittPatt, respectively Demographic filtering substantially improves retrieval rankings compared to the blind search, with an improvement generally proportional to the reduction in gallery size.

Table 2-38 Filtered Search Rankings (Source: [P-26])

NeoFace 3.1	Filtered	1x	1y	1z
1a+1b	No	217,761	**48,982**	122,325
1a+1b	Yes	36,666	**8,009**	20,290
		2x	2y	2z
2a+2c	No	74	**3**	1,798
2a+2c	Yes	15	**2**	179
PittPatt 5.2.2	Filtered	2x	2y	2z
2a+2c	No	493	527	10,048
2a+2c	Yes	69	75	1,660

3. Fused Search

In the fused search, match scores using different probe images of the same suspect are summed up without weighting before ranking the gallery images. Table 2-39 shows the gallery retrieval rankings for fused probes with and without demographic filtering. In general, fusion improves retrieval rates for gallery images ranked similarly by each of the probes, but degrades performance for gallery images ranked differently across the fused probes.

Table 2-39 Fused Search Rankings (Source: [P-26])

NeoFace 3.1	Filtered	1x	1y	1z
1a+1b	No	217,761	**48,982**	122,325
1a+1b	Yes	36,666	**8,009**	20,290
		2x	2y	2z
2a+2c	No	74	**3**	1,798
2a+2c	Yes	15	**2**	179
PittPatt 5.2.2	Filtered	2x	2y	2z
2a+2c	No	493	527	10,048
2a+2c	Yes	69	75	1,660

2.5.3.4.5 Key Conclusions

This paper conducted a limited case study to simulate the identification scenario presented by the investigation using two state-of-the-art commercial face recognition systems and gauge the maturity of face recognition technology in matching low quality face images of uncooperative subjects, in real-world surveillance and security threat application operation. Based on the generated experimental results, the key conclusions of the paper include:

- While the Boston Marathon bombings case offers only a small number of published face images for automatic matching, the authors believe there is still valuable insight to be gained from an interpretation of the results.
- Even with NeoFace, the matching accuracy is likely not yet accurate enough for "lights out" deployment in law enforcement applications.
- More progress must be made in overcoming challenges such as pose, resolution, and occlusion in order to increase the utility of unconstrained facial imagery.
- Still, with demographic filtering, multiple probes, and a human in the loop, state-of-the-art face matchers can potentially assist law enforcement in apprehending suspects in a timely fashion.
- The notable rank-one hit for Dzhokhar Tsarnaev is an illustrative example of this potential. However, the hit was against a graduation photograph with similar pose that was tweeted after he had been publicly identified, and not a conventional mug shot from a prior arrest.

2.6 Gender Identification

As illustrated in Table 2-40, we review literature one paper, which examines face detection and gender identification.

Table 2-40 Selected Paper on Gender Identification

#	Paper Title	Authors	Source	Year	Type
P-27	An Automatic Face Detection and Gender Classification from Color Images using Support Vector Machine	Md. Hafizur Rahman, Suman Chowdhury, Md. Abul Bashar	Journal of Emerging Trends in Computing and Information Sciences, Vol. 4, pp. 5-11, No. 1 Jan 2013	2013	End-to-End FR System

2.6.1.1 Paper # [P-27]

2.6.1.1.1 Abstract

This paper presents combined face detection and gender classification method of discriminating between faces of men and women. This is done by detecting the human face area in image given and detecting facial features based on the measurements in pixels. The proposed algorithm converts the RGB image into the YCbCr color space to detect the skin regions from the facial image. But in order to detect facial features the color image is converted into gray scale image. This paper presents appearance-based approach with Gabor filter and SVM classifier. Gabor filter banks are used to extract important facial features, SVM classifier is then used to recognize the facial features. It is proved that SVM can provide superior performance. Different kernel functions have been useful in cases where the data are not linearly separable. These kernel functions transform data to higher dimensional space where they can be separated easily.

2.6.1.1.2 Motivation

Human's face is a prominent feature in machine learning and computer vision system. A face conveys various information including gender, age, ethnicity etc. Face information is applicable in many sectors like biometric authentication and intelligent human-computer interface. Extracting two sets of data for both male and female and separate them accurately is a challenging job. Hence, the classifier selection is a key step the gender classification performance. The paper proposes the use of support vector machine (SVM) classification method, providing a sound theoretic basis for constructing classification models with high generalization ability.

2.6.1.1.3 Method

The proposed automatic face detection and gender classification method is described in Figure 2-89. Some of the modules of the proposed system are described next.

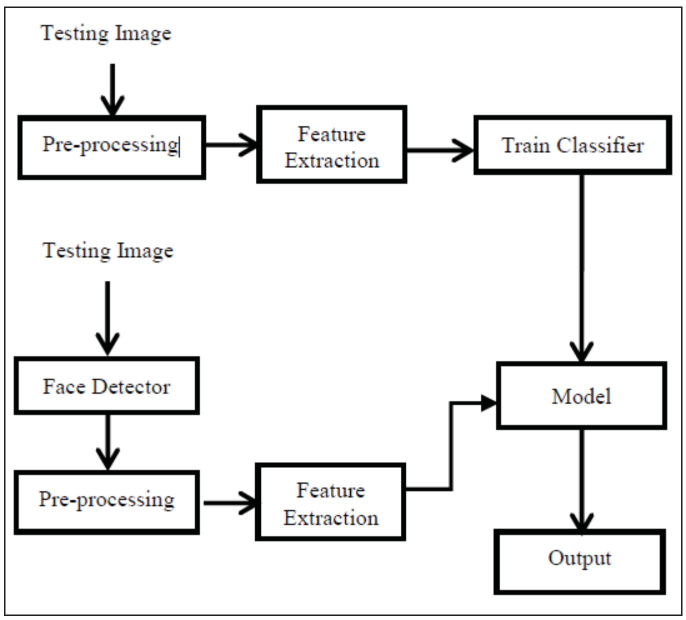


Figure 2-89 General Approach for Proposed Gender Classification System (Source: [P-27])

2.6.1.1.3.1 Face Detection

The approach on this paper will use mainly the color based algorithm with the technique of color space transformation from RGB (red, green and blue) to YCbCr (luminance, chrominance blue and red). The proposed method first detects the face region using skin-color from image. The input image is then converted to binary image. To remove small areas that have been obtained in previous stage, geometric operations, using the available filters, will be done on this area. These processes (Dilation, Erosion, and Hole filling) remove many of unacceptable areas from the face area. The final output after segmentation of skin area for example is shown in Figure 2-90.

Figure 2-90 Original and Skin Area Segmentation Images (Source: [P-27])

2.6.1.1.3.2 Preprocessing

The problem of elimination of non-standard illumination is one of the most complicated problems in the area of computer vision, due to the complex illuminated environment in the real world. In face detection and gender recognition problems, non-standard illumination effects become severe. The accuracy on detecting skin color in complex background is difficult to increase. It is because the appearance of skin-tone color depends on lighting condition.

A Lighting Compensation (LC) is applied for robust skin-tone color detection, and Figure 2-91 illustrates some examples of images and the result images after applying the LC algorithm.

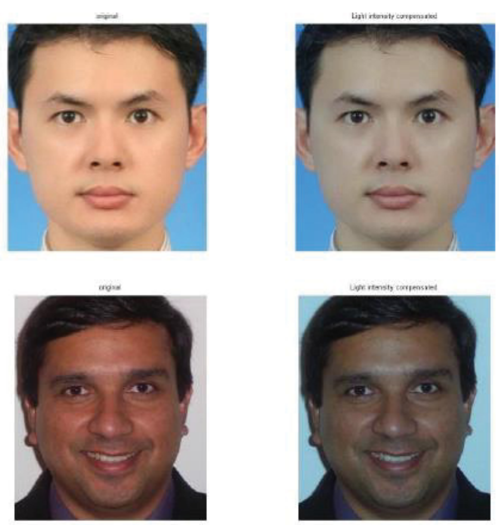


Figure 2-91 Original and Skin Area Segmentation Images (Source: [P-27])

2.6.1.1.3.3 Feature Extraction

Since face recognition is not a difficult task for human beings, selection of biologically motivated Gabor filters is well suited to this problem. A 2D form of Gabor wavelet consists of a planer sinusoid multiplied by a two dimensional Gaussian is used for image processing. 2D

Gabor wavelet highlights and extracts local features from an image, and it has the tolerance of changes in location, shape, scale and light.

An image is converted into 40 images with 5 scales and 8 orientations and the features are the individual Gabor filters coefficients. Figure 2-92 shows how single image is convolved with Gabor Filter of five scales and eight orientations.

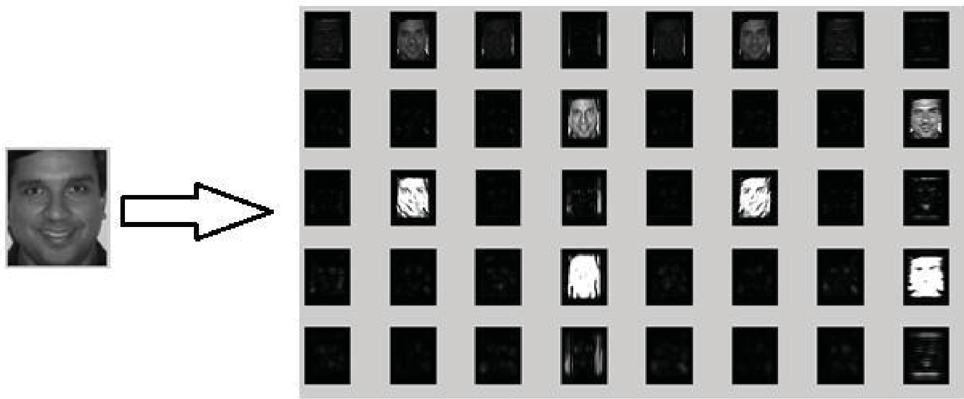


Figure 2-92 Gabor Features of Single Face Image (Source: [P-27])

2.6.1.1.3.4 Support Vector Machine

The SVM is a learning algorithm for classification. It tries to find the optimal separating hyper plane such that the expected classification error for unseen patterns is minimized. For linearly non-separable data the input is mapped to high-dimensional feature space where they can be separated by a hyper plane. This projection into high-dimensional feature space is efficiently performed by using kernels. By using different kernel functions, SVM can implement a wide variety of learning algorithms. It is well known that the SVM has a great potential to perform well.

2.6.1.1.4 Experimental Results

Since the focus of this research is gender recognition, frontal images with slight expression and illumination variations were selected. Since the focus of this research is gender recognition, frontal images with slight expression and illumination variations were also selected. It contains a total of 995 faces which includes 515 male and 480 female faces. These are the faces used for training. The male faces are indexed as +1, while the female faces as -1.

The performance of the proposed method is compared to different cited face detection and gender classification algorithms.

2.6.1.1.4.1 Database

To evaluate the performance of gender classification algorithm, a database was prepared by combining images from several existing databases with different ethnicity and nationalities and referred as the mixed database. The mixed database is composed of images from the following facial data databases: CMU PIE, AR, FERET, Indian Face, and Chinese Face, as illustrated in Table 2-41.

Table 2-41 Details of the Mixed Database (Source: [P-27])

Database	No. of male face images	No. of female face images
CMU PIE [12]	66	56
AR [13]	156	168
Indian Face	150	132
Chinese Face	80	62
FERET [3]	79	62
Total (995)	**515**	**480**

2.6.1.1.4.2 FR Algorithm(s)

This is not a FR system. The aim of the proposed work is to detect the face and identify the gender of the subject and not the subject himself/herself.

2.6.1.1.4.3 Sample Results

Figure 2-93 illustrates sample output from the implemented face detector. A comparison between the gender recognition results generated using a selected set of cited gender identification algorithms tested against the proposed method are shown in Table 2-42. Figure 2-94 shows sample output results of the implemented face detection and gender identification method.

Figure 2-93 Sample Output from the Implemented Face Detector (Source: [P-27])

Table 2-42 Comparison Results of Different Gender Identification Algorithms (Source: [P-27])

Methods	Male Detection Rate (%)	Female Detection Rate (%)
Neural Network	62.31	65.2
Threshold Adaboost	75.26	72.45
LUT Adaboost	75.78	76.71
Mean Adaboost	71.84	73.23
LSVM	78.2	75.8
SVM+Pol	86.72	84.51
SVM+RBF	87	85.5

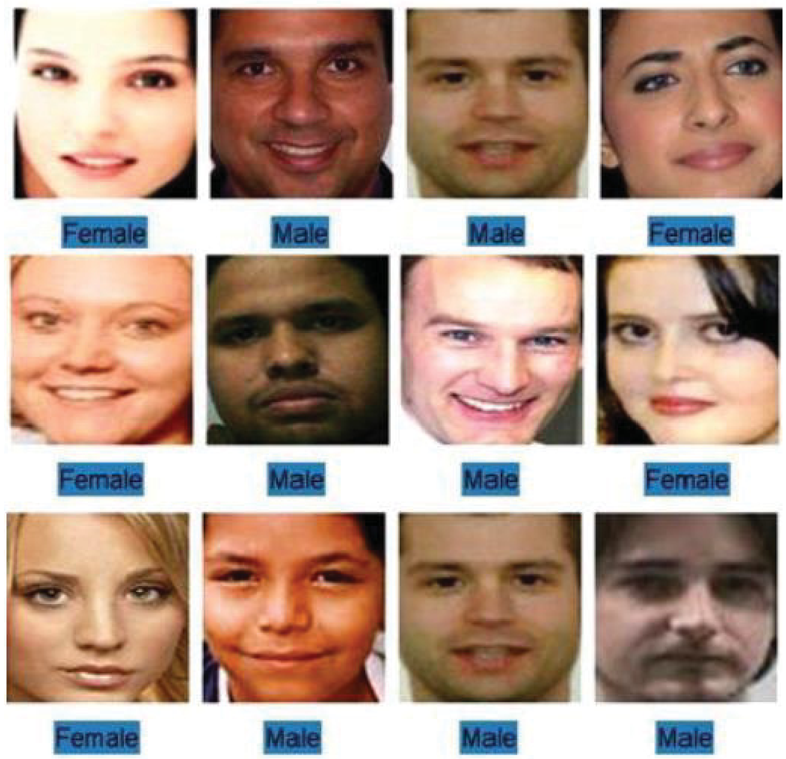


Figure 2-94 Sample Output from the Implemented Gender Classification Method (Source: [P-27])

2.6.1.1.5 Key Contributions

The paper presented a color model conversion algorithm based on chrominance color information, which is applied for face region detection and gender recognition. Its key contributions are listed as follows:

- By applying the threshold measurements in pixel, face area is detected from color image. This achieves a great deal of accuracy.

- Output of the face detector is cropped and used as input to the gender classifier.
- The SVM classifier is trained by Gabor features and applied for gender classification.
- The proposed approach was tested on random images from internet and photos taken by digital camera, and achieved great deal of accuracy of about 88%.

Note that, one way to exploit gender classification in face recognition application is as follows: Suppose we are seeking to identify a male subject. Then we first conduct a gender classification of the database, if it is not done already, and then only match the probe images of the subject to the database images, which belong to males.

2.7 Predicted Battle Field Face Recognition Technology

As illustrated in Table 2-43, we review a paper discussing possible future FR technology, in the battlefield.

Table 2-43 Selected Paper on Predicted Future FR Battlefield Technology

#	Paper Title	Authors	Source	Year	Type
P-28	Can You See Me Now? Visualizing Battlefield Facial Recognition Technology in 2035	Thomas C. Westbrook, Major, USAF	A Research Report Submitted to the Faculty In Partial Fulfillment of Graduation Requirements, Air Command and Staff College, Air University, Maxwell Air Force Base, Alabama, April 2010.	2010	Academic Research Paper

This paper ([P-28]) discusses the technical aspects of FR technologies at a very high level and it does not present any technical details or experimental results. It focuses on the following topics:

- Identifying the applicability and limitations of the current state-of-the-art FR technologies
- Envisioning the future applicability of FR technology in the battle field
- Proposing general research directions to address the current limitations of FR technology so that it becomes viable and meets the suggested military applications in a real-world battle field environment.

Since the objectives of this paper differ vastly form the previously reviewed papers, it will be reviewed differently. We begin with an abstract of the paper.

A. Abstract

In paper # [P-28], the author examines the state-of-the art in facial recognition technology in the areas of access control and law enforcement. The technologies are then projected forward to examine a notional facial recognition system in year 2035 and examine its components. In discussions with biometric experts from the United States and the United Kingdom, the viability of such a system is examined. Challenges, both in the technology and political/bureaucratic realms, are examined and possible solutions are posted.

The paper poses the following three scenarios in which a facial recognition system could be of use to coalition forces operating in the counter-insurgency uncontrolled environment:

1. Passive identification in a crowd
2. Intelligent source verification
3. Sector control.

These scenarios show the range of applications for facial recognition on the battle field.

The paper concludes with a roadmap to assist in focusing of research and allocation of resources as facial recognition technology evolves. With appropriate levels of efforts, a viable technology that is currently capable only in very controlled conditions will evolve into a viable system for positive identification on the battle field by year 2035.

B. The State of the Current State-of-the-Art FR Technology

The paper presents a high-level survey of the current FR technology advances and its real-world applications. It also presents the notional framework of a general FR system, as illustrated in Figure 2-95.

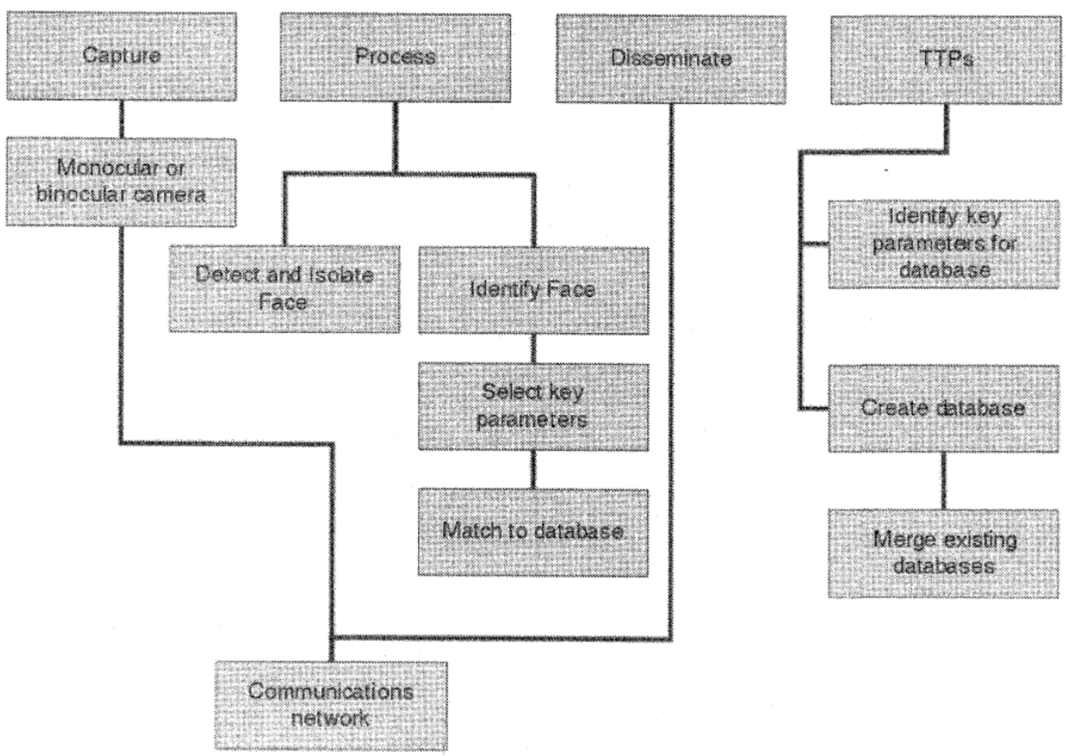


Figure 2-95 Framework of the Predicted Future FR System (Source: [P-28])

C. Long-term FR Research and Development

The paper discusses the current state of the art techniques available for each module of the FR system, focusing on their limitations and suggested research and developments to overcome these limitations, so that the overall FR system can become more robust and applicable to real-world situation, especially military battle field applications.

Some of the stated challenges of the current FR technologies and the requirements of the envisioned future robust FR system that need to be addressed include:

- In the future, a FR system will have access to a much larger stream of data:
    - High resolution and definition video imagery
    - The acquisition system will be fully autonomous
    - Multiple cameras will be able to track individuals as they move
    - Networked sensors will be able to capture images from multiple angles.
- Database: This aspect of facial recognition is not actually a technology component, but the tactics and procedures that guide the use of the system collection of the databases. Today, in the United States, the databases used are spread throughout the US government and beyond. There are so many different and apparently independently collected and managed FR databases. The author recommends that it is essential that all these database collection efforts be synchronized into one centralized data collection component, which is compatible across all FR systems.
- The paper identifies a data communication challenge to future FR technology. Due to the immense amount of inherent high resolution data expected to be processed by such a system, the current communication network would overloaded if the full database of raw data is transmitted to a central processing facility in order to run facial recognition algorithms on it. Also, the security of the transmission itself will have to be addressed, as current communication network are not secure enough, at least for sensitive military information.
- Current raw computing power can provide some boost to the performance of current FR technology, as would distributed computing. In the end, the number of faces in the database would begin to cause existing algorithms to return too many false positives. Essentially, what is required is not more or faster computer power, but better algorithms that can deal with both a large database and facial image that may not be ideal.
- The largest technical challenge in developing a future facial recognition system in reality comes from the core of a facial recognition system, the hardware and software that actually correlate the data sets. With current technology, facial images have to be taken from certain angles or under controlled conditions to allow the software to operate correctly. Future FR systems must be robust in real-world uncontrolled environments.
  The processing component will be the core of any future facial recognition system. However, the ability to identify an individual in a crowd will require an image acquisition source. The acquisition will require high capacity path to the processor. In order to optimize the balance between processing and network saturation, it is important that a fair share of pre-processing be done by the image acquisition system. For instance, super-resolution and image enhancement algorithms could be implemented within the acquisition system to ensure that image date transmitted to the FR system has good quality.

These are some of the challenges to the current FR systems and research in view of the expected requirements and envisioned robust FR systems presented in this paper.

D. Conclusions

The author concludes that if the trend in FR technical advances over the past two decades continues, the realization of robust FR systems, which is operational in real-world uncontrolled environment, is possible. According to the author, the biggest roadblock is the geo-political aspect. That is, whether or not disparate countries and organizations can work together to develop an interoperable standard for their facial recognition databases.

THIS PAGE INTENTIONALLY LEFT BLANK

3 CONCLUDING REMARKS AND RECOMMENDATIONS

The literature review presented in this report indicates the existence of various promising approaches for FR systems applicable to FSAR, which is expected to have uncontrolled probe imagery data acquisition conditions, long range and low resolution, moving camera video acquisition and real-time performance requirements. We make the following observations regarding the applicability of some of the FR reviewed techniques to the FSAR system, in view of its requirements:

Uncontrolled Environment

The FSAR application involves a rifle-mounted video camera operated in a typically uncontrolled outdoor battle field. Thus, all the reviewed challenges encountered by FR systems in uncontrolled environments will be present and are expected to pose a significant challenge to the FR sub-system, integrated within the FSAR system. We have only reviewed a limited number of papers addressing the typical FR challenges, including pose, illumination, occlusion, facial expression, aging and resolution. For each challenges, the proposed solutions can be vastly different and achieve vastly different results on different databases. Hence, it is difficult to compare the reviewed solutions for each challenge and make a recommendation for the best solution. Besides, a limited number (4) of research papers were reviewed for each challenge and this is by no means a comprehensive survey of all possible solutions to each FR challenge found in the literature. However, we note that, for the various proposed solutions for each FR challenge, these solutions are implemented according to one of two approaches:

1. Preprocessing step: In this approach, the proposed solution is applied as pre-processing step to the input probe imagery in order to reduce or eliminate the effects of the FR challenge under consideration. The output of this pre-processing step can then be fed into any suitable FR system. For example, in reviewed paper [P-12], a new approach is proposed to transform facial expressions to neutral-face like images, just like the training images. A suitable FR system can then be applied on the preprocessed probe images, which now have the same facial expression as the training images. Hence, this improves robustness of the FR systems when the database and testing face images differ in facial expression. Similarly, many other reviewed solutions for the different FR challenges fall into this category.

2. Standalone FR system: In this approach, the proposed solution to the considered challenge is integrated within the proposed facial recognition framework and is not done separately as a pre-processing step.

Low Resolution

The ATC system is expected to perform short to medium range facial recognition for personnel identification at standoff distance below 100m. For the higher range values (20m or higher), the resolution and quality and resolution of the acquired imagery is expected to be degraded quickly and inversely proportionally to the acquisition range. Thus, the probe imagery resolution and quality are expected to pose significant challenges to the proposed FR sub-system integrated within the FSAR system. In Section 2.3.5, we reviewed several papers, which propose effective pre-processing operations in order to derive super resolution images from their low resolution counterparts. Image enhancements for the reduction of blur and magnification artifacts have also been proposed in some of the reviewed papers, producing significantly enhanced SR images, with resolution comparable to the gallery images. Using the SR images results in significantly better recognition rates than when using the initial poor resolution input images acquired at long range. Some of the reviewed techniques are quite promising that the low resolution challenge due to long range acquisition may be overcome.

Real Time Performance

As the FSAR involved rifle-mounted video camera streaming data, the FR sub-system must have real-time performance. In section 2.4, we reviewed end-to-end real-time FR systems applicable to uncontrolled environment. Also, as illustrated in Table 2-34, there are already many COTS FR systems, which operate in real-time. However, it should be mentioned that for the FR sub-system of the FSAR system, it will have limited computing power if it operates and processes all data locally. One way to address this challenge is to develop an FSAR system that exploits hardware acceleration, such as Field Programmable Gate Array (FPGA), Graphics Processing Unit (GPU) or even cloud storage and computing.

Moving Camera Video Imagery

The input data for the FSAR FR system consists of streaming video from a rifle-mounted camera, hence a moving camera. Most of reviewed FR systems assume still camera input imagery or video, except for one system reviewed in paper [P-26]. Many of the FR techniques developed for the fixed-camera imagery data can be extended to the moving camera imagery after applying suitable pre-processing steps. These steps will have to be performed on the input mobile-camera video imagery in order to stabilize the video and enhance the image resolutions beforehand.

Recommendations

In view of the above observations, we propose the following work plan towards the development of the FSAR facial recognition system in the future:

1. System specification and design: Conduct a high-level FSAR FR system specification and design, which satisfies the following requirements:
    - The system should be highly modular.

- Each of the previously discussed uncontrolled FR challenges should have its own module.
- Each of these challenges-related modules should be designed as pre-processing steps with multiple options of algorithms to use for processing.
- The output of the pre-processing steps is fed into a matching and recognition module, which should have multiple choices of real-time face recognition algorithms
- A training module, which interfaces with the database
- The output module

2. Database: Explore and plan the creation of a database, which satisfies the following requirements:
    - Multiple training images or video sequences for each subject in the database, if possible. More accurate and reliable face recognition results can be achieved when we are matching multiple probe images to multiple gallery images and especially probe video to gallery video imagery.
    - Also, a database which contains variations of the FR challenges (pose, illumination, etc.) for each subject would be preferred..
3. Image acquisition system: Explore the possible image/video acquisition systems, which will be mounted on the rifle. There are many issues to be investigated and there is a trade-off between many factors including the camera system quality, resolution, cost, size, weight, durability, zoom, range, field of view, power consumption, etc.
4. Computing Engine: Just like the image acquisition system there is a trade-off between many factors including computer power/speed, storage, RAM, size, weight, cost, etc. As mentioned above, hardware acceleration, such as FPGA and GPU should be explored to increase the computing capability of the FSAR system. Also, cloud computing has already empowered mobile communication devices to perform many services in real-time. This option should be explored for the FSAR system as a whole and the FR sub-system in particular. Several issues such as signal and internet availability and secure communication in the battle field will have to be investigated. These technical issues may be resolved over time with technological advances in telecommunications.
5. Comprehensive literature review: As mentioned earlier, due to the limited budget, only a high-level FR literature review has been conducted and documented in this report. For each unconstrained FR challenge, it is recommended that a more comprehensive literature review is to be conducted, while focusing on modular approaches that can be applied as pre-processing operations.
6. Evaluation of promising COTS FR packages with representative FSAR data to assess whether their performances are good enough for operational use.
7. Develop a new FR system based on the promising approaches described, making use of suitable SDKs identified. A comprehensive evaluation using representative FSAR data is necessary to assess the performance, in particular for uncontrolled environments.

THIS PAGE INTENTIONALLY LEFT BLANK

4 REFERENCES

R-1 Automatic Target Cueing (ATC), Task 1 Report – Literature Survey on ATC, MDA Systems Ltd., Issue 1/0, Ref. RX-RP-53-5690, Oct. 30, 2013.

R-2 S. Ohlyan and S. Sangwan, "A Survey on Various Problems & Challenges In Face Recognition," International Journal of Engineering Research & Technology (IJERT), Vol. 2 Issue 6, June – 2013.

R-3 "Committing a crime? Your driver's licence could be a witness against you", accessed on June 1, 2013 from http://www.driving.ca/news/metro/Committing+crime+Your+driver+licence+could+witness+against/5003044/story.html.

R-4 "Police using facial recognition software to help ID G20 suspects", accessed on June 1, 2013 from http://news.nationalpost.com/2010/07/15/police-using-facial-recognition-software-to-help-id-g20-suspects/.

R-5 Passport Canada, "Facial Recognition Application Project – Passport Canada", from http://www.international.gc.ca/department-ministere/atip-aiprp/publications/facial-faciale.aspx?lang=eng [Accessed 28 October 2013].

R-6 Information Technology Laboratory (NITS), http://www.nist.gov/itl/iad/ig/feret.cfm [Accessed 28 October 2013].

R-7 "ACSYS Biometrics – The evolution of security", accessed on June 1, 2013 from http://www.acsysbiometrics.com/index.html.

R-8 "ICBC's facial recognition technology protects customers by identifying fraud", http://www.icbc.com/news/2011feb23-03 [Accessed 28 October 2013].

R-9 FaceFirst, http://www.facefirst.com [Accessed 28 October 2013].

R-10 IntelliVision Intelligent Video Analytics, http://www.intelli-vision.com/products/intelligent-video-analytics [Accessed 28 October 2013].

R-11 Avalon Biometrics - A Gemalto Company, http://www.avalonbiometrics.com/details/faves_facial_automated_verification_solution.html [Accessed 30 October 2013].

R-12 Ex-Sight.Com LTD, http://www.ex-sight.com/cms/c/2/face-recognition-facial-recognition [Accessed 30 October 2013].

R-13 Cognitec – The Face Recognition Company, http://www.cognitec-systems.de/FaceVACS-VideoScan.20.0.html [Accessed 30 October 2013].

R-14 herta Security, http://www.hertasecurity.com/wp-content/uploads/2013/05/Herta_Datasheet_BioSurveillance.pdf [Accessed 30 October 2013].

R-15 SAFRAN Morpho, http://www.morpho.com/IMG/pdf/MorphoFace_Investigate_en-2.pdf [Accessed 30 October 2013].

R-16 iOmniscient Detection System, http://www.iomniscient.com/index.php?option=com_content&view=article&catid=35:product&id=76:iq-facial-recognition [Accessed 30 October 2013].

R-17 ANiMETRiCS, http://animetrics.com/wp-content/uploads/2012/09/animetrics_fims_datasheet.pdf [Accessed 30 October 2013].

R-18 OpenCV, http://www.opencv.org/ [Accessed 30 October 2013].

R-19 2D3 Tungsten Media Toolkit, http://www.2d3.com/product/?v=21 [Accessed 30 October 2013].

R-20 Cognitec – The Face Recognition Company, http://www.cognitec-systems.de/FaceVACS-SDK.19.0.html [Accessed 30 October 2013].

R-21 Kee-square intelligent sensing for safety and security, http://www.keesquare.com/htmldoc/sites/default/files/documents/morpheus_fr_brochure.pdf [Accessed 30 October 2013].

R-22 Face Recognition Homepage, http://www.face-rec.org/databases/ [Accessed 30 October 2013].

A Face recognition Databases

Table A-1 lists many of the FR databases found in the literature, their descriptions and their usability.

Table A-1 Face Recognition Databases (Source: [R-22])

#	Name	Description	Acquisition Setting Variations
1	FERET (Facial Recognition Technology)	The FERET program set out to establish a large database of facial images that was gathered independently from the algorithm developers. Dr. Harry Wechsler at George Mason University was selected to direct the collection of this database. The database collection was a collaborative effort between Dr. Wechsler and Dr. Phillips. The images were collected in a semi-controlled environment. To maintain a degree of consistency throughout the database, the same physical setup was used in each photography session. Because the equipment had to be reassembled for each session, there was some minor variation in images collected on different dates. The FERET database was collected in 15 sessions between August 1993 and July 1996. The database contains 1564 sets of images for a total of 14,126 images that includes 1199 individuals and 365 duplicate sets of images. A duplicate set is a second set of images of a person already in the database and was usually taken on a different day. For some individuals, over two years had elapsed between their first and last sittings, with some subjects being photographed multiple times. This time lapse was important because it enabled researchers to study, for the first time, changes in a subject's appearance that occur over a year.	Age. Other conditions are mainly controlled.
2	SCface - Surveillance Cameras Face Database	SCface is a database of static images of human faces. Images were taken in uncontrolled indoor environment using five video surveillance cameras of various qualities. Database contains 4160 static images (in visible and infrared spectrum) of 130 subjects. Images from different quality cameras mimic the real-world conditions and enable robust face recognition algorithms testing, emphasizing different law enforcement and surveillance use case scenarios. SCface database is freely available to research community	Camera quality and resolution
3	Multi-PIE	A close relationship exists between the advancement of face recognition algorithms and the availability of face databases varying factors that affect facial appearance in a controlled manner. The PIE database, collected at Carnegie Mellon University in 2000, has been very influential in advancing research in face recognition across pose and illumination. Despite its success the PIE database has several shortcomings: a limited number of subjects, a single recording session and only few expressions captured. To address these issues researchers at Carnegie Mellon University collected the Multi-PIE database. It contains 337 subjects, captured under 15 view points and 19 illumination conditions in four recording sessions for a total of more than 750,000 images.	Pose, illumination, and facial expression

#	Name	Description	Acquisition Setting Variations
4	The Yale Face Database	Contains 165 grayscale images in GIF format of 15 individuals. There are 11 images per subject, one per different facial expression or configuration: center-light, w/glasses, happy, left-light, w/no glasses, normal, right-light, sad, sleepy, surprised, and wink.	Facial expression, illumination, disguise and occlusion.
5	PIE Database, CMU	A database of 41,368 images of 68 people, each person under 13 different poses, 43 different illumination conditions, and with 4 different expressions.	Illumination and facial expression
6	Project - Face In Action (FIA) Face Video Database, AMP, CMU	Capturing scenario mimics the real world applications, for example, when a person is going through the airport check-in point. Six cameras capture human faces from three different angles. Three out of the six cameras have smaller focus length, and the other three have larger focus length. Plan to capture 200 subjects in 3 sessions in different time period. For one session, both in-door and out-door scenario will be captured. User-dependent pose and expression variation are expected from the video sequences.	Camera quality, resolution and focus, pose and facial expression
7	AT&T "The Database of Faces" (formerly "The ORL Database of Faces")	Ten different images of each of 40 distinct subjects. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open / closed eyes, smiling / not smiling) and facial details (glasses / no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement).	Illumination, facial expressions (open / closed eyes, smiling / not smiling), facial details (glasses / no glasses).
8	Cohn-Kanade AU Coded Facial Expression Database	Subjects in the released portion of the Cohn-Kanade AU-Coded Facial Expression Database are 100 university students. They ranged in age from 18 to 30 years. Sixty-five percent were female, 15 percent were African-American, and three percent were Asian or Latino. Subjects were instructed by an experimenter to perform a series of 23 facial displays that included single action units and combinations of action units. Image sequences from neutral to target display were digitized into 640 by 480 or 490 pixel arrays with 8-bit precision for grayscale values. Included with the image files are "sequence" files; these are short text files that describe the order in which images should be read.	Gender, ethnicity, facial expression.
9	MIT-CBCL Face Recognition Database	The MIT-CBCL face recognition database contains face images of 10 subjects. It provides two training sets: 1. High resolution pictures, including frontal, half-profile and profile view; 2. Synthetic images (324/subject) rendered from 3D head models of the 10 subjects. The head models were generated by fitting a morphable model to the high-resolution training images. The 3D models are not included in the database. The test set consists of 200 images per subject. The illumination, pose (up to about 30 degrees of rotation in depth) and the background have been varied.	Resolution, pose, illumination, pose, background.

#	Name	Description	Acquisition Setting Variations
10	Image Database of Facial Actions and Expressions - Expression Image Database	24 subjects are represented in this database, yielding between about 6 to 18 examples of the 150 different requested actions. Thus, about 7,000 color images are included in the database, and each has a matching gray scale image used in the neural network analysis.	Pose and state of the subject's body/head/face, facial expression, color and gray-scale images
11	Face Recognition Data, University of Essex, UK	395 individuals (male and female), 20 images per individual. Contains images of people of various racial origins, mainly of first year undergraduate students, so the majority of individuals are between 18-20 years old but some older individuals are also present. Some individuals are wearing glasses and beards.	Ethnicity, young and old subjects, disguise and occlusion.
12	NIST Mug Shot Identification Database	There are images of 1573 individuals (cases) 1495 male and 78 female. The database contains both front and side (profile) views when available. Separating front views and profiles, there are 131 cases with two or more front views and 1418 with only one front view. Profiles have 89 cases with two or more profiles and 1268 with only one profile. Cases with both fronts and profiles have 89 cases with two or more of both fronts and profiles, 27 with two or more fronts and one profile, and 1217 with only one front and one profile.	Gender and pose.
13	NLPR Face Database	450 face images. 896 x 592 pixels. JPEG format. 27 or so unique people under with different lighting/expressions/backgrounds.	Lighting, facial expression and picture backgrounds.
14	M2VTS Multimodal Face Database (Release 1.00)	Database is made up from 37 different faces and provides 5 shots for each person. These shots were taken at one week intervals or when drastic face changes occurred in the meantime. During each shot, people have been asked to count from '0' to '9' in their native language (most of the people are French speaking), rotate the head from 0 to -90 degrees, again to 0, then to +90 and back to 0 degrees. Also, they have been asked to rotate the head once again without glasses if they wear any.	Pose, expression, disguise, and occlusion.
15	The Extended M2VTS Database, University of Surrey, UK	Contains four recordings of 295 subjects taken over a period of four months. Each recording contains a speaking head shot and a rotating head shot. Sets of data taken from this database are available including high quality color images, 32 KHz 16-bit sound files, video sequences and a 3D model.	Facial expression, pose, resolution, color, sound and video and 3D models.
16	The AR Face Database, Purdue University, USA	4,000 color images corresponding to 126 people's faces (70 men and 56 women). Images feature frontal view faces with different facial expressions, illumination conditions, and occlusions (sun glasses and scarf).	Gender, pose, expression, illumination and occlusion.

#	Name	Description	Acquisition Setting Variations
17	The University of Oulu Physics-Based Face Database	Contains 125 different faces each in 16 different camera calibrations and illumination conditions, an additional 16 if the person has glasses. Faces in frontal position captured under Horizon, Incandescent, Fluorescent and Daylight illuminant. Includes 3 spectral reflectance of skin per person measured from both cheeks and forehead. Contains RGB spectral response of camera used and spectral power distribution of illuminants.	Camera calibration/resolution/ quality, illumination, occlusion and disguise, color/grayscale.
18	CAS-PEAL Face Database	The CAS-PEAL face database has been constructed under the sponsors of National Hi-Tech Program and ISVISION. The goals to create the PEAL face database include: providing the worldwide researchers of FR community a large-scale Chinese face database for training and evaluating their algorithms; facilitating the development of FR by providing large-scale face images with different sources of variations, especially Pose, Expression, Accessories, and Lighting (PEAL); advancing the state-of-the-art face recognition technologies aiming at practical applications especially for the oriental.	Ethnicity, pose, expression, accessories and illumination.
19	Japanese Female Facial Expression (JAFFE) Database	The database contains 213 images of 7 facial expressions (6 basic facial expressions + 1 neutral) posed by 10 Japanese female models. Each image has been rated on 6 emotion adjectives by 60 Japanese subjects.	Facial expression, emotions, Japanese females.
20	BioID Face DB - HumanScan AG, Switzerland	The dataset consists of 1521 gray level images with a resolution of 384x286 pixel. Each one shows the frontal view of a face of one out of 23 different test persons. For comparison reasons the set also contains manually set eye positions	Controlled conditions.
21	Psychological Image Collection at Stirling (PICS)	This is a collection of images useful for research in Psychology, such as sets of faces and objects. The images in the database are organised into SETS, with each set often representing a separate experimental study.	Different sets have different acquisition characteristics.
22	The Sheffield Face Database (previously: The UMIST Face Database)	Consists of 564 images of 20 people. Each covering a range of poses from profile to frontal views. Subjects cover a range of race/sex/appearance. Each subject exists in their own directory labelled 1a, 1b, ... 1t and images are numbered consecutively as they were taken. The files are all in PGM format, approximately 220 x 220 pixels in 256 shades of grey.	Pose, ethnicity, gender and appearance.

#	Name	Description	Acquisition Setting Variations
23	Face Video Database of the Max Planck Institute for Biological Cybernetics	This database contains short video sequences of facial Action Units recorded simultaneously from six different viewpoints, recorded in 2003 at the Max Planck Institute for Biological Cybernetics. The video cameras were arranged at 18 degrees intervals in a semi-circle around the subject at a distance of roughly 1.3m. The cameras recorded 25 frames/sec at 786x576 video resolution, non-interlaced. In order to facilitate the recovery of rigid head motion, the subject wore a headplate with 6 green markers. The website contains a total of 246 video sequences in MPEG1 format.	Video sequences with variable pose and facial expression.
24	Caltech Faces	450 face images. 896 x 592 pixels. JPEG format. 27 or so unique people under with different lighting/expressions/backgrounds.	Lighting, expressions and backgrounds.
25	EQUINOX HID Face Database	Human identification from facial features has been studied primarily using imagery from visible video cameras. Thermal imaging sensors are one of the most innovative emerging technologies in the market. Fueled by ever lowering costs and improved sensitivity and resolution. These sensors provide exciting new opportunities for biometric identification. As part of their involvement in this effort, Equinox is collecting an extensive database of face imagery in the following modalities: co-registered broadband-visible/LWIR (8-12 microns), MWIR (3-5 microns), SWIR (0.9-1.7 microns). This data collection is made available for experimentation and statistical performance evaluations.	Thermal imaging sensors and modalities.
26	VALID Database	With the aim to facilitate the development of robust audio, face, and multi-modal person recognition systems, the large and realistic multi-modal (audio-visual) VALID database was acquired in a noisy "real world" office scenario with no control on illumination or acoustic noise. The database consists of five recording sessions of 106 subjects over a period of one month. One session is recorded in a studio with controlled lighting and no background noise. The other 4 sessions are recorded in office type scenarios. The database contains uncompressed JPEG Images at resolution of 720x576 pixels.	Modality, illumination, acoustic noise.
27	The UCD Color Face Image Database for Face Detection	The database has two parts. Part one contains color pictures of faces having a high degree of variability in scale, location, orientation, pose, facial expression and lighting conditions, while part two has manually segmented results for each of the images in part one of the database. These images are acquired from a wide variety of sources such as digital cameras, pictures scanned using photo-scanner, other face databases and the World Wide Web. The database is intended for distribution to researchers.	Scale, location, orientation, pose, facial expression and lighting conditions, modality, resolution.
28	Georgia Tech Face Database	The database contains images of 50 people and is stored in JPEG format. For each individual, there are 15 color images captured between 06/01/99 and 11/15/99. Most of the images were taken in two different sessions to take into account the variations in illumination conditions, facial expression, and appearance. In addition to this, the faces were captured at different scales and orientations.	Illumination conditions, facial expression, and appearance.

#	Name	Description	Acquisition Setting Variations
29	Indian Face Database	The database contains a set of face images taken in February, 2002 in the IIT Kanpur campus. There are eleven different images of each of 40 distinct subjects. For some subjects, some additional photographs are included. All the images were taken against a bright homogeneous background with the subjects in an upright, frontal position. The files are in JPEG format. The size of each image is 640x480 pixels, with 256 grey levels per pixel. The images are organized in two main directories - males and females. In each of these directories, there are directories with name as a serial numbers, each corresponding to a single individual. In each of these directories, there are eleven different images of that subject, which have names of the form abc.jpg, where abc is the image number for that subject. The following orientations of the face are included: looking front, looking left, looking right, looking up, looking up towards left, looking up towards right, looking down. Available emotions are: neutral, smile, laughter, sad/disgust.	Pose and facial expressions.
30	VidTIMIT Database	The VidTIMIT database is comprised of video and corresponding audio recordings of 43 people, reciting short sentences. It can be useful for research on topics such as multi-view face recognition, automatic lip reading and multi-modal speech recognition. The dataset was recorded in 3 sessions, with a space of about a week between each session. There are 10 sentences per person, chosen from the TIMIT corpus. In addition to the sentences, each person performed a head rotation sequence in each session. The sequence consists of the person moving their head to the left, right, back to the center, up, then down and finally return to center. The recording was done in an office environment using a broadcast quality digital video camera. The video of each person is stored as a numbered sequence of JPEG images with a resolution of 512 x 384 pixels. The corresponding audio is stored as a mono, 16 bit, 32 kHz WAV file.	Pose and facial expression.
31	Labeled Faces in the Wild	Labeled Faces in the Wild is a database of face photographs designed for studying the problem of unconstrained face recognition. The database contains more than 13,000 images of faces collected from the web. Each face has been labeled with the name of the person pictured. 1680 of the people pictured have two or more distinct photos in the database. The only constraint on these faces is that they were detected by the Viola-Jones face detector.	Unconstrained - All acquisition setting can change.
32	The LFWcrop Database	LFWcrop is a cropped version of the Labeled Faces in the Wild (LFW) dataset, keeping only the center portion of each image (i.e. the face). In the vast majority of images almost all of the background is omitted. LFWcrop was created due to concern about the misuse of the original LFW dataset, where face matching accuracy can be unrealistically boosted through the use of background parts of images (i.e. exploitation of possible correlations between faces and backgrounds). As the location and size of faces in LFW was determined through the use of an automatic face locator (detector), the cropped faces in LFWcrop exhibit real-life conditions, including mis-alignment, scale variations, in-plane as well as out-of-plane rotations.	Similar to Labeled Faces in the Wild, above.

#	Name	Description	Acquisition Setting Variations
33	Labeled Faces in the Wild-a (LFW-a)	The "Labeled Faces in the Wild-a" image collection is a database of labeled, face images intended for studying Face Recognition in unconstrained images. It contains the same images available in the original Labeled Faces in the Wild data set, however, here they are provided after alignment using a commercial face alignment software. Some of the results were produced using these images. This alignment is shown to improve the performance of face recognition algorithms. The same directory structure has been maintained as in the original LFW data set, and so these images can be used as direct substitutes for those in the original image set. Note, however, that the images available here are grayscale versions of the originals.	Similar to Labeled Faces in the Wild, above.
34	3D_RMA database	The 3D_RMA database is a collection of two sessions (Nov 1997 and Jan 1998) consisting of 120 persons. For each session, three shots were recorded with different (but limited) orientations of the head. Details about the population and typical problems affecting the quality are given in the referred link. 3D was captured thanks to a first prototype of a proprietary system based on structured light (analog camera!). The quality was limited but sufficient to show the ability of 3D face recognition. For privacy reasons, the texture images are not made available. In the period 2003-2008, this database has been downloaded by about 100 researchers. A few papers present recognition results with the database (like, of course, papers from the author).	Pose, camera quality and resolution.
35	GavabDB: 3D face database, GAVAB research group, Universidad Rey Juan Carlos, Spain	GavabDB is a 3D face database. It contains 549 three-dimensional images of facial surfaces. These meshes correspond to 61 different individuals (45 male and 16 female) having 9 images for each person. The total of the individuals are Caucasian and their age is between 18 and 40 years old. Each image is given by a mesh of connected 3D points of the facial surface without texture. The database provides systematic variations with respect to the pose and the facial expression. In particular, the 9 images corresponding to each individual are: 2 frontal views with neutral expression, 2 x-rotated views (ą30o, looking up and looking down respectively) with neutral expression, 2 y-rotated views (ą90o, left and right profiles respectively) with neutral expression and 3 frontal gesture images (laugh, smile and a random gesture chosen by the user, respectively).	Pose, facial expression.
36	FRAV2D Database	This database is formed by up to 109 subjects (75 men and 34 women), with 32 color images per person. Each picture has a 320 x 240 pixel resolution, with the face occupying most of the image in an upright position. For one single person, all the photographs were taken on the same day, although the subject was forced to stand up and sit down again in order to change pose and gesture. In all cases, the background is plain and dark blue. The 32 images were classified in six groups according to the pose and lighting conditions: 12 frontal images, 4 15o-turned images, 4 30o-turned images, 4 images with gestures, 4 images with occluded face features and 4 frontal images with a change of illumination. This database is delivered for free exclusively for research purposes.	Pose, gesture, lighting, background.

#	Name	Description	Acquisition Setting Variations
37	FRAV3D Database	This database contains 106 subjects, with approximately one woman every three men. The data were acquired with a Minolta VIVID 700 scanner, which provides texture information (2D image) and a VRML file (3D image). If needed, the corresponding range data (2.5D image) can be computed by means of the VRML file. Therefore, it is a multimodal database (2D, 2.5D y 3D). During all time, a strict acquisition protocol was followed, with controlled lighting conditions. The person sat down on an adjustable stool opposite the scanner and in front of a blue wall. No glasses, hats or scarves were allowed. A total of 16 captures per person were taken in every session, with different poses and lighting conditions, trying to cover all possible variations, including turns in different directions, gestures and lighting changes. In every case only one parameter was modified between two captures. This is one of the main advantages of this database, respect to others. This database is delivered for free exclusively for research purposes.	Gender, modality, disguise, occlusion, illumination.
38	BJUT-3D Chinese Face Database	The BJUT-3D is a three dimension face database including 500 Chinese persons. There are 250 females and 250 males in the database. Everyone has a 3D face data with neutral expression and without accessories. Original high-resolution 3D face data is acquired by the CyberWare 3D scanner in given environment, Every 3D face data has been preprocessed, and cut the redundant parts. Now the face database is available for research purpose only. The Multimedia and Intelligent Software Technology Beijing Municipal Key Laboratory in Beijing University of Technology is serving as the technical agent for distribution of the database and reserves the copyright of all the data in the database.	Ethnicity, gender, facial expression and resolution.
39	The Bosphorus Database	The Bosphorus Database is a new 3D face database that includes a rich set of expressions, systematic variation of poses and different types of occlusions. This database is unique from three aspects: (1) The facial expressions are composed of judiciously selected subset of Action Units as well as the six basic emotions, and many actors/actresses are incorporated to obtain more realistic expression data; (2) A rich set of head pose variations are available; (3) Different types of face occlusions are included. Hence, this new database can be a very valuable resource for development and evaluation of algorithms on face recognition under adverse conditions and facial expression analysis as well as for facial expression synthesis.	Facial expression, pose, occlusion, action.
40	PUT Face Database	PUT Face Database consists of almost 10000 hi-res images of 100 people. Images were taken in controlled conditions and the database is supplied with additional data including: rectangles containing face, eyes, nose and mouth, landmarks positions and manually annotated contour models. Database is available for research purposes.	Controlled conditions.

#	Name	Description	Acquisition Setting Variations
41	The Basel Face Model (BFM)	The Basel Face Model (BFM) is a 3D Morphable Face Model constructed from 100 male and 100 female example faces. The BFM consists of a generative 3D shape model covering the face surface from ear to ear and a high quality texture model. The model can be used either directly for 2D and 3D face recognition or to generate training and test images for any imaging condition. Hence, in addition to being a valuable model for face analysis it can also be viewed as a meta-database which allows the creation of accurately labeled synthetic training and testing images. To allow for a fair comparison with other algorithms, both the training data set (the BFM) and the model fitting results are provided for several standard image data sets (CMU-PIE, FERET) obtained with a fitting algorithm. The BFM web page additionally provides a set of registered scans of ten individuals, together with a set of 270 renderings of these individuals with systematic pose and light variations. These scans are not included in the training set of the BFM and form a standardized test set with a ground truth for pose and illumination.	Gender, pose, light.
42	Plastic Surgery Face Database	The plastic surgery face database is a real world database that contains 1800 pre and post-surgery images pertaining to 900 subjects. Different types of facial plastic surgeries have different impact on facial features. To enable the researchers to design and evaluate face recognition algorithms on all types of facial plastic surgeries, the database contains images from a wide variety of cases such as Rhinoplasty (nose surgery), Blepharoplasty (eyelid surgery), brow lift, skin peeling, and Rhytidectomy (face lift). For each individual, there are two frontal face images with proper illumination and neutral expression: the first is taken before surgery and the second is taken after surgery. The database contains 519 image pairs corresponding to local surgeries and 381 cases of global surgery (e.g., skin peeling and face lift).	Before-after plastic surgery.
43	The Iranian Face Database (IFDB)	The Iranian Face Database (IFDB), the first image database in middle-east, contains color facial imagery of a large number of Iranian subjects. IFDB is a large database that can support studies of the age classification systems. It contains over 3,600 color images. IFDB can be used for age classification, facial feature extraction, aging, facial ratio extraction, percent of facial similarity, facial surgery, race detection and other similar researches.	Ethnicity, age, surgery.

#	Name	Description	Acquisition Setting Variations
44	The Hong Kong Polytechnic University NIR Face Database	The Biometric Research Centre at The Hong Kong Polytechnic University developed a real time NIR face capture device and used it to construct a large-scale NIR face database. The NIR face image acquisition system consists of a camera, an LED light source, a filter, a frame grabber card and a computer. The camera used is a JAI camera, which is sensitive to NIR band. The active light source is in the NIR spectrum between 780nm - 1,100 nm. The peak wavelength is 850 nm. The strength of the total LED lighting is adjusted to ensure a good quality of the NIR face images when the camera face distance is between 80 cm - 120 cm, which is convenient for the users. By using the data acquisition device described above, NIR face images are collected from 335 subjects. During the recording, the subject was first asked to sit in front of the camera, and the normal frontal face images of him/her were collected. Then the subject was asked to make expression and pose changes and the corresponding images were collected. To collect face images with scale variations, the subjects were asked to move near to or away from the camera in a certain range. At last, to collect face images with time variations, samples from 15 subjects were collected at two different times with an interval of more than two months. In each recording, About 100 images were collected from each subject, and in total about 34,000 images were collected in the PolyU-NIRFD database	Pose, scale, age.
45	The Hong Kong Polytechnic University Hyperspectral Face Database (PolyU-HSFD)	The Biometric Research Centre at The Hong Kong Polytechnic University established a Hyperspectral Face database. The indoor hyperspectral face acquisition system was built which mainly consists of a CRI's VariSpec LCTF and a Halogen Light, and includes a hyperspectral dataset of 300 hyperspectral image cubes from 25 volunteers with age range from 21 to 33 (8 female and 17 male). For each individual, several sessions were collected with an average time space of 5 month. The minimal interval is 3 months and the maximum is 10 months. Each session consists of three hyperspectral cubes - frontal, right and left views with neutral-expression. The spectral range is from 400 nm to 720 nm with a step length of 10 nm, producing 33 bands in all. Since the database was constructed over a long period of time, significant appearance variations of the subjects, e.g. changes of hair style and skin condition, are presented in the data. In data collection, positions of the camera, light and subject are fixed, which allows us to concentrate on the spectral characteristics for face recognition without masking from environmental changes.	Pose, illumination, appearance, age.

#	Name	Description	Acquisition Setting Variations
46	MOBIO - Mobile Biometry Face and Speech Database	The MOBIO database consists of bi-modal (audio and video) data taken from 152 people. The database has a female-male ratio or nearly 1:2 (100 males and 52 females) and was collected from August 2008 until July 2010 in six different sites from five different countries. This led to a diverse bi-modal database with both native and non-native English speakers. In total 12 sessions were captured for each client: 6 sessions for Phase I and 6 sessions for Phase II. The Phase I data consists of 21 questions with the question types ranging from: Short Response Questions, Short Response Free Speech, Set Speech, and Free Speech. The Phase II data consists of 11 questions with the question types ranging from: Short Response Questions, Set Speech, and Free Speech. The database was recorded using two mobile devices: a mobile phone and a laptop computer. The mobile phone used to capture the database was a NOKIA N93i mobile while the laptop computer was a standard 2008 MacBook. The laptop was only used to capture part of the first session, which consists of data captured on both the laptop and the mobile phone.	Gender, camera modality/quality, ethnicity/diversity.
47	Texas 3D Face Recognition Database (Texas 3DFRD)	Texas 3D Face Recognition database (Texas 3DFRD) contains 1149 pairs of facial color and range images of 105 adult human subjects. The images were acquired at the company Advanced Digital Imaging Research (ADIR), LLC (Friendswood, TX), formerly a subsidiary of Iris International, Inc. (Chatsworth, CA), with assistance from research students and faculty from the Laboratory for Image and Video Engineering (LIVE) at The University of Texas at Austin. This project was sponsored by the Advanced Technology Program of the National Institute of Standards and Technology (NIST). The database is being made available by Dr. Alan C Bovik at UT Austin. The images were acquired using a stereo imaging system at a high spatial resolution of 0.32 mm. The color and range images were captured simultaneously and thus are perfectly registered to each other. All faces have been normalized to the frontal position and the tip of the nose is positioned at the center of the image. The images are of adult humans from all the major ethnic groups and both genders. For each face, is also available information about the subjects' gender, ethnicity, facial expression, and the locations 25 anthropometric facial fiducial points. These fiducial points were located manually on the facial color images using a computer based graphical user interface. Specific data partitions (training, gallery, and probe) that were employed at LIVE to develop the Anthropometric 3D Face Recognition algorithm are also available.	Modality, ethnicity, gender, facial expression.
48	Natural Visible and Infrared facial Expression database (USTC-NVIE)	The database contains both spontaneous and posed expressions of more than 100 subjects, recorded simultaneously by a visible and an infrared thermal camera, with illumination provided from three different directions. The posed database also includes expression images with and without glasses.	Pose and illumination.

#	Name	Description	Acquisition Setting Variations
49	FEI Face Database	The FEI face database is a Brazilian face database that contains a set of face images taken between June 2005 and March 2006 at the Artificial Intelligence Laboratory of FEI in Sao Bernardo do Campo, Sao Paulo, Brazil. There are 14 images for each of 200 individuals, a total of 2800 images. All images are colorful and taken against a white homogenous background in an upright frontal position with profile rotation of up to about 180 degrees. Scale might vary about 10% and the original size of each image is 640x480 pixels. All faces are mainly represented by students and staff at FEI, between 19 and 40 years old with distinct appearance, hairstyle, and adorns. The number of male and female subjects are exactly the same and equal to 100.	Pose, scale, ethnicity and gender.
50	ChokePoint	ChokePoint video dataset is designed for experiments in person identification/verification under real-world surveillance conditions using existing technologies. An array of three cameras was placed above several portals (natural choke points in terms of pedestrian traffic) to capture subjects walking through each portal in a natural way. While a person is walking through a portal, a sequence of face images (i.e. a face set) can be captured. Faces in such sets will have variations in terms of illumination conditions, pose, sharpness, as well as misalignment due to automatic face localisation/detection. Due to the three camera configuration, one of the cameras is likely to capture a face set where a subset of the faces is near-frontal. The dataset consists of 25 subjects (19 male and 6 female) in portal 1 and 29 subjects (23 male and 6 female) in portal 2. In total, the dataset consists of 54 video sequences and 64,204 labelled face images.	Illumination conditions, pose, sharpness.
51	UMB database of 3D occluded faces	The University of Milano Bicocca 3D face database is a collection of multimodal (3D + 2D color images) facial acquisitions. The database is available to universities and research centers interested in face detection, face recognition, face synthesis, etc. The UMB-DB has been acquired with a particular focus on facial occlusions, i.e. scarves, hats, hands, eyeglasses and other types of occlusion which can occur in real-world scenarios.	Occlusion/disguise.
52	new VADANA: Vims Appearance Dataset for facial Analysis	The primary use of VADANA is for the problems of face verification and recognition across age progression. The main characteristics of VADANA, which distinguish it from current benchmarks, is the large number of intra-personal pairs (order of 168 thousand); natural variations in pose, expression and illumination; and the rich set of additional meta-data provided along with standard partitions for direct comparison and bench-marking efforts.	Pose, expression and illumination.
53	MORPH Database (Craniofacial Longitudinal Morphological Face Database)	MORPH database is the largest publicly available longitudinal face database. The MORPH database contains 55,000 images of more than 13,000 people within the age ranges of 16 to 77. There is an average of 4 images per individual with the time span between each image being an average of 164 days. This data set was comprised for research on facial analytics and facial recognition.	Age, appearance, expression.

#	Name	Description	Acquisition Setting Variations
54	Long Distance Heterogeneous Face Database (LDHF-DB)	LDHF database contains both visible (VIS) and near-infrared (NIR) face images at distances of 60 m, 100 m and 150 m outdoors and at a 1 m distance indoors. Face images of 100 subjects (70 males and 30 females) were captured; for each subject one image was captured at each distance in daytime and nighttime. All the images of individual subjects are frontal faces without glasses and collected in a single sitting.	Modality, range/distance, resolution.
55	PhotoFace: Face recognition using photometric stereo	This unique 3D face database is amongst the largest currently available, containing 3187 sessions of 453 subjects, captured in two recording periods of approximately six months each. The Photoface device was located in an unsupervised corridor allowing real-world and unconstrained capture. Each session comprises four differently lit color photographs of the subject, from which surface normal and albedo estimations can be calculated (photometric stereo Matlab code implementation included). This allows for many testing scenarios and data fusion modalities. Eleven facial landmarks have been manually located on each session for alignment purposes. Additionally, the Photoface Query Tool is supplied (implemented in Matlab), which allows for subsets of the database to be extracted according to selected metadata e.g. gender, facial hair, pose, expression.	Data modality, gender, disguise, facial hair, pose, and expression.
56	The EURECOM Kinect Face Dataset (EURECOM KFD)	The Dataset consists of multimodal facial images of 52 people (14 females, 38 males) acquired with a Kinect sensor. The data is captured in two sessions at different intervals (of about two weeks). In each session, 9 facial images are collected from each person according to different facial expressions, lighting and occlusion conditions: neutral, smile, open mouth, left profile, right profile, occluded eyes, occluded mouth, side occlusion with a sheet of paper and light on. An RGB color image, a depth map (provided both as a bitmap depth image and a text file containing the original depth levels sensed by Kinect) as well as the associated 3D data are provided for all samples. In addition, the dataset includes 6 manually labeled landmark positions for every face: left eye, right eye, tip of the nose, left side of mouth, right side of mouth and the chin. Other information, such as gender, year of birth, ethnicity, glasses (whether a person wears glasses or not) and the time of each session are also available.	Facial expressions, lighting and occlusion conditions.
57	YouTube Faces Database	The data set contains 3,425 videos of 1,595 different people. All the videos were downloaded from YouTube. An average of 2.15 videos are available for each subject. The shortest clip duration is 48 frames, the longest clip is 6,070 frames, and the average length of a video clip is 181.3 frames. In designing video data set and benchmarks following the example of the 'Labeled Faces in the Wild' LFW image collection. Specifically, the goal is to produce a large scale collection of videos along with labels indicating the identities of a person appearing in each video. In addition, published benchmark tests, intended to measure the performance of video pair-matching techniques on these videos. Finally, the descriptor encodings are provided for the faces appearing in these videos, using well established descriptor methods.	Labeled ground truth.

#	Name	Description	Acquisition Setting Variations
58	YMU (YouTube Makeup) Dataset	The dataset consists of 151 subjects, specifically Caucasian females, from YouTube makeup tutorials. Images of the subjects before and after the application of makeup were captured. There are four shots per subject: two shots before the application of makeup and two shots after the application of makeup. For a few subjects, three shots each before and after the application of makeup were obtained. The makeup in these face images varies from subtle to heavy. The cosmetic alteration is mainly in the ocular area, where the eyes have been accentuated by diverse eye makeup products. Additional changes are on the quality of the skin due to the application of foundation and change in lip color. This dataset includes some variations in expression and pose. The illumination condition is reasonably constant over multiple shots of the same subject. In few cases, the hair style before and after makeup changes drastically.	Ethnicity, makeup disguise, facial expression, pose.
59	VMU (Virtual Makeup) Dataset	The VMU dataset was assembled by synthetically adding makeup to 51 female Caucasian subjects in the FRGC dataset. Makeup was added by using a publicly available tool from Taaz. Three virtual makeovers were created: (a) application of lipstick only; (b) application of eye makeup only; and (c) application of a full makeup consisting of lipstick, foundation, blush and eye makeup. Hence, the assembled dataset contains four images per subject: one before-makeup shot and three after makeup shots	Ethnicity, disguise via makeup.
60	MIW (Makeup in the "wild") Dataset	The MIW dataset contains 125 subjects with 1-2 images per subject. Total number of images is 154 (77 with makeup and 77 without makeup). The images are obtained from the internet and the faces are unconstrained.	Disguise and possibly all other factors.
61	3D Mask Attack Database (3DMAD)	The 3D Mask Attack Database (3DMAD) is a biometric (face) spoofing database. It currently contains 76500 frames of 17 persons, recorded using Kinect for both real-access and spoofing attacks. Each frame consists of: (1) a depth image (640x480 pixels – 1x11 bits); (2) the corresponding RGB image (640x480 pixels – 3x8 bits); (3) manually annotated eye positions (with respect to the RGB image). The data is collected in 3 different sessions for all subjects and for each session 5 videos of 300 frames are captured. The recordings are done under controlled conditions, with frontal-view and neutral expression. The first two sessions are dedicated to the real access samples, in which subjects are recorded with a time delay of ~2 weeks between the acquisitions. In the third session, 3D mask attacks are captured by a single operator (attacker	Pose, facial expression, disguise and occlusion.