THE HONG KONG
POLYTECHNIC UNIVERSITY
香港理工大學
Pao Yue-kong Library
包玉剛圖書館

Copyright Undertaking

This thesis is protected by copyright, with all rights reserved.

By reading and using the thesis, the reader understands and agrees to the following terms:

1. The reader will abide by the rules and legal ordinances governing copyright regarding the use of the thesis.
2. The reader will use the thesis for the purpose of research or private study only and not for distribution or further reproduction or any other purpose.
3. The reader agrees to indemnify and hold the University harmless from and against any loss, damage, cost, liability or expenses arising from copyright infringement or unauthorized usage.

IMPORTANT

If you have reasons to believe that any materials in this thesis are deemed not suitable to be distributed in this form, or a copyright owner having difficulty with the material being included in our database, please contact lbsys@polyu.edu.hk providing details. The Library will look into your claim and consider taking remedial action upon receipt of the written requests.

Pao Yue-kong Library, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

http://www.lib.polyu.edu.hk

FACIAL IMAGE ANALYSIS AND ITS APPLICATIONS TO FACE RECOGNITION

HUILING ZHOU

Ph.D

The Hong Kong Polytechnic University

2017

THE HONG KONG POLYTECHNIC UNIVERSITY

DEPARTMENT OF ELECTRONIC AND INFORMATION ENGINEERING

Facial Image Analysis and Its Applications to Face Recognition

Huiling Zhou

A thesis submitted in partial fulfillment of the requirements

for the degree of Doctor of Philosophy

July 2016

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____(Signed)

Huiling Zhou (Name of student)

Abstract

The human face, which carries various modalities like identity, expression, age, ethnicity, gender, etc., has been more intensely studied in the past few decades than any other part of the human body. Great advances have been made into some of the face-related research areas, and face recognition is one of them, being the task used to identify or verify a person from images and videos. With the development of big data and better computational power, general face recognition methods dealing with real-life variations like poses, expressions, lighting, etc. have achieved superior performances. Latest recognition rates on the most difficult face dataset at present, i.e. the Labeled Faces in the Wild dataset (LFW), have been improved to over 99.5%, which is reported to be even better than human performance. However, face recognition on low-resolution (LR) and aging face images still remains as some of the most challenging problems. On the other hand, currently, only a few research works have investigated high-resolution (HR) face recognition, where more distinctive information on faces (e.g. mark-scale features like moles and scars, and pore-scale features like pores), can be extracted to improve face recognition accuracy. Besides, an efficient scheme for facial feature localization is useful for better face alignment, which is a preprocessing step for all face-related applications. Therefore, the objective of this thesis is to develop efficient methods for facial image analysis and face recognition from low resolution to high resolution, and across ages.

In this thesis, we have first proposed a shape-appearance-correlated Active Appearance Model for facial feature localization, which aims to deal with the poor performance of current methods in generic situations, where the images of a query do not exist in the training set. We first introduce a fast initialization scheme, which retrieves the

most similar faces to a test face in terms of both poses and textures. Based on the idea of locality constraint, these nearest neighbors form a locally linear subspace. Then, the shape and appearance of the selected images are analyzed, and their correlation is maximized by applying Canonical Correlation Analysis (CCA). Our approach can increase the correlation between the principal components learned for face appearances and shapes, as well as the respective projection coefficients. This can improve the convergence speed and the fitting accuracy, while almost no additional computational cost is necessary.

Having aligned face images to be compared based on the facial feature points, more reliable and accurate face recognition can be carried out. In this thesis, we investigate efficient face recognition algorithms for low-resolution, high-resolution, and aging face images. It has been found that, when the resolution of a face image is lower than 24×24 pixels, the performances of existing face recognition methods degrade significantly. One way to solve this problem is to perform image super-resolution, which increases image resolution by recovering the missing high-frequency information in the input LR face images. We have proposed a two-step face super-resolution (also known as face-hallucination) framework based on orthogonal Canonical Correlation Analysis (OCCA) and linear mapping. In the global face reconstruction, the OCCA is applied to increase the correlation between the PCA coefficients of the LR and the HR face pairs. In the residual face compensation, a linear-mapping method is proposed to include both the inter- and intra-information about the manifolds of different resolutions. Both contributions improve the visual quality of super-resolved face images. When these super-resolved face images are used for face recognition, a higher recognition rate can be obtained, compared to that using the original LR face images.

When the resolution of face images increases, it has also been found that the performances of existing face recognition methods have slight improvements. In order to explore more distinctive information from HR faces, we have proposed to use pore-scale facial features for HR face recognition. To compare two HR face images, the pore-scale facial features are detected and extracted from these two images, and matched keypoints between them are established. In our proposed algorithm, the matched keypoints are converted to block matches, which are further aggregated to eliminate the outliers. During face verification, the performance is determined, based on the maximum density of the local, aggregated, matched blocks on the face images. In this way, face verification, based on pore-scale facial features, can deal with the expression and pose variations well, and it has also been proven to be more robust against alignment error.

In the last part of this thesis, we focus on age-invariant face recognition, which is a very challenging task and performances of existing algorithms are still unsatisfactory. We have proposed two methods for face recognition across ages. In our first work, we predict facial features at different ages by linear mapping on the local features. As the facial features of a person at two different ages should be correlated with each other, CCA is applied to the two sets of features, at different ages, to generate more coherent features for face recognition. In our second work, we consider the fact that age progression is quite personalized, so a person may look younger or older than another person, even though their ages are the same. In our algorithm, we have proposed to use the appearance-age labels so that computers can learn more effectively and consistently, than that with the real-age labels. In our method, an identity inference model is designed based on age-subspace learned from appearance ages. We first model human identity and aging variables simultaneously using probabilistic Linear Discriminant Analysis (PLDA). The

aging subspace is learnt independently using the appearance-age labels from a recently proposed aging dataset. Then, the identity subspace is determined iteratively with the Expectation-Maximization (EM) algorithm. In this way, the face recognition becomes simpler as identity inference no longer needs age labels. Besides, different identity features are further combined using CCA, where their correlations are maximized for face recognition.

All the methods proposed in this thesis have been evaluated and compared to existing state-of-the-art methods. Experimental results and analyses show that our algorithms can achieve convincing and consistent performances.

List of Publications

[1]. Dong Li, Huiling Zhou and Kin-Man Lam, "High-resolution face verification using pore-scale facial features," *IEEE transactions on image processing*, *24*(8), pp.2317-2327, 2015.

[2]. Huiling Zhou, Kin-Man Lam and Xiangjian He, "Shape-appearance-correlated active appearance model," *Pattern Recognition*, *56*, pp.88-99, 2016.

[3]. Huiling Zhou and Kin-Man Lam, "Face hallucination using orthogonal canonical correlation analysis," *Journal of Electronic Imaging*, 25(3), pp. 033005-033005, 2016.

[4]. Huiling Zhou and Kin-Man Lam, "Age-invariant face recognition based on identity inference from appearance age," *submitted to IEEE transactions on image processing*, 2016.

[5]. Huiling Zhou, Jiwei Hu and Kin-Man Lam, "Can ambiguous words be helpful in image-understanding systems?" Proceedings, *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC'2013),* pp. 1-4, 2013.

[6]. Ke Sun, Huiling Zhou and Kin-Man Lam, "An adaptive-profile active shape model for facial-feature detection," Proceedings, *IEEE International Conference on Pattern Recognition (ICPR),* pp. 2849-2854, 2014.

[7]. Huiling Zhou, Jiwei Hu and Kin-Man Lam, "Global face reconstruction for face hallucination using orthogonal canonical correlation analysis," Proceedings, *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC'2015),* pp. 537-542, 2015.

[8]. Huiling Zhou, Kwok-Wai Wong and Kin-Man Lam, "Feature-aging for age-invariant face recognition," Proceedings, *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC'2015),* pp. 1161-1165, 2015.

Acknowledgement

First and foremost I want to thank my supervisor Prof. Kin-Man Lam for the continuous support of my Ph.D study and research, for his patience, motivation and immense knowledge. He is not only a mentor, but also a friend who understood me when I felt unsure about this Ph.D journey and encouraged me to get through the tough time. His guidance helps me in all the time of research and his wisdom of life will stay with me for the rest of my life.

My sincere thanks also goes to Prof. Xiangjian He, Prof. Guoping Qiu and Prof. Man-Wai Mak for their extensive discussion and insightful suggestions. They have incented me to widen my research from various perspectives.

I am grateful to all the DSP lab members I worked with in the past four years: Jiwei Hu, Kwok-Wai Wong, Kuong-Hon Pong, Muwei Jian, Dong Li, Chensheng Sun, Hailiang Li, Xiaolong Ma, Wentao Liu, Hao Wu, Kwok-Wai Hung, Junjie Huang, Meng Yao, Khaled W. Aldebei, Mohammed Ambusaidi, Hongbin Zhang, Cigdem Turan, Tingtian Li and Engr Saad Shakeel Chughtai. I am thankful to their share of knowledge and thoughts at work, and the life we have enjoyed together.

Moreover, I feel indebted to many people outside research who always care about me and cheer me up when I feel down. I wouldn't have the chance for Ph.D without Yan Liu and couldn't finish it with joy and peace without Daniel Roxby. Last but not least, I could never go this far in life and grow into an independent and strong girl without my one and only mother, Hui Zhang.

Table of Contents

Abstract i
List of Publications v
Acknowledgement vii
List of Abbreviations xiii
List of Figures xv
List of Tables xix
Chapter 1. Introduction 1
1.1 Research background 1
1.2 Motivation 2
1.3 Statements of originality 3
1.4 Outline of the thesis 5
Chapter 2. Literature review 9
2.1 Review on Face Recognition Algorithms 9
2.1.1 Problem statement 9
2.1.2 Development of face recognition algorithms 12
2.1.3 Review on current challenges of face recognition 17
2.2 Review on model-based facial feature localization 24
2.3 Review on related methods 27
2.3.1 Principal Component Analysis 27
2.3.2 Probabilistic Linear Discriminant Analysis 28
2.3.3 Canonical Correlation Analysis 30
2.3.4 Neighbor embedding 31
2.3.5 Linear Mapping 32
2.3.6 Locality Preserving Projections 33
2.4 Conclusions 34
Chapter 3. Shape-appearance-correlated active appearance model for facial feature localization 35
3.1 Introduction 35
3.2 Shape-appearance-correlated active appearance model 36
3.2.1 Review on active appearance model 36

3.2.2 Proposed model 39
3.3 Experimental results 47
3.3.1 Performance on a controlled dataset 49
3.3.2 Performance on a semi-controlled dataset 50
3.3.3 Performance on an in-the-wild dataset 54
3.3.4 Visual performance on faces in the wild 55
3.4 Conclusions 59
Chapter 4. Low-resolution face recognition based on face hallucination 61
4.1 Introduction 61
4.2 Face hallucination based on orthogonal CCA 64
4.2.1 Global face reconstruction based on orthogonal CCA 64
4.2.2 Residual face compensation using linear mapping 67
4.3 Experimental results 70
4.3.1 Global face reconstruction 71
4.3.2 Comparison of face-hallucination methods 73
4.3.3 Comparison of low-resolution face recognition methods 77
4.3.4 Analysis of the parameter setting 78
4.3.5 Impact of blurring effect 82
4.4 Conclusions 83
Chapter 5. High-resolution face verification using pore-scale facial features 85
5.1 Introduction 85
5.2 Pore-scale facial feature for face verification 88
5.2.1 Pore-scale facial-feature detection 88
5.2.2 Pore-scale facial-feature description 90
5.2.3 Pore-scale facial-feature matching and robust fitting 92
5.2.4 Similarity measurement 97
5.3 Experimental results 98
5.3.1 Preprocessing 99
5.3.2 Face verification with pose variations 101
5.3.3 Face verification under different expressions 104
5.3.4 Face verification on faces captured in different time sessions 105
5.3.5 Face verification under large time span and different expressions 106

5.3.6 Robustness to alignment error 107
5.4 Conclusions 109
Chapter 6. Feature-aging for age-invariant face recognition 111
6.1 Introduction 111
6.2 Age-invariant face recognition based on feature-aging 112
6.2.1 Feature extraction 113
6.2.2 Forward and backward feature-aging 114
6.2.3 Face recognition based on feature-aging 115
6.3 Experimental results 116
6.3.1 Experimental setup 116
6.3.2 Experimental results and analysis 117
6.4 Conclusions 120
Chapter 7. Age-invariant face recognition based on identity inference from appearance age 121
7.1 Introduction 121
7.2 Ageing-guided identity inference model for age-invariant face recognition 122
7.2.1 Identity inference model 123
7.2.2 Independent aging subspace learning 126
7.2.3 Face recognition based on identity inference model 130
7.3 Experimental results 135
7.3.1 Face recognition on the FGNET dataset 136
7.3.2 Face recognition on the MORPH dataset 139
7.3.3 Face verification on the CACD dataset 141
7.4 Conclusions 143
Chapter 8. Conclusions and future work 145
8.1 Summary and conclusions 145
8.2 Future work 147
Reference 149

List of Abbreviations

AAM	Active Appearance Model
ASM	Active Shape Model
AOM	Active Orientation Model
CCA	Canonical Correlation Analysis
CLM	Constrained Local Model
DOG	Difference of Gaussians
EBGM	Elastic Bunch Graph Matching
EER	Equal Error Rate
EM	Expectation-Maximization
FAR	False Acceptance Rate
Fast-SIC	Fast Simultaneous Inverse Composition
FRR	False Rejection Rate
GCCA	Generalized Canonical Correlation Analysis
HMM	Hidden Markov Model
HOG	Histograms of Oriented Gradients
HR	High Resolution
HVS	Human Visual System
ICA	Independent Component Analysis
***K*-NN**	*K* Nearest Neighbors
LBP	Local Binary Pattern
LDA	Linear Discriminant Analysis
LLE	Locally Linear Embedding

LM	Linear Mapping
LPP	Locality Preserving Projection
LR	Low Resolution
NE	Neighbor Embedding
OCCA	Orthogonal Canonical Correlation Analysis
OLPP	Orthogonal Locality Preserving Projection
PCA	Principal Component Analysis
PLDA	Probabilistic Linear Discriminant Analysis
POIC	Project-Out Inverse Compositional
ROC	Receiver Operating Characteristic
SIFT	Scale Invariant Feature Transform
SVM	Support Vector Machine

List of Figures

Fig. 2-1 Typical frame from a surveillance video. 18

Fig. 2-2 Some face examples from the aging dataset FGNET with real-age labels. 22

Fig. 3-1 Comparison of the faces selected by LC-AAM and our proposed scheme. 40

Fig. 3-2 A cropped face for extracting the LBP features. 41

Fig. 3-3 Illustration of the proposed fast initialization scheme. 43

Fig. 3-4 Sample face images from the selected datasets. 47

Fig. 3-5 Fitting results of different methods on the IMM dataset. 49

Fig. 3-6 Fitting results of the different methods on the Bosphorus dataset with pose variations. 51

Fig. 3-7 Fitting results of the different methods on the Bosphorus dataset with expression variations. 52

Fig. 3-8 Visual fitting results and the corresponding mean point-to-point errors of different methods on the IMM dataset and Bosphorus dataset. 53

Fig. 3-9 Fitting results of the different methods on the LFPW dataset in the wild. 54

Fig. 3-10 Visual fitting results and the corresponding mean point-to-point errors of different methods on the PubFig dataset. 56

Fig. 3-11 Enlarged visual fitting results of selected results from Fig. 3-10. 57

Fig. 3-12 Visual fitting results of different methods training on the LFPW dataset and testing on the PubFig dataset. 58

Fig. 3-13 Enlarged visual fitting results of selected results from Fig. 3-12. 59

Fig. 4-1 (a) Original HR training faces, (b) the corresponding hallucinated global faces, and (c) the corresponding residual faces. 67

Fig. 4-2 Illustration of the proposed face-hallucination framework. 70

Fig. 4-3 Global face-reconstruction results. 73

Fig. 4-4 Comparison of the final hallucinated faces based on different methods. 76

Fig. 4-5 Boxplots for different face-hallucination methods. 77

Fig. 4-6 (a) The average PSNRs, and (b) average SSIMs of the reconstructed global faces based on our method under different training-set sizes and numbers of nearest neighbors. 79

Fig. 4-7 Globally reconstructed faces with different training-set sizes N. 79

Fig. 4-8 SSIM (and normalized PSNR) values of final hallucinated faces, based on our method under different training-set sizes and numbers of nearest neighbors. 81

Fig. 4-9 Hallucinated faces reconstructed from LR face images under different amounts of blurring. 83

Fig. 5-1 The face-verification framework based on the PPCASIFT feature. 86

Fig. 5-2 The size of a cropped skin region, whose pore features are to be extracted. 88

Fig. 5-3 Visualization of keypoints on the skin images of 4 distinct subjects. 89

Fig. 5-4 Initial matching results for face images of pose R10 and pose R45. 94

Fig. 5-5 Illustration of the parallel block aggregation scheme. 95

Fig. 5-6 Selection of a local region with the maximized local-matching density. 97

Fig. 5-7 Examples of preprocessed faces from the different datasets. 100

Fig. 5-8 EER of the different methods for face images under different poses. 102

Fig. 5-9 ROC curves of the different methods for face images under all poses. 102

Fig. 5-10 Distance matrices of the PPCASIFT, PSIFT and Gabor+PCA methods. 103

Fig. 5-11 EER of the different methods under different expressions. 104

Fig. 5-12 ROC curves of the different methods under different expressions. 104

Fig. 5-13 EER of the different methods through different sessions.....106
Fig. 5-14 ROC curves of the different methods through different sessions.....106
Fig. 5-15 ROC curves of the different methods through different sessions.....107
Fig. 5-16 Face images with a displacement vector added.....109
Fig. 6-1 Faces used for experiments are cropped and aligned with two eyes, with 53 facial feature points denoted as yellow dots. 112
Fig. 7-1 Some face examples from the aging dataset Chalearn with appearance-age labels. 127
Fig. 7-2 2-D age manifold visualization. 129
Fig. 7-3 Visualization of the identity inference model. 130
Fig. 7-4 The overall framework of the proposed age-invariant face recognition algorithm based on identity inference with independent aging subspace learning. 131
Fig. 7-5 Cropped face for weight LBP feature extraction.....132
Fig. 7-6 Sample face images from the comparison datasets. 135
Fig. 7-7 Some examples of the recognition failure cases on the FGNET dataset.....139
Fig. 7-8 ROC curves for face verification on the CACD-VS dataset. 143

List of Tables

Table 4-1 Global-face reconstruction performances in terms of the means and standard deviations of the PSNR and SSIM of the different methods. 73

Table 4-2 Average recognition rates (%) of different face recognition methods with the LR faces of size 32×32, 24×24 and 16×16 pixels, respectively. 78

Table 5-1 EER(%) of different face-verification methods for all the experiments.....108

Table 5-2 EER(%) of the different face-verification methods with different alignment errors. 109

Table 6-1 Face recognition rates (%) of the Gabor methods and Age-invariant methods with face images from different age groups.....118

Table 6-2 Face recognition rates (%) of different conventional face recognition methods and our proposed method with face images from different age.....119

Table 7-1 The partitioning of ages into groups for the Chalearn dataset.....128

Table 7-2 Statistics of the face aging datasets. 136

Table 7-3 Parameter settings based on the FGNET dataset.....136

Table 7-4 Rank-1 recognition rates on the FGNET dataset.....138

Table 7-5 Rank-1 recognition accuracies on the MORPH dataset. 140

Table 7-6 Verification accuracies on the CACD-VS dataset.....142

Chapter 1. Introduction

The objective of this chapter is to introduce the general research background of human face analysis techniques, from the characteristics of faces to some of the useful applications. Based on the development and some existing problems of the state-of-the-art methods, we discussed our research motivation and original works proposed in this thesis, along with the outline of the content.

1.1 Research background

The human face, which carries various modalities like identity, expression, age, ethnicity, gender, etc., has been more intensely studied in the past few decades than any other part of the human body. The mechanism of the human visual system (HVS) has received attention from philosophers to scientists, and from psychology, neuroscience to engineering. Historia Animalium, which is regarded as the encyclopedic "inventory of renaissance zoology" by Aristotle, mentioned that "straight eyebrows are a sign of softness of disposition; such as curve in towards the nose, of harshness; such as curve out towards the temples, of humor and dissimulation; such as are drawn in towards one another, of jealousy". Face appearance has always been used as a tool to access more sophisticated analysis on discovering intellectual or character qualities of a person. As an early study [10] pointed out, information received from human faces is more useful for people identification, compared to human bodies or gaits. There is a common saying "I am not very good at names, but I never forget a face"; this also suggests the powerful discernment aided by our visual system. Besides, unlike other biometric personal evidences, such as fingerprints, retinas or iris scans, which reply on the participants' cooperation, faces are resources that are easy to obtain and non-intrusive to use, without

need for direct contact. Therefore, this has naturally inspired researchers to develop desirable computer-aided systems which can mimic the HVS to perform face-related tasks.

Nowadays, great efforts have been made into some of the face-related research areas, such as face recognition [11-17], facial expression analysis [18-22], 2D or 3D face modeling [23-27], face animation [28-32] etc. Among them, face recognition is actively researched in the areas of image processing, pattern recognition, neural networks, and computer vision, mainly due to its extensive range of real-life applications related to video surveillance, law enforcement, security control, smart cards, etc. Thus, in this thesis, we mainly consider the techniques for face recognition, including a facial feature localization scheme, which helps with the face alignment before applying recognition. Actually, most of the face-related applications also required the faces concerned to be aligned as a pre-processing step.

1.2 Motivation

Face recognition, as the task to identify or verify a person from images and videos, has been intensively studied for decades. With the development of big data and better computational power, general face recognition methods dealing with real-life variations like poses, expressions, lighting, etc. have achieved superior performances. Latest recognition rates [13, 33, 34] on the most difficult face dataset at present, i.e. the Labeled Faces in the Wild dataset (LFW) [35] have been improved to be over 99.5%, which is reported to be even better than human performance. However, face recognition on low-resolution (LR) [1, 36, 37] and aging face images [38-40] still remains as some of the most challenging problems. On the other hand, currently, only a few research works have investigated high-resolution (HR) face recognition, where more distinctive information

(e.g., mark-scale features like moles and scars, and pore-scale features like pores and hair), can be extracted to improve face recognition accuracy. All these facts have encouraged us to dig deeper into the core of current recognition approaches, and to extend our research scope to more challenging applications. As a result, we have proposed algorithms for recognition or verification of low-resolution, high-resolution, and aging faces. Besides, we have also investigated and devised an efficient facial feature localization scheme for better alignment for preprocessing.

1.3 Statements of originality

The following contributions reported in this thesis are claimed to be original.

a. An Active Appearance Model (AAM) is proposed to localize important facial feature points on human faces, especially under the generic environment, where a probe face is a novel face, unseen in the gallery. Our algorithm consists of a fast face-model initialization scheme to retrieve the nearest neighbors, which have similar poses and textures to a probe face, and an efficient fitting scheme, where the correlation between shape features and appearance features is increased by using the orthogonal Canonical Correlation Analysis (OCCA).

b. A two-step face-hallucination framework for LR face recognition is proposed, which reconstructs a HR version of a face from an input LR face, based on learning from LR-HR example face pairs. The first step is to perform global face reconstruction, where the OCCA is applied to increase the correlation between the Principal Component Analysis (PCA) coefficients of the LR and the HR face pairs. The second step is to perform residual face compensation, where a linear-mapping method,

incorporating both the inter- and intra-information about manifolds of different resolutions, is proposed.

c. A pose-invariant face-verification method is proposed for HR face recognition, which is robust to alignment errors, using the HR information based on pore-scale facial features. A new keypoint descriptor, namely, pore-PCA Scale Invariant Feature Transform (PPCASIFT) – adapted from PCA-SIFT – is devised for the extraction of a compact set of distinctive pore-scale facial features. Having matched the pore-scale features of two-face regions, an effective robust-fitting scheme is proposed for the face-verification task.

d. An age-invariant face recognition framework based on both forward and backward prediction of aging features is proposed. Based on the fact that the age of the query input should be older than that of the gallery face, backward prediction is performed for the Gabor features of the query face image, so that the Gabor features at a younger age are generated. Similarly, the Gabor features of the gallery face image are forward predicted to generate the corresponding features at an older age. Then Canonical Correlation Analysis (CCA) is employed to the two sets of features, at different ages, to generate more coherent features for face recognition.

e. An identity-inference model for age-invariant face recognition, based on independently aging subspace learning is proposed. Firstly, both human identity and aging variables are modeled simultaneously using probabilistic Linear Discriminant Analysis (PLDA). Then, the aging subspace is learnt independently from another dataset with appearance-age labels. After putting the learned aging subspace into the original model, the identity subspace is obtained iteratively with the Expectation-maximization (EM) algorithm. What's more, different representations of identity

features are further combined using CCA, where their correlations are maximized for face matching.

1.4 Outline of the thesis

This thesis is organized into seven chapters and each of them is outlined as follows.

Chapter 2 gives an overview of face recognition and facial feature analysis techniques. The general concepts are first introduced, then the history and development of both tasks are discussed, where the related methods are categorized and compared with each other. Some existing challenges are concluded for each of the tasks, which serve as the driving force for recent research. Furthermore, some subspace learning methods are briefly reviewed, such as Principal Component Analysis (PCA), probabilistic Linear Discriminant Analysis (PLDA), Canonical Correlation Analysis (CCA), Neighbor Embedding (NE), Linear Mapping (LM), and Locality Preserving Projections (LPP). All these methods are also applied to the frameworks proposed in this thesis, which will be described in the subsequent chapters.

Chapter 3 presents an Active Appearance Model (AAM) to localize important facial feature points on human faces, which can provide crucial information about face structure and help with face alignment needed for applications like face recognition and face animation, etc. The proposed framework is able to perform facial-feature localization in the wild, especially under the generic environment where the probe face is an unseen face in the gallery. Firstly, a fast face-model initialization scheme is proposed, based on the idea that the local appearance of feature points can be accurately approximated with locality constraints. Nearest neighbors, which have similar poses and textures to a test face, are retrieved from a training set for constructing the initial face model. To further

improve the fitting of the initial model to the test face, the orthogonal CCA (OCCA) is employed to increase the correlation between shape features and appearance features represented by PCA. With these two contributions, a novel AAM is proposed, namely the shape-appearance-correlated AAM (SAC-AAM), and the optimization is solved by using the recently proposed fast simultaneous inverse compositional (Fast-SIC) algorithm.

Chapter 4 introduces a two-step face-hallucination framework, where a HR version of a face is generated from an input LR face, based on learning from LR-HR example face pairs using OCCA and linear mapping. It can be applied to various real-life tasks such as video surveillance, face recognition, image-based rendering, etc. In the global face reconstruction, face images are first represented using PCA. CCA with the orthogonality property is then employed, to maximize the correlation between the PCA coefficients of the LR and the HR face pairs, so as to improve the hallucination performance. In the residual face compensation, a linear-mapping method is proposed to include both the inter- and intra-information about manifolds of different resolutions.

Chapter 5 presents a pose-invariant face-verification method for HR face recognition, which is robust to alignment errors, using the HR information based on pore-scale facial features. Firstly, the pore-scale facial features are detected and extracted from a query image. Then, initial keypoint matches between the testing image and that of the claimed identity in the gallery are established; the initial keypoint matches are then converted to block matches, which are further aggregated to eliminate the outliers. Finally, the verification result is determined based on the maximum density of the local, aggregated, matched blocks on the face images.

Chapter 6 introduces a face recognition framework across ages. The facial features at different ages are predicted by linear mapping on the local Gabor features. Based on the

fact that the age of the query input should be older than that of the gallery face, both the forward and backward prediction of ageing features are proposed for recognition. Backward prediction is performed for the Gabor features of the query face image, so that the Gabor features at a younger age are generated. Similarly, the Gabor features of the gallery face image are forward predicted to generate the corresponding features at an older age. As the facial features of a person at two different ages should be correlated with each other, CCA is applied to the two sets of features, at different ages, to generate more coherent features for face recognition.

Chapter 7 presents an identity inference model for age-invariant face recognition, where the appearance-age labels, instead of real-age labels, are used. Firstly, both human identity and aging variables are modeled at the same time using PLDA. The aging subspace is learnt independently with the appearance-age labels, and the identity subspace is determined iteratively with the Expectation-Maximization (EM) algorithm. We found that the learned aging subspace is insensitive to the training face images used, and is independent of the identity model. Consequently, the recognition of aging faces becomes simpler as identity inference no longer needs to consider age labels. Besides, different identity features learnt from the identity model are further combined using CCA, where their correlations are maximized for face recognition.

Finally, we give the conclusions of our proposed work in Chapter 8, where some suggestions for further development are also provided.

Chapter 2. Literature review

In this chapter, general concepts and development of both face recognition and facial feature analysis techniques are introduced. The related methods are categorized and compared with each other, while some existing challenges are also discussed. As well, some related subspace learning methods and similarity measurement methods, which are used in the algorithms proposed in this thesis, are also reviewed.

2.1 Review on Face Recognition Algorithms

2.1.1 Problem statement

Face recognition, defined as the task to identity or verify a person from a set of images and videos, is one of the most intensively studied research topics in the past few decades. Compared to other biometrics, such as fingerprint and iris, the human face is easy and non-intrusive to obtain, and can be applied to various real-life applications. It has been widely used in the areas of information security, video surveillance, law enforcement, smart cards, access control systems, and so forth.

In general, face recognition involves two interlinked tasks: (a) face verification, which is a one-to-one matching problem and where the input query face is verified whether it has the same identity as the gallery face, and (b) face identification, which is a one-to-many matching problem and where the identity of the input probe face is obtained by comparing it with a set of gallery images of different individuals. Both tasks are generally performed in three functional modules: (a) face detection, (b) feature extraction, and (c) face recognition. In the face detection stage, the locations and sizes of the faces will be found in a given image. Crucial facial features, such as the eyes, nose, mouth, facial contour, etc., are localized in this process. As the detection result serves as an

important pre-processing step for face recognition later, it must perform well under variations, like illumination, pose, expression, and occlusion. As a result, face detection has been intensively studied in the past decades as an important topic, and many methods have been proposed, from model-based approaches [41-44] to feature-based approaches [45-48]. The next module performs feature extraction from those located facial-feature points of the faces, which are usually normalized, to obtain salient information that can be used for distinguishing faces of different individuals. This process needs to be robust against geometric and photometric variations. The extracted facial features are then used for matching by the face recognition module, where the features are matched against one or more face images stored in the gallery dataset. In this thesis, an efficient facial feature localization scheme is proposed to assist with face detection and landmark localization in Chapter 3, while the rest of the chapters (i.e. Chapter 4 to Chapter 7) mainly focus on the last modules of a face recognition system, where reliable features are extracted and applied to face recognition or verification.

As a pattern-recognition system, face recognition has some measurements for its performance. For face identification, the recognition accuracy, which is the percentage of probe images that are correctly identified, is often used. The false acceptance rate (FAR) and false rejection rate (FRR) are used to evaluate the performance of face verification. An ideal verification system should have both low FAR and FRR, but a practical system needs to make a trade-off, where the equal error rate (EER) between these two rates is applied as a suitable measurement. For a more comprehensive performance evaluation, the Receiver Operating Characteristic (ROC) curve is used for verification, and a cumulative match score (e.g., rank-1 to rank-5 recognition accuracy) for identification task.

In real applications, face recognition systems make use of various source formats ranging from static, controlled photos to uncontrolled video sequences, all of which have been produced under different environments. As a result, they are expected to achieve stable performance and be robust to the possible variations caused by one or more factors as listed below:

Illumination: the appearance of a face can change dramatically under the variation of illumination. It is even found that the difference between two images of the same person under various illumination is greater than that between two images of different persons under the same illumination condition.

Pose: faces captured under various poses are more difficult to be identified due to the fact that most gallery images available in a database are of frontal view only.

Expression: the effect of expression is similar to that of pose, which increases the difference between the probe and the gallery faces.

Occlusion: occlusion effect can lead to the missing of certain facial parts (e.g., eyes caused by sunglasses, lower facial part caused by wearing scarf, etc). This may decrease the discrimination information available from faces and affect the classification process.

Low resolution: faces captured from video surveillance cameras are often of resolution lower than 24×24 pixels. The performances of face recognition will degrade significantly due to the loss of information.

High resolution: with higher resolutions, faces can provide more distinctive information. However, most of the current face recognition methods are devised for normal resolution, and the improvement is only slight with the increase in the resolution.

2.1.2 Development of face recognition algorithms

The earliest work on automatic face recognition dates back to the 1950s in psychology [49] and to the 1960s in engineering literature [50]. But the fully automatic face recognition system emerged in the 1970s [51] and after the seminal work of Kanade [52]. Ever since then, research on face recognition has been intensively studied and improved in areas of psychology, neuroscience and engineering, etc. In this chapter, we give a review on computer-based face recognition methods widely studied in computer vision and pattern recognition. The methods are classified into different approaches, based on the way they identify or verify faces: (a) subspace-based methods; (b) feature-based methods, which use all-inclusive texture features; (c) neural network methods, and (d) other methods such as those exploring correlation between facial features and Support Vector Machine (SVM).

2.1.2.1 Subspace-based methods

Subspace methods generally make use of a set of images as training samples so as to obtain a new coordinate subspace, where face images are projected with lower dimensions, while maintaining maximum variance. All these methods can further be divided into linear subspaces and non-linear subspaces.

Linear-subspace approaches construct the projection feature subspace based on a linear combination of bases, and perform dimension reduction for computational efficiency. Principal Component Analysis (PCA) [12, 53] is one of the earliest methods, and has been widely used in face representation and dimensionality reduction. It is derived from the Karhunen-Loeve's transformation [54] and its objective is to find a lower dimensional subspace whose bases correspond to the maximum variance directions in the

original image space. The corresponding basis images are also called eigenfaces. Later on, Linear Discriminant Analysis (LDA) [55] was proposed, which aims to maximize the between-class scatters of different faces, while minimizing the within-class scatters of the same person when performing recognition. It has shown that LDA can capture more discriminant information than PCA can, while PCA is optimal for reconstruction due to its orthogonal property, and PCA can outperform LDA when the training data set is small [56]. Independent Component Analysis (ICA) [57] is introduced as a generalization of PCA, but allows for better characterization of data and provides more discriminant features with the high-order statistics. Locality Preserving Projections (LPP) [58] is a liner approximation to nonlinear Laplacian Eigenmap [59] which seeks to preserve the local information of the image space. LPP can have more discriminating power than PCA and LDA, while it is less sensitive to outliers. To represent a face image in a better way with the original 2D matrix without vectorization, 2D-PCA [60], 2D-LDA [61] and 2D-LPP [62] were also proposed as variants. However, due to the fact that human face images reside in a high dimensional nonlinear space, recognition methods using linear approximation can only obtain limited performances, especially when large variations appear.

A lot of non-linear manifold methods, which capture complex nonlinearity of face images, have been proposed for face recognition. It has been shown that a face image can become linearly separable if it is nonlinearly projected onto a high-dimensional feature space. One way to achieve this goal is by using kernel-based methods, such as Kernel PCA (KPCA) [63-65], Kernel LDA (KLDA) [66-68] and Kernel LPP (KLPP) [69-71]. The advantage of the kernel-based methods is that it increases the discriminating power of the input data, which is nonlinear in the original space but linear in the feature space.

Another way to analyse nonlinear data is to perform other manifold learning techniques such as Isomap [72], Locally Linear Embedding (LLE) [73, 74] and Laplacian Eigenmap [59] which are proposed to focus on the preservation of local neighbour structure.

2.1.2.2 **Feature-based methods**

Face recognition methods, based on the subspace analysis introduced in the previous section, extract facial information from the holistic face region. As their basic assumption is that each pixel in an image is equally important, they need a higher degree of correlation between the testing and training images. In this way, they may be not only computationally expensive but also ineffective in performance when the training faces have large variations.

Unlike holistic methods, feature-based methods first locate several distinctive facial features, such as the eyes, nose, mouth, etc. Then, the geometric relationship among these fiducial points and their corresponding local visual features are analysed and transformed into a feature vector from the whole face. Standard statistical pattern recognition techniques can then be applied to match these features for recognition purpose. Some of the early work on facial-feature extraction can be found in [75-77]. Later on, statistic models like Active Shape Models (ASMs) [44] and Active Appearance Models (AAMs) [43] are introduced to localize the facial features based on shape or both shape and texture information. They have been intensively studied, and various improvements [41, 78-84] have been proposed to address the real application problems, which will be discussed in detail in Section 2.2. Another well-known feature-based approach is the Elastic Bunch Graph Matching (EBGM) [85-90] method, based on the Dynamic Link Structures [91]. The method first computes the Gabor jets of face images, and then a stack-like structure called a face bunch is combined from the responses of Gabor filters. Recognition of a new face image is performed by comparing its image graph to those of known gallery face

images. EBGM-based methods are proved to be more efficient in preserving local geometry and texture information, but are more computationally expensive.

The main advantage of feature-based methods is that they are more robust to illumination, pose and expression variations, because they mainly rely on the local facial features. Another advantage is that they can produce a more compact representations compared to holistic methods, and achieve higher computational efficiency. However, these approaches heavily depend on facial-feature localization, which may lead to inaccurate and less discriminative recognition.

2.1.2.3 **Neural network methods**

Another non-linear solution to the face recognition problem is given by neural networks. It is a powerful and robust classification technique, which can be used for predicting the unknown data. The earliest work on face image analysis using artificial neural networks dates back to 1990s [92, 93]. Two main neural-network topologies, namely feed-forward neural networks (FNN) [94-97] and recurrent neural networks (RNN) [98-101], are designed to be either single-layered or multi-layered. Based on these topologies, many methods, based on neural networks, have been proposed for face recognition. Examples include deep convolution neural networks [102], PCA with artificial neural networks [103], bilinear CNNs [104], back propagation networks and radial basis function network [105], Gabor wavelet faces with artificial neural networks [106], etc. Among them, the deep convolution neural networks (also known as deep-learning methods) have demonstrated outstanding performances in many different vision tasks, with the fast development in computer power and big-data techniques. They are composed of multiple levels of representations, starting with the raw input to a slightly more abstract higher level. With sufficient transformations, very complex functions can

be learned for classification. As mentioned in Section 1.2, recent performances reported by deep-learning methods [33, 34, 102, 107-109], on the general face recognition tasks, are close to and even better than humans. Deep-learning methods will be further improved and have more success in the future, as it allows little engineering by hand and easier access to data.

2.1.2.4 **Some other face recognition methods**

If we view face recognition as a *K*-class problem, where *K* is the number of distinct subjects in a gallery dataset, SVM-based classifiers are suitable for the task. SVM was first proposed for pattern classification [110, 111], where it finds the hyperplane that separates the largest possible fraction of points of the same class on the same side, while maximizing the distance from the either class to the hyperplane. By modifying the decision surface learned by SVM, a similarity metric between faces can be generated, which is learnt from examples of face differences during training [112, 113].

Correlation-based face recognition often refers to those methods based on template matching. The templates, which represent significant facial features of testing images, are compared with those of the gallery images, returning a vector of matching scores computed by normalized cross correlation [76, 90, 114]. The matching templates can also be allowed to translate, scale, and rotate, so the methods can achieve better recognition accuracy, but are more expensive in computation.

For face recognition, numerous representations of global and local facial image features have been proposed to provide discriminative information. The commonly used features include gray-level intensities [115, 116], Gabor wavelet [89, 117], Local Binary Pattern [11, 118], Histograms of Oriented Gradients (HOG) [119, 120], Hidden Markov Models (HMM) [121, 122], SIFT and SURF descriptors [123-126], and so on. They are

used for different applications, and combined with different recognition frameworks for improved performance. It is also common to apply the features at multi-scales, or to combine the different feature descriptors for recognition, as it has been proven that high-dimensional versions of feature descriptors can achieve significant improvements over their low-dimensional ones [127].

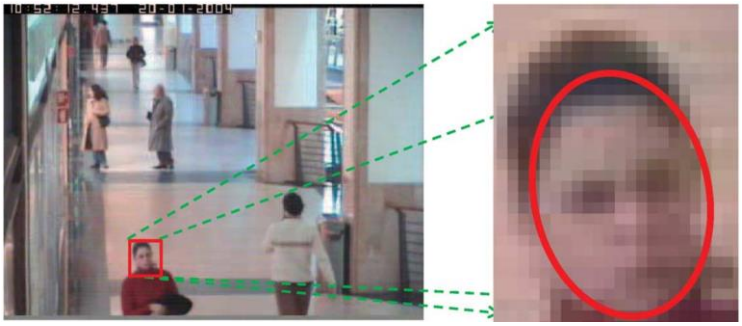
2.1.3 Review on current challenges of face recognition

As introduced in the previous sections of this chapter, general face recognition methods can now achieve near-human performances, even under the variations like illumination, pose, expression, etc. However, it is still challenging for current face recognition systems to deal with low-resolution, high-resolution, and aging images. Thus, in this section, we will review the state-of-the-art methods in these three areas, in particular those low-resolution face recognition methods based on face super-resolution, high-resolution face verification based on facial feature analysis, and age-invariant face recognition based on generative statistic models.

2.1.3.1 Face recognition on low-resolution images

Low-resolution face recognition refers to the process of recognizing faces from small size (usually less than 24×24 pixels) or poor quality images [37], as shown in Fig. 2-1. This happens in long distance video surveillance applications, and is a challenging task. In general, low-resolution face recognition can be categorized into face super-resolution-based methods [1, 128-130] and resolution-robust feature representation methods [36, 131-133]. For face super-resolution (also known as face hallucination) methods, they aim to generate a high-resolution (HR) face image from one or multiple input low-resolution (LR) face images. On the other hand, resolution-robust feature representation methods

attempt to directly extract those features which are invariant to resolution change or to construct the relationship between HR and LR face images in a discriminative subspace for direct comparison and classification. In this review, we focus on the face-hallucination techniques for low-resolution face recognition, due to its long history of development and broader applications, in addition to face recognition.



(a) (b)

Fig. 2-1: Typical frame from a surveillance video: (a) video frame and (b) extracted face region. (Image adopted from paper [1]).

In general, face-hallucination techniques can be divided into two categories: reconstruction-based methods [134-136] and learning-based methods [3-9, 137-140]. The reconstruction-based methods reconstruct a HR image based on its LR counterpart only, without reference to any additional, external information. The learning-based methods explore the correlation between a set of HR face images and their corresponding LR counterparts in the reconstruction of the HR version of a LR face image. Due to the fact that human faces are highly structurally symmetrical, learning-based hallucination methods can generally achieve superior performances to reconstruction-based hallucination methods, especially when the magnification factor is large; say, 4-8 times [140]. Most of the current learning-based face-hallucination methods have two steps: global face reconstruction and residual face compensation. The first step is to model the

global face appearance using subspace methods. However, the global faces, reconstructed in this step, always look blurred and lacked facial details. Therefore, the second step is to compensate for the reconstruction errors [3-5, 7, 138, 139].

It has been shown in a recent survey paper [37] that face-hallucination based recognition methods can achieve better performance when the input LR faces are of smaller sizes (like 24×24 pixels). However, one of their common problems is that they are sensitive to different variations, such as pose and expression. One way to solve this problem is to learn from a large number of training samples at the expense of computational efficiency.

2.1.3.2 Face recognition on high-resolution images

Face recognition methods, which usually represent face images using holistic or local facial features, rely heavily on alignment. Their performances will be greatly degraded when images are under variations in expression or pose, especially when there is one gallery per subject only. With the development of multimedia hardware, such as HDTV and digital cameras, it has become easy to access HR images. This enables us to analyze more sophisticated features, in addition to the traditional facial features like face shape, eyes, nose, and mouth. HR face recognition is a relatively recent topic; it extracts subtle and detailed information, such as mark-scale features (e.g. moles, scars) and pore-scale features (e.g. pores, hair), which contains more distinctive information than LR face images do.

An analysis from macrocosm to microcosm was proposed [141] to solve HR face recognition. This method uses Gabor filters to extract pore-scale features (namely skin texton) to form a texton histogram (a similar idea to bag-of-words [142]). Then, regularized LDA is applied to preserve intrinsic information and reduce interference. This

method based on Gabor-based skin texture can achieve recognition rates of between 38.5% and 57.3%. Thus, the skin texture can be used as an auxiliary feature only. Also, since the method uses LDA, more than one HR face is required for training. This may not be feasible for some real-world applications.

Another HR face-recognition method was proposed based on facial-marker detection [143]. This method uses LoG blob detection for marker extraction after applying the Active Appearance Model (AAM) [43] to detect and remove facial features, such as the eyes, nose, mouth, etc. However, only a very limited quantity of marker-scale features can be extracted from human faces, so the features only complement the traditional methods. Similarly, in [144], facial marks, which are manually annotated by multiple observers, are used as biometric signatures to distinguish between identical twins. That work paid more attention to particular biometric traits, like facial markers, than to the overall facial appearance. However, there is no guarantee that a face image has a sufficient number of traits (e.g. scars, moles, freckles, etc.) for recognition. How to effectively make use of the finer details of human faces remains as a challenging task.

2.1.3.3 **Face recognition on aging images**

With the development of big data and better computational power, general face recognition methods dealing with real-life variations, like poses, expressions, lighting, etc., have achieved superior performances. Latest recognition rates [13, 33, 34, 109] on the most difficult face dataset at present, i.e. Labeled Faces in the Wild dataset (LFW) [35], have been improved to over 99.5%, which is reported to be even better than human performance. However, face recognition, under age progression, still remains as one of the most challenging problems; the best performance, to date, [145] on the most challenging age dataset FGNET [146] stays around 76%. Among the previous related

studies on human aging effects [147-149], it has been widely accepted that facial aging is a complex process, which affects both the shape and texture of a face. In the early growth of a face, from birth to teenager, the greatest change of age progression is in the craniofacial growth (shape change). As people grow older from adulthood to old age, progression of age mainly appears as skin aging (texture change).

There are several reasons why face recognition under age variation is more challenging than other variations: (a) Age progression through life cannot be modeled using a simple progression, as mentioned before; (b) Aging effects are quite specific to different individuals, so it is almost impossible to precisely define the cause of age progression. For example, healthy people who reach their old age will probably look quite different from those who have suffered from accidents or diseases in their lives; (c) Collecting suitable training data for studying the aging effects is also difficult since it requires a much longer time period and greater effort. Aging datasets, collected from photos of different age stages, may undergo more serious distortion than other variations, as shown in Fig. 2-2; and (d) Last but not least, almost all the previous age-related research work is based on datasets where a real-age label is given to each individual. This makes the recognition task by machines incredibly tough, because most of the existing methods can only teach machines to learn from the facial-appearance information. Two people with similar real ages may look very different in appearance, as shown in Figs. 2-2(b) and 2-2(d). It will inevitably make the learning or classification process less accurate.

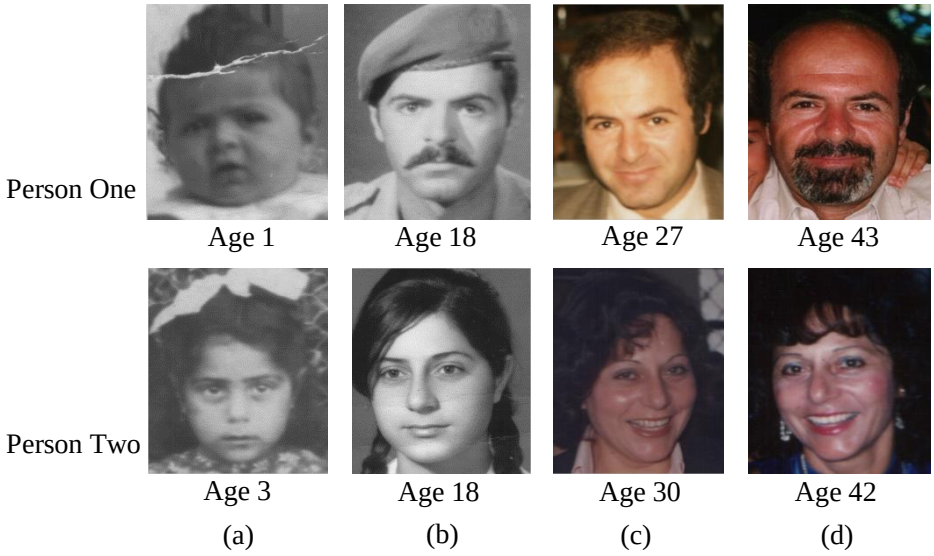


Fig. 2-2. Some face examples from the aging dataset FGNET [6] with real-age labels.

In recent years, many age-related works have been proposed on age estimation [150-154], age simulation [155-158], age-invariant face recognition or verification [39, 40, 145, 159, 160], etc. While they serve different application goals, the underlying theories and methods overlap and correlate extensively. Generally, all these approaches can be categorized into two groups. The first is the generative approaches [40, 154, 158, 160], which construct 2D or 3D generative models to compensate for the aging process, and synthesize face images that match the age of query face images. These methods, however, often rely on strong parametric assumptions and high complexity in computation, and also require sufficient training samples for learning the relationships between a face at two different ages. The second approach is based on discriminative models [39, 159, 161-163], which use robust facial features and discriminative learning methods to reduce the gap between face images captured at different ages. However, lacking the underlying mechanism for capturing facial structures across different ages may limit their generalization performances.

Age estimation and simulation both use similar approaches to age-invariant face recognition tasks. However, age estimation and simulation mainly focus on manipulating the aging information that varies with age progression, while age-invariant face recognition aims to seek the identity information that is stable for the same individual over age progression. This substantial difference inspires a new approach that attempts to separate a face into its aging factor and identity factor [40, 156, 164]. One of the earliest works on face recognition that describes a face with its within-individual and between-individual variations was introduced in [164-166]. Probabilistic Linear Discriminant Analysis (PLDA) [167] was employed to establish a generative linear model, and the optimal latent identity variable was iteratively derived by using the Expectation-Maximization (EM) [168] algorithm. This method was further applied to age-invariant face recognition in [40], where the within-individual variance was suitable for using the aging information, while the between-individual variation was suitable for using the identity information. Again, the EM algorithm is used to obtain both the latent variables simultaneously, and the identity factor is then used for recognition. Experiments showed that this method outperforms other existing methods. Later on, this idea was also applied to render aging faces, by modeling the aging layer as a linear combination of age-progression patterns while keeping the personalized layer invariant through time [156]. All these methods generate the aging subspace and the identity subspace using a single model at the same time. However, this approach has a high demand on the training datasets, because both the identity and the aging information must be learned as thoroughly as possible. Unfortunately, it is a great challenge to obtain suitable datasets for age-invariant face recognition. For the three most well-known datasets for this task, they either suffer from lack of training samples (FGNET dataset [146]) or lack of samples with long time

periods for learning aging patterns (MORPH [169] dataset and CACD dataset [170]). What's worse, all the previous learning frameworks were based on real-age labels, which may be inconsistent with the corresponding appearance ages (people with the same real age may look different in age due to differences in individual skin care or health conditions). This means that the existing methods achieve limited performances on face recognition with age variations.

2.2 Review on model-based facial feature localization

Facial-feature detection and localization is a crucial process for various applications, such as facial-expression recognition, face animation, 3D face reconstruction, etc. In general, facial-feature localization methods can be categorized into model-based methods and texture-based methods. Model-based methods consider a face image and the ensemble of facial landmarks as a whole shape. They learn shape information from labeled training images, and attempt to fit the proper shape to an unknown face in the testing stage. On the other hand, texture-based methods aim to find each facial landmark or local groups of landmarks independently, without the guidance of a model. For a recent complete survey on facial-feature localization techniques, readers can refer to [171].

Among all the competitive techniques, model-based algorithms have been proven to be most effective in automatic facial-information learning [171]. The earliest work of such algorithms includes the deformable template method in [172] and the active contour model in [173]. These approaches aim to extract facial features and locate face boundaries by studying the feature points individually, and hence have limited robustness and accuracy. Most recently, more efficient methods, including the Active Shape Model (ASM) [44] and the Active Appearance Model (AAM) [43], have been proposed. ASM considers the

facial-shape information (based on manually annotated facial-feature points) from a holistic perspective, while AAM also includes texture information (usually in terms of the pixel intensities within a face region). Due to these models' efficiency and accuracy, many variant ASM and AAM methods have been proposed in the past few decades, and they improve the localization performance. However, both ASM and AAM have problems in three different aspects, namely, insufficient robustness to variations, sensitivity to face-model initialization, and poor performance in generic situations. In the following, the challenges in these three aspects and those existing methods, which address these challenges, are discussed.

Insufficient robustness to variations. Since both ASM and AAM rely on global parametric models, they can work well for faces available in a training set with small variations in illumination, pose and expression. However, when these variations become greater, their performances usually degrade dramatically. One way to solve this problem is to integrate ASM and AAM [79, 174]. In [174], a texture-constrained shape model was used to prevent the local-minima problem, and it can achieve a robust performance under illumination variations. In [79], the profile-search step in ASM is changed into a gradient-based optimization problem to more accurately localize feature points. Recently, improved ASM models using 2-D profiles were proposed to achieve pose-adaptive localization [41, 175, 176]. It has been proven that the 2-D profiles can capture more information around each landmark than the original 1-D profiles. By properly setting the initial face model and using an optimization method, these methods can achieve accurate results, and thus have become popular model-based localization methods.

Sensitivity to initialization. In the process of refining the feature-point locations, both ASM and AAM usually perform gradient-descent optimization over a whole face, so

their performances are sensitive to the initial face model. This issue has drawn much attention, and can be improved in two major steps, namely, constructing a more representative initial face model and using a robust feature-point refining scheme. For the first step, several frameworks [177-179] reformulate the original AAM as a sparse representation problem [180] and approximate the local appearance of feature points with locality constraints. After the shape and appearance priors are learned, the *K* nearest neighbors with similar patterns to the test face in terms of pose, expression, etc. are searched from a training set, and are used to model the face in a locally linear sub-space. It has been shown that this pre-processing step helps to reach faster convergence and to obtain better fitting results. Similarly, [176, 181] pre-define the number of face clusters and classify the test face into one of the clusters based on a statistical analysis. For the second step, in order to refine the face model, a stacking strategy is usually employed to search, in series, for a better location for each feature point in the face model iteratively [41, 182, 183].

Poor performance in generic situations. In the survey work of [184], statistical evaluation has shown that person-specific active models (i.e. images of a query also exist in the training set) are both easier to build and more robust to fitting than generic ones (i.e. no images of a query in the training set). To solve the generalization problem, frameworks [185, 186] based on AAM were proposed to learn a discriminative fitting function and establish a mapping between the facial appearance and the face shape in order to improve the alignment accuracy. Unlike AAMs – which model a whole facial region – the family of Constrained Local Models (CLMs) [187-189] extracts templates around each landmark and matches them to new instances of an object using a shape-constrained search and

iterative template generation. This process always relies on the response surfaces generated by fitting the current feature templates using normalized correlation at each point. Recently, an approach which can handle unseen faces and variations was proposed, and is known as the Active Orientation Model (AOM) [81]. It establishes a generative deformable appearance model based on the principal components of images' gradient orientations, and it uses the project-out inverse compositional algorithm to optimize the results. An improved AAM model [78] using more efficient optimization algorithms was also proposed for generic situations.

2.3 Review on related methods

In this section, we will briefly review some techniques that are related to our methods proposed in this thesis, including subspace learning methods and similarity measurement methods for face-image analysis.

2.3.1 Principal Component Analysis

Principal Component Analysis (PCA) is one of the most popular unsupervised techniques for human face representation and recognition. It aims to project the original data onto a lower dimensional, linear feature subspace, which captures the maximum variance. Suppose that there are a set of *N*-dimensional training samples with zero mean, denoted as X_i, where i=1,2,…,M, $X_i \in \boldsymbol{R}^N$ and $\sum_{i=1}^{M} X_i = 0$. The covariance matrix of the input data can be estimated as follows:

$$\Sigma = \frac{1}{M}\sum_{i=1}^{M} X_i X_i^T . \tag{2.1}$$

The PCA solves the following eigen problem:

$$\lambda v = \Sigma v , \tag{2.2}$$

where $\boldsymbol{v}$ are the eigenvectors of Σ, and λ are the corresponding eigenvalues. Normally the first L ($L \square N$) eigenvectors, corresponding to the first L largest eigenvalues, are selected as the basis vectors, which are also known as eigenfaces. These eigenfaces with large eigenvalues represent the global, overall structure of the training images, while the eigenfaces with small eigenvalues represent the local, detail structure. After projection, most of the variance is preserved while the dimension is much lower than the original data. Besides, the sensitivity to local noise is also reduced, which enables PCA to achieve good performance under blurring, partial occlusion, and changes in facial expression. However, when the variations are caused by global components, such as lighting or pose, the performance of PCA will decrease dramatically.

2.3.2 Probabilistic Linear Discriminant Analysis

Linear Discriminant Analysis (LDA), which aims to find the underlying subspace that best discriminates among classes, provide more discriminative class information for facial analysis. The main goal of LDA is to maximize the discrimination between different classes, while minimizing the within class distance. For face recognition, the Fisherface method [55] first projects face data onto a PCA subspace, thus eliminating singularities, and then derives an LDA subspace for representations. It is found that fisherfaces can capture more discriminative information than eigenfaces when there are enough training face images with class labels, and achieve better recognition accuracy. However, according to an early studies [164, 166], the LDA projection obtained can only be used to classify examples of the classes represented in the training data, but not novel classes. To

solve this problem, a probability model is required for this purpose, and the method is known as probabilistic Linear Discriminant Analysis (PLDA).

Suppose that the *n*th image of individual *m* is denoted as $\boldsymbol{x}_{mn}$, then the identity inference model can be presented as follows:

$$\boldsymbol{x}_{mn} = \boldsymbol{\mu} + \mathbf{E}\boldsymbol{u}_m + \mathbf{A}\boldsymbol{v}_{mn} + \boldsymbol{\varepsilon}_{mn}, \tag{2.3}$$

where the first two terms are comprised of the signal components $\boldsymbol{\mu}$ and $\mathbf{E}\boldsymbol{u}_m$, which depend only on the identity of the person, while the last two terms are comprised of the noise components $\mathbf{A}\boldsymbol{v}_{mn}$ and $\boldsymbol{\varepsilon}_{mn}$, which are different for images of the same individual and represent the within-individual noise.

Generally, $\boldsymbol{\mu}$ represents the overall mean of the training set. The matrix **E** is called the between-individual subspace, whose columns are the bases for cross-identity variations, and $\boldsymbol{u}_m$ can be viewed as the position of $\boldsymbol{x}_{mn}$ in this subspace. Similarly, the matrix **A** is the within-individual subspace and $\boldsymbol{v}_{mn}$ is the position in this subspace. The term $\boldsymbol{\varepsilon}_{mn}$ represents the remaining residual noise caused by other variations, and can be modeled as a Gaussian function with diagonal covariance Σ. The goal of establishing this PLDA model is to compute the likelihood that two face images are generated from the same underlying identity factor $\boldsymbol{u}_m$ for recognition.

The models in Eqn. (2.3) can be re-written in terms of conditional probabilities as follows:

$$Pr(\boldsymbol{u}_m) = G_u[0, \mathbf{I}], \tag{2.4}$$

$$Pr(\boldsymbol{v}_{mn}) = G_v[0, \mathbf{I}], \tag{2.5}$$

$$Pr(\boldsymbol{x}_{mn} \mid \boldsymbol{u}_m, \boldsymbol{v}_{mn}) = G_x[\boldsymbol{\mu} + \mathbf{E}\boldsymbol{u}_m + \mathbf{A}\boldsymbol{v}_{mn}, \Sigma], \tag{2.6}$$

where $G_a[\boldsymbol{\beta}, \boldsymbol{\Gamma}]$ is a Gaussian distribution with mean $\boldsymbol{\beta}$ and covariance $\boldsymbol{\Gamma}$. Both the latent variables $\boldsymbol{u}_m$ and $\boldsymbol{v}_{mn}$ are specified with priors as well. The objective of the learning stage is to estimate the parameters $\theta = \{\mathbf{E}, \mathbf{A}, \boldsymbol{\mu}, \Sigma\}$, based on training data $\mathbf{X} = \{\boldsymbol{x}_{mn} \in \boldsymbol{R}^d \mid m = 1, ..., M, n = 1, ..., N_m\}$, where d is the dimension of $\boldsymbol{x}$ and N_m is the total number of images for identity m. Since both the model parameters θ and latent variables are unknown, the problem is solved by using the EM algorithm, where the parameters and variables are jointly estimated until convergence.

2.3.3 Canonical Correlation Analysis

Canonical Correlation Analysis is a learning method which seeks basis vectors for two sets of variables, say $\boldsymbol{x}$ and $\boldsymbol{y}$, such that their projections onto the basis vectors have a maximized correlation. Denote $\mathbf{X}$ and $\mathbf{Y}$ as the matrices whose columns are the sets of variables $\boldsymbol{x}$ and $\boldsymbol{y}$ with zero mean, respectively. Suppose that $\mathbf{M}$ and $\mathbf{N}$ are the respective direction matrices for $\mathbf{X}$ and $\mathbf{Y}$, and the corresponding canonical variate matrices of the projection coefficients are denoted as $\mathbf{U}$ and $\mathbf{V}$, i.e. $\mathbf{U}=\mathbf{M}^T \cdot \mathbf{X}$ and $\mathbf{V} = \mathbf{N}^T \cdot \mathbf{Y}$. Then, CCA maximizes the following correlation:

$$\rho = \frac{E[\mathbf{UV}]}{\sqrt{E[\mathbf{U}^2]\mathrm{E}[\mathbf{V}^2]}} = \frac{\mathbf{M}^T \mathbf{C}_{XY} \mathbf{N}}{\sqrt{\mathbf{M}^T \mathbf{C}_{XX} \mathbf{M} \cdot \mathbf{N}^T \mathbf{C}_{YY} \mathbf{N}}}. \tag{2.7}$$

where T represents the transpose operation; $\mathbf{C}_{XX}$ and $\mathbf{C}_{YY}$ denote the within-set covariance matrices of $\mathbf{X}$ and $\mathbf{Y}$, respectively; and $\mathbf{C}_{YY}$ denotes the covariance matrix of $\mathbf{X}$ and $\mathbf{Y}$. It can be shown that the optimal direction matrices $\mathbf{M}$ and $\mathbf{N}$ are the eigenvectors of $\mathbf{R}_1 = \mathbf{C}_{XX}^{-1} \mathbf{C}_{XY} \mathbf{C}_{YY}^{-1} \mathbf{C}_{YX}$ and $\mathbf{R}_2 = \mathbf{C}_{YY}^{-1} \mathbf{C}_{YX} \mathbf{C}_{XX}^{-1} \mathbf{C}_{XY}$, respectively.

In [190], the original CCA is extended to orthogonal CCA (OCCA). The orthogonality property is crucial for data reconstruction, and can make the PCA projections more consistent. Therefore, in the proposed face super-resolution framework, we also apply OCCA to impose extra constraints on the original CCA. The orthogonal direction matrices **M** and **N** can be computed in an iterative way as follows:

$$\underset{\boldsymbol{m}_k,\boldsymbol{n}_k}{\operatorname{argmax}} \boldsymbol{m}_k^T \mathbf{C}_{XY} \boldsymbol{n}_k$$

$$\text{subject to } \begin{cases} \boldsymbol{m}_1^T \boldsymbol{m}_k = \boldsymbol{m}_2^T \boldsymbol{m}_k = \cdots = \boldsymbol{m}_{k-1}^T \boldsymbol{m}_k = 0, \\ \boldsymbol{n}_1^T \boldsymbol{n}_k = \boldsymbol{n}_2^T \boldsymbol{n}_k = \cdots = \boldsymbol{n}_{k-1}^T \boldsymbol{n}_k = 0, \\ \boldsymbol{m}_k^T \mathbf{C}_{XX} \boldsymbol{m}_k = 1, \\ \boldsymbol{n}_k^T \mathbf{C}_{YY} \boldsymbol{n}_k = 1, \end{cases} \tag{2.8}$$

where $\boldsymbol{m}_k$ and $\boldsymbol{n}_k$ are the *k*th column vector of the direction matrices **M** and **N**, respectively. The first two constraints are designed for the orthogonal property, while the last two are additional constraints on the norm of $\boldsymbol{m}_k$ and $\boldsymbol{n}_k$. The details of deriving the direction vectors can be found in [190]. Having computed the orthogonal-direction matrices, they can be further normalized to become orthonormal.

2.3.4 **Neighbor embedding**

The neighbor embedding technique is originated from manifold learning, which aims to find a low-dimensional space for describing high-dimensional data. Based on the assumption that both low- and high-dimensional data lie on a manifold with similar local structures, Locally Linear Embedding (LLE) [74] has been frequently used in face super-resolution frameworks. The method seeks a vector of weights $\boldsymbol{w}$ for the low-dimensional neighbors to approximate the data under consideration in terms of a linear relationship. This process is presented as a least square problem with sum-to-one constraint as follows:

$$\boldsymbol{w} = \arg\min \left| \boldsymbol{o}_l - \sum_{i=1}^{K} \boldsymbol{w}_i \boldsymbol{o}_{Xi} \right|^2 , \text{ subject to } \sum_{i=1}^{K} \boldsymbol{w}_i = 1, \tag{2.9}$$

where $\boldsymbol{o}_l$ is the patch laying in the low-dimensional manifold represented by its *K* nearest neighbors $\{\boldsymbol{o}_{Xi}\}_{i=1}^{K}$, and the term is the reconstruction error, which is to be minimized. This can be solved efficiently by using the methods described in [73]. Having computed the weight vector $\boldsymbol{w}$, the corresponding patch $\boldsymbol{o}_h$ lying in the high-dimensional manifold can be reconstructed by linearly combining its corresponding *K* nearest neighbors, i.e.

$$\boldsymbol{o}_h = \sum_{i=1}^{K} w_i \boldsymbol{o}_{Yi} \ . \tag{2.10}$$

As discussed in [191], the neighbor embedding methods work well when both the two manifolds share similar structures. However, the weights are learnt solely within one manifold, which are then applied to the other one. It may fail to describe the inter-space relationships between the two separate manifolds, and may lead to a poor performance if there is a large dissimilarity between the manifolds.

2.3.5 **Linear Mapping**

Instead of relying on the assumption of manifold similarity, direct mapping methods seek to find a model, which can project data from a low-dimensional space to a high-dimensional space, and vice versa:

$$f : \boldsymbol{o}_l \rightarrow \boldsymbol{o}_h . \tag{2.11}$$

For facial images, the mapping can be complex and non-linear, but approximating it by a simple linear mapping function still shows promising results in natural image super-resolution [191]. Normally, the linear-mapping function can be represented as:

$$\boldsymbol{o}_h = \boldsymbol{\Psi} \boldsymbol{o}_l , \tag{2.12}$$

where $\boldsymbol{\Psi}$ is a linear operator and can be computed as follows:

$$\boldsymbol{\Psi} = \boldsymbol{o}_h \boldsymbol{o}_l^T (\boldsymbol{o}_l \boldsymbol{o}_l^T + \lambda I)^{-1}, \tag{2.13}$$

where λ is a regularization parameter.

Previous work [192] has shown that predicting the HR Gabor features from LR ones directly, using linear mapping, produces a better performance on LR face recognition, compared to that using the Gabor features from super-resolved face images.

2.3.6 Locality Preserving Projections

It has been shown in [153, 193] that Locality Preserving Projection (LPP) [70] and its orthogonal variant OLPP [194] are able to project faces onto a more discriminative subspace, and characterize the age manifold better than PCA and LLE.

LPP aims to preserve local structure based on the assumption that a nearest-neighbor search in the low-dimensional space will yield similar results to that in the high-dimensional space. In the LPP theory, the objective function is defined as follows:

$$\sum_{ij} (\boldsymbol{x}_i - \boldsymbol{x}_j)^2 w_{ij} . \tag{2.14}$$

The weight w_{ij} is based on the heat kernel, and is defined as $w_{ij} = \exp(-\left\| \boldsymbol{x}_i - \boldsymbol{x}_j \right\|^2 / t)$ when the two face features $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are the K nearest neighbors of each other, otherwise $w_{ij} = 0$. The weight matrix $\mathbf{W}=[w_{ij}]$ is symmetric. A diagonal matrix $\mathbf{D}=[d_{ij}]$, whose entries are the column sums of $\mathbf{W}$, i.e. $d_{ii} = \sum_j w_{ij}$, and the corresponding Laplacian matrix $\mathbf{L}=\mathbf{D}-\mathbf{W}$, can be computed. Then, the optimal projections can be obtained by solving the following eigenproblem:

$$\mathbf{X}\mathbf{L}\mathbf{X}^T \boldsymbol{a} = \lambda \mathbf{X}\mathbf{D}\mathbf{X}^T \boldsymbol{a}, \tag{2.15}$$

where the solutions are the column vectors $\{\boldsymbol{a}_0,...,\boldsymbol{a}_n\}$, which are the eigenvectors of $(\mathbf{XDX}^T)^{-1}\mathbf{XLX}^T$, with their eigenvalues in ascending order, i.e. $\lambda_0 < ... < \lambda_n$.

2.4 Conclusions

This chapter serves as a survey of the principles and development of face recognition and facial feature analysis techniques. Some well-known face analysis techniques have also been reviewed, such as PCA, PLDA, CCA, Neighbor Embedding, Linear Mapping and LPP. In the following chapters, the methods proposed in this thesis will be presented, and compared to some of the state-of-the-art methods presented in this chapter.

Chapter 3. Shape-appearance-correlated active appearance model for facial feature localization

3.1 Introduction

As discussed in some survey work [171, 195, 196], Active Appearance Model (AAM) takes advantage of all grey-level information across faces to build a convincing model with a relatively small number of landmarks, while Active Shape Model (ASM) is a special case of AAM. Therefore, in this chapter, we focus on establishing a shape-appearance-correlated AAM (SAC-AAM) framework to tackle the three challenges described in Section 2.2 at the same time, especially under a generic localization environment where the images of a query do not exist in the training set.

The contributions of this framework are given as follows. In order to fulfill the goals, we first propose a fast initialization scheme, which retrieves the most similar faces to a test face in terms of both poses and textures. Based on the idea of locality constraint, these nearest neighbors form a locally linear subspace. Then, the shape and appearance of the selected images are analyzed, and their correlation is maximized by applying Canonical Correlation Analysis (CCA) [197]. Actually, the orthogonal CCA (OCCA) [190] is employed in our framework due to its superior data reconstruction property. We will show that our approach can increase the correlation between the principal components learned for face appearances and shapes, as well as the respective projection coefficients. This can improve the convergence speed and the fitting accuracy, while almost no additional computational cost will be added. By conducting experiments on different face datasets and comparing our proposed framework with state-of-the-art model-based methods, experimental results show that our framework can achieve a great improvement in terms

of fitting accuracy, especially for faces under large pose, expression, and occlusion variations, as well as for unseen faces.

This chapter is organized as follows. Section 3.2 reviews the active appearance models and presents the proposed SAC-AAM framework; the details of generating initial face models and obtaining more correlated principal components are described. Experiment results and analysis are given in Section 3.3, and conclusions are provided in Section 3.4.

3.2 Shape-appearance-correlated active appearance model

3.2.1 Review on active appearance model

As mentioned in the previous section, unlike ASM – which only deals with shape information – AAM also takes texture information into consideration. The shape vector is usually presented by concatenating the position coordinates of labeled landmarks, while texture is modeled in terms of the demeaned pixel intensities or colors within the convex hull of a facial shape. When given a training set of face images with corresponding labeled landmarks, the shape model is established from $2N$ fiducial points denoted as $\boldsymbol{s} = (x_1, y_1, x_2, y_2, ..., x_N, y_N)^T$. The shapes are normalized by using the Procrustes analysis [198], which is a commonly used method to align shapes to a common coordinate system (usually, the mean shape of the training objects). Then, the Principal Component Analysis (PCA) is applied to project the normalized and aligned shapes onto the shape subspace. Thus, the shape instance $\boldsymbol{s}$ can be presented as a linear combination of principal shapes as follows:

$\hat{\boldsymbol{s}} = \bar{\boldsymbol{s}} + \mathbf{P}_s \cdot \boldsymbol{\alpha}$, and (3.1)

$$\boldsymbol{\alpha} = \mathbf{P}_s^T (\boldsymbol{s} - \bar{\boldsymbol{s}}), \tag{3.2}$$

where $\bar{\boldsymbol{s}}$ is the mean shape, $\mathbf{P}_s$ is the matrix whose columns form a set of orthonormal base vectors, and the weight vector $\boldsymbol{\alpha}$ (also known as projection parameters) is used to control the shape variations.

The appearance model of a face image $\boldsymbol{I}$ is learned by first warping it into a "shape-free" model, usually the mean shape $\bar{\boldsymbol{s}}$. This is represented as a warping function $W(\boldsymbol{x};\boldsymbol{\alpha})$, where $\boldsymbol{x}$ denotes a set of pixels inside the mean shape $\bar{\boldsymbol{s}}$. Then, PCA is again applied to project the "shape-free" appearance of the image $I(W(\boldsymbol{x};\boldsymbol{\alpha}))$ on to the appearance subspace. The appearance instance $\boldsymbol{r}$ can be represented as a linear combination of principal appearances as follows:

$$\hat{\boldsymbol{r}} = \bar{\boldsymbol{r}} + \mathbf{P}_r \cdot \boldsymbol{\beta}, \text{ and} \tag{3.3}$$

$$\boldsymbol{\beta} = \mathbf{P}_r^T (\boldsymbol{r} - \bar{\boldsymbol{r}}), \tag{3.4}$$

where $\bar{\boldsymbol{r}}$ is the mean appearance, $\mathbf{P}_r$ is the matrix whose columns form a set of orthonormal base vectors, and the weight vector $\boldsymbol{\beta}$ is used to control the appearance variations. It should be noted that, in this chapter, we focus on the AAM, which models the shape and appearance information independently, rather than combining shape and appearance with a single set of linear parameters as in [199].

With an appropriate initialized face, the fitting process for AAM aims to find the optimal shape and appearance parameters, which minimize the discrepancy between the synthesized image and the observed facial image. Various cost functions and optimization algorithms have been proposed to estimate $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, among which the l_2-norm error

minimization and the Inverse Compositional (IC-AAM) algorithm [199] are widely used, represented as follows:

$$\{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0\} = \arg\min_{\{\boldsymbol{\alpha},\boldsymbol{\beta}\}} \left\| \boldsymbol{I}(W(\boldsymbol{x};\boldsymbol{\alpha})) - \bar{\boldsymbol{r}} - \mathbf{P}_r \cdot \boldsymbol{\beta} \right\|^2. \tag{3.5}$$

As discussed in a current work named Locality-constrained AAM (LC-AAM) [179], conventional AAMs assume a linear relationship across a whole data set, which is not always held, especially under a large variation of pose and expression. One efficient way to solve this problem is to explore the local linear subspace by modeling AAM as a sparsity-regularized problem. In [179], the original sparsity problem is approximated by adding locality constraints, as follows:

$$\{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0\} \approx \arg\min_{\{\boldsymbol{\alpha},\boldsymbol{\beta}\}} \left\{ \sum_{\boldsymbol{x}\in\bar{\boldsymbol{s}}} \left[\boldsymbol{I}(W(\boldsymbol{x};\boldsymbol{\alpha})) - \sum_{i=1}^{K} \beta_i \cdot \mathbf{P}_{ri} \right]^2 + \lambda_1 \left\| \mathrm{d}\square\ \boldsymbol{\alpha} \right\|^2 + \lambda_2 \left\| \mathrm{d}\square\ \boldsymbol{\beta} \right\|^2 \right\}, \tag{3.6}$$

where the synthesized appearance image is represented as a linear combination of all the training faces, λ_1 and λ_2 are the regularization coefficients, and $\left\| \mathrm{d}\square\ \bullet \right\|^2$ denotes the distances between the input image and the respective appearance bases. In practice, Eqn. (3.6) can be computed efficiently by directly selecting the *K* nearest neighbors of the input face image to form the shape and appearance bases, as shown in Eqn. (3.7). With a smaller but similar training dataset, LC-AAM transforms the original non-linear problem into a locally linear one, and utilizes the popular project-out inverse compositional algorithm [199] to solve the optimization problem.

$$\{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0\} \approx \arg\min_{\{\boldsymbol{\alpha}_K,\boldsymbol{\beta}_K\}} \left\{ \boldsymbol{I}(W(\boldsymbol{x};\boldsymbol{\alpha}_K)) - \bar{\boldsymbol{r}} - \mathbf{P}_r \cdot \boldsymbol{\beta}_K \right\}. \tag{3.7}$$

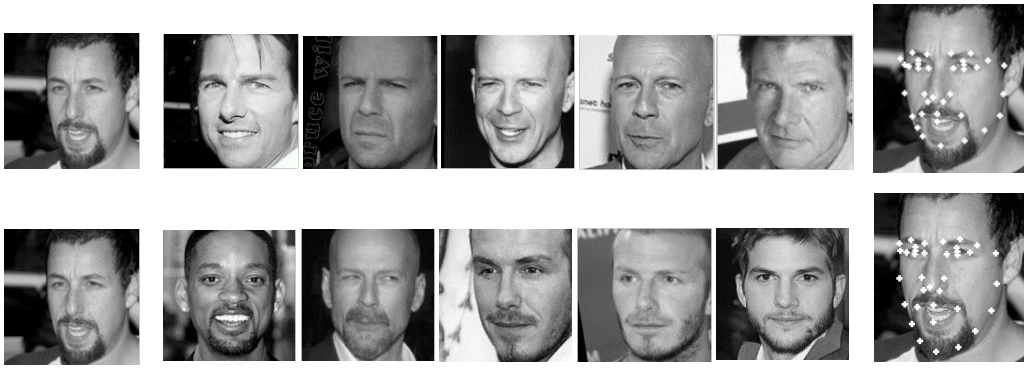
This approach can achieve a good performance on face images with pose and expression variations when images of the same subject are included in a training dataset

(i.e. in a person-specific environment). However, if no images of the query face exist in the training set (i.e. in a generic environment) and the query face is partially occluded by facial hair, the performance of fitting the initial model to the query face deteriorates dramatically, as shown in Fig. 3-1(a). In our proposed framework, the sample faces, to be used to form the initial face, are selected by a weighted *K*-nearest neighbor (*K*-NN) searching scheme, which considers both pose and texture information. Compared to LC-AAM, the faces selected using our approach not only have similar poses, but also have similar facial textures (in particular, in the lower part of a face, e.g. the mouth and chin areas) to the ones in a query face. Giving a testing face, the corresponding top five similar faces selected by the method in LC-AAM and by using our proposed initialization scheme are shown in Fig. 3-1(b). Those faces selected by LC-AAM have similar poses to the query, but the appearance around the mouth area is different. Using our proposed scheme, the selected faces have greater similarity around the mouth areas. Hence, they can provide more useful information for learning the correlation between the face shape and the complex texture around mouth regions. Furthermore, some of the selected faces still have similar poses to the query. Consequently, our proposed scheme can improve the learning and the correlation of the principal components of the shape and appearance information. This can improve the fitting results by avoiding being trapped in local minima, as shown in Fig. 3-1(c).

3.2.2 **Proposed model**

In this section, we will present our proposed Shape-Appearance-Correlated Active Appearance Model (SAC-AAM) in detail. Our method follows the concept in the previous work shown in [179], which reformulates the conventional AAM as a sparsity-regularized

AAM problem. However, in this chapter, we propose a more efficient initialization scheme to approximate sparsity regularization by retrieving the *K* nearest neighbors in terms of both pose and texture. Then, OCCA is employed to enhance the correlation between the shape features and the appearance features represented by using PCA. We will show that this can generate more correlated principal components for the shape and the appearance features, which allows optimization to be solved efficiently by using the fast simultaneous inverse compositional algorithm.



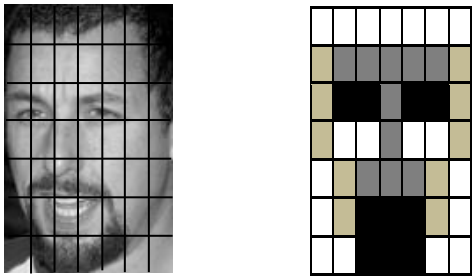
(a) (b) (c)

Fig. 3-1. Comparison of the faces selected by LC-AAM and our proposed scheme: (a) an input face is cropped and normalized based on the position of the two eyes; (b) the faces in the upper row are selected by LC-AAM, which only exhibit similar poses to the input face, while – as shown in the lower row – the faces selected using our scheme have similar poses and texture appearance to the input face; and (c) the final fitting results based on LC-AAM (upper row) and on our proposed scheme (bottom row).

3.2.2.1 **Efficient face-model initialization scheme**

In the literature of face detection, recognition and facial-expression analysis, various types of facial features are employed. In our proposed framework, we use two efficient and effective features, namely the Histogram of Oriented Gradients (HOG) [120] and

Local Binary Patterns (LBP) [11], for searching example face images with a similar pose and texture appearance to the query face, respectively.



(a) (b)

Fig. 3-2. (a) A cropped face partitioned into 7×7 windows for extracting the LBP features. (b) The weights used in the Chi square distance measure, where black, dark grey, light grey, and white represent the weights of 4, 3, 2, and 1, respectively.

Each face image is cropped to the size of 160×160 and normalized based on the positions of two eyes' centers [197]. Then, a *K*-NN search based on the HOG features and Euclidean distance is used to select the most similar faces from a dataset. As shown in Fig. 3-1.(b), the *K* nearest neighbors have similar poses to the test face, due to the fact that HOG captures the edges' orientations and hence an object's shape. However, retrieving faces with similar poses only is insufficient, because some parts of a test face image may be occluded by facial hair or hair shading, in particular when the test face has no images in the training dataset. In order to achieve a more efficient and accurate subspace learning based on the selected samples, the weighted LBP features are also considered in the search, which aims to select faces having a similar texture to the test face. In the search, faces are cropped and normalized in the same way as the training faces, and are also divided into 7×7 windows, which can achieve the best performance by experiment. Because each of the windows has a different degree of importance, different weights are set for them, as shown in Fig. 3-2.

With the LBP feature histogram for each block of a face image, the weighted Chi square distance is used to measure the similarity between the test face and all the faces in the training dataset as follows:

$$\chi_w^2(f,g) = \sum_{j,i} w_j \frac{(f_{i,j} - g_{i,j})^2}{f_{i,j} + g_{i,j}}, \tag{3.8}$$

where f and g are the normalized histograms of the test and training face images, respectively; i and j are the indices representing the ith bin in the histogram of the jth block; and w_j is the weight predefined for block j.

Fig. 3-1(b) shows the top five faces selected from the training dataset using the LBP feature. We can see that the selected face images have a similar appearance around the mouth regions. However, these selected faces may have poses that are different from the test face. Having retrieved the similar-pose faces and similar-texture faces using the HOG and LBP features, respectively, the mean shape of the similar-pose faces is computed. In order to use the similar-texture faces more efficiently in learning, they are wrapped to the mean shape by using Procrustes warp. In this way, the initial face model is more similar to the test face in terms of shape and appearance, and thence helps to establish a more locally linear subspace for representation. It should also be noted that, for normal faces without large occlusion or pose variation, our initialization scheme does not affect the efficiency and can achieve a slightly better performance compared to using either one of the two features to search the dataset, as shown in Fig. 3-10. The overall initialization scheme is illustrated in Fig. 3-3.

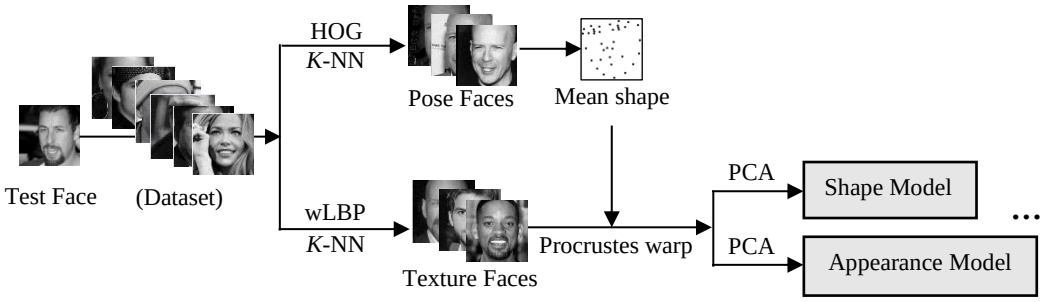


Fig. 3-3. Illustration of the proposed fast initialization scheme (FIS).

In experiments, we have found that using only the top five pose faces and the top twenty texture faces is sufficient to achieve a good performance. More experimental results will be shown in subsequent sections.

3.2.2.2 **Orthogonal CCA for SAC-AAM**

With the *K* texture faces selected from training samples (after wrapping to the mean shape of the retrieved pose faces), PCA is applied to both the shape matrix $\mathbf{S} = [\boldsymbol{s}_1, \boldsymbol{s}_2, ..., \boldsymbol{s}_K]$ and the appearance matrix $\mathbf{R} = [\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_K]$, where the columns of the matrices are the landmark coordinates and grey-level intensities, respectively, within the shape hull of the respective training faces. We compute the mean shape vector $\bar{s}$ and the mean appearance vector $\bar{\boldsymbol{r}}$. Then, matrices $\mathbf{P}_s$ and $\mathbf{P}_r$ are composed of the orthonormal eigenvectors of the shape and appearance training vectors, respectively. The corresponding projection coefficients of the shape and appearance vectors are denoted by $\mathbf{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, ..., \boldsymbol{a}_K] \in \boldsymbol{R}^{m \times K}$ and $\mathbf{B} = [\boldsymbol{b}_1, \boldsymbol{b}_2, ..., \boldsymbol{b}_K] \in \boldsymbol{R}^{n \times K}$. 95% of the total energy of both the shape and appearance information is retained. The number of eigenvectors used for shape and appearance are denoted as m and n, respectively, of which both are smaller than *K* (*m* and *n* are usually less than 10, while *K* is set at 20 in our algorithms). Similar to Eqn. (3.2) and Eqn. (3.4), the projection coefficients can be computed by:

$\boldsymbol{a}_i = \mathbf{P}_s^T(\boldsymbol{s}_i - \bar{\boldsymbol{s}})$, and

$$\boldsymbol{b}_i = \mathbf{P}_r^T(\boldsymbol{r}_i - \bar{\boldsymbol{r}}). \quad (3.9)$$

Since the shape and appearance information of a person possesses an intrinsic correlation, it can be explored and enhanced by applying OCCA to the demeaned coefficients matrices $\hat{\mathbf{A}} = [\hat{\boldsymbol{a}}_1, \hat{\boldsymbol{a}}_2, ..., \hat{\boldsymbol{a}}_K]$ and $\hat{\mathbf{B}} = [\hat{\boldsymbol{b}}_1, \hat{\boldsymbol{b}}_2, ..., \hat{\boldsymbol{b}}_K]$. In fact, it is proven that this is equivalent to applying OCCA to matrices **A** and **B** which are already demeaned. By solving the optimization problem in Eqn. (2.8) in Section 2.3.3, we can obtain two projection matrices with orthonormal column vectors $\mathbf{W}_s$ and $\mathbf{W}_r$, and two canonical variate matrices $\mathbf{C}_s = \mathbf{W}_s^T\mathbf{A}$ and $\mathbf{C}_r = \mathbf{W}_r^T\mathbf{B}$, where the correlation coefficient $\rho = \frac{E[\mathbf{C}_s\mathbf{C}_r]}{\sqrt{E(\mathbf{C}_s^2)E(\mathbf{C}_r^2)}}$ is maximized. Then, we rewrite $\mathbf{C}_s$ and $\mathbf{C}_r$ as follows:

$$\mathbf{C}_s = \mathbf{W}_s^T\mathbf{A} = \mathbf{W}_s^T\mathbf{P}_s^T(\mathbf{S} - \bar{\mathbf{S}}) = \tilde{\mathbf{P}}_s^T\hat{\mathbf{S}} \text{ and}$$

$$\mathbf{C}_r = \mathbf{W}_r^T\mathbf{B} = \mathbf{W}_r^T\mathbf{P}_r^T(\mathbf{R} - \bar{\mathbf{R}}) = \tilde{\mathbf{P}}_r^T\hat{\mathbf{R}}, \quad (3.10)$$

where $\hat{\mathbf{S}} = \mathbf{S} - \bar{\mathbf{S}}$ and $\hat{\mathbf{R}} = \mathbf{R} - \bar{\mathbf{R}}$ are the demeaned shape and appearance matrices, respectively, and $\tilde{\mathbf{P}}_s = \mathbf{P}_s\mathbf{W}_s$ and $\tilde{\mathbf{P}}_r = \mathbf{P}_r\mathbf{W}_r$ are the corresponding eigen-matrices after the OCCA transformation.

As shown in Eqn. (3.10), the multiplication of an original eigen-matrix ($\mathbf{P}_s$ or $\mathbf{P}_r$) and the corresponding OCCA projection matrix ($\mathbf{W}_s$ or $\mathbf{W}_r$) forms a new eigen-matrix ($\tilde{\mathbf{P}}_s$ or $\tilde{\mathbf{P}}_r$), which can project the shape or appearance vector on to a more correlated subspace. In addition, these two new eigen-matrices are orthonormal, which can be proven as follows:

$\tilde{\mathbf{P}}_s^T \tilde{\mathbf{P}}_s = (\mathbf{P}_s \mathbf{W}_s)^T (\mathbf{P}_s \mathbf{W}_s) = \mathbf{I}$ and

$$\tilde{\mathbf{P}}_r^T \tilde{\mathbf{P}}_r = (\mathbf{P}_r \mathbf{W}_r)^T (\mathbf{P}_r \mathbf{W}_r) = \mathbf{I}, \tag{3.11}$$

where **I** is an identity matrix. Therefore, the new eigen-matrices are applied in the model-fitting process of the feature points. Since the matrices of PCA projection coefficients **A** and **B** are both small, applying OCCA will only increase the computational cost slightly, but can improve the accuracy of final fitting as shown in the experimental results.

3.2.2.3 **Fitting scheme for SAC-AAM**

Fitting an AAM usually involves estimating the model parameters so that the distance between the model instance and the given image is minimized. Typically, this process is presented as the optimization of a least-square problem, as shown in Eqn. (3.5). In our framework, with training faces resembling the test face in terms of shape and appearance and being used for initialization, and the use of more correlated eigen-matrices and the corresponding projection coefficients, we refine the optimization in Eqn. (3.6) with far fewer shape and appearance eigenvectors. In the related work [179], the optimization problem is solved by using the project-out inverse compositional (POIC) algorithm. However, as illustrated in [78], the POIC algorithm is efficient but does not work well for unseen variations, so it is unsuitable for generic situations. In contrast to POIC, the simultaneous inverse compositional (SIC) algorithm [200] has been proven to perform robustly in the case of generic fitting but is extremely complex computationally. To tackle this problem, the fast simultaneous inverse compositional (Fast-SIC) algorithm is employed to achieve relatively accurate fitting results while greatly reducing the computation time. Instead of concatenating the warped shape parameters and appearance parameters, and optimizing them as a whole, Fast-SIC first optimizes the fitting with

respect to the appearance parameters, and the solution is then used for optimization with respect to the warped parameters in each iteration. The cost required is slightly more than that of POIC, which is only an approximation to Fast-SIC (and hence to SIC), but achieves better fitting results.

Algorithm 3-1: Fast-SIC for SAC-AAM

Pre-compute:

(3) Evaluate the gradients $\nabla \bar{\boldsymbol{r}}$ and $\nabla \mathbf{P}_{ri}$ for $i = 1, \ldots, K$

(4) Evaluate the Jacobian $\frac{\partial W}{\partial \boldsymbol{\alpha}_K}$ at $(\boldsymbol{x};0)$

Iterate:

(1) Warp $\boldsymbol{I}$ with $W(\boldsymbol{x};\boldsymbol{\alpha}_K)$ to compute $\boldsymbol{I}(W(\boldsymbol{x};\boldsymbol{\alpha}_K))$

(2) Compute the error image $E_{Fsic}(\boldsymbol{x}) = \bar{\boldsymbol{r}} + \tilde{\mathbf{P}}_r \cdot \boldsymbol{\beta}_K - \boldsymbol{I}(W(\boldsymbol{x};\boldsymbol{\alpha}_K))$

(5) Compute the steepest descent image $\mathbf{J} = \nabla \bar{\boldsymbol{r}} \frac{\partial W}{\partial \boldsymbol{\alpha}_K}$

(6) Project out appearance from $\mathbf{J}$ to obtain $\mathbf{J}_{Fsic} = \boldsymbol{\alpha}_K[\mathbf{P}_{rx}\boldsymbol{\beta}'_K, \mathbf{P}_{ry}\boldsymbol{\beta}'_K]\frac{\partial W}{\partial \boldsymbol{\alpha}_K}$

(7) Compute the Hessian Matrix $\mathbf{H}_{Fsic} = \mathbf{J}_{Fsic}^T \mathbf{J}_{Fsic}$ and invert it

(8) Compute $\mathbf{J}_{Fsic}^T E_{Fsic}(\boldsymbol{x})$

(9) Compute $\Delta\boldsymbol{\alpha}_K = \mathbf{H}_{Fsic}^{-1} \mathbf{J}_{Fsic}^T E_{Fsic}(\boldsymbol{x})$

(10) Update $W(\boldsymbol{x};\boldsymbol{\alpha}_K) \leftarrow W(\boldsymbol{x};\boldsymbol{\alpha}_K) \circ W(\boldsymbol{x};\Delta\boldsymbol{\alpha}_K)^{-1}$ and $\boldsymbol{\beta}_K \leftarrow \boldsymbol{\beta}_K + \Delta\boldsymbol{\beta}_K$

Until $\|\Delta\boldsymbol{\alpha}_K\| \le \varepsilon$

In our algorithm, we also fit our model into the Fast-SIC optimization framework which firstly linearizes the appearance model and then projects it out. However, in contrast to Fast-SIC, which directly uses the raw pixel intensities as the features without applying any priors, our algorithm can show a further improvement on the fitting performance. Another advantage of our algorithm is that we solve the standard generic AAM dilemma (the number of appearance parameters is at least one order of magnitude greater than the

number of shape parameters, i.e. $n \square m$ for matrices **A** and **B**) using a simple fitting process, where both the numbers of shape and appearance parameters are small, and also smaller than the number of nearest neighbors. We refine the fitting model as in Eqn. (3.12):

$$\{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0\} \approx \arg \min_{\{\boldsymbol{\alpha}_K, \boldsymbol{\beta}_K\}} \left\{ \boldsymbol{I}(W(\boldsymbol{x}; \boldsymbol{\alpha}_K)) - \overline{\boldsymbol{r}} - \tilde{\mathbf{P}}_r \cdot \boldsymbol{\beta}_K \right\}, \tag{3.12}$$

where $\boldsymbol{\alpha}_K$ and $\boldsymbol{\beta}_K$ are the PCA projection parameters in the more correlated shape and appearance eigen-spaces constructed by using the retrieved *K* nearest face neighbors. The Fast-SIC algorithm used in our proposed fitting model is summarized in Algorithm 3-1. More details of Fast-SIC can be found in [78].

(a) (b) (c) (d) (e)

Fig. 3-4. Sample face images from the selected datasets: (a) IMM dataset, (b) Bosphorus dataset with pose variations, (c) Bosphorus dataset with expression variations, (d) LFPW dataset, and (e) PubFig dataset.

3.3 Experimental results

To evaluate the performance of our proposed generic AAM framework, we compare it with several state-of-the-art methods on different datasets, namely the IMM dataset [201]

under controlled variations of pose and expression, the Bosphorus dataset [202] with cropped face images under semi-controlled variations of pose and expression, and the labeled faces in the wild datasets LFPW [187] and PubFig [203] with uncontrolled variations of pose and expression, as well as with occlusion. Some faces of these datasets are shown in Fig. 3-4. All experiments were conducted under Matlab R2010b environment on an Intel i7 3.5 GHz CPU with 16GB RAM PC.

Numerous measurement metrics have been proposed for different face analysis methods, such as the ROC curves for face recognition and the F1-score, precision, and recall for face classification and retrieval, etc. To measure the performances of facial feature localization, we use the ground-truth-based localization error, as in [171, 176], which is the point-to-point error (PtP Error), normalized by the eye distance. Given the ground-truth landmarks, the localization error e_i^k is computed as follows:

$$e_i^k = \frac{d[(x_i^k, y_i^k), (\tilde{x}_i^k, \tilde{y}_i^k)]}{IOD}, \tag{3.13}$$

where $d(.,.)$ is the Euclidean distance between the *k*th landmark of the ground-truth face and the corresponding detected landmark of the *i*th test face; and *IOD* is the Inter-Ocular Distance, which is the distance between the two eye pupils. According to the measurement in [171], $e_i^k < 0.1$ can be taken as an acceptable error criterion under a controlled environment. In other words, a landmark is considered to be detected correctly if its normalized error is below the threshold. In our framework, we employ the cumulative curve corresponding to the percentage of test images for which the mean localization error of all the landmarks (also called normalized root-mean-squared error (NRMSE)) is less than a specified threshold. In the following subsections, we will elaborate on the

experimental setup on each dataset, and compare our proposed method with several state-of-the-art methods both statistically and visually.

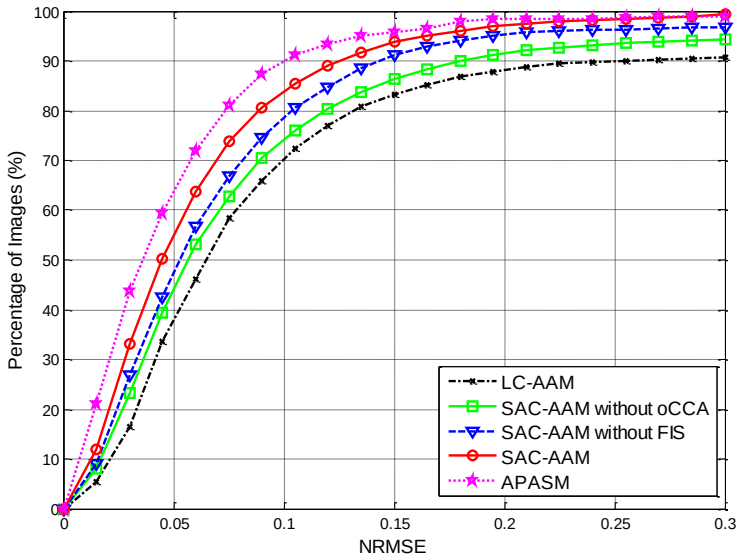


Fig. 3-5. Fitting results of different methods on the IMM dataset.

3.3.1 **Performance on a controlled dataset**

In this experiment, 156 gray-scale face images of 39 distinct subjects in the IMM dataset [201] are selected. Each subject is sized 640×480 and has 4 images with neutral-frontal, smiling-frontal, neutral-left, and neutral-right views, respectively. We use the re-annotated faces with 58 landmarks similar to our previous work [176]. Since it is a relatively small and simple dataset, we select one subject for testing and others for training each time. We mainly examine the efficiency of our proposed SAC-AAM framework together with each of the contributions of our proposed framework, i.e. the fast initialization scheme (denoted by SAC-AAM without OCCA) and using OCCA to increase the correlation between shape and appearance (denoted as SAC-AAM without FIS). We compare them with the recent locality-constrained AAM (LC-AAM) [179] and the adaptive-profile ASM (APASM) [176]. For the IMM dataset, we retrieve the top five

pose faces and twenty texture faces using the *K*-NN search for our proposed SAC-AAM, and the top twenty nearest neighbors for LC-AAM as described in [179]. The experimental setup is the same for all the methods compared, and the experiment results are presented in Fig. 3-5 in terms of the cumulative curves.

From the results, we can see that by using the proposed face-model initialization scheme and OCCA to improve the correlation of appearance and shape, our proposed method can achieve a more accurate fitting performance than LC-AAM. Each of the contributions can make some improvements to our proposed framework, and our SAC-AAM achieves detection accuracy of higher than 80 percent when the error criterion equals 0.1, and is at least 10 percent higher than LC-AAM. However, for the IMM dataset, our previous work, APASM, achieves the best performance because it is based on the ASM model which locally searches for the best position for each landmark and works well under controlled environments. Nevertheless, it is the slowest among the methods compared. For our proposed SAC-AAM, it takes 10-12s to process and localize each query image in the IMM dataset. LC-AAM requires a similar runtime, but APASM needs about 20s.

3.3.2 Performance on a semi-controlled dataset

To further examine each of the contributions of our proposed framework, and to compare them with the methods mentioned in Section 3.3.1, the Bosphorus dataset [202] is employed. It contains the high-resolution images of 105 people with larger variations in pose and expression than the IMM dataset, as illustrated in Figs. 3-4(b) and 3-4(c). These face images have been cropped to include only the face, such that the landmarks on their face contours cannot be localized. We further divide the dataset into faces with pose

variations and faces with expression variations. For those faces with pose variations, each subject has four poses: frontal, right10, right20, and right30, respectively, with a total of 32 landmarks. For those with expression variations, each subject has five different expressions: angry, happy, disgusted, surprised, and eye-closed, with a total of 22 landmarks. The image size is reduced to 280×340 for faster computation. The same experimental settings and parameter selection are used as in Section 3.3.1, and one subject is selected for testing, while the others are selected for training each time. Fig. 3-6 and Fig. 3-7 show the corresponding cumulative curves with pose variations and expression variations.

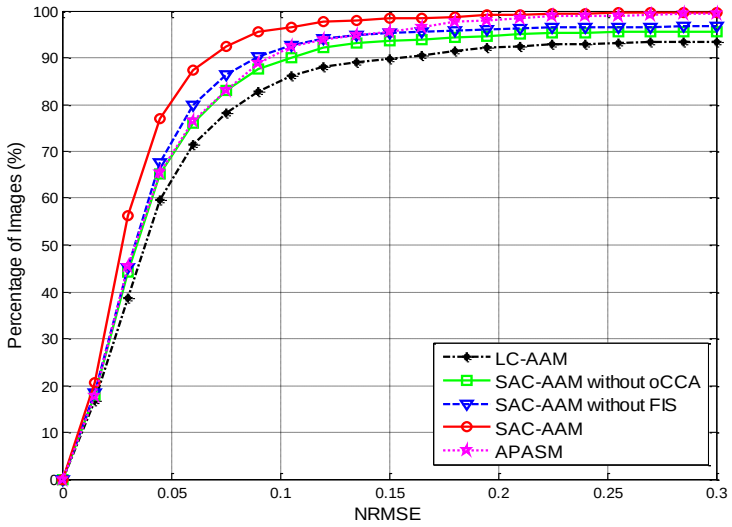


Fig. 3-6. Fitting results of the different methods on the Bosphorus dataset with pose variations.

From the results, we can see that refining the initial face model, by adding local constraints, can improve the overall performance of the AAM models on the cropped face images from the semi-controlled dataset. Even if those points lying along the face contour are excluded, our proposed AAM framework with each contribution achieves better performance than LC-AAM and our previous work APASM, with about 5-10 percent higher in detection accuracy. For APASM, we can also observe that it deals with pose

variations better than expression variations due to the use of the HOG feature being able to select training images with similar poses. Compared with our proposed AAM framework, which increases the correlation between shape and texture under pose and expression variations, when the variations become larger, the performance of ASM-based methods deteriorates because they determine the final location of each feature point separately. The average runtime of the proposed SAC-AAM method on the Bosphorus dataset is about 5.5s. To give a better illustration, some of the visual fitting results based on APASM, LC-AAM, and our proposed SAC-AAM on the IMM and Bosphrus datasets are shown in Fig. 3-8.

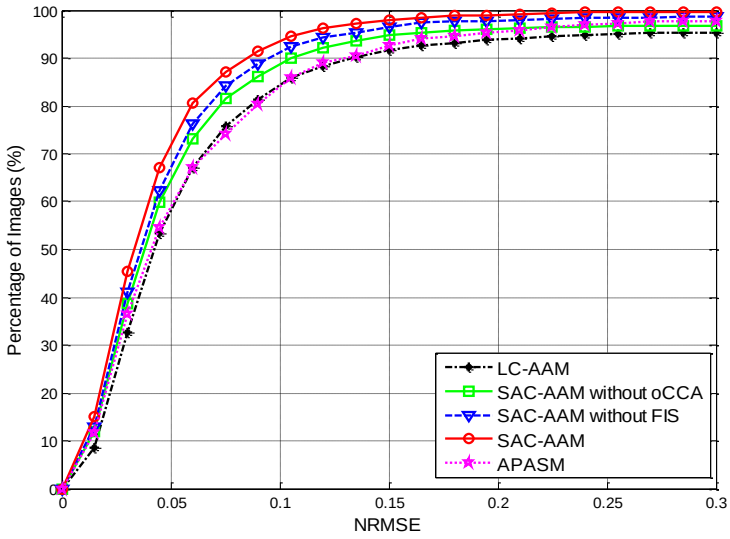


Fig. 3-7. Fitting results of different methods on the Bosphorus dataset with expression variations.

For each individual face image, we calculate the mean point-to-point error (MPtP error) between the estimated landmarks and the ground-true landmarks for all the feature points. This measurement shows the overall localization performance in a straightforward way. Although all the methods can achieve a good performance on this semi-controlled dataset, we can still observe an obvious improvement using our proposed SAC-AAM

framework compared to LC-AAM, as highlighted by the yellow circles and the corresponding enlarged regions shown in the yellow rectangles.

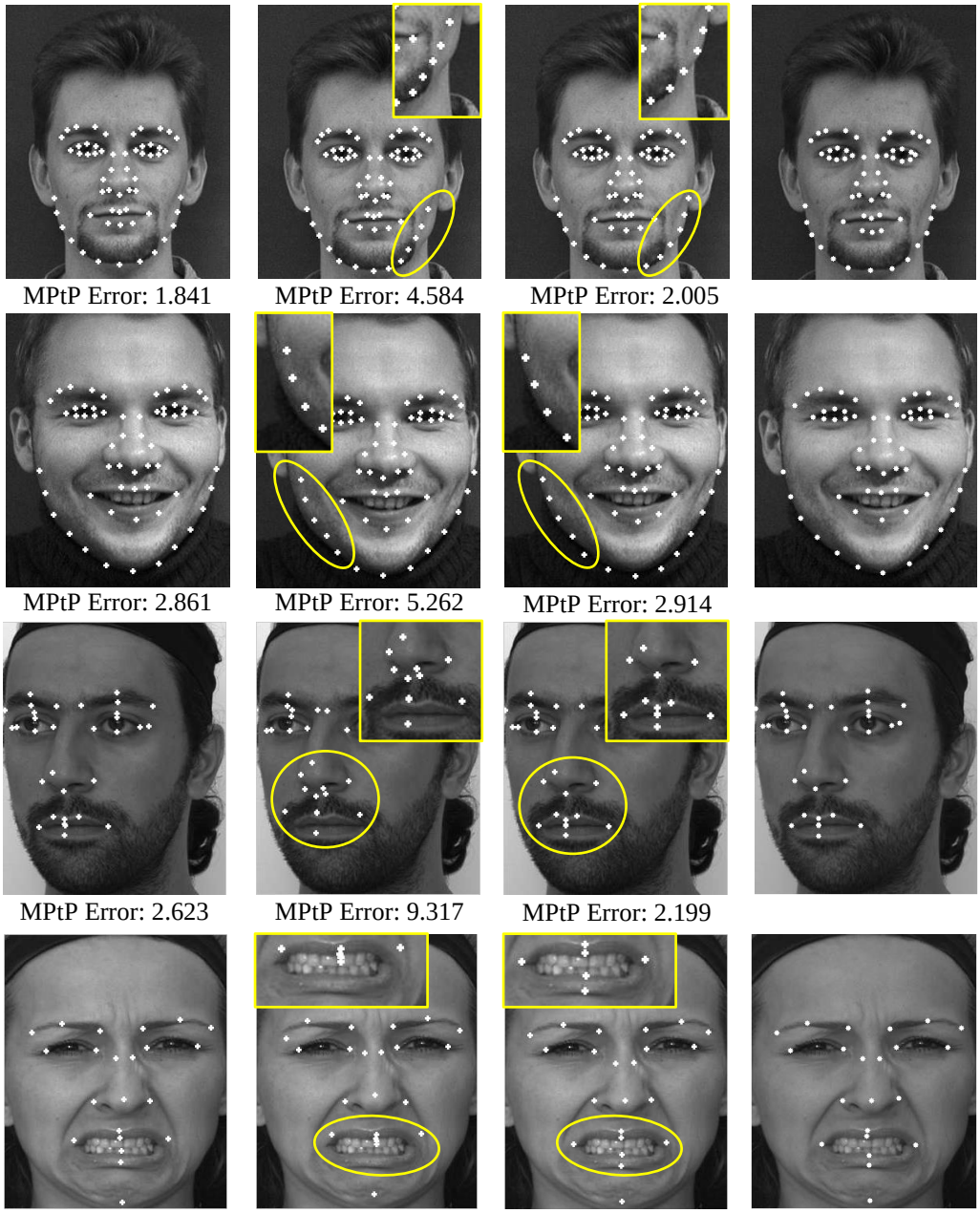


Fig. 3-8. Visual fitting results and the corresponding mean point-to-point errors of different methods on the IMM dataset (the first two rows) and Bosphorus dataset (the last two rows): (a) APASM, (b) LC-AAM, (c) our proposed SAC-AAM framework, and (d) the corresponding face image with ground-true landmarks.

3.3.3 **Performance on an in-the-wild dataset**

Nowadays, with the rapid improvement of facial-feature localization techniques, as well as the availability of new face datasets, the ultimate goal of recently proposed methods is to localize facial points accurately on faces in the wild, especially under unseen variations. Therefore, in this experiment our proposed AAM framework is evaluated on a famous in-the-wild dataset, namely the re-annotated LFPW dataset [204]. To have a fair evaluation, we compare our method (together with each of the contributions) with several state-of-the-art AAM methods, namely LC-AAM, Active Orientation Models (AOMs) [81], and AAM with fast simultaneous inverse compositional algorithm (AAM-FSIC) [78].

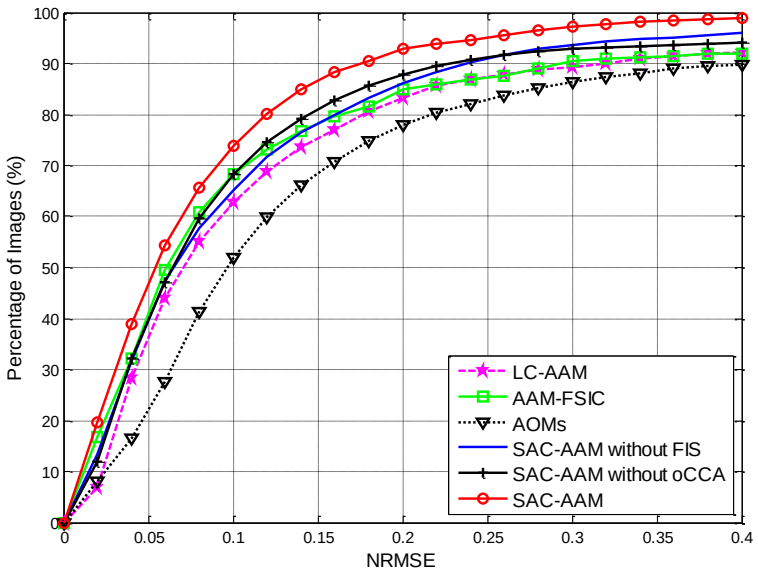


Fig. 3-9. Fitting results of the different methods on the LFPW dataset in the wild.

The re-annotated LFPW dataset is an improved version of the original LFPW dataset [204], where each face is labeled with 68 points. All images in the dataset were downloaded from the web with large variations in pose, expression, and lighting conditions, as shown in Fig. 3-4(c). The resolutions of the images also vary hugely from 100×100 to 400×400. We select 800 face images for training and 222 face images for

testing, without the subjects in the training set included in the testing set. The fitting results of the different methods are shown in Fig. 3-9. We can observe that the performance of all methods decreases with the in-the-wild faces, while our methods, together with each contribution, achieve superior results compared to other state-of-art methods, with an average of 10 percent higher detection accuracy.

3.3.4 **Visual performance on faces in the wild**

In this section, we conducted two experiments based on face images in the wild. The fitting results are illustrated visually, based on the PubFig dataset [203]. This dataset is similar to the LFPW dataset, but each face image has 36 feature points.

In the first experiment, the PubFig dataset was used for both training and testing. Because our framework is based on LC-AAM, we visually compare the fitting results based on LC-AAM, SAC-AAM with one of the two contributions, i.e. SAC-AAM without OCCA and SAC-AAM without FIS, and SAC-AAM with both OCCA and FIS. Some selected fitting results, as well as their corresponding mean point-to-point errors (MPtP errors), are shown in Fig. 3-10.

We can observe that, for those generic cases under simple pose and illumination variations (shown in the first two rows), all methods can work well. However, for the faces with strong variations in illumination or with occlusion, our proposed SAC-AAM framework, as well as SAC-AAM with one of the two contributions, achieves much better performance, which are highlighted with yellow circles in Fig. 3-10.

For better visualization, we have also illustrated in Fig. 3-11 the improvements of the fitting results by enlarging the regions marked by the yellow circles in Fig. 3-10. With both of the proposed contributions, our SAC-AAM method can achieve much better

localization performances in the occluded mouth regions, facial contours, and eye regions, where most existing AAM methods cannot achieve accurate results.

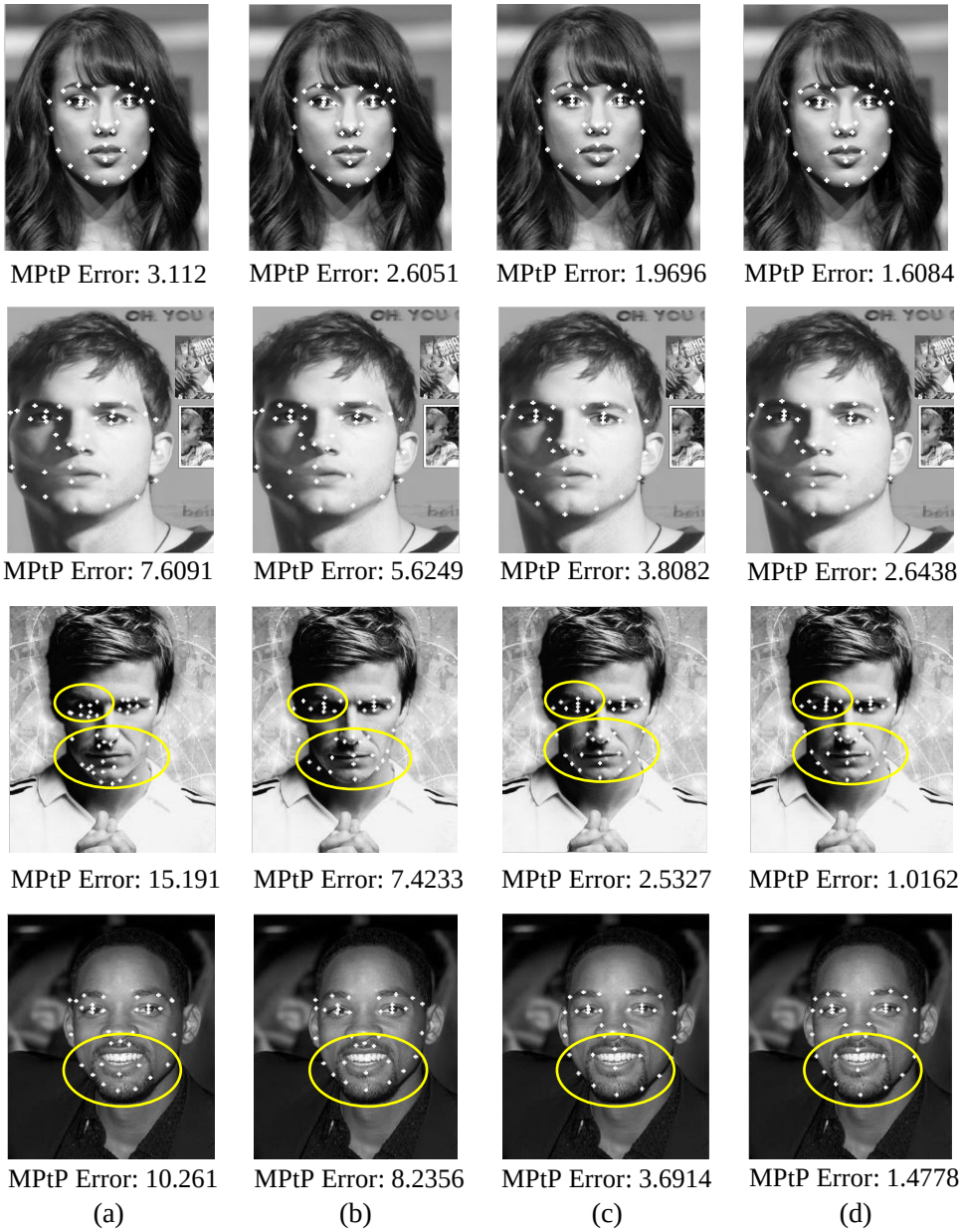
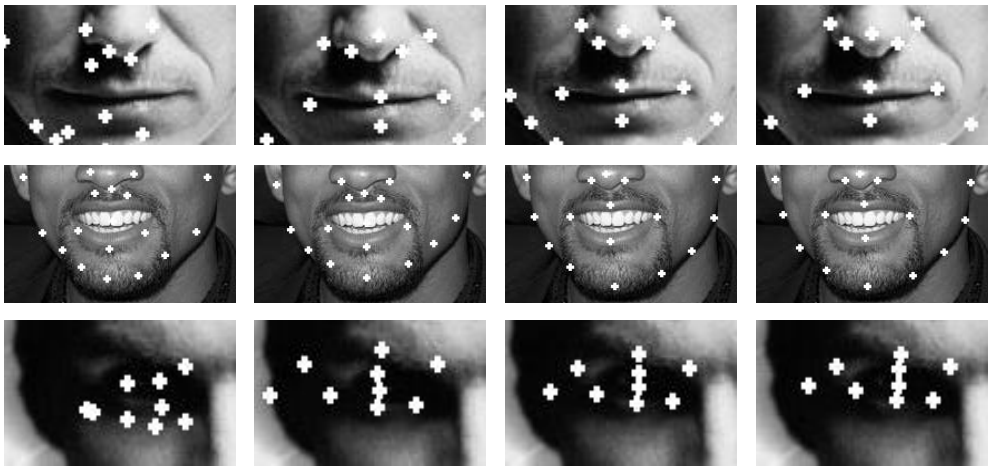


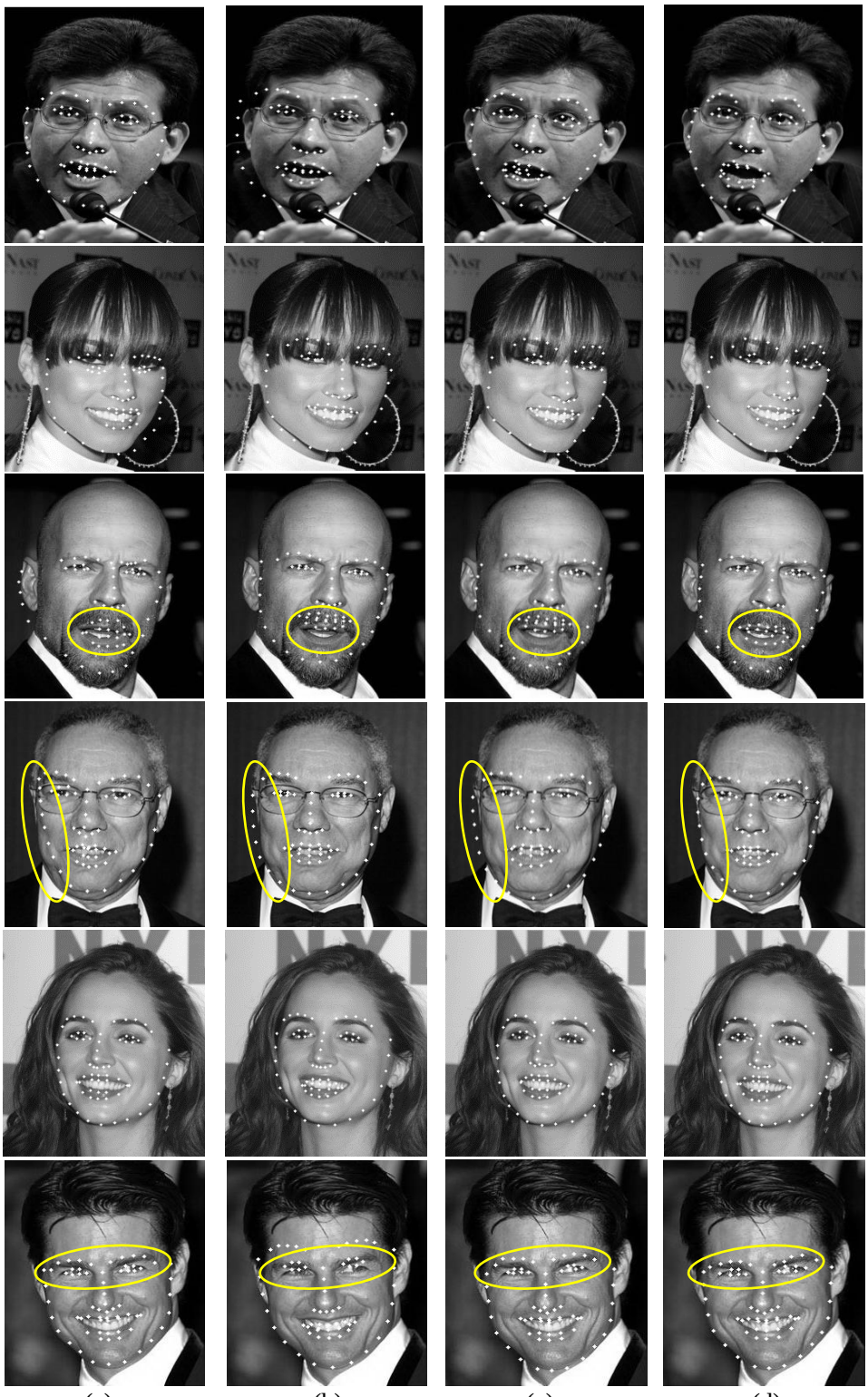
Fig. 3-10. Visual fitting results and the corresponding mean point-to-point errors of different methods on the PubFig dataset: (a) LC-AAM, (b) SAC-AAM without OCCA, (c) SAC-AAM without fast initialization scheme, and (d) our proposed SAC-AAM framework.



(a) (b) (c) (d)

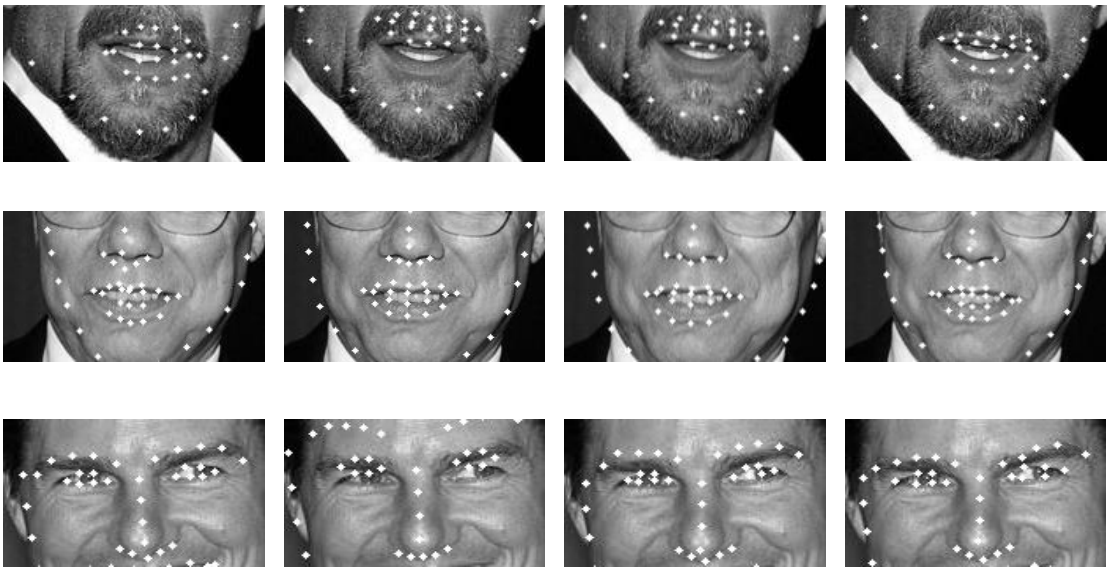
Fig. 3-11. Enlarged visual fitting results of selected results from Fig. 3-10. The first row is the mouth region, the second row is the face contour, and the third row is the eye region: (a) LC-AAM, (b) SAC-AAM without OCCA, (c) SAC-AAM without fast initialization scheme, and (d) our proposed SAC-AAM.

In the second experiment, we evaluated the generalization capability of different methods using training and testing data from two different datasets. Similar to Section 3.3.3, we compare our proposed SAC-AAM framework with LC-AAM, AOMs, and AAM-FSIC. For AOMs, the source code provided uses the Multi-PIE dataset as the training set. For the other methods, the LFPW dataset is the training dataset, while PubFig is the testing dataset. Some visual results are illustrated in Fig. 3-12, and some highlighted regions (e.g. the mouth region, facial contour, and eye region) are also enlarged and illustrated in Fig. 3-13 for better visualization. Though the MPtP errors are not available as PubFig dataset does not have landmark information, it still can be observed that our proposed SAC-AAM again generalizes better for unseen faces than other recent AAM variants.



(a) (b) (c) (d)

Fig. 3-12. Visual fitting results of different methods training on the LFPW dataset and testing on the PubFig dataset: (a) LC-AAM, (b) AOMs, (c) AAM-FSIC, and (d) our proposed SAC-AAM.



(a) (b) (c) (d)

Fig. 3-13. Enlarged visual fitting results of selected results from Fig. 3-12. The first row is the mouth region with mustache, the second row is the face contour, and the third row is the eye region: (a) LC-AAM, (b) AOMs, (c) AAM-FSIC, and (d) our proposed SAC-AAM.

3.4 **Conclusions**

In this chapter, we have proposed a shape-appearance-correlated Active Appearance Model (SAC-AAM) for generic facial-feature localization. Based on the idea of approximating the local appearance of feature points with locality constraints to improve face-model initialization, we have proposed an efficient initialization scheme which retrieves *K* nearest neighbors from a training set with similar poses and textures to a test face. With a small number of representative samples, the correlation between the shape and the appearance models can be learned more efficiently and this can better represent the test face images. To further improve the fitting performance of AAM, we have applied OCCA to increase the correlation between the shape features and the appearance features represented by PCA. With these two main contributions, we have devised our AAM model

and solved the optimization using the recently proposed fast simultaneous inverse compositional (Fast-SIC) algorithm. With only a small number of training images selected for learning and the fast optimization algorithm used, our proposed framework is efficient and accurate. Experimental results on different datasets have shown better performances, in terms of the statistical and visual performances, achieved by our proposed framework, as well as each of the two contributions, i.e. fast initialization scheme and OCCA to increase correlation between shape and appearance. The fitting results have also demonstrated that our method can achieve superior performance compared to other state-of-the-art AAM models, especially under generic environments.

Chapter 4. Low-resolution face recognition based on face hallucination

4.1 Introduction

As mentioned in Section 2.1.3.1, the most efficient way to establish a learning-based face-hallucination model is using both global reconstruction and detail compensation. In the global-face reconstruction phase, a subspace representation based on Principal Component Analysis (PCA) [139], Kernel PCA [138], Locality Preserving Projection (LPP) [7], etc. is usually applied. However, directly applying these subspace transformations will lead to substantial differences between the projected low-resolution (LR) and high-resolution (HR) coefficients [4]. So, how to increase the correlation of the LR and HR manifolds remains to be studied. Recently, it has been found that the position-patch method [3] can further improve the two-step scheme, where the global face is reconstructed using overlapped patches at the same position in the face images. Unlike previous methods whose compensation in the second stage relies on one model only, the method proposed a unified regularization framework which combines a global reconstruction model, local sparsity model, and pixel correlation model to achieve excellent performance. However, the complicated compensation technique inevitably increases its computational complexity, which is not desirable for real-time applications. Another way to bridge the gap between the HR and LR manifolds is based on statistical analysis. Among all the different methods, face hallucination based on Canonical Correlation Analysis (CCA) has received promising results [4, 5]. In the global reconstruction step, CCA is applied to project the PCA coefficients of both the LR and the HR training images into a coherent subspace, where the correlation between them is

maximized. When given a novel LR face input, the corresponding reconstruction coefficients and weights are computed in the subspace to form the global HR face. However, regardless of whether 1D or 2D CCA is used in these works, the direction vectors learned are not orthogonal; this makes the reconstruction with pseudo-inverse lose some correlation, and thus makes the final global reconstruction less precise. We call this kind of CCA the 'original CCA'. Later on, some constraints are added to the original CCA so as to obtain orthonormal direction matrices for perfect data reconstruction [190]. Experiment results will show its superior performance on feature fusion.

In the residual-face compensation phase, patch-based methods are often applied to further narrow the gaps between reconstructed global faces and the ground-truth HR faces. One way to assist the patch-based reconstruction is based on neighbor embedding [9], which assumes that both HR and LR patches lie on manifolds with similar local structures. In [8], sparse coding is used to represent image patches as a sparse linear combination of the atoms from an over-complete dictionary. Ma et al. [6] proposed a one-step face-hallucination framework based on position patches, which are defined as the patches at the same position in different face images. In this approach, local information from overlapped patches is used to reconstruct HR faces. However, this method requires very accurate alignment of the face images in order to achieve a good performance. To further boost the performance, face hallucination in the wavelet domain based on eigen-transformation was proposed in [137]. Another example-based face super-resolution method with class-specific predictors [190] was also proposed to achieve better computational complexity.

Among those two-step face super-resolution methods, many of them apply neighbor embedding to compensate for the detail loss [3-5, 7]. However, in real cases, the

relationship established among LR manifolds may not always hold in the corresponding HR manifolds, thus it decreases the final hallucination performance. Recently, another way to model the relationship between LR and HR patches was proposed, known as direct mapping [191, 205]. Instead of relying on the manifold similarity assumption, it directly learns a function to map a given LR patch into its HR version. The inter-space relationship between different resolutions is studied, and this works well on single-image super-resolution, especially with self-examples.

In this chapter, a two-step face-hallucination framework is proposed for low-resolution face recognition, where a HR version of a face from an input LR face is reconstructed, based on learning from LR-HR example face pairs using orthogonal Canonical Correlation Analysis (OCCA) and linear mapping. The global face reconstruction phase is inspired by the previous CCA-based face-hallucination methods and the orthogonal variant of the original CCA method. At the same time, neighbor embedding combined with direct mapping is proposed for the residual face compensation phase so as to achieve better performance. The contributions of this framework are in the following aspects:

- A two-step face-hallucination framework, based on orthogonal CCA, is proposed. We will show that, after including the orthogonality constraint on the original CCA, the direction vectors derived will be more correlated and will also facilitate reconstruction. Compared to [8], which simply maximizes the correlation between the HR and LR coefficients, our method can further improve the quality of the global reconstruction result.
- In the residual-face compensation phase, a simple and efficient method is proposed, based on Neighbor Embedding and Linear Mapping. It considers both the intra-

space information between the LR and HR residual patches of the same person and the inter-space information between LR and HR residual patches of different people. It can achieve a superior performance in HR face hallucination compared to the current state-of-the-art methods.

- The robustness of the proposed method is also evaluated against different parameter settings and blurring effects. Experiment results show that our method has a strong potential for low-resolution face recognition task.

This chapter is organized as follows. Section 4.2 describes the details of our proposed method, including global face reconstruction and the detail compensation process. Experiment results and analysis are given in Section 4.3, and conclusions are provided in Section 4.4.

4.2 Face hallucination based on orthogonal CCA

4.2.1 Global face reconstruction based on orthogonal CCA

In this section, we will present our global face reconstruction framework based on OCCA. PCA is first applied to both the LR and HR demeaned training faces. This is necessary because some noise can be removed, and the dimensionality of the face samples can also be reduced so that the matrices involved in the OCCA become non-singular and the direction matrices can be computed more efficiently.

Suppose that the LR and HR training face images are represented as $\mathbf{I}_X$ and $\mathbf{I}_Y$, respectively, which are in the form of matrices, with each column representing one face. The corresponding mean faces of the LR and HR training face images are denoted as μ_X and μ_Y, respectively. Applying PCA, the orthonormal eigenvector matrices $\mathbf{E}_X$ and $\mathbf{E}_Y$

are obtained, with the leading eigenvalues containing 98% of the total variation. The projection coefficients $\mathbf{B}_X$ and $\mathbf{B}_Y$ can be computed as follows:

$\mathbf{B}_X = \mathbf{E}_X^T(\mathbf{I}_X - \boldsymbol{\mu}_X)$, and

$$\mathbf{B}_Y = \mathbf{E}_Y^T(\mathbf{I}_Y - \boldsymbol{\mu}_Y). \tag{4.1}$$

It can be proven that the two coefficient matrices are also zero-centered. Thus, we can directly apply the OCCA to $\mathbf{B}_X$ and $\mathbf{B}_Y$, as described in Section 2.3.3. After rescaling, two orthonormal direction matrices $\mathbf{P}_X$ and $\mathbf{P}_Y$, as well as two projected coefficient matrices $\mathbf{O}_X$ and $\mathbf{O}_Y$ with increased correlation, can be obtained, i.e.

$$\begin{aligned}\mathbf{O}_X &= \mathbf{P}_X^T \cdot \mathbf{B}_X = \mathbf{P}_X^T \cdot \mathbf{E}_X^T(\mathbf{I}_X - \boldsymbol{\mu}_X) \\ &= (\mathbf{E}_X\mathbf{P}_X)^T(\mathbf{I}_X - \boldsymbol{\mu}_X)\end{aligned}, \tag{4.2}$$

$$\begin{aligned}\mathbf{O}_Y &= \mathbf{P}_Y^T \cdot \mathbf{B}_Y = \mathbf{P}_Y^T \cdot \mathbf{E}_Y^T(\mathbf{I}_Y - \boldsymbol{\mu}_Y) \\ &= (\mathbf{E}_Y\mathbf{P}_Y)^T(\mathbf{I}_Y - \boldsymbol{\mu}_Y)\end{aligned}. \tag{4.3}$$

These equations show that, after applying orthogonal CCA to the PCA coefficients, the eigenvectors are projected on to subspaces, which are more correlated, by multiplying the orthonormal projection matrices on their right-hand side. Denote $\tilde{\mathbf{E}}_X = \mathbf{E}_X\mathbf{P}_X$ and $\tilde{\mathbf{E}}_Y = \mathbf{E}_Y\mathbf{P}_Y$. Then, it can be verified that

$\tilde{\mathbf{E}}_X^T \cdot \tilde{\mathbf{E}}_X = (\mathbf{E}_X\mathbf{P}_X)^T \cdot (\mathbf{E}_X\mathbf{P}_X) = \mathbf{I}$, and

$$\tilde{\mathbf{E}}_Y^T \cdot \tilde{\mathbf{E}}_Y = (\mathbf{E}_Y\mathbf{P}_Y)^T \cdot (\mathbf{E}_Y\mathbf{P}_Y) = \mathbf{I}, \tag{4.4}$$

where $\mathbf{I}$ is an identity matrix. The projected eigenvector matrices remain orthonormal, so they can be used directly for reconstruction. To reconstruct the global face, we compute the PCA coefficients $\boldsymbol{b}_l$ of the input LR face image $\boldsymbol{I}_l$. Then, these coefficients are projected on to a more correlated subspace using OCCA:

$$\boldsymbol{o}_l = \mathbf{P}_X^T \cdot \boldsymbol{b}_l . \tag{4.5}$$

Since human faces have a similar structure and texture, the projected coefficients of the input LR face can be represented as a linear combination of the projected coefficients of its *K* nearest neighbors $\left\{\boldsymbol{o}_{Xj}\right\}_{j=1}^{K}$ from the training LR face images $\mathbf{O}_X$. This is realized by Locally Linear Embedding (LLE), which minimizes the reconstruction error as follows:

$$\varepsilon = \left| \boldsymbol{o}_l - \sum_{j=1}^{K} w_j^g \boldsymbol{o}_{Xj} \right|^2 , \text{ subject to } \sum_{j=1}^{K} w_j^g = 1, \tag{4.6}$$

where $\boldsymbol{o}_{Xj}$ is the *j*th column of $\left\{\boldsymbol{o}_{Xj}\right\}_{j=1}^{K}$ and w_j^g is the corresponding weight for $\boldsymbol{o}_{Xj}$. This can be solved as described in [73].

As mentioned previously, the basic assumption of the learning-based face-hallucination approach is that the same neighborhoods are preserved in both the HR and the LR manifolds. Based on this assumption, the corresponding projected PCA coefficients of a desired global HR face can be represented as a linear combination of its *K* nearest neighbors $\left\{\boldsymbol{o}_{Yj}\right\}_{j=1}^{K}$ in $\mathbf{O}_Y$, using the same weight contributions as follows:

$$\boldsymbol{o}_h = \sum_{j=1}^{K} w_j^g \boldsymbol{o}_{Yj} . \tag{4.7}$$

Similar to Eqns. (4.2) and (4.3), $\boldsymbol{o}_h$ is also related to the global HR face as follows:

$$\boldsymbol{o}_h = \mathbf{P}_Y^T \cdot \boldsymbol{b}_h = \mathbf{P}_Y^T \cdot \mathbf{E}_Y^T (\boldsymbol{I}_h^g - \boldsymbol{\mu}_Y) = \tilde{\mathbf{E}}_Y^T (\boldsymbol{I}_h^g - \boldsymbol{\mu}_Y) , \tag{4.8}$$

where $\boldsymbol{b}_h$ is the PCA coefficients of the global HR face computed by projecting its demeaned face onto the HR eigenvector matrix. Thus, the global HR face can be obtained as follows:

$$\boldsymbol{I}_h^g = \tilde{\mathbf{E}}_Y \cdot \boldsymbol{o}_h + \boldsymbol{\mu}_Y . \tag{4.9}$$

4.2.2 **Residual face compensation using linear mapping**

In our method, a holistic approach is employed first to reconstruct HR faces. The global faces, constructed in the previous stage, usually lack those details represented by high-frequency information. Thus, we propose a residual-face compensation step, based on linear mapping, so as to make the final hallucinated faces more realistic, possessing more characteristics.

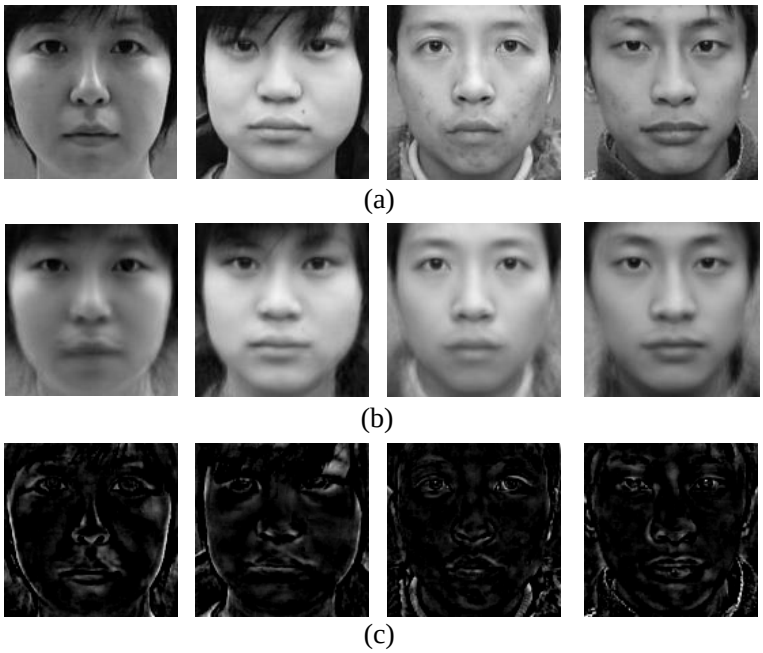


Fig. 4-1. (a) Original HR training faces, (b) the corresponding hallucinated global faces, and (c) the corresponding residual faces. The gray-scale values of the residual faces are normalized to the range [0, 255] for better visualization.

First of all, the leave-one-out method is used to compute the global HR face of each of the LR training-set images using the OCCA. Then, the HR residual face is computed by subtracting the reconstructed global face from the corresponding original HR training face. The corresponding LR residual faces are computed by subtracting the down-sampled version of the reconstructed global faces from the original LR training faces. As shown in

Fig. 4-1, significant residual details appear mainly around the mouth and the face boundary regions due to a rough alignment being used.

Similar to the idea in [192], the features extracted from the LR residual images are first super-resolved for predicting the corresponding features of the HR residual images, based on the retrieved nearest neighbors. Unlike the previous work which directly uses the same neighbors as for HR residual face reconstruction, we first predict the HR features for the residual image from its LR counterpart via linear mapping, and then perform nearest neighbor searching in the HR space. In this way, we consider both the intra-space information between the LR and HR residual patches of the same person and the inter-space information between LR and HR residual patches of different people.

To save the computation cost, we conduct an *N*-nearest-neighbor search using the LR residual face images. Furthermore, as illustrated in Fig. 4-1, unlike the original face images, residual faces possess lots of edges, which can be best represented in terms of gradient information. As a result, instead of only using gray-level intensity as the feature for searching, the first-order gradient is also employed. The gradient feature for each pixel in a patch can be computed based on the intensity values of its four neighbors, i.e. $\nabla u = \begin{bmatrix} u_{right} - u_{left} \\ u_{down} - u_{up} \end{bmatrix}$, where u_{left}, u_{right}, u_{up}, and u_{down} represent the pixel intensities on the left, right, top and bottom, respectively, of the pixel under consideration. Then, the intensity and gradient values can be concatenated directly to form a feature vector $\boldsymbol{f}_l^r$, to be used in the *N* nearest-neighbor search. For HR residual faces, only the demeaned intensity values are used as the feature $\boldsymbol{f}_h^r$. Since the features used for both manifolds are different, directly applying neighbor embedding and using the same weight vector will

lead to substantial distortion. Thus, we firstly apply linear mapping to learn the interrelationship between the features of the two manifolds, then reconstruct the HR residual faces based on the intrarelationship within the manifold.

After retrieving the N nearest neighbors to the input LR residual face image, a linear mapping function $\Psi: \boldsymbol{f}_l^r \rightarrow \boldsymbol{f}_h^r$ is learnt between those LR neighbors and their corresponding HR counterparts, using the method described in Eqns (2.12) and (2.13). In this way, the mapping function will model the relationship between those pairs of residual images, which are most similar to the input LR residual face image. Then, in the HR residual face image space, we search *N* nearest neighbors again that best represent the predicted HR features, i.e.

$$\varepsilon = \left| \boldsymbol{f}_h^r - \sum_{j=1}^{N} w_j^r \boldsymbol{f}_{Yj}^r \right|^2, \tag{4.10}$$

subject to $\sum_{j=1}^{N} w_j^r = 1,$ where $\boldsymbol{f}_{Yj}^r$ is the feature vector of the *j*th nearest neighbors selected from the HR residual training faces, and w_j^r is the corresponding weight for $\boldsymbol{f}_{Yj}^r$. Then, the corresponding HR residual face can be reconstructed as follows:

$$\boldsymbol{I}_h^r = \sum_{j=1}^{M} w_j^r \boldsymbol{f}_{Yj}^r + \boldsymbol{\mu}_Y^r, \tag{4.11}$$

where $\boldsymbol{\mu}_Y^r$ is their mean intensity vector.

It should also be noted that the residual face compensation is a patch-wise process. Eqns (4.10) and (4.11) describe the operations on a patch, and the patches in a face image overlap each other. For those overlapped areas in a reconstructed HR residual face, the

mean pixel value at each position is calculated for its final representation. Thus, the hallucinated HR version of the input LR face image can be represented as follows:

$$\boldsymbol{I}_h = \boldsymbol{I}_h^g + \boldsymbol{I}_h^r . \tag{4.12}$$

The overall structure of our proposed face-hallucination framework is illustrated in Fig. 4-2.

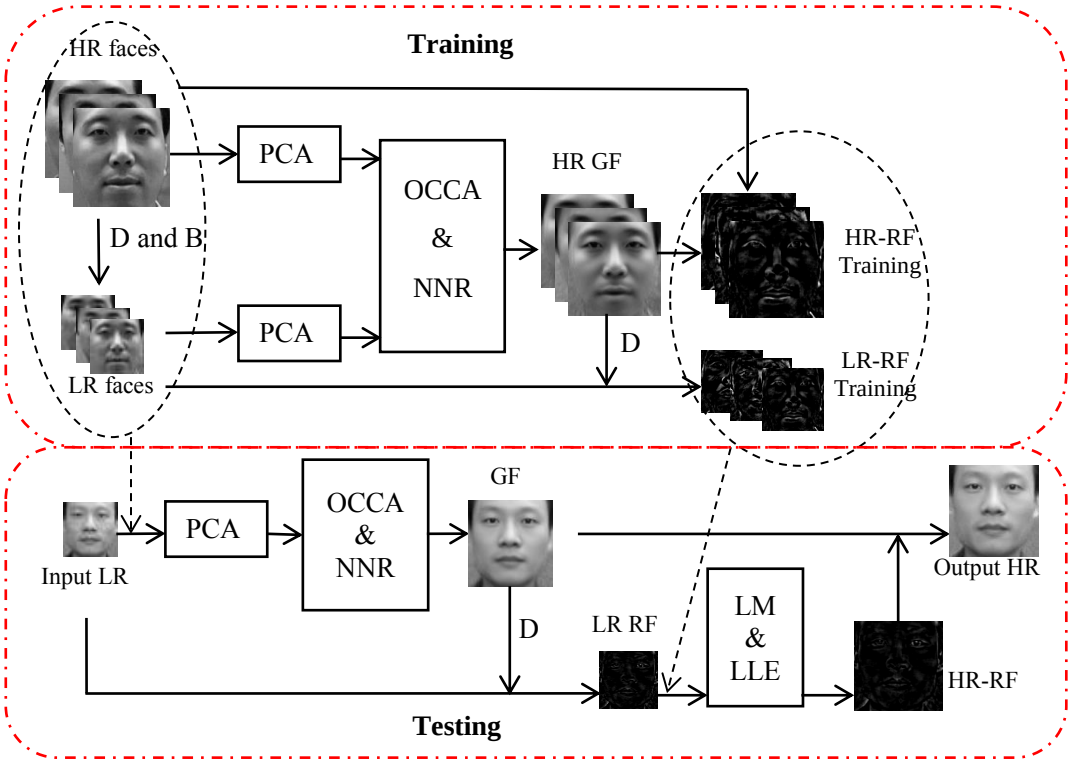


Fig. 4-2. Illustration of the proposed face-hallucination framework, where D represents 'down-sampling'; B is 'blurring'; OCCA is 'orthogonal CCA'; NNR is 'nearest-neighbor reconstruction'; LM is 'linear mapping'; LLE is 'locally linear embedding'; and GF and RF denote 'global face' and 'residual face', respectively.

4.3 Experimental results

To compare our proposed framework with other face-hallucination methods, the CAS-PEAL-R1 dataset [206] was used. This dataset contains 1,040 individuals (595

males and 445 females) with different poses, illumination, backgrounds, etc. In our experiments, we chose all 1,040 frontal faces, with neutral expressions and normal lighting, from the dataset. All of these faces are cropped to include the face region only and are aligned based on the two eye centers to form HR face images of size 128×128 pixels. Then, all the HR faces are blurred with low-pass Gaussian filtering, and are down-sampled to the size of 32×32 pixels. It should be noted that all the face images are at the same scale. For experiment evaluation, we randomly selected 1,000 face images as a training set, and the remaining 40 face images as a testing set in each experiment. The experiment is repeated five times, and the average results are measured.

In this section, we firstly compare the global face reconstruction and final face-hallucination results of the proposed method with other state-of-the-art face-hallucination methods. Then, some of the important parameters used in our method will be discussed. We will also evaluate the effect of blurring on our method so as to verify its robustness.

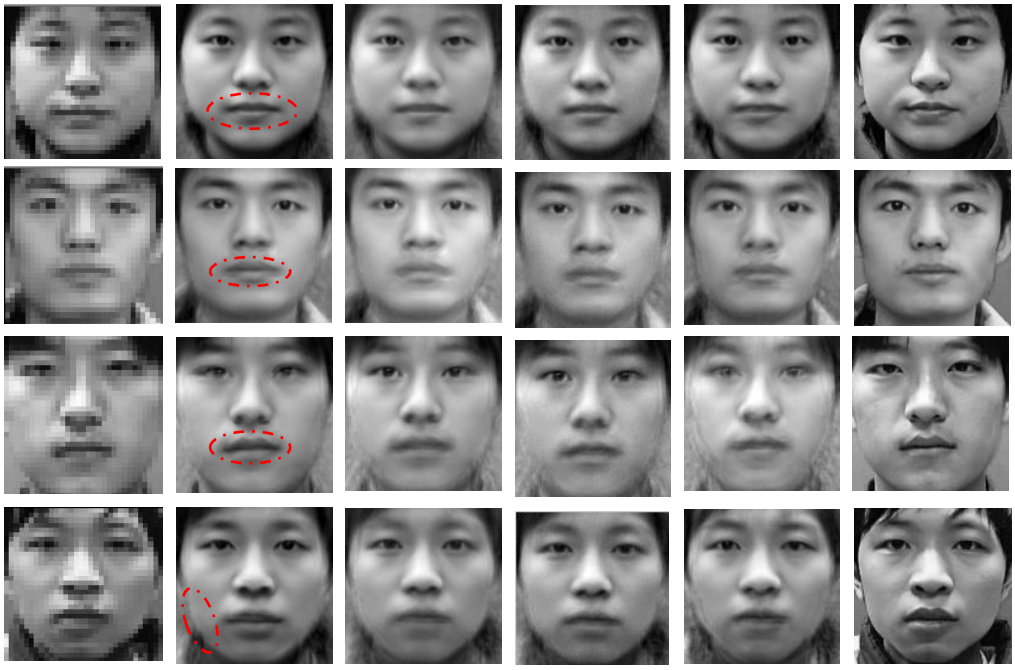
4.3.1 Global face reconstruction

As mentioned in Section 4.2.1, our proposed orthogonal CCA method delivers a good reconstruction performance, with the correlation between the HR and LR coefficients improved. In order to evaluate the OCCA method in terms of global face reconstruction, we compare it with another three related methods: original CCA [4], 2D CCA [5] and LPH [7]. In [4], the OCCA algorithm is applied to increase the correlation of the HR and LR coefficients without considering the reconstruction issue. In [5], 2D CCA is directly applied to the training and testing image data without performing vectorization and PCA. In [7], global faces are reconstructed based on the LPP method, followed by the use of the radial-basis-function regression. For the original CCA and LPH methods, all the

parameters are set the same as in [4]. We also follow the same setting of 2D CCA in [5], which partitions faces into three parts and applies 2D CCA to each part for reconstruction. For the proposed method, leading eigenvectors of 98% of the total variation are used, and the number of nearest neighbors K is set at 200 in the nearest-neighbor searching.

Figure 4-3 shows the global face-reconstruction results based on the different methods. It can be seen that the proposed method can achieve the best performance in terms of visual quality. It should be noted that we have aligned the face images roughly based on their eye positions only. This leads to the reconstruction results based on the original CCA, 2D CCA, and LPH suffering from a severe jagged effect around the mouth and the chin regions in the face images. However, the global faces reconstructed using OCCA exhibit much less distortion in these regions and a better visual quality, as indicated by the red-dashed circles in Fig. 4-3(b).

We also compare our proposed OCCA with the original CCA and LPH for global-face reconstruction in terms of the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) [207]. In total, 200 reconstructed global faces, after five trials, are obtained. Then, the averages and standard deviations of the PSNR and SSIM of the three methods are measured, and the results are tabulated in Table 4-1. From the results, we can see an obvious improvement in both measurements using our method. Besides, the results show that the OCCA can reconstruct global faces with a more stable and uniform performance, as the standard deviations based on the proposed method are smaller; this further shows the advantage of including the orthogonality property in reconstruction. Global faces reconstructed using 2D CCA achieve better visual results, but the average PSNR is lower because applying projection directly onto face images inevitably introduces noises.



(a) (b) (c) (d) (e) (f)

Fig. 4-3. Global face-reconstruction results: (a) input LR faces, (b) global faces produced by our proposed orthogonal CCA, (c) global faces reconstructed using original CCA in [4], (d) global faces reconstructed using 2D CCA in [5], (e) global faces reconstructed using LPH in [7], and (e) the original HR faces.

Table 4-1 Global-face reconstruction performances in terms of the means and standard deviations (mean±standard deviation) of the PSNR and SSIM of the different methods.

	Orthogonal CCA	Original CCA	2D CCA	LPH
PSNR	25.57 ± 1.08	23.05 ± 2.66	23.66 ± 2.64	23.87 ± 2.61
SSIM	0.831 ± 0.02	0.792 ± 0.04	0.812 ± 0.04	0.804 ± 0.04

4.3.2 **Comparison of face-hallucination methods**

After applying the simple and efficient locally linear embedding to generate a compensated face, the final hallucinated face can be obtained by combining the global face and the compensated face. In this section, we compare the quality of the final

hallucinated faces reconstructed using the proposed method with those using other state-of-the-art methods, including regularization-based hallucination [3], 2D CCA-based hallucination [5], CCA-based hallucination [4], sparse-coding-based hallucination [8], position-patch-based hallucination [6], and LLE-based hallucination [9]. For the LLE-based method in [9], both HR natural images and face images are used for learning so as to reconstruct the edge information more accurately. When implementing the position-patch method [6], 1,000 and 40 face images are used for training and testing, respectively. We attempt to have the same experimental setup for all of the different methods, and all the parameters set the same as in the original papers. For the compensation step in our proposed framework, the patch size used is 8×8 pixels, with 4 pixels overlapped. We randomly selected 300 pairs of HR and LR face image to form the training samples, and set the number of nearest neighbors M at 180 for searching. To achieve more reliable results, this experiment was also repeated five times, and the average results are computed. Thus, 200 hallucinated faces are generated for analysis.

The hallucinated faces reconstructed using the different methods are shown in Fig. 4-4. It can be seen that applying LLE directly to reconstruct HR faces will lead to significant blurring and blocky artifacts, since face images are usually much smoother than natural images, and contain fewer edges. The position-patch method can estimate high-frequency details more accurately than the LLE method, and produce sharper HR face images. However, ringing artifacts can be seen around the face contours. Hallucination results using PCA on overlapped patches without dimensionality reduction appear blurred. This demonstrates that only using a one-step scheme cannot improve the overall performance. Hallucinated faces by sparse coding and the original CCA can achieve a better visual

performance, but still have some ringing artifacts around the unaligned mouth and face contours. Applying 2D CCA to partitioned faces generates faces that are clear but suffering from noise.

The results from the unified regularization-based method produce the sharpest faces, but the method requires runtimes about 20 times more than the proposed method. Of the different methods, the proposed method, i.e. OCCA for global face reconstruction and linear mapping for compensation, achieves the best performance, in terms of visual quality and computational efficiency. The proposed method saves a lot of computation in the compensation stage because the training set used is smaller and it does not need to search within the neighborhood of each patch. In general, the proposed framework can super-resolve an input LR image to produce an output in 1.9s, using MatLab 2010b with i7 CPU at 3.5GHz.

Fig. 4-5(a) shows the boxplot of the PSNR values of all the different methods, while Fig. 4-5(b) shows the corresponding SSIM values. We can see that the proposed hallucination method can achieve almost the same performance as the unified regularization method, while outperforming other methods in both measurements. In particular, when compared to the original CCA, the proposed method can achieve a more stable and reliable performance, because the two boxes of the two measurements, based on our method are much more compact.

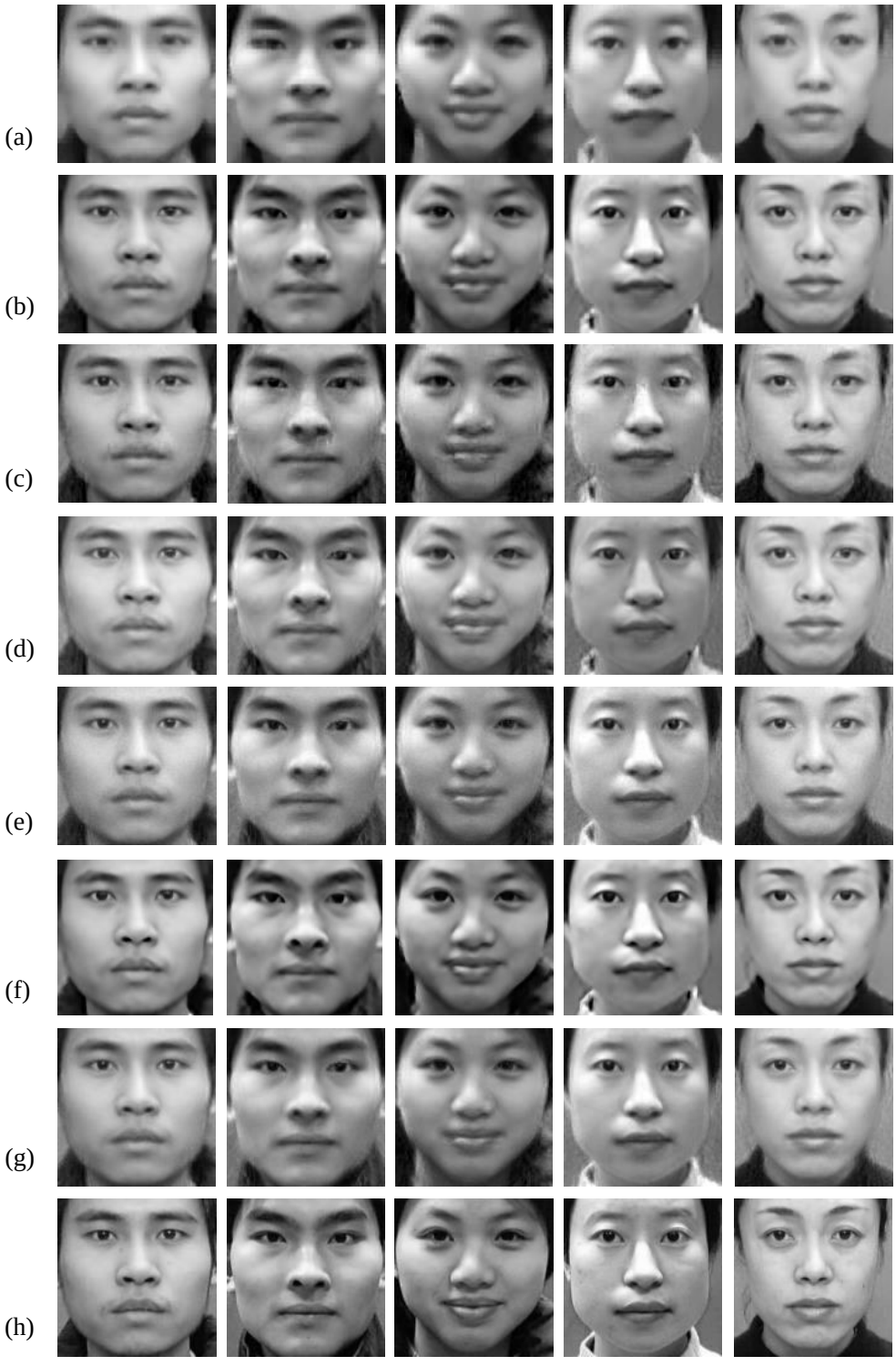


Fig. 4-4. Comparison of the final hallucinated faces based on different methods: (a) LLE [12], (b) position-patch [14], (c) sparse-coding [13], (d) original CCA [8], (e) 2D CCA [10], (f) the unified regularization method [9], (g) the proposed orthogonal CCA, and (h) the original HR faces.

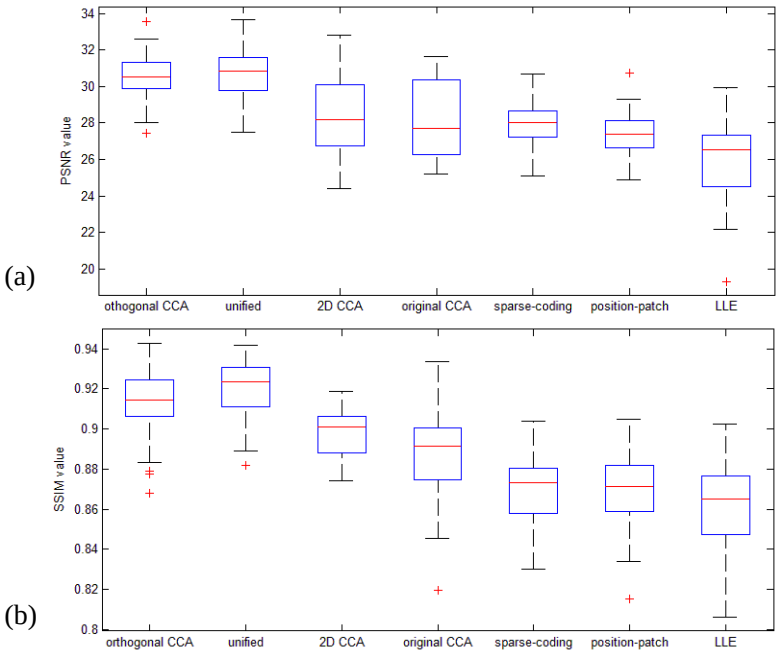


Fig. 4-5. Boxplots for different face-hallucination methods: (a) PSNR, and (b) SSIM. The seven methods compared are: (1). proposed orthogonal CCA, (2). unified regularization [3], (3). 2D- CCA [5], (4). original CCA [4], (5). sparse-coding [8], (6). position-patch [6], and (7). LLE [9].

4.3.3 Comparison of low-resolution face recognition methods

In this section, extensive experiments on low-resolution face recognition using super-resolved faces are conducted. In the experiments, LR faces of three different resolutions are considered: 32×32, 24×24 and 16×16 pixels, respectively. As introduced in Section 2.1.3.1 for super-resolution-based methods, we apply face hallucination and face recognition in two steps. In the face hallucination step, we super-resolve face images with a magnification factor 4 using our proposed framework. In the recognition step, we denote three methods as SR+Eigenface, SR+LBP, and SR+Gabor, respectively, where the most famous and standard recognition methods Eigenface [53], LBP [11] and Gabor [63] are used with the super-resolved faces. We perform face recognition, based on six methods, on all the face images in the same dataset, and the average recognition rates of the different

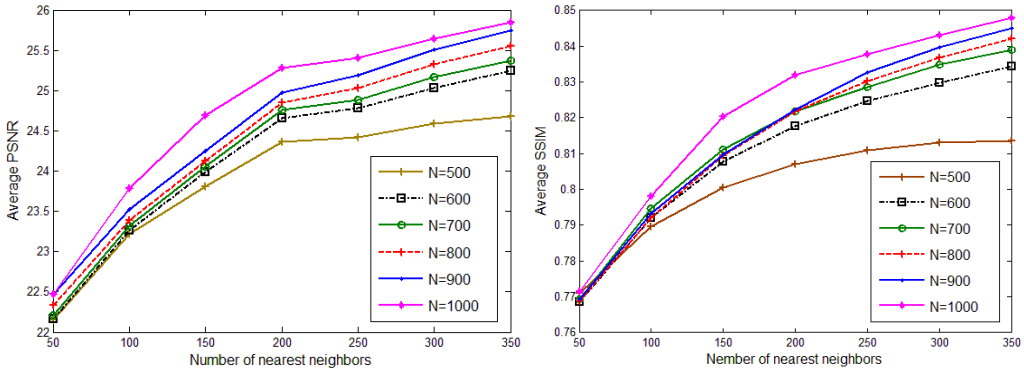
methods are tabulated in Table 4-2. From the results, it can be seen that the performances of different face recognition methods will decrease dramatically as the face-image resolution decreases, especially from 24×24 to 16×16 pixels. With the proposed super-resolution method, we can see a great improvement in terms of recognition rates, which proves the potential of our proposed framework. In our experiment, it is observed that when the face images are down sampled to be size of 12×12, the recognition rate with face hallucination will decrease to be under 50% which is normally considered as unacceptable. How to further improve the recognition performance on such lower-resolution faces will remain a challenging task in the further.

Table 4-2 Average recognition rates (%) of different face recognition methods with the LR faces of size 32×32, 24×24 and 16×16 pixels, respectively.

Face size / Methods	32×32	24×24	16×16
Eigenface [53]	85.2	63.3	38.5
LBP [11]	89.7	65.2	41.6
Gabor [63]	88.6	66.1	42.7
SR+Eigenface	92.3	72.8	67.4
SR+LBP	95.2	74.5	68.8
SR+Gabor	94.9	74.9	68.5

4.3.4 **Analysis of the parameter setting**

In this section, the influence of the parameters in the proposed framework will be discussed. We will analyze the performance of our proposed method in terms of global reconstruction and residual face compensation, i.e. the first and the second stages of the proposed method.



(a) (b)

Fig. 4-6. (a) The average PSNRs, and (b) average SSIMs of the reconstructed global faces based on our method under different training-set sizes and numbers of nearest neighbors.

(a) (b) (c) (d) (e) (f)

Fig. 4-7. Globally reconstructed faces with different training-set sizes N, and K=200: (a) N=600, (b) N=700, (c) N=800, (d) N=900, (e) N=1,000, and (f) the original HR faces.

4.3.4.1 **Parameters for global face reconstruction**

For the global face-reconstruction phase, we investigate the effect of the training-set size and the number of nearest neighbors searched on the reconstruction performance. The size of the training set N is changed from 500 to 1,000, with an interval of 100; and for each training-set size, the number of nearest neighbors K varies from 50 to 350, with an

interval of 50. The average PSNR and SSIM values of the 40 testing face images for each case are computed and displayed in Fig. 4-6. It can be seen that, at each training-set size, increasing the number of nearest neighbors will lead to a higher PSNR and SSIM. However, if a larger number of nearest neighbors are searched, the computational cost will increase. From the experiment results, there is a slight improvement in performance when K is larger than 200, thus K is set at 200 in our framework. On the other hand, increasing the training-set size will constantly improve the reconstruction performance of our method. In the training phase, learning the OCCA direction matrices iteratively requires most of the computation. However, once the matrices have been learned, they can be used directly for the reconstruction of novel LR face images. More training samples can help to learn more discriminative features for OCCA. As a result, we keep using the largest training set, i.e. N=1,000, in our method. Fig. 4-7 shows some reconstructed global faces, with K=200 and different training-set sizes.

4.3.4.2 **Parameters for residual face compensation**

In this section, we explore the influence of the parameters used in the residual face-compensation phase, i.e. the training-set size T and the nearest neighbors N for locally linear embedding. Since the residual images are much simpler than the original face images, and since they contain mainly edges and contours, it is reasonable to reduce the training-set size. Thus, in this experiment the training-set size T is varied from 100 to 600 by randomly selecting samples from the image dataset, with an interval of 100, while the number of nearest neighbors searched N is set at $T/5$, $2T/5$, $3T/5$, $4T/5$, and T accordingly. Similar to the previous section, we compute the average PSNR and SSIM values of the final hallucinated results for 200 testing face images. After computing the corresponding values, we find that the residual-compensation stage does not always improve the

performance with an increase in the training-set size and the number of nearest neighbors searched. Also, the results based on the different parameter settings are not as impressive as those for global reconstruction. For a better visualization, we normalize all the PSNRs according to the minimum and maximum SSIM values, then we display both experiment results together, as shown in Fig. 4-8.

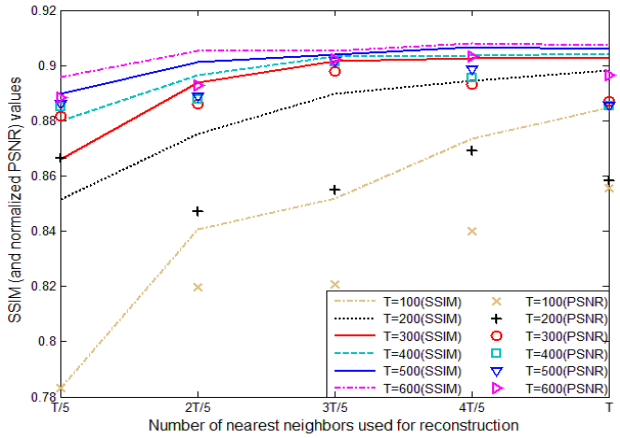


Fig. 4-8. SSIM (and normalized PSNR) values of final hallucinated faces, based on our method under different training-set sizes and numbers of nearest neighbors. The lines represent SSIM values, and the discrete symbols represent normalized PSNR values.

From the results, it can be seen that when the training-set size T reaches 300 and the number of nearest neighbors is larger than $3T/5$, the final PSNR and SSIM values become stable. This can be explained thus: with a certain number of training samples, which contain sufficient variances for HR reconstruction, further increasing the number of nearest neighbors to be searched will introduce more noise to the linear regression and reconstruction results. Considering both performance and efficiency, the best result can be achieved when T =300 and N =180, as indicated by the red line and red circle in Fig. 4-8.

4.3.5 **Impact of blurring effect**

Face hallucination in real-world applications imposes further challenges on research study because of the effect of added noise and blurring. As also mentioned in [6], patch-based methods will inevitably retain most of these distortions, while a global image can preserve the characteristics from the training samples and remove the local distortion. To further verify the robustness of our proposed method, in this section we test the impact of blurring, which is the main concern of all face super-resolution techniques. Three face-hallucination methods are chosen which have achieved the best performances in the previous experiments on global face reconstruction, i.e. the unified regularization method which uses position-patch in the first stage, the proposed OCCA framework, and the original CCA (2D CCA applied to the whole face is not as good as the original 1D CCA). In this experiment, the global faces reconstructed using the three methods are shown in Fig. 4-9. To blur an image, a 3×3 Gaussian kernel with a standard deviation of δ is convolved with the LR face images.

From the results, it can be seen that the patch-based method is sensitive to blurring, while the global-construction-based methods can maintain their good performances. Compared to the original CCA method, the proposed OCCA method is able to produce HR images of better fidelity, especially around the mouth region and the face contour. We have also observed that the original CCA method results in an increase of the overall pixel intensity in the reconstructed face images; this distorts the super-resolved results. What's more, this also implies that, under distortions like blurring, patch-based methods are not desirable for use in the second stage of the reconstruction. How to make full use of the

global face, and propose an efficient detail-compensation method, are yet to be studied in depth.

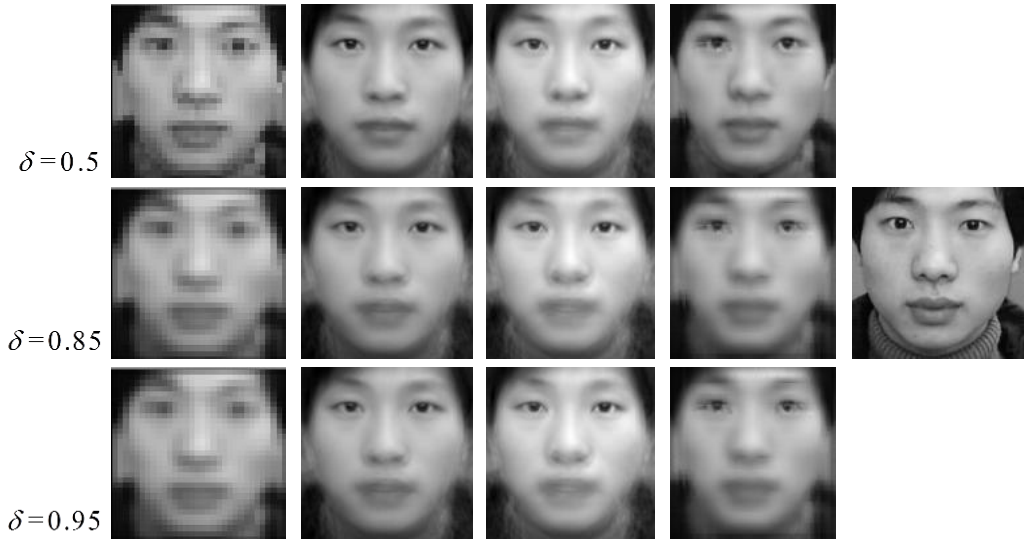


Fig. 4-9. Hallucinated faces reconstructed from LR face images under different amounts of blurring: (a) input LR face images blurred by a Gaussian kernel with standard deviation δ, (b) results for our orthogonal CCA, (c) results for the original CCA [4], (d) results for position-patch [6], and (e) the original HR face.

4.4 **Conclusions**

In this chapter, a two-step face-hallucination framework based on orthogonal CCA and linear mapping has been proposed. By applying orthogonal transformation to the original CCA, the orthogonal CCA proposed becomes more efficient for data reconstruction. Experiments have shown a great improvement in global face reconstruction using our method. To further enhance performance, an efficient linear mapping technique for residual face compensation has also been proposed, which can make use of both inter and intra-information between the LR and HR datasets. The final hallucination results, based on the proposed method, demonstrate a comparable

performance with other state-of-the-art face-hallucination methods. Besides, results on low-resolution face recognition have also demonstrated that our hallucinated faces followed by other feature extraction methods can achieve better performances. Experiments on the impact of parameter settings and blurring on our method have also been conducted, and the results show its robustness and reliability.

Chapter 5. High-resolution face verification using pore-scale facial features

5.1 Introduction

As mentioned in Section 2.1.3.2, many of the face recognition algorithms are based on holistic facial features, which project the lexicographic ordering of raw pixels onto a certain subspace. They suffer significant degradation in performances when the face images considered are under pose, expression, and/or illumination variations. Local features, extracted from local regions or parts of the images only, can be used to achieve better performances under the different variations. However, feature representations and face recognition algorithms always require the face images to be normalized and aligned to achieve a satisfactory accuracy level. In addition, the pose, expression, and illumination variations will cause non-linear distortions on the 2D face images, due to the fact that the facial features (eyes, nose, mouth, etc.) do not appear on a planar surface. With the easy access to high-resolution (HR) face images nowadays, some HR face databases have recently been developed. However, few studies have tackled the use of HR information for face recognition or verification.

In human biology, it is impossible for two people, even identical twins [208], to have an identical skin appearance. Inspired by this idea, a novel pore-scale facial feature has been proposed in [209]. By adapting the Scale Invariant Feature Transform (SIFT) detector and descriptor to the pore-scale facial-feature framework, and using a candidate-constrained matching scheme, the algorithm [209] can establish a large number of reliable correspondences of keypoints between two face images of the same subject which may

have a big difference in pose. Such pore-scale facial features are dense and distinguishable, which are the desirable aspects for face verification.

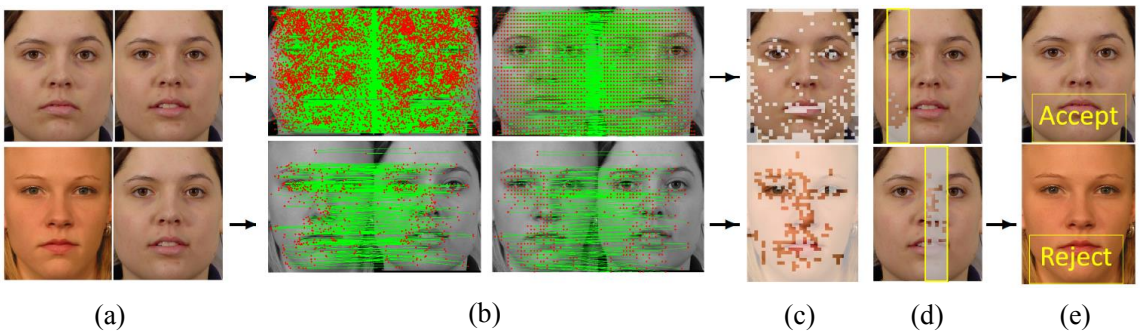


Fig. 5-1: The face-verification framework based on the PPCASIFT feature: (a) two examples of query pairs for face verification - the top row is a client and the claimed identity, and the bottom row is an impostor and the claimed identity; (b) the pore-scale facial-keypoint-matching results and the initial block-matching results between the query and the claimed-identity images; (c) the refined block-matching results (the color regions represent those aggregated, matched blocks); (d) the local region (in the yellow box) of the gallery image which has the maximum local-matching density; and (e) the verification results.

In this chapter, we propose a face-verification algorithm based on the pore-scale facial features to take advantage of the HR information. One of the major advantages of the proposed approach is that the facial-skin regions under consideration are usually more linear - i.e. approximate to a planar surface - than other facial features, so the recognition performance will be very robust to pose, expression, and illumination variations, etc. Furthermore, only one gallery sample per subject is needed, and an accurate face alignment is not necessary to achieve a good performance. An overview of our proposed framework is shown in Fig. 5-1. Firstly, the porescale facial features are detected and extracted from a testing or a query image. Then, initial keypoint matches between the

testing image and that of the claimed identity in the gallery are established; the initial keypoint matches are then converted to block matches, which are further aggregated to eliminate the outliers, as shown in Fig. 5-1(b) and (c). Finally, the verification result is determined based on the maximum density of the local, aggregated, matched blocks on the face images, as illustrated in Fig. 5-11(d). In our experiments, we will show the superior performance of our face-verification algorithm compared to the standard face-verification methods.

The contributions of the proposed framework and the novel aspects of the proposed method are listed as follows:

- An alignment-error-insensitive and pose-invariant face verification approach is proposed. In other words, only the approximate locations of facial features such as the eyes and mouth are necessary. Non-frontal-view face images do not need to be included in the gallery. These make our method suitable for practical and real applications.
- To the best of our knowledge, our method is the first to perform face verification using pore-scale facial features rather than landmark-features (e.g. contours, eyes, nose, mouth) or marker-scale features (e.g. moles, scars).
- A new descriptor is proposed, namely Pore-Principal Component Analysis (PCA)-SIFT (PPCASIFT), which can achieve a similar performance to the Pore-SIFT (PSIFT) [209] descriptor but which requires only 9% of the PSIFT descriptor's computation time in the matching stage.
- A fast and robust fitting method is proposed to establish the block matching of two faces based on matched keypoints, which considers the non-rigid structure of faces and which can also remove outliers at the same time.

- A pose-invariant similarity measure, namely the maximized local-matching density, is proposed to provide a normalized similarity measure for pose-invariant face verification. Based on the maximized local-matching density, no prior knowledge of pose information is needed.

This chapter is organized as follows. Section 5.2 describes the details of our proposed method, including pore-scale facial-feature detection, description, matching and fitting, along with the introduced similarity measurement scheme. Experiment results and analysis are given in Section 5.3, and conclusions are provided in Section 5.4.

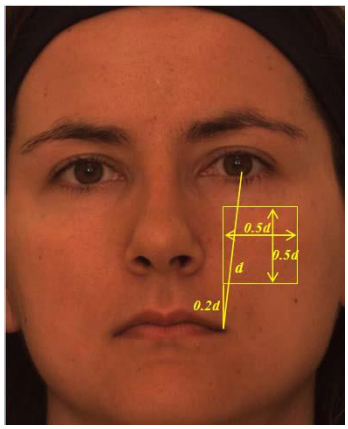


Fig. 5-2: The size of a cropped skin region, whose pore features are to be extracted.

5.2 Pore-scale facial feature for face verification

5.2.1 Pore-scale facial-feature detection

Pore-scale facial features include pores, fine wrinkles and hair, which commonly appear in the whole face region. Most of the pore-scale facial features are blob-shaped features. Hence, the PSIFT detector [209] employs the Hessian-Laplace detector on the multiscale Difference of Gaussians (DoG) for blob detection.

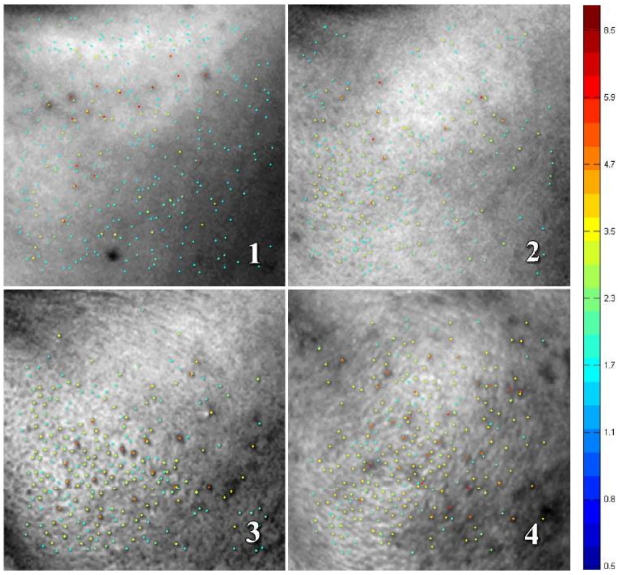


Fig. 5-3: Visualization of keypoints on the skin images of 4 distinct subjects (different colors for the keypoints indicate their scales as represented by the color bar).

In order to generate a sufficient number of correspondences between two face images, a large number of reliable feature points should first be detected on each of the two faces. Meanwhile, from the biological viewpoint, different people should have a similar quantity of pores in their facial skin. Hence, the PSIFT detector employs a quantity-driven approach via an adaptive threshold. To avoid the extremely dense DoG responses at hairy (e.g. bearded) areas, the PSIFT detector estimates the adaptive threshold based on a cropped skin region in the cheek instead of the whole face image, as shown in Fig. 5-2. With a consideration of both the robustness and the completeness of matching, the number of detected pore-scale keypoints N_k in the cropped region is specified to be within the range [210, 240] by an adaptive threshold. To determine the value of the adaptive threshold, the binary search method [210] is performed on a threshold list, which is from 0 to 0.0025. Also, the quantity of inliers is used to evaluate the matching performance by experiments at different sampling frequencies and different priors of smoothing,

respectively. Thus, the number of DoG octaves is set at 3, and 8 scales are sampled in each DoG octave. In addition, the prior smoothing, which is applied to each image level before building the scale-space representation, is set at 1. Usually, about 4,500 keypoints are detected for a whole face region. Fig. 5-3 illustrates the keypoint-detection results on the cropped region of four distinct subjects, where keypoints with different scales are represented using different colors.

5.2.2 Pore-scale facial-feature description

Most pore-scale facial features are similar to each other when they are observed individually, because most of them are blob-shaped, and the surrounding region of each keypoint has almost the same color. However, the spatial distribution of pores on the skin is distinctive. Based on this biological observation, designing a distinctive pore-scale facial-feature descriptor becomes possible.

PSIFT [209], adapted from SIFT [126], was proposed using the gradient information of neighboring keypoints to describe the textures and the spatial distribution of pores. The number of sub-regions and the support size of each sub-region are expanded in PSIFT so as to extract the relative-position information about the neighboring keypoints. In this way, a PSIFT descriptor is constructed from the gradient orientations within a region containing 8×8 sub-regions with each sub-region represented by a histogram of 8 orientation bins. Therefore, PSIFT is represented as a 512-element feature vector for each keypoint description. In addition, the keypoints are not assigned a main orientation because most of them are blobshaped and do not have a coherent orientation. Furthermore, as the rotation of a face image is usually not large, generating a rotation-free description of the pore-scale facial feature is not necessary.

However, matching two keypoints using descriptors of 512 dimensions is computationally expensive. To improve the method's efficiency, we propose a more compact keypoint descriptor in this chapter, namely Pore-PCA-SIFT (PPCASIFT), which is adapted from PCA-SIFT [125] and which uses PCA to reduce the dimensionality of the descriptor. We extract the description of a keypoint using a patch/sampling size of 41×41, with the keypoint at the center, at a given window size (48 times the scale determined by the PSIFT detector). The parameters are determined by a Powell method [211] based on the verification rates on a small dataset. The initial setting was chosen so that the PPCASIFT descriptor window has the same size as PSIFT. Unlike PCA-SIFT, the patches of porescale facial keypoints do not need to be aligned or assigned their main orientations. However, PCA can still represent these patches. The main reason for this is that the patches contain the relative spatial information about the neighboring keypoints, and the patterns of patches are relatively simple. If a sufficient number of patches is available for learning the principal components, the patches can then be represented efficiently in a much lower subspace. We selected 16 face images from 4 distinct subjects with different skin conditions in order to extract about 90,000 patches (after removing the keypoints near the borders). The horizontal and vertical gradient maps in the 41×41 patch are computed, and are represented by a vector containing 2×39×39 = 3,042 elements. The vectors are normalized to unit magnitude, and then PCA is applied to these training vectors. The 72 leading eigenvectors are used to form the projection matrix for PPCASIFT, which is of dimension 3,042×72. To generate the PPCASIFT descriptor for a given keypoint, its normalized gradient vector is computed and is then projected onto the eigenspace formed by the 72 eigenvectors. Compared to the PSIFT descriptor which has a dimension of 512,

the dimension of the PPCASIFT descriptor is 72 only. In other words, the PPCASIFT feature is much more compact and computationally efficient in the matching stage than PSIFT.

5.2.3 Pore-scale facial-feature matching and robust fitting

In [209], a double-matching scheme (namely *candidate-constrained matching*) was proposed to narrow the matching of keypoints between two face images and to achieve accurate face matching, based on both intra- and inter-scale facial information. RANSAC [212], as a robust fitting method, is then applied to the matched keypoints to remove the outliers reliably. To perform face verification efficiently, we modify the candidate-constrained matching from two passes to one pass only. In addition, a new robust fitting scheme, namely *parallel-block aggregation*, is proposed to refine the candidate-constrained matching results. As the keypoint/block matching may result in one-to-many or many-to-many matches, matching from gallery faces may differ from that from testing faces. In our experiments, we only consider the block matching from a testing face image to a gallery face image.

5.2.3.1 Feature matching

For verifying whether or not two face images are of the same identity, correspondences are established from the query image to the gallery image of the claimed identity. Suppose that the position and the scale of a keypoint in the query image are (x^q, y^q) and σ^q , respectively, while the position and the scale of the *i*th keypoint in the claimed image are (x_i^c, y_i^c) and σ_i^c ,respectively. Assume that the height of the gallery image is *H*.

First, the spatial information of the face image is considered in feature matching. Considering that the poses of faces are limited to within a certain range, and the *y* coordinates of the two keypoints from the two face images at different poses are close, then the position of the matched keypoint in the gallery image should satisfy the following constraint:

$$\left| y_i^c - y^q \right| < \lambda H , \tag{5.1}$$

where λ is a factor and is set at 0.2 in our experiments.

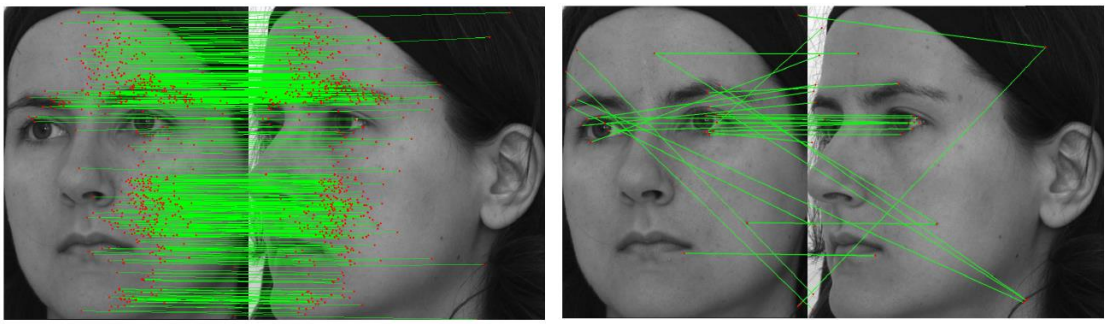
Second, the scales of the two keypoints, one from the query and the other from the gallery, should be close to each other. Therefore, the ratio of the scale of the keypoint in the query image and the *i*th keypoint in the gallery image should be close to 1, and is defined to be within the range as follows:

$$1/\mu \le \left| \sigma_i^c / \sigma^q \right| \le \mu , \tag{5.2}$$

where μ is a constant larger than 1. When μ is close to 1, the scales of the two keypoints are similar. In our experiments, μ is set at 2.

Based on these two constraints, the number of keypoint candidates is narrowed to about 30% of all the keypoints in the gallery image. Then, the distances between a keypoint in the query image and the remaining keypoints in the gallery image are computed. The distance between two keypoints is measured using the Euclidean distance between their corresponding pore-scale feature descriptors. The best-matched keypoint in the gallery image is the one with the smallest Euclidean distance. We define the distance ratio, which is the ratio between the distances of a keypoint from the query image to its nearest keypoint and to its second-nearest keypoint in the gallery image. We accept the match if the distance ratio is smaller than 0.85, which is set empirically by experiments.

The initial matching results using our method are shown in Fig. 5-4; this shows the effectiveness of using the pore-scale facial features.



(a) (b)

Fig. 5-4: Initial matching results for face images of pose R10 and pose R45: (a) PPCASIFT, and (b) SIFT. The matching results of PSIFT are much denser and more structurally accurate than those of SIFT.

5.2.3.2 **Robust fitting**

After the detection, description, and initial matching of the keypoints, a large number of matched keypoint pairs between a query/testing face and a gallery/claimed face have been established. However, the matching results still include many outlier pairs. Consequently, further refinement is necessary to improve the matching accuracy. In [209], RANSAC [212] is used to refine the matching results by fitting to the epipolar constraint. However, this process is of high complexity and requires a large number of iterations, which is not desirable for real-time face verification. In addition, the number of matches cannot be used directly as a similarity measure for verification, because the number depends on the degree of variation between the two faces to be matched. In this chapter, we propose a more efficient and effective scheme to refine the matches and provide a normalized measure of the correspondences for face verification. Our basic idea is to transfer the keypoint correspondences to block-based correspondences. The line

connecting two correctly matched blocks in the two face images should be approximately parallel to the other lines of the corresponding neighboring blocks.

First, matched keypoint pairs are transformed into matched block pairs, which can further remove some outliers. All the face images are divided into non-overlapping blocks of size $W' \times H'$. Assume that a query face image, I_q, establishes keypoint correspondences to a gallery face image, I_g. If a keypoint resides in a block, denoted as $\boldsymbol{B}$, in I_q and is matched to a keypoint in I_g which resides in block $\boldsymbol{B}'$, the block pair $\boldsymbol{B}$ and $\boldsymbol{B}'$ is considered to be initially matched. As shown in Fig. 5-1(b), the initial block-matching result of an impostor is much sparser than that of a genuine subject; this characteristic is useful for face verification.

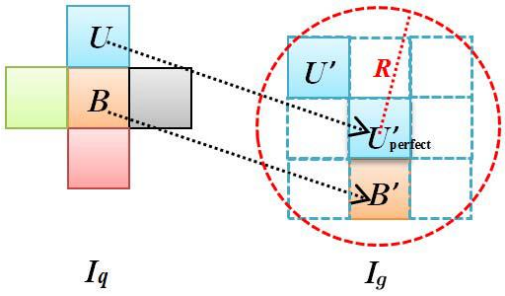


Fig. 5-5: Use block $\boldsymbol{B}$ and its upper neighboring block $\boldsymbol{U}$ in a query image ($\boldsymbol{I_q}$) to illustrate the parallel-block-aggregation scheme. $\boldsymbol{B}'$ and $\boldsymbol{U}'$ in a gallery image ($\boldsymbol{I_g}$) are the corresponding matched blocks of $\boldsymbol{B}$ and $\boldsymbol{U}$, respectively. $\boldsymbol{U_{perfect}}$ is the perfect location for $\boldsymbol{U}$ such that $\boldsymbol{UU_{perfect}}$ is exactly parallel to $\boldsymbol{BB}'$. However, human faces are non-rigid and may have expressions. Block $\boldsymbol{B}'$ is aggregated if $\boldsymbol{U}'$ is inside the neighborhood of $\boldsymbol{U_{perfect}}$, which is represented as a circle with the red, dashed line.

However, some of these initially matched block pairs are still outliers. In order to achieve a robust and accurate face verification performance, we propose a new, robust fitting scheme, namely parallel-block aggregation. Algorithm 5-1 shows the pseudocode

of the parallel-block-aggregation algorithm, and Fig. 5-5 illustrates the robust fitting scheme. Block $\boldsymbol{B}$ and one of its n_u neighboring blocks, $\boldsymbol{U}$, in $\boldsymbol{I_q}$ are initially matched to $\boldsymbol{B'}$ and $\boldsymbol{U'}$ in $\boldsymbol{I_g}$, respectively. For the perfect matching of an inlier, $\boldsymbol{U}$ should match $\boldsymbol{U_{perfect}}$. However, due to the fact that faces are non-rigid and may have local changes caused by facial expressions, the matching from $\boldsymbol{U}$ to any block in the neighborhood of $\boldsymbol{U_{perfect}}$ within a certain radius, R, is considered valid, i.e. the line joining $\boldsymbol{U}$ and $\boldsymbol{U'}$, and that joining B and $\boldsymbol{B'}$, are considered to be parallel. R is the threshold of the distance between $\boldsymbol{U'}$ and $U_{perfect}$, which forms an acceptable, circular region for block matching. As illustrated in Fig. 5-5, the block $\boldsymbol{U'}$ inside the circular region is called a parallel-supporting block of $\boldsymbol{B'}$. The block $\boldsymbol{B'}$ is aggregated if the number of its parallel-supporting blocks is larger than or equal to n_t. By experiments, we set $n_u = 4$, the distance threshold $R = 1.5 \times$ block size, and $n_t = 1$, which can produce the best performance. Fig. 5-1(c) and Fig. 5-6(c) illustrate the parallel-block-aggregation results.

Algorithm 5-1 Parallel-Block-Aggregation Algorithm

Transform the matched keypoint pairs to initially-matched block pairs
for each block B' of the $W' \times H'$ blocks **do**
  for each initially-matched block B **do**
    $N_B = 0$
    for each of the n_u neighbors U **do**
      Compute the location of U'_{perfect} based on B, B', and U
      if distance($U'U'_{\text{perfect}}$)$< R$ **then**
        $n_B = n_B + 1$
        U' is a parallel-supporting block
      end if
    end for
    if $n_B \geq n_t$ **then**
      B' is aggregated
    end if
  end for
end for

5.2.4 **Similarity measurement**

The area of the aggregated blocks (or the number of aggregated blocks) is variant to poses, due to the fact that the areas of a corresponding region in two faces with different poses are not the same. In this section, we propose a new normalized similarity measure, namely the maximized local matching density, for face verification with pose variations.

5.2.4.1 **Matching density**

Denote **R**(**x**) as a face region of a particular size located at x. The total number of blocks in **R**(**x**) is counted, and is denoted as N_{total} (**R**(**x**)). After robust fitting, the number of matched blocks in the region **R**(**x**) is denoted as $N_{matched}$ (**R**(**x**)). Then, the matching density ρ of the region **R**(**x**) is defined as follows:

$$\rho(\mathrm{R(x)}) = N_{matched}(\mathrm{R(x)}) / N_{total}(\mathrm{R(x)}), \tag{5.3}$$

where the value is within the range [0,1].

(a) (b) (c) (d)

Fig. 5-6: Selection of a local region with the maximized local-matching density: (a) the frontal-view gallery face image, (b) two query face images at pose “right 10°” (the 1-st row) and pose “right 45°” (the 2-nd row), (c) the corresponding block-aggregation results on the gallery face for the two query images at different poses, and (d) the selected local regions in the two images with maximum local-matching density.

5.2.4.2 **Local-region selection**

The well-matched regions of two faces of the same subject are always unoccluded and more planar areas, as illustrated in Fig. 5-6(c). These regions have significantly fewer non-linear distortions when the faces to be matched have different poses. To ensure that the matching density of a well-matched local region is robust to poses, the size of such a local region **R**(**x**) should always be smaller than the common area $\mathbf{R}'(\theta)$ of the two faces, where θ is the pose difference between the fontal-view gallery face image and the query image. Thus, the matching density of the local region $\rho(\mathbf{R})$ is more invariant to the pose difference θ. Generally, we define the size of the local region R as in Fig. 5-6, by considering the following conditions:

$$N_{total}(\mathbf{R}(\mathbf{x})) \le \min_{\theta} N_{total}(\mathbf{R}'(\mathbf{x})) \approx N_{total}(\mathbf{R}'(45^{\circ})) \approx 20\% W \times H \,. \tag{5.4}$$

Determined by experiments, the size of the local region **R** is set at 15%W×H, as shown in Fig. 5-6.

Then, the location **x**, where the local-matching density $\rho(\mathbf{R}(\mathbf{x}))$ is a maximum, is searched as follows:

$$P = \max_{\mathbf{x}} \rho(\mathbf{R}(\mathbf{x})), \tag{5.5}$$

where P is the maximized local-matching density of the local region **R** (represented by the yellow box in Fig. 5-6(d)), which is insensitive to pose variations and which represents the similarity between the claimed gallery image and the query image.

5.3 Experimental results

The performances of our proposed face-verification methods (based on PSIFT and PPCASIFT features) are evaluated using images under pose variations, expression variations, different capture times, and alignment errors. In all the experiments, only a

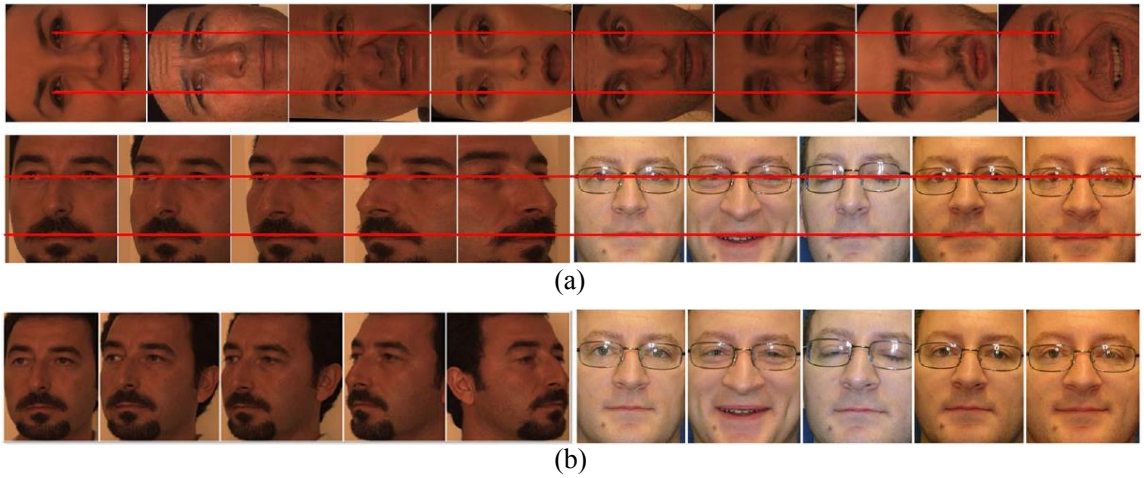
single frontal-view face of each subject is in the gallery set. To compare the performances of the different methods, we measure the receiver-operating characteristic (ROC) curve by varying a threshold to produce different false-rejection rates (FRR) and false-acceptance rates (FAR). The equal-error rate (EER), where the above two rates are equal, is also measured.

5.3.1 **Preprocessing**

The performance of the pore-scale face-verification algorithm is evaluated on three public databases: the Bosphorus dataset [202], the Multi-PIE dataset [213], and the FRGC v2.0 dataset [214]. All the face images used in the experiments are converted to gray-scale images. For each database, a single neutral, frontal-view facial image of each subject is taken for the gallery.

The Bosphorus dataset [202] contains 4,666 HR face images of 105 subjects. The Multi-PIE database [213] contains 755,370 images of 337 subjects, which were recorded over a span of 6 months. Individual attendance at sessions where the HR images were captured varies from 203 to 249 subjects. Overall, 129 subjects appear in all four sessions, which are used in the experiment. The third dataset used is the FRGC v2.0 database [214]. It contains approximately 50,000 images of over 200 subjects, which were collected about once a week from 2002 (Fall) to 2004 (Spring). In the experiments, we use the landmark information to retrieve high-quality images from the FRGC v2.0 database. A total of 9,844 images, whose number of pixels between the centers of the two eyes is larger than 280, are selected. Then, these high-quality images are divided into five sessions according to the capture time.

Since our proposed method is robust to alignment errors, we simply crop the images to include the faces only. In order to further improve efficiency, we down-sample all the cropped facial images to a resolution of about 560×670. The impact of resolution on keypoint matching was discussed in [209], which has shown that down-sampling within a certain range has a slight effect on the matching result. For feature-block matching, we partition each face image into 30×45 blocks uniformly, which is experimentally determined.



(a)

(b)

Fig. 5-7: Examples of preprocessed faces from the different datasets: (a) Facial images with different poses and expressions, which are aligned both horizontally and vertically for the alignment-required methods; (b) Facial images cropped for our proposed face-verification method, which is robust to alignment errors.

To compare our method with other standard face-verification methods wherein face alignment must be performed, all the face images are manually aligned according to the centers of the two eyes and the outer corners of the lips. All of the eyes and lips are aligned to a corresponding vertical and a horizontal line, as illustrated in Fig. 5-7. Then, the aligned faces are normalized to the same size, 560×670. Some example face images used for verification are shown in Fig. 5-7. All of these aligned images are then down-sampled

by a factor of 0.2 (i.e. to the size of 112×134), which can result in a performance that is better than or similar to using either the resolution 560×670 or down-sampled images corresponding to the factors set at 0.5 and 0.1.

Our proposed method is compared with the Eigenface method (PCA), the Gabor feature with PCA (Gabor+PCA), the LBP method [11] and the LBP feature with PCA (LBP+PCA). For the Gabor+PCA face-verification method, Gabor filters of eight orientations (0, π/8, ... , 7π/8) and five scales (π/2, π / 2$\sqrt{2}$, ... , π/8) are employed to extract the features, which are concatenated and then normalized to zero mean and unit variance. Since the Gabor features are extremely huge, PCA is applied to reduce the feature dimensionality. To retain as much information as possible, N −1 components are used, where N is the number of training samples. For LBP-based face verification, the $\mathrm{LBP}^{u2}_{(8,2)}$ operator is used, and the images are divided into 7 × 7 non-overlapping windows. All the images are down-sampled to the size of 112×134, which is similar to the image resolution used in [11]. For the LBP+PCA method, the LBP features from non-overlapping windows are concatenated to a long feature vector. Then, PCA is applied to reduce the feature dimensionality to $N - 1$. The similarity metric for the LBP method is the weighted Chi-square distance where the same weight matrix as [11] is used. For PCA, Gabor+PCA and LBP+PCA, similarity metric used is the l_2 distance.

5.3.2 Face verification with pose variations

To evaluate the robustness of the different face-verification methods to pose variations, the 105 frontal-view faces from the Bosphorus dataset were selected to form the gallery set, while images of the 5 poses (R10, R20, R30, R45, L45) form 5 testing sets, respectively.

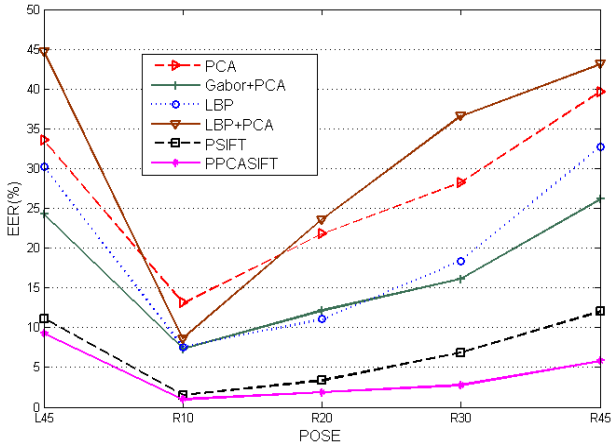


Fig. 5-8: EER of the different methods for face images under different poses.

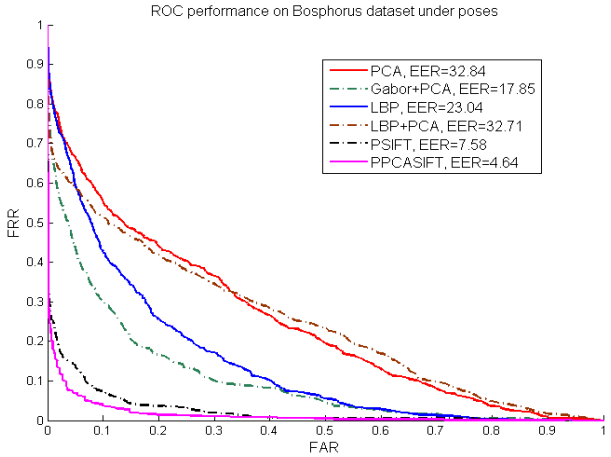
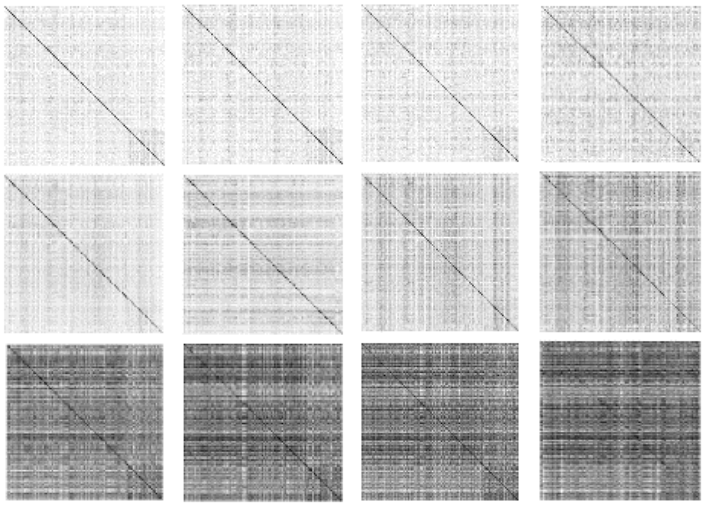


Fig. 5-9: ROC curves of the different methods for face images under all poses.

Fig. 5-8 and Fig. 5-9 show the EER results for each pose and the ROC curves under different poses, respectively. We can see that the performances of all the other methods degrade significantly when the pose variation becomes larger. Note that the performances of the LBP-based face-verification methods degrade significantly under large pose variations. This may be because the histograms representing the texture information about the frontal and non-frontal faces become more uncorrelated when the pose difference increases. The LBP method using a constant weight matrix is not suitable for face images with different poses; i.e. the prior knowledge of pose for adaptive weights is necessary. In

contrast, both the PPCASIFT- and PSIFT-based face-verification methods can maintain their performances with a lower EER under a large pose difference, such as 45 degrees. In particular, PPCASIFT achieves a slightly better performance than PSIFT, and has a feature dimension of 72 only. The result shows that PSIFT and PPCASIFT are robust to large pose variations.



(a) (b) (c) (d)

Fig. 5-10: Distance matrices of the PPCASIFT, PSIFT and Gabor+PCA methods (from top to bottom) under different poses: (a) R10, (b) R20, (c) R30, and (d) R45. PPCASIFT, which has its diagonal line the darkest, performs the best in distinguishing clients from impostors.

We also transform the PSIFT and PPCASIFT similarity metrics into distance metrics by subtracting each similarity score or matching density from one, respectively. Fig. 5-10 shows the distance matrices of the three face-verification methods with the best performance under pose variations: these are PPCASIFT, PSIFT, and Gabor+PCA, respectively. This can provide a more intuitive way of illustrating their verification performances. Both the PPCASIFT and PSIFT methods can effectively distinguish clients from impostors, and they can achieve a better performance than the Gabor+PCA method, as their results show a much darker diagonal line than the Gabor+PCA method for all four

poses. In addition, PPCASIFT performs better than PSIFT, especially under large pose variations such as R30 and R45.

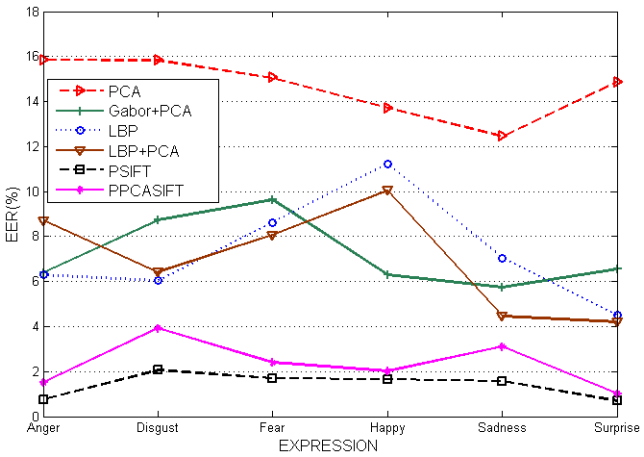


Fig. 5-11: EER of the different methods under different expressions.

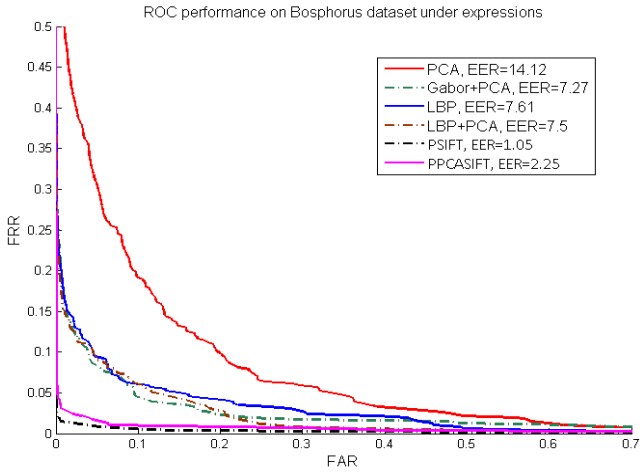


Fig. 5-12: ROC curves of the different methods under different expressions.

5.3.3 Face verification under different expressions

The verification of faces under facial-expression variations is another hot topic for real applications. We evaluate the robustness of our proposed method using images expressing 6 different emotions (anger, disgust, fear, happiness, sadness, and surprise) from the 105 subjects in the Bosphorus dataset. We compare our proposed method with

the other methods in terms of the EER for each expression and the ROC curves under expressions, as shown in Fig. 5-11 and Fig. 5-12, respectively.

From the results, all the verification methods can achieve a better EER than the results in the previous section, since all the testing faces are frontal view. The Gabor+PCA, the LBP and the LBP+PCA methods are more effective than PCA. Both the PSIFT and the PPCASIFT methods outperform the other four methods, and achieve lower EERs in all cases. However, in this expression-variation case, the PPCASIFT method with local-matching density falls a little behind PSIFT. One reason for this may be that the PPCASIFT subspace is learned only from images with a neutral expression rather than with large expression variations.

5.3.4 Face verification on faces captured in different time sessions

Our method relies on matching facial-skin regions, which involves the challenge of skin conditions changing with time. Therefore, in this section, we will evaluate the robustness of our proposed algorithm to face images captured at different times.

In this experiment, faces in the Multi-PIE dataset, which appear in all four sessions, are used. The longest time interval between the photos captured is 6 months. We select faces with a neutral expression in Session 0 to form the gallery set, while the faces captured in Session 0 with expressions, and those captured in the other three sessions, form the testing sets. The EERs and ROC curves of each face-verification method in sessions are shown in Fig. 5-13 and Fig. 5-14, respectively. The results show that our method still outperforms the other three methods for face images captured at different sessions. The superior performance of our facial-skin-based method may be due to the fact that, over time, facial appearances can change in ways other than just their skin

condition. In addition, the geometric relations between the pores in a skin region should be very stable over time.

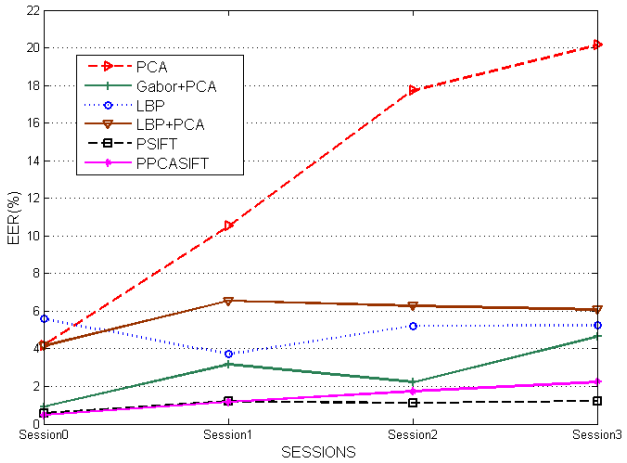


Fig. 5-13: EER of the different methods through different sessions.

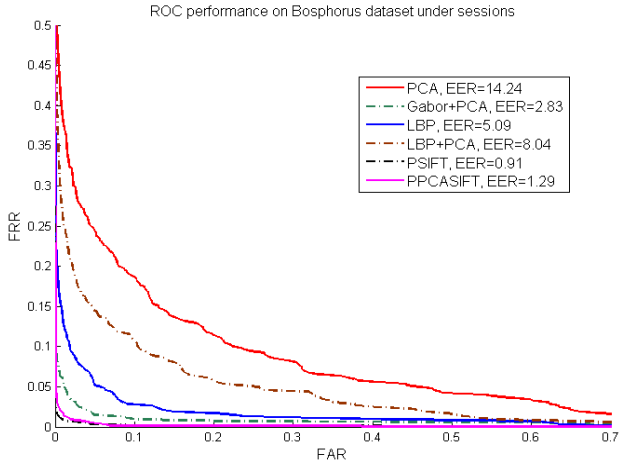


Fig. 5-14: ROC curves of the different methods through different sessions.

5.3.5 Face verification under large time span and different expressions

In order to further examine the robustness of our proposed PPCASIFT and PSIFT methods, we use the FRGC v2.0 database for an extensive face-verification experiment. After eliminating the low-quality images in uncontrolled environments (e.g. images taken in the wild with a resolution of less than 400×400), we use the remaining 9,844 images of 362 subjects for our verification task. According to the capture-time duration of each

image, a new experiment protocol is proposed. The total time span is about 400 days (58 weeks), as shown in Fig. 5-15. We used one randomly selected frontal face of each subject, taken at the beginning of this time span, as the gallery image, and we divided the remaining images into five groups (0-2 weeks, 3-10 weeks, 11-18 weeks, 19-26 weeks, and more than 26 weeks) for testing. We also compared the EERs of the different face-verification methods. The results are tabulated in Table 5-1, together with those of the previous three experiments. In general, these methods using pore-scale features can achieve much better performances when faces are under variations of pose, expression, and capture time. In particular, by extracting the PPCASIFT features, we can achieve greater efficiency while still maintaining a similar performance to PSIFT.

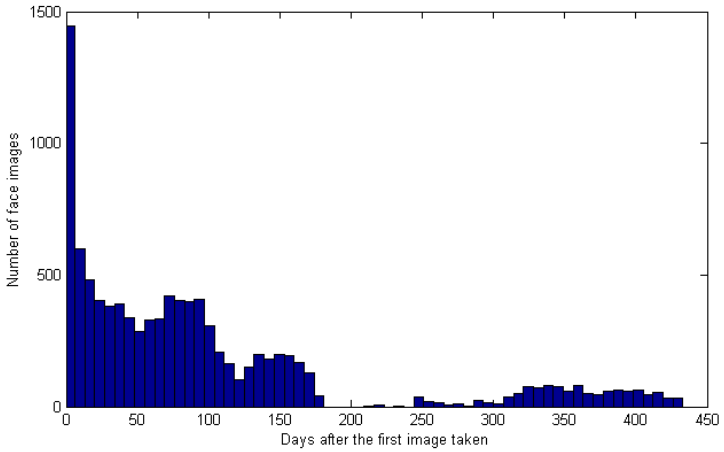


Fig. 5-15: ROC curves of the different methods through different sessions.

5.3.6 **Robustness to alignment error**

In this section, we will evaluate the performances of PSIFT and PPCASIFT in cases of different alignment errors. Those well-aligned frontal face images with neutral expression among the 105 subjects in the Bosphorus dataset are chosen to form the gallery set. The other 453 face images, expressing 6 different emotions (anger, disgust, fear, happiness, sadness, and surprise), form the testing set. All these images are normalized to

size 560×670. A random displacement vector (Δx , Δy) is added to the location of each face in the testing set, where Δx and Δy are uncorrelated and normally distributed with zero mean and a standard deviation of σ. Fig. 5-16(a) shows some samples of the testing images when σ = 50, and Fig. 5-16(b) shows the corresponding cropped regions for the PSIFT detector used to estimate the adaptive threshold for pore keypoint detection. It is obvious that the estimated thresholds will also be distorted. For PCA, Gabor+PCA, LBP and LBP+PCA, these images are further down-sampled by a factor of 0.2. Six experiments were conducted with the testing images distorted by the displacement vector with six different σ values.

Table 5-1. EER(%) of different face-verification methods for all the experiments.

Variations	Equal Error Rate(%)					
	PCA	Gabor+PCA	LBP	LBP+PCA	PSIFT	PPCASIFT
P-R10	13.11	7.34	7.60	8.16	1.49	1.00
P-R20	21.79	12.13	11.06	23.6	3.37	1.90
P-R30	28.21	16.14	18.4	6.88	36.61	2.77
P-R45	39.61	26.13	32.76	43.14	12.01	5.82
P-L45	33.54	24.25	30.2	44.75	11.14	9.28
P-All	32.84	17.85	23.04	32.71	7.58	**4.64**
E-Anger	15.85	6.39	6.3	8.72	0.76	1.51
E-Disgust	15.83	8.74	6.03	6.42	2.06	3.92
E-Fear	15.05	9.63	8.63	8.08	1.7	2.4
E-Happy	13.71	6.3	11.23	10.06	1.65	2.02
E-Sadness	12.46	5.75	7.04	4.46	1.58	3.12
E-Surprise	14.86	6.55	4.5	4.21	0.7	1.01
E-All	14.12	7.27	7.61	7.5	**1.05**	2.25
T-Session0	4.17	0.91	5.59	5.44	0.55	0.47
T-Session1	10.51	3.18	3.71	8.56	1.2	1.18
T-Session2	17.73	2.22	5.22	8.28	1.11	1.74
T-Session3	20.13	4.64	5.24	8.09	1.21	2.24
T-All	14.24	2.83	5.09	8.04	**0.91**	1.29
FRGC-0-2W	15.6	4.21	8.28	8.02	2.01	3.3
FRGC-3-10W	27.11	9.67	12.87	12.2	4.25	6.75
FRGC-11-18W	33.63	12.05	16.23	15.54	6.32	8.21
FRGC-19-26W	28.94	9.62	14.2	12.76	5.6	7.7
FRGC-Aft26W	28.75	14.26	18.09	16.14	8.11	10.55
FRGC-All	27.57	10.02	14.12	13.16	**5.51**	7.36

The EERs of PSIFT and PPCASIFT, as well as other face-verification algorithms, with different alignment errors are summarized in Table 5-2. It can be seen that PSIFT

and PPCASIFT are robust to different alignment errors, and perform well even when suffering from a large alignment error. Gabor+PCA, LBP and LBP+PCA work well only when there is a small or no alignment error; their performances drop significantly when the alignment error increases. Of these methods, PCA is sensitive to alignment errors even with small σ values.

(a)

(b)

Fig. 5-16: Face images with a displacement vector added: (a) Samples of testing images when σ = 50, and (b) the corresponding cropped regions for the PSIFT detector used to estimate the adaptive threshold.

Table 5-2. EER(%) of the different face-verification methods with different alignment errors.

σ of $(\Delta x, \Delta y)$	Equal Error Rate(%)					
	PCA	Gabor+PCA	LBP	LBP+PCA	PSIFT	PPCASIFT
0	14.12	7.27	7.61	7.50	**0.77**	1.94
10	15.23	7.35	7.89	7.96	**0.73**	2.20
20	19.49	9.38	13.88	11.26	**0.87**	2.12
30	21.86	14.06	21.98	14.38	**0.79**	2.59
40	24.90	16.97	25.22	17.79	**0.63**	2.30
50	29.13	22.97	31.06	21.59	**0.97**	2.60

5.4 Conclusions

In this chapter, we have addressed the problem of HR face verification based on pore-scale facial features. The proposed method is robust to alignment errors and pose variations, while the gallery set requires only a single image per subject. The PSIFT and PPCASIFT features are highly distinctive, and PPCASIFT can efficiently reduce the

computational time of the matching process to about 9% of that of PSIFT, while a similar performance level can be maintained. For each query in the feature-matching stage, PSIFT needs 1.45 seconds, while PPCASIFT needs only 0.13 seconds on an Intel i7 3.4GHz CPU with 8 threads and 8GB Ram PC under the MATLAB R2014a programming environment. These runtimes can be further reduced by using GPU parallel computing techniques. The runtime of the robust-fitting stage is less than 0.01 seconds. Furthermore, the proposed parallel-block-aggregation and matching-density schemes can be applied to other image analysis tasks such as object recognition, image annotation, since they provide an approach to transforming point matching into similarity measurement. Experimental results have shown that our method can achieve a superior performance under a range of variations, especially under large pose variations. To the best of our knowledge, this is the first work on HR face verification that uses pore-scale facial features and establishes such a large number of correspondences between faces. In addition, our proposed face-verification method can tackle pose, expression, and capture-time variations simultaneously.

Chapter 6. Feature-aging for age-invariant face recognition

6.1 Introduction

A lot of research on human face recognition has been conducted in the past three decades. Various face-recognition algorithms which can deal with faces under different facial expressions, lighting conditions, and poses have been proposed, and can achieve satisfactory performances. However, the changes in face appearance caused by age progression have received limited attention to date; this effect has a significant impact on the face-recognition algorithms.

In this chapter, we consider both the forward and backward prediction of aged features for face recognition. With a live query input, the age difference between the query face image and a gallery face in a face database can be computed, if the time, when the gallery face was captured, is known. Otherwise, age estimation is needed to find out the age difference between the query and the gallery face images. For face recognition, different features can be used, such as Gabor wavelets [215], Local Binary Pattern (LBP) [11], Locality Preserving Projections (LPP) [58], etc. In general, the age of the query input should be older than that of the gallery face. In our algorithm, we extract the Gabor features at a number of landmarks in the face images. Based on the estimated age difference between the two faces, backward prediction is performed for the Gabor features of the query face image, so that the Gabor features at a younger age are generated. Similarly, the Gabor features of the gallery face image are forward predicted to generate the corresponding features at an older age. In other words, to compare the query and gallery faces, we use both the features at the two different ages. As the facial features of

a person at two different ages should be correlated with each other, Canonical Correlation Analysis [216] (CCA) is employed to the two sets of features, at different ages, to generate more coherent features for face recognition.

This chapter is organized as follows. Section 6.2 describes the details of our proposed method, including the learning of the forward and backward prediction, and the use of CCA to generate coherent features. Experiment setup and results are given in Section 6.3, and conclusions are provided in Section 6.4.

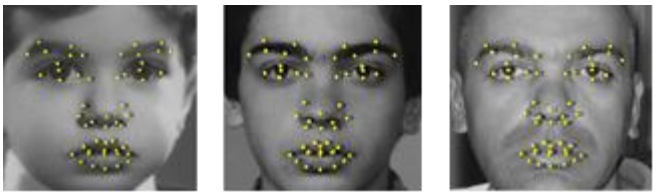


Fig. 6-1: Faces used for experiments are cropped and aligned with two eyes, with 53 facial feature points denoted as yellow dots.

6.2 Age-invariant face recognition based on feature-aging

In this section, we will first present the features used in our algorithm. Then, based on the corresponding features from two sets of training face images having a similar age, linear regression is employed to learn the forward prediction and the backward prediction of the facial features from a younger age to an older age and from an older age to a younger age, respectively. For the face images used in our experiments, 68 feature points have been labeled. As those feature points located on the face contour are not reliable for recognition, only 53 feature points are used in our algorithm, as shown in Fig. 6-1. The forward and backward prediction for each of these feature points will be learned. Having predicted the facial features of a face image at another age, CCA is then introduced to

combine the features, at two different ages, to produce a coherent feature for face recognition.

6.2.1 Feature extraction

Gabor wavelets (GW) have been commonly used as local features for many pattern recognition and computer vision applications, such as texture retrieval, object detection, recognition, etc. It was found in [217, 218] that the Gabor functions are similar to the response of the two-dimensional receptive field profiles of the mammalian simple cortical cell. The Gabor features exhibit the desirable characteristics of capturing salient visual properties such as spatial localization, orientation selectivity, and spatial frequency selectivity. In the spatial domain, a GW is a complex exponential modulated by a Gaussian function, which is defined as follows [219]:

$$\psi_{\omega,\theta}(x,y)=\frac{1}{2\pi\sigma^2}e^{-\frac{(x\cos\theta+y\sin\theta)^2+(-x\sin\theta+y\cos\theta)^2}{2\sigma^2}}\cdot[e^{i(\omega x\cos\theta+\omega y\sin\theta)}-e^{-\frac{\omega\sigma^2}{2}}], \tag{6.1}$$

where (x, y) represent the pixel position in the spatial domain, ω is the radial center frequency of the complex exponential, θ is the orientation of the GW, and σ is the standard deviation of the Gaussian function. By using different center frequencies and orientations, a family of Gabor kernels can be produced, which can then be used to extract features from an image. In [63, 89], GWs have been employed for face recognition, and a promising performance can be achieved.

In the Gabor feature, only the Gabor magnitudes are used because the Gabor phases change linearly with small displacements. Gabor filters at five different scales and eight different orientations are adopted in our algorithm. Consequently, at each of the 53 facial

feature points, its Gabor feature is formed by concatenating the outputs of the 40 (5×8) filters.

6.2.2 Forward and backward feature-aging

In our algorithm, we predict the Gabor features of a face image from one age to another age. In our training set, we assign training face images into different groups according to their ages. Consider two different age groups, which are denoted as AG1 and AG2, respectively. Here, we assume that images in AG1 are younger than those in AG2. The training samples are denoted as $\mathbf{U}^1 = [\boldsymbol{u}_1^1, ..., \boldsymbol{u}_M^1]$ and $\mathbf{U}^2 = [\boldsymbol{u}_1^2, ..., \boldsymbol{u}_M^2]$ for AG1 and AG2, respectively, where *M* is the number of training pairs and $\boldsymbol{u}_j^i$ represents the Gabor features of the *j*th training sample of the *i*th age group at a particular landmark position in the face image. Linear regression is employed in the prediction. If the Gabor features of AG2 are predicted from AG1, "forward prediction" is performed and the Gabor features are under forward aging. Similarly, "backward prediction" is performed when we generate features from AG2 to age AG1.

For forward prediction, we assume that there is a linear mapping from the features $\mathbf{U}^1$ to $\mathbf{U}^2$, as follows:

$$f : \mathbf{U}^1 \rightarrow \mathbf{U}^2 . \tag{6.2}$$

The mapping function should be complicated and nonlinear. However, to simplify the learning of the mapping function, linear mapping is adopted to obtain the approximated aged features. Therefore, the mapping can be written as follows:

$$\mathbf{U}^2 = \mathbf{A}_f \mathbf{U}^1 , \tag{6.3}$$

where $\mathbf{A}_f$ is the forward mapping matrix, which can be computed by solving (6.3) as $\mathbf{U}^1$ and $\mathbf{U}^2$ are known. The dimension of $\mathbf{U}^1$ and $\mathbf{U}^2$ are $40\times M$, while the dimension of $\mathbf{A}_f$ is 40×40. $\mathbf{A}_f$ is computed as follows:

$$\mathbf{A}_f = \mathbf{U}^2(\mathbf{U}^1)^+ \tag{6.4}$$

where $(\mathbf{U}^1)^+ = (\mathbf{U}^1)^T(\mathbf{U}^1(\mathbf{U}^1)^T)^{-1}$ is the pseudo inverse of $\mathbf{U}^1$ and T represents the transpose operation. Having learned the linear mapping function $\mathbf{A}_f$ based on the training samples, a given Gabor feature can be aged from AG1 to AG2, as follows:

$$\boldsymbol{u}_j^2 = \mathbf{A}_f \boldsymbol{u}_j^1 . \tag{6.5}$$

Similarly, when Garbor features at AG2 are available, we can predict the corresponding features at AG1 by learning the backward mapping function $\mathbf{A}_b$, as follows:

$$\mathbf{U}^1 = \mathbf{A}_b\mathbf{U}^2 , \tag{6.6}$$

where $\mathbf{A}_b = \mathbf{U}^1(\mathbf{U}^2)^T(\mathbf{U}^2(\mathbf{U}^2)^T)^{-1}$.

6.2.3 Face recognition based on feature-aging

Two face images are matched based on the difference between their local Gabor features. Assume that p feature points are located in each face image, where the Gabor features are extracted. Therefore, p forward and backward prediction matrices are learned, as described in Section 6.2.2, to change the ages of the Gabor features, according to the age difference between the query and gallery faces.

Assume that a gallery face is in AG1, while the query face is in AG2. Denote the Gabor feature of the gallery face and query face at the ith feature point as $\boldsymbol{u}_i^{g1}$ and $\boldsymbol{u}_i^{q2}$, respectively. For the gallery face, its respective Gabor features are subject to the forward mapping so as to produce corresponding features of AG2, as follows:

$$\boldsymbol{u}_i^{g2} = \mathbf{A}_f^i \boldsymbol{u}_i^{g1}, \tag{6.8}$$

where $\mathbf{A}_f^i$ is the forward mapping function for the *i*th feature point, and $\boldsymbol{u}_i^{g2}$ is the predicted feature at AG2. Similarly, for the Gabor features of the query input, backward mapping is applied, as follows:

$$\boldsymbol{u}_i^{q1} = \mathbf{A}_b^i \boldsymbol{u}_i^{q2}, \tag{6.9}$$

where $\mathbf{A}_b^i$ is the backward mapping function for the *i*th feature point, and $\boldsymbol{u}_i^{q1}$ is the predicted feature at AG1. Finally, these features at different ages are projected into corresponding subspaces to form coherent features, as follows:

$$\boldsymbol{v}_i^g = \begin{pmatrix} \boldsymbol{\alpha}^T \boldsymbol{u}_i^{g1} \\ \boldsymbol{\beta}^T \boldsymbol{u}_i^{g2} \end{pmatrix} \text{ and } \boldsymbol{v}_i^q = \begin{pmatrix} \boldsymbol{\alpha}^T \boldsymbol{u}_i^{q1} \\ \boldsymbol{\beta}^T \boldsymbol{u}_i^{q2} \end{pmatrix}, \tag{6.10}$$

where $\boldsymbol{v}_i^g$ and $\boldsymbol{v}_i^q$ are the features for the gallery face and the query face, respectively. $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the projection directions learnt from the generalized CCA [220]. Euclidean distance between the two feature vectors is computed, and the nearest neighbor rule is used for face recognition.

6.3 Experimental results

6.3.1 Experimental setup

To compare our proposed age-invariant face recognition method with other standard methods, we conducted our experiment using the FGNet Aging Database [146]. It contains 1,002 face images of 82 people with ages ranging from 0 to 69, which is one of the most widely used datasets having the largest age range. However, some of the images in the database are of low quality, and suffer from blurring, camera artifacts, lighting variations, etc. We have further reduced the number of images, and have only retained 126 images of

42 people, which have relatively better quality and less variations. Each person has three images, each at a different age stage. All of these images are cropped to contain the face region only, with a size of 240×240, and are aligned based on both eyes, as shown in Fig. 6-1.

In order to learn the progression between different ages, we divide the images into three age groups, namely AG1 for ages from 0 to 6, AG2 from 7 to 18, and AG3 from 19 to 69. In the training phase, local Gabor features, at each of the 53 facial feature points in the training images, are extracted. Then, the linear mapping function and CCA projection matrices are learned for each pair of age groups. After training, the mapping function and the projection matrices are used to generate Gabor features from one age to another age and to form a coherent feature for face recognition. In our experiment, we randomly chose 12 people, with their images in three different age groups, for training each time, and the remaining people with different ages are used for testing. The experiments were repeated five times, so the performance is measured based on 150 images.

6.3.2 Experimental results and analysis

Firstly, we compare our proposed age-invariant face recognition method with the Gabor and the local Gabor methods so as to evaluate the improvement after combining aged feature prediction by Linear Mapping (LM) and generalized CCA (GCCA). The comparison results, in terms of recognition rates, are summarized in Table 6-1, where AgeInvLPG is our proposed method. We have also examined the improvement due to the use of aged feature prediction without performing fusion by GCCA, and the method is denoted as AgeInvLP. For the local Gabor method, only those Gabor features at the 53 feature points are considered. For the Gabor method, the Gabor features of the whole faces

are used for representation. All the methods, compared in the experiment, have the same experimental setup, as described in Section III.A. Both the Gabor and local Gabor features are extracted using Gabor filters of five different scales and eight orientations.

Table 6-1. Face recognition rates (%) of the Gabor methods and Age-invariant methods with face images from different age groups.

Gallery vs. Query	Gabor	Local Gabor	AgeInvLP	AgeInvLPG
AG1 vs. AG2	4	8.7	10.5	**15.3**
AG1 vs. AG3	12.7	12	15.2	**19.3**
AG2 vs. AG1	16	13.3	33.3	**50.7**
AG2 vs. AG3	28.7	34	36.8	**58.7**
AG3 vs. AG1	12.7	10.7	14.3	**16.7**
AG3 vs. AG2	10	6	22.5	**32.7**

From Table 6-1, we can observe that the method based on local Gabor features, which only considers the features at the facial-feature points, can achieve a similar performance to the method based on all the Gabor features. The local Gabor method has a much smaller size than the Gabor method, so it is much more computationally efficient. However, we can see that both of these Gabor-based methods cannot recognize faces accurately due to the fact that the Gabor features have significant differences when faces have a great age difference. With our proposed age-invariant Gabor framework, a significant increase in recognition rates can be achieved, especially when the gallery and query face images come from two neighboring age groups. For the AgeInvLP method, i.e. aged features are predicted but no fusing by GCCA is performed, the recognition rate is increased by 1.8% to 20%, depending on the age difference between the face images. When compared to AgeInvLP, the AgeInvLPG method, i.e. the two features at different ages are fused by using GCCA, the improvement in terms of recognition rate is between 2.4% and 21.9%.

This shows that the coherent feature, generated by projecting two features at different ages, is effective for age-invariant face recognition.

To give more comprehensive analysis, we have also compared our proposed age-invariant face recognition method, i.e. AgeInvLPG, with some conventional face recognition methods, including eigenface, LBP [11], and LPP [58]. For eigenface, all the principal components are selected for recognition, i.e. $M-1$ principal components. For LPP, the number of Laplacian faces used is also $M-1$. We also follow the same nearest-neighbor classifier approach for recognition, and set the number of nearest neighbors at 7. The comparison results, in terms of recognition rates, are summarized in Table 6-2.

Table 6-2. Face recognition rates (%) of different conventional face recognition methods and our proposed method with face images from different age.

Gallery vs. Query	PCA	LBP	LPP	AgeInvLPG
AG1 vs. AG2	6	8.5	9.9	**15.3**
AG1 vs. AG3	8	12.7	8.3	**19.3**
AG2 vs. AG1	18	40.4	24.6	**50.7**
AG2 vs. AG3	10.7	49.9	36.3	**58.7**
AG3 vs. AG1	3.3	11.8	5.5	**16.7**
AG3 vs. AG2	6.7	28.8	22.6	**32.7**

From the results, we can see that our proposed algorithm can achieve better performance in terms of recognition rates. The eigenface and the LPP methods cannot work well when the face images, to be compared, have a large age difference, in particular, when either the query or gallery image involved belongs to the youngest age group, i.e. in the range of 0-6 years old. This may be due to the fact that the appearances of babies, and young children, change a lot in the first few years after birth. The LBP method can achieve better recognition performance than both PCA and LPP because it can capture more local

texture information about faces. However, the performance of LBP still falls behind our proposed age-invariant Gabor-based method. When our method is compared to [39], which uses age-insensitive features for face recognition, the improvement in terms of the rank-1 recognition rate is similar. However, our approach can be applied to [39] to further improve its performance. This means that our proposed approach is effective and efficient for age-invariant face recognition.

6.4 **Conclusions**

In this chapter, we have proposed a novel approach for age-invariant face recognition by predicting facial features at different ages. Since features at different ages should be correlated to each other, a coherent feature is formed by fusing the two features at different ages by the generalized CCA. By comparing it to those existing face recognition methods on face images with age differences, our algorithm can improve the recognition rate significantly.

A challenge of the research is that only a limited number of training samples are available. Therefore, in our experiments, we simply divide images into three age groups. Nevertheless, experiment results have proven the effectiveness of our proposed algorithm.

Chapter 7. Age-invariant face recognition based on identity inference from appearance age

7.1 Introduction

As mentioned in Section 1.2, current methods on age-invariant face recognition are based on real-age labels, which may be inconsistent with the corresponding appearance ages, thus can only achieve limited performances. One way to solve the age-gap problem is to seek the underlying sequential patterns [154, 221], and then apply manifold learning to analyze the age characteristics [153, 193]. It has been shown that applying the orthogonal Locality Preserving Projections (OLPP) [194] to an aging database, with ages ranged from 0 to 93 years, yields better statistical age-estimation results.

With the increasing interest in age-related topics, the corresponding tasks have become more and more demanding in recent years. The apparent age estimation challenge on the public ChaLearn dataset [2], with face images in the wild and labeled with the appearance age, is one of the great sources for learning age progression.

In this chapter, we propose an age-invariant face recognition framework, namely aging-guided identity inference model (AG-IIM), where the feature gap between two face images of the same person, captured at different ages, can be reduced. Similar to the method that deals with the aging and identity information separately [40, 156, 164], we establish a generative model based on probabilistic Linear Discriminant Analysis (PLDA). Unlike those previous works, which learned and derived the aging and identity subspaces at the same time, we propose to learn aging spaces separately by using manifold learning on an aging dataset with appearance-age labels. We empirically show that our method can obtain a more discriminative identity subspace. Besides, our method is the first to tackle

the age-invariant face recognition problem based on appearance age, so that computers can both learn more effectively and more consistently. A byproduct of this framework is to provide a much easier way to collect aging photos for face recognition, where only identity labels are required. The aging characteristics can be automatically learnt by any aging dataset with appearance-age labels. Having obtained the identity and aging subspaces, as well as the underlying identity factors based on different features, an effective fusion mechanism based on Canonical Correlation Analysis (CCA) [197] is utilized to further boost the recognition performance. Extensive experiments on three different aging datasets show that our framework can achieve a great improvement in terms of the rank-1 recognition accuracy compared to other state-of-the-art methods, especially when the faces undergo large age variation.

This chapter is organized as follows. Section 7.2 describes the details of our method, where the proposed aging-guided identity inference model and age-invariant face recognition framework will be presented. Experiment results and analysis of face recognition on different aging datasets are given in Section 7.3, and conclusions are provided in Section 7.4.

7.2 Ageing-guided identity inference model for age-invariant face recognition

In this section, the proposed identity inference model, based on PLDA and independent aging subspace learning, is presented. Unlike the previous work that jointly learns the subspaces for aging and identity at the same time, we first derive a discriminative aging subspace by preserving the locality of the appearance-age information, and then the identity subspace with the assistance of the aging subspace

through the PLDA model. This strategy makes optimization on the latent variables more efficient, and the collection of aging face images easier, as only the identity label is required.

7.2.1 **Identity inference model**

Similar to [40, 164], which use PLDA for face recognition, we also model a face by incorporating both the within-individual variation (aging) and the between-individual variation (identity) by fitting in the model introduced in Eqns (2.3) to (2.6):

$$\boldsymbol{x}_{mn} = \boldsymbol{\mu} + \mathbf{E}\boldsymbol{u}_m + \mathbf{A}\boldsymbol{v}_{mn} + \boldsymbol{\varepsilon}_{mn}, \tag{7.1}$$

where the first two terms are comprised of the identity components $\boldsymbol{\mu}$ and $\mathbf{E}\boldsymbol{u}_m$, which depend only on the identity of the person, while the last two terms are comprised of the aging components $\mathbf{A}\boldsymbol{v}_{mn}$ and $\boldsymbol{\varepsilon}_{mn}$, which are different for images of the same individual and represent the within-individual noise.

Before applying the EM algorithm, we can rewrite the model in (7.1) using matrix form:

$$\begin{aligned} \boldsymbol{x}_{mn} &= \boldsymbol{\mu} + [\mathbf{E} \quad \mathbf{A}] \begin{bmatrix} \boldsymbol{u}_m \\ \boldsymbol{v}_{mn} \end{bmatrix} + \boldsymbol{\varepsilon}_{mn} \\ &= \boldsymbol{\mu} + \quad \mathbf{B} \quad\quad \boldsymbol{z}_{mn} \quad + \boldsymbol{\varepsilon}_{mn} \end{aligned}. \tag{7.2}$$

In the E-step, the model statistics of the first two moments of the Gaussian function are computed as follows [164]:

$$E[\boldsymbol{z}_m] = (\mathbf{B}^T \boldsymbol{\Sigma}'^{-1} \mathbf{B} + \mathbf{I})^{-1} \mathbf{B}^T \boldsymbol{\Sigma}'^{-1} (\boldsymbol{z}_m - \boldsymbol{\mu}'), \tag{7.3}$$

$$E[\boldsymbol{z}_m \boldsymbol{z}_m{}^T] = (\mathbf{B}^T \boldsymbol{\Sigma}'^{-1} \mathbf{B}^T + \mathbf{I})^{-1} + E[\boldsymbol{z}_m] E[\boldsymbol{z}_m]^T, \tag{7.4}$$

where $\boldsymbol{\mu}' = [\boldsymbol{\mu}, \boldsymbol{\mu}, ..., \boldsymbol{\mu}]^T$, and $\boldsymbol{\mu}$ is the mean face vector of all the N_m images belonging to the same identity, and $\boldsymbol{\Sigma}'$ is the diagonal matrix of $\boldsymbol{\Sigma}$, where

$$\boldsymbol{\Sigma}' = \begin{pmatrix} \boldsymbol{\Sigma} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{\Sigma} \end{pmatrix}. \tag{7.5}$$

In the M-step, we update the model parameters using the following rules:

$$\boldsymbol{\mu} = \frac{1}{\sum_m N_m} \sum_{N_m} \boldsymbol{x}_{mn}, \tag{7.6}$$

$$\mathbf{B} = \left(\sum_{m,n} (\boldsymbol{x}_{mn} - \mu) E[\boldsymbol{z}_m]^T \right) \left(\sum_{m,n} E[\boldsymbol{z}_m \boldsymbol{z}_m{}^T] \right)^{-1}, \text{ and} \tag{7.7}$$

$$\boldsymbol{\Sigma} = \frac{1}{\sum_m N_m} \sum_{m,n} \mathbf{Diag}\left[(x_{mn} - \mu)(x_{mn} - \mu)^T - \mathbf{B} E[z_m](x_{mn} - \mu)^T \right]. \tag{7.8}$$

where **Diag**(*) represents the operation of retaining only the diagonal elements from a matrix, and the updated **E** and **A** are computed from the new **B** using the equivalence between Eqns (7.3) and (7.4).

When looking at the model in (7.1) again, both **E** and **A** can be viewed as the corresponding identity and aging subspaces. Thus, solving the probabilistic problem in can obtain the two subspaces at the same time [40]. In order to recognize face images accurately, the training samples are required to have both correct identity labels and aging labels at the same time; this imposes great difficulty in labeling the collected face samples. What's more, the desired identity and aging subspaces should be as independent of each other as possible, so as to separate the identity and aging factors as much as possible. However, using the identity and aging labels from the same dataset will inevitably lead to some correlation. These two practical problems push us to raise a bold question – can we obtain an ideal aging subspace independently from other aging datasets? The answer is

yes! We will show the details of establishing the aging subspace, and then compare it with the aging subspace, jointly learnt from the PLDA model in Section 7.2.2.

Algorithm 7-1 Identity inference model learning

Input: The independent aging dataset **Z** with age group label **L**, $\{\boldsymbol{z}_l \,|\, l = 1,2,...,L\}$ based on appearance age and training dataset **X** with identity label **M**, $\{\boldsymbol{x}_m \,|\, m = 1,2,...,M\}$.

Output: Independent aging subspace **A** and dataset-specific identity subspace **E**.

% Independent aging subspace learning %

1) Construct the adjacency graph within each age group using (7.9), where $w_{ij} = \exp(-\|\boldsymbol{z}_i - \boldsymbol{z}_j\|^2 / t)$ if *i* is among the *k* nearest neighbors of *j* in the same age group, or if *j* is among the *K* nearest neighbors of *i* in the same age group, otherwise $w_{ij} = 0$.

2) Compute the eigenvectors and eigenvalues for the generalized eigenvector problem in (13).

3) Select the *p* eigenvectors, with the corresponding smallest eigenvalues, to form the aging subspace **A** in (7.1), where $\mathbf{A} = [\boldsymbol{a}_0, \boldsymbol{a}_1, ..., \boldsymbol{a}_{p-1}]$.

% Identity subspace learning %

1) Initialize the model parameter $\theta = \{\mathbf{E}, \mathbf{A}, \boldsymbol{\mu}, \Sigma\}$, where

$\mathbf{E} = \text{rand}(*)$, $\mathbf{A} = [\boldsymbol{a}_0, \boldsymbol{a}_1, ..., \boldsymbol{a}_{p-1}]$, $\boldsymbol{\mu} = \dfrac{1}{\sum_m N_m} \sum_{N_m} \boldsymbol{x}_{mn}$ and $\sigma^2 = 0.1$.

Iterate:

2) E-step: update the latent variables $\boldsymbol{z}_m$ using (7.3) and (7.4).

3) M-step: update the model parameters $\theta = \{\mathbf{E}, \Sigma\}$ using (7.7) and (7.8).

Until convergence

After obtaining the desired aging subspace, we then aim at optimizing the model in (7.1) by finding the parameters $\theta' = \{\mathbf{E}, \boldsymbol{\mu}, \Sigma\}$ and the latent variables $\boldsymbol{u}_m$ and $\boldsymbol{v}_{mn}$. As the aging subspace **A** is kept unchanged during the learning of the identity model, the EM algorithm can converge faster, with five to ten iterations only in our algorithm. We call

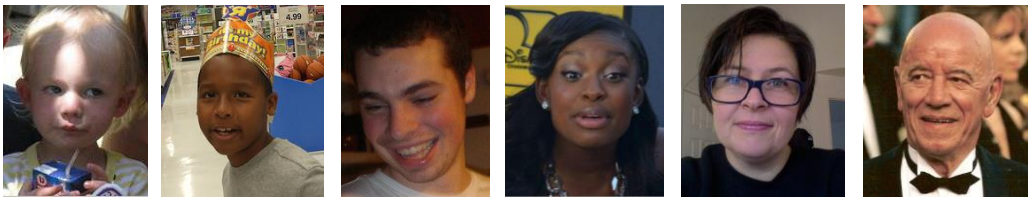
this revised model as an identity inference model. The algorithm for this part of model learning is summarized in Algorithm 7-1.

7.2.2 Independent aging subspace learning

As mentioned in the previous biometric studies [221], faces can be considered as points in a high-dimensional space, where aging is reflected by the distance of the face from the average of all face samples. It has been proved in [153, 193] that human aging effects can be projected onto a discriminant subspace using manifold learning, where it has a significant trend for sequential patterns. These works also used this subspace for image-based human age estimation on their own aging dataset, with ages ranged from 0 to 93 years. The findings enlighten us on the aging subspace learning with manifold, which can be later used in the identity inference model. However, how to select a suitable aging face dataset and aging features for learning remains a difficult task.

7.2.2.1 Aging dataset for aging subspace learning

Recently, a new age-related dataset named Chalearn [2], whose face images are labeled with appearance ages, has been released. It is known to be the first dataset labeled with the appearance age instead of the real age. It contains 8,000 images, where the age of the face in each image was labeled by multiple individuals, and the average is taken as the appearance age. All the images are in the wild environments with real-life variations, and some of them are shown in Fig. 7-1. At the time we conducted our experiments, only the training set with 4,113 images of ages ranged from 1 to 86 years old had been released. Therefore, we used all these images, labeled with appearance ages, for the aging subspace learning.



Age 2 Age 10 Age 20 Age 30 Age 45 Age 65

Fig. 7-1: Some face examples from the aging dataset Chalearn [2] with appearance-age labels.

7.2.2.2 **Aging subspace learning**

It has been shown in [153, 193] that Locality Preserving Projection (LPP) [70] and its orthogonal variant OLPP [194] are able to project faces onto a more discriminative subspace, and characterize the age manifold better than Principal Component Analysis (PCA) and Locally Linear Embedding (LLE). Thus, OLPP is employed in our algorithm, which aims to preserve local structure based on the assumption that a nearest-neighbor search in the low-dimensional space will yield similar results to that in the high-dimensional space.

In the LPP theory, the objective function is defined as:

$$\sum_{ij}(\mathbf{z}_i - \mathbf{z}_j)^2 w_{ij} . \tag{7.9}$$

The weight w_{ij} is based on the heat kernel, and is defined as $w_{ij} = \exp(-\left\|\mathbf{z}_i - \mathbf{z}_j\right\|^2 / t)$ when the two face features $\mathbf{z}_i$ and $\mathbf{z}_j$ are the K nearest neighbors of each other, otherwise $w_{ij} = 0$. The weight matrix $\mathbf{W}=[w_{ij}]$ is symmetric, a diagonal matrix $\mathbf{D}=[d_{ij}]$, whose entries are the column sums of W, i.e. $d_{ii} = \sum_j w_{ij}$, and the corresponding Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{W}$, can be computed. Then, the optimal projections can be obtained by solving the following eigenproblem:

$$\mathbf{ZLZ}^T\boldsymbol{a} = \lambda\mathbf{ZDZ}^T\boldsymbol{a}, \tag{7.10}$$

where the solutions are the column vectors $\{\boldsymbol{a}_0,...,\boldsymbol{a}_n\}$, which are the eigenvectors of $(\mathbf{ZDZ}^T)^{-1}\mathbf{ZLZ}^T$, with their eigenvalues in ascending order, i.e. $\lambda_0 < ... < \lambda_n$.

In our algorithm, we apply OLPP to derive the orthogonal basis functions using (7.10) iteratively, as described in [194]. Since the appearance-age labels are given, we can further improve the learned manifold with supervised learning by utilizing the age-group label information. The weight w_{ij} is non-zero only for two face samples being the K nearest neighbors of each other and within the same age group, otherwise it is zero. The ages are partitioned into 12 groups, with the groups for the younger and older ages having smaller age intervals, as shown in Table 7-1.

Table 7-1. The partitioning of ages into groups for the Chalearn dataset.

Group	Age range	#Image	Group	Age range	#Image
1	1-3	130	**7**	26-35	1222
2	4-6	147	**8**	35-45	602
3	7-10	96	**9**	46-55	367
4	11-15	89	**10**	56-60	142
5	16-20	396	**11**	61-65	71
6	21-25	772	**12**	66-89	79

For each face image in the Chalearn dataset, we first locate the two eyes and align the face, based on the eye positions as in [222]. We crop the faces to include the face regions only, and normalize them to the size 126×126 pixels. In order to alleviate the illumination impact, we normalize all the faces to have zero mean and unit variance. Previous work on age estimation [153, 193] used the whole face for feature extraction, which is not suitable for images under real-life variations. What's more, only using one feature is not sufficient for recognition tasks in the wild. In order to find the best features, supervised OLPP is applied to several state-of-the-art features, and the best two of these features, namely the

multi-scale weighted Local Binary Patterns (wLBP) [11] and the Histogram of Oriented Gradients (HOG) [120] (specific configurations are given in the experimental session), are selected by using the same empirical mechanism as in [153]. As explained in [80], the LBP feature is able to capture the local texture information about a face, while the HOG feature represents the edge structure of a face well. In this sense, they are complementary to each other, and they are simple and fast to implement.

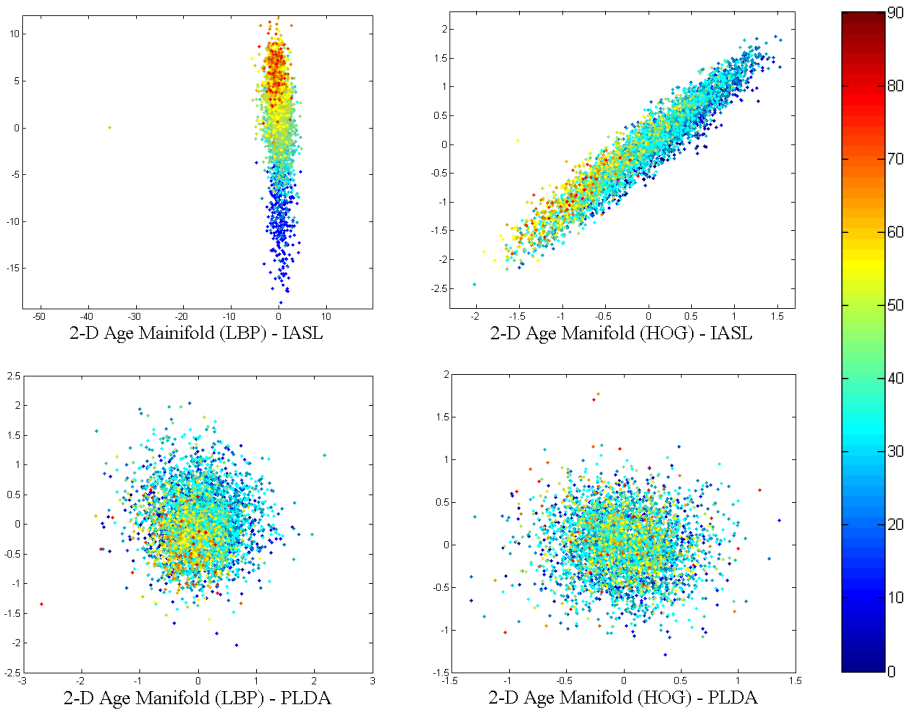


Fig. 7-2: 2-D age manifold visualization. (a) The two manifolds based on the wLBP features by using our Independent Aging Subspace Learning (IASL) and PLDA, respectively; and (b) the two manifolds based on the HOG features by using IASL and PLDA, respectively.

We have studied the 2-D age manifolds on the different features to determine whether they can provide a distinct age progression. We have also applied PLDA on the Chalearn dataset to derive the aging subspace (where its basis are the columns of the matrix **D** in

(7.1)). Fig. 7-2 illustrates the corresponding 2-D age manifolds, which shows that our aging subspaces learnt from two features have much more distinctive patterns of aging progression than those obtained from the PLDA model. Fig. 7-3 gives further visualization of how the identity subspace captures the faces with different appearances (**E** varies while **A** stays constant) and how the aging subspace illustrates the faces with different ages (**A** varies while **E** stays constant). More experiment results are given in Section 7.3.

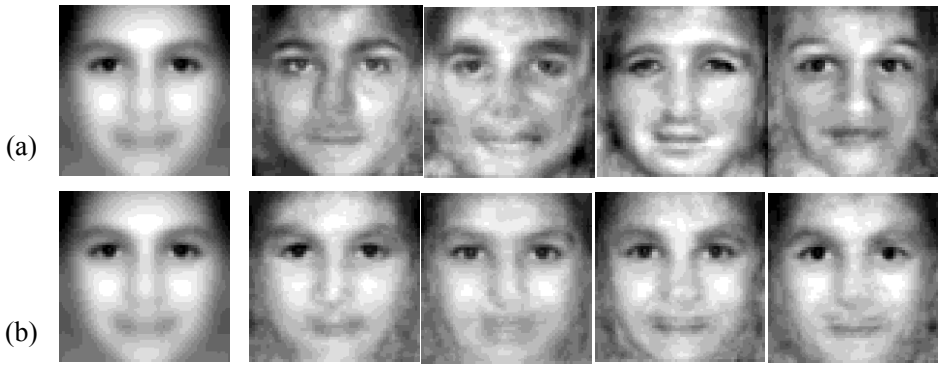


Fig. 7-3: Visualization of the identity inference model. (a) The mean face of all the face images and the faces in four directions in the identity subspace, where all the images look like different persons; and (b) the mean face and the faces in four directions in the aging subspace, where all images look like the same person but at different ages.

7.2.3 Face recognition based on identity inference model

In this section, the overall framework of the age-invariant face recognition, based on the proposed identity inference model, is presented. Fig. 7-4 shows the overall framework, which includes pre-processing steps, such as feature extraction and dimension reduction, face recognition after obtaining the identity subspace, and face matching based on different feature-fusion schemes.

7.2.3.1 **Feature extraction**

As mentioned in previous section, local features, such as LBP and HOG, have been proven to have more discriminative power and are widely used in face recognition. Furthermore, they are easy and fast to implement. In our independent age subspace learning, we have also found that these two local features could achieve the best performance in terms of aging progression representation, as illustrated in Fig. 7-2. Thus, we use both the weighted LBP (wLBP) and HOG as the feature descriptors in our experiments throughout this chapter.

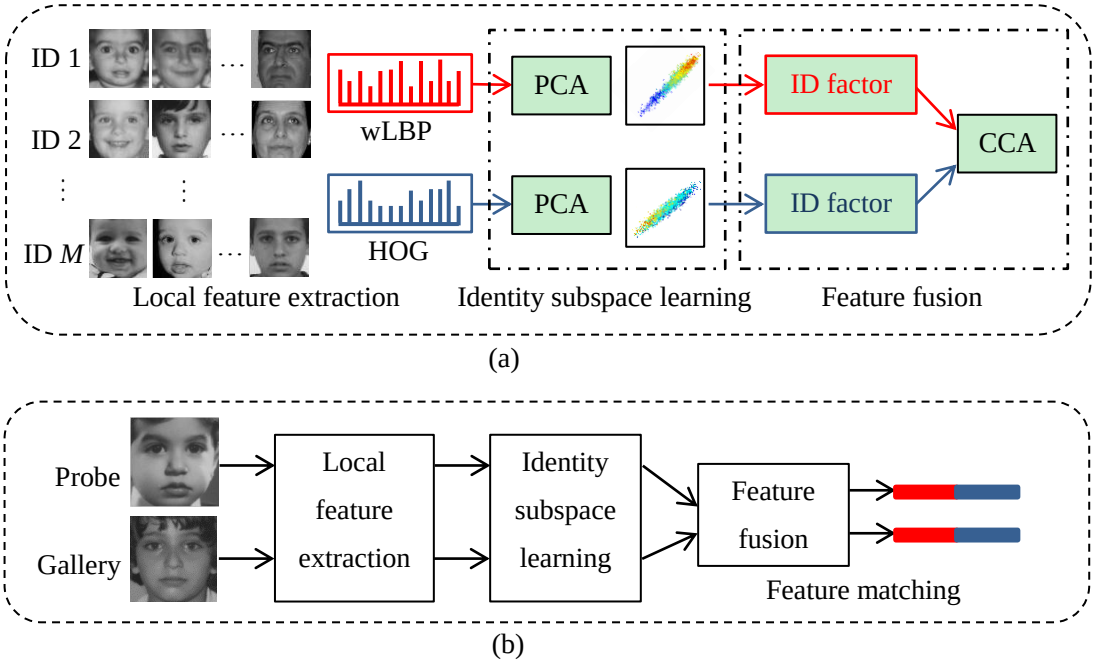
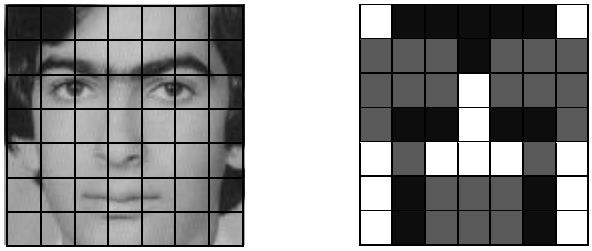


Fig. 7-4: The overall framework of the proposed age-invariant face recognition algorithm based on identity inference with independent aging subspace learning: (a) the training phase and (b) the recognition phase.

For all the face images, we perform the same preprocessing as mentioned in Section 7.2.2.2. The weighted LBP features are extracted with 7×7 windows, at three different radii {1, 3, 5} (due to the limited size of images), which has been found to achieve the

best performance. As each of the windows has a different degree of importance, different weights are assigned to them, as illustrated in Fig. 7-5. Unlike the previous LBP features designed for face recognition [11] and face retrieval tasks [80], the weight mask used in our framework places greater importance on those regions that are more easily influenced by aging effects, such as the forehead, cheeks, and mouth corners [162]. For the HOG feature, by experiment, the best performance can be achieved if the patch size is 21×21 pixels, with an overlapping factor of 0.5 and 4 orientations. In order to facilitate dimensionality reduction for a small-size dataset, like FGNET, we use all the face images in the Chalearn dataset to determine the PCA subspace, with 95% variances retained.



(a) (b)

Fig. 7-5: (a) A cropped face partitioned into 7×7 windows for extracting the weighted MLBP features, and (b) the weights used for the MLBP features in each partition, where black, gray, and white represent the weights of 3, 2, and 1, respectively.

7.2.3.2 **Face recognition based on identity inference model**

As mentioned in Section 2.2.2, applying the supervised OLPP we can independently learn a discriminative aging subspace that can better represent the aging progression. After obtaining the aging subspace **A** of the model shown in Eqn. (7.5), we plug it into the model for the computation of the identity subspace **E**. It should be noticed that the composite matrix $\mathbf{B} = [\mathbf{E} \;\; \mathbf{A}]$ needs to be updated as a whole in order to optimize the model statistics $E[\boldsymbol{z}_m]$ and $E[\boldsymbol{z}_m \boldsymbol{z}_m{}^T]$. Our experiments show that fixing the aging subspace **A** can produce

a better identity subspace **E** for face recognition. This is due to the fact that our proposed aging subspace is learnt by using face images with appearance ages. Thus, in the matrix **B**, only the identity subspace is updated, i.e. $\mathbf{B}' = [\mathbf{E}' \ \ \mathbf{A}]$.

In the recognition stage, we calculate the predictive distributions between the input probe image $\boldsymbol{x}_p$ and each of the gallery images, i.e. $p_r(\boldsymbol{x}_p | \boldsymbol{x}_1)$, $p_r(\boldsymbol{x}_p | \boldsymbol{x}_2)$, …, and $p_r(\boldsymbol{x}_p | \boldsymbol{x}_M)$, and then evaluate the likelihood for each of these distributions. As proved in [165], the likelihood between the probe and a gallery image takes in a Gaussian form, and can be simplified by projecting the data onto a subspace similar to the original LDA method. Suppose that the probe image $\boldsymbol{x}_p$ is put into the identity inference model in (7.2), the output feature vector can then be computed as follows:

$$\boldsymbol{f}_p = \mathbf{E}^T(\mathbf{B}\mathbf{B}^T + \boldsymbol{\Sigma}')^{-1}(\boldsymbol{x}_p - \boldsymbol{\mu}) = \mathbf{E}^T(\mathbf{E}\mathbf{E}^T + \mathbf{A}\mathbf{A}^T + \boldsymbol{\Sigma}')^{-1}(\boldsymbol{x}_p - \boldsymbol{\mu}). \quad (7.11)$$

In this way, it is equivalent to projecting the input probe image into a discriminative subspace, where a distance metric can be used for face recognition.

7.2.3.3 **Feature fusion scheme for face matching**

In order to further improve the face recognition performance, both the wLBP and HOG features are used and fused in our framework. One way to perform feature fusion is to compute the z-score, where both feature vectors are normalized and then concatenated to form a long feature vector. This is the simplest way, but does not take the correlation between the two features into consideration. What's more, concatenating two different types of features directly may cancel out their discriminative power, which leads to an even lower recognition rate. As the two features are extracted from the same identity, they should be correlated. Therefore, as in [192], CCA is used to project the two features into

a coherent subspace, where the correlation between them is maximized. The projected features are then combined to form a single coherent feature vector for age-invariant face recognition.

With the pairs of output features computed with the identity inference model (7.11), denoted as $\mathbf{F}_{wLBP}$ and $\mathbf{F}_{HOG}$, we apply CCA to learn the pairs of directions $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ that maximize the correlation between the projected features, i.e. $\boldsymbol{g}_{wLBP} = \boldsymbol{\alpha}^T \boldsymbol{f}_{wLBP}$ and $\boldsymbol{g}_{HOG} = \boldsymbol{\beta}^T \boldsymbol{f}_{HOG}$, with the correlation between gwLBP and gHOG maximized. The direction matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ can be derived by maximizing the following criterion function:

$$\rho = \frac{E[\boldsymbol{g}_{wLBP}\boldsymbol{g}_{HOG}]}{\sqrt{E[\boldsymbol{g}_{wLBP}{}^2]E[\boldsymbol{g}_{HOG}{}^2]}} = \frac{\boldsymbol{\alpha}^T \mathbf{C}_{12}\boldsymbol{\beta}}{\sqrt{\boldsymbol{\alpha}^T \mathbf{C}_{11}\boldsymbol{\alpha} \cdot \boldsymbol{\beta}^T \mathbf{C}_{22}\boldsymbol{\beta}}}, \tag{7.12}$$

where $\mathbf{C}_{11}$ and $\mathbf{C}_{22}$ denote the covariance matrices of $\boldsymbol{g}_{wLBP}$ and $\boldsymbol{g}_{HOG}$, respectively, and $\mathbf{C}_{12}$ is the covariance matrix of $\boldsymbol{g}_{wLBP}$ and $\boldsymbol{g}_{HOG}$.

In the training stage, after $\mathbf{F}_{wLBP}$ and $\mathbf{F}_{HOG}$, whose columns are the feature vectors $\boldsymbol{f}_{wLBP}$ and $\boldsymbol{f}_{HOG}$, are computed by using Eqn. (7.11), we normalize them to have zero mean and unit variance. Then, the projection matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are computed by using Eqn. (7.12). It can be shown that the optimal direction matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the eigenvectors of $\mathbf{R}_1 = \mathbf{C}_{11}^{-1}\mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{21}$ and $\mathbf{R}_2 = \mathbf{C}_{22}^{-1}\mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{C}_{12}$, respectively. When the pair of features, $\boldsymbol{f}_{p\text{-}wLBP}$ and $\boldsymbol{f}_{p\text{-}HOG}$, of an input probe image are obtained in the testing stage, we further project them into the corresponding CCA subspaces to form coherent features, as follows:

$$\boldsymbol{g}_p = [\boldsymbol{g}_{p\text{-}wLBP}\ \boldsymbol{g}_{p\text{-}HOG}], \tag{7.13}$$

where $\boldsymbol{g}_{p\text{-}wLBP} = \boldsymbol{\alpha}^T \boldsymbol{f}_{p\text{-}wLBP}$ and $\boldsymbol{g}_{p\text{-}HOG} = \boldsymbol{\beta}^T \boldsymbol{f}_{p\text{-}HOG}$. Then, the Euclidean distance is computed, and the nearest-neighbor rule is used for face recognition.

Fig. 7-6: Sample face images from the comparison datasets: (a) FGNET dataset, (b) MORPH dataset and (c) CACD dataset.

7.3 Experimental results

To evaluate the performance of our proposed aging-guided identity inference model (AG-IIM), for age-invariant face recognition, we compare it with several state-of-the-art methods on different datasets, namely the FGNET dataset [146], the MORPH dataset [169], and the CACD dataset [170]. FGNET is known to be the first popular face aging dataset and has been widely used for evaluating age-related facial image analysis tasks. It contains 1,002 images of 82 individuals, and the images were collected at ages ranging from 0 to 69. The MORPH dataset was proposed later, and contains a larger number of subjects. It has two sections, namely MORPH album one and MORPH album two. As album one is small (only 1,690 face images in total), most recent works use album two for

experiments, as it has 55,134 facial images of 13,617 persons. The CACD dataset is the latest aging dataset, which contains 163,446 images of 2,000 celebrity individuals retrieved from the Internet. Some statistics and sample facial images are given in Table 7-2 and Fig. 7-6, respectively. It can be seen that FGNET is the most challenging dataset, as it has the smallest number of images but the largest age gap, while all the photos are also taken under large variations.

Table 7-2. Statistics of the face aging datasets.

Dataset	#Image	#Identity	Age range	Age gap	In the wild
FGNET	1,002	82	0-69	0-45	Yes
MORPH	55,134	13,617	16-77	0-5	No
CACD	163,446	2,000	16-62	0-10	Yes

7.3.1 Face recognition on the FGNET dataset

To fully evaluate our proposed model, we first present the parameters for the wLBP and HOG features used in feature preprocessing, the aging subspace learning, and the identity inference, in Table 7-3.

Table 7-3. Parameter settings based on the FGNET dataset.

Parameters		wLBP	HOG
Feature Configuration	PCA variance	95%	95%
	Feature dimension	1,473	1,371
Aging Learning Model (Supervised OLPP)	# Nearest neighbors	6	5
	Heat kernel coef. t	1	1
Identity Inference Model in (7.1)	#Aging eigenvectors	600	150
	#Identity eigenvectors	80	350

One of the biggest advantages of our proposed method is that, during experiments, age labels of training samples are no longer needed, because we have independently

learned the aging subspace for the identity inference model. Furthermore, as the FGNET dataset only has 1,002 images in total, while the feature dimensions are much higher, we have applied a simple but effective approach to solving the overfitting problem. Unlike the previous strategies [39, 40], which applied random subspaces and feature slicing, we use the ChaLearn images, together with the FGNET images, to learn the PCA subspace, with 95% of the variance retained. In this way, the more discriminative power of the training features can be preserved, while projecting to the same PCA subspace with aging images from the ChaLearn dataset also improves the aging pattern learning.

Our algorithm is fine-tuned using the FGNET dataset. We will later show that the model, trained by using FGNET, can also be applied to recognize faces from other datasets. The performances are similar irrespective of whether or not the training and testing images are from the same dataset. We conducted a thorough evaluation and comparison with some recent, state-of-the-art age-invariant face recognition methods. These include: (a). one of the earliest frameworks on age-invariant face recognition [160], which establishes a 3D aging modeling scheme for age correction for recognition; (b). a discriminative model proposed in [39]; (c). a hidden factor analysis framework [40], which separates aging and identity at the same time; (d). a feature-aging model, which uses local Gabor feature and linear mapping to predict the aging of facial features [223], and (e). a two-step framework [145], based on a maximum entropy feature descriptor, and identity factor analysis matching, which has achieved the best recognition performance on FGNET previously. Following the same experiment set-up, we evaluate all these methods, in leave-one-person-out fashion, based on the rank-1 recognition rates, as shown in Table 7-4. It should be noted that, except [223], all the results of the compared methods, reported in this chapter, are based on the best results as reported in their respective papers.

From Table 7-4, we can see that the proposed AG-IIM, based on independent aging subspace learning, achieves better performances than other methods. Furthermore, the HOG feature is more effective than the LBP feature for representing the aging factors. By fusing the two features using CCA, the recognition rate can be further improved significantly. To the best of our knowledge, this is the highest recognition accuracy that has ever been achieved on the FGNET dataset, which is known to be the most challenging dataset for age-invariant face recognition.

Table 7-4. Rank-1 recognition rates on the FGNET dataset.

Algorithms	**Recognition Rates**
3D aging model (2010) [160]	37.4%
Discriminative aging model (2011) [39]	47.5%
Hidden factor analysis model (2013) [40]	69.0%
Feature-aging model (2015) [223]	71.3%
Maximum entropy model (2015) [145]	76.2%
Proposed AG-IIM with the MLBP feature only	**80.8%**
Proposed AG-IIM with the HOG feature only	**84.14%**
Proposed AG-IIM with feature fusion by CCA	**88.23%**

In the experiments, we have also studied some of the failure cases, in which our proposed AG-IIM could not find the correct subject in the gallery. Among most of these cases, the failure is mainly due to the similar appearances of the query and the gallery faces, at similar ages (appearance ages), as shown in the first three columns in Fig. 7-7. We have also observed some interesting results, for which the input query images were matched to training images with a large age difference, as shown in the last three columns in Fig. 7-7. After a close-up analysis, some underlying similarities between these incorrect retrieved pairs can be observed, such as noses, mouths, facial structures, etc. This further

shows that the proposed framework is seeking the substantial identity information, instead of being tricked by the superficial likeness.

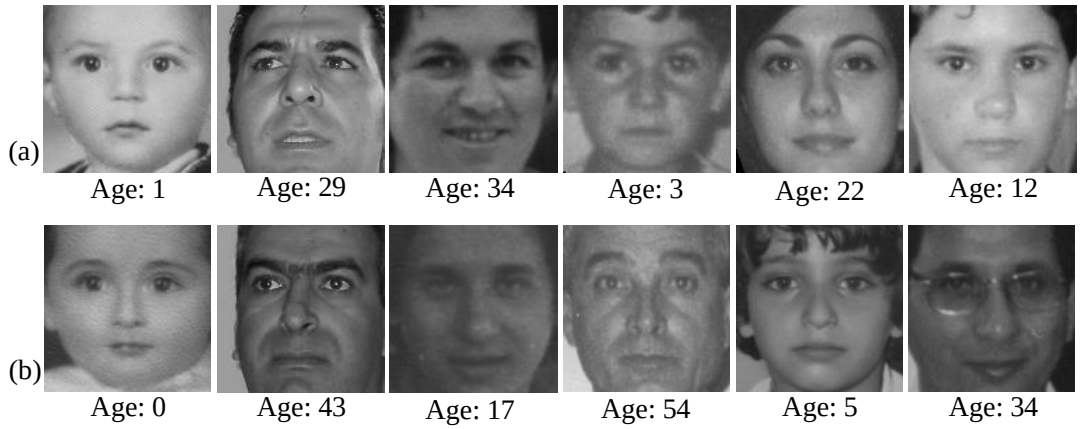


Fig. 7-7. Some examples of the recognition failure cases on the FGNET dataset: (a) input query images, and (b) the corresponding retrieved images, based on Euclidean distance.

7.3.2 Face recognition on the MORPH dataset

In this section, we extend the experiment on the MORPH dataset to examine the efficiency of the proposed method. Unlike FGNET, the MORPH dataset does not have a large age gap for each of its subjects. However, all its face images are still under large pose, lighting, and expression variations. For this dataset, we follow the same split rule, as the previous method [39], where 10,000 individuals are randomly selected. Then, the youngest face images of the selected subjects are used to form the gallery set, while the corresponding oldest images are used to form the probe set. In this way, both the gallery and probe sets have 10,000 images from the different individuals. All the images of the remaining 3,617 subjects will be used for training the identity subspace based on the identity inference model, where the aging subspace is still learnt independently. Same as FGNET, only the identity labels of the training images are used during the experiment, which makes the training process much simpler.

For the proposed AG-IIM method, the same pre-processing steps, including eye detection, face alignment, and feature extraction, are applied to all the images in the MORPH dataset, as described in the previous section. Furthermore, in addition to using the MORPH images for training, we also use the identity subspace learnt from the FGNET dataset, i.e. no MORPH images are used for training, so that the generalization ability of our proposed algorithm can be tested. To fully evaluate the recognition performance of the proposed algorithm, we compare the rank-1 recognition rate with some of the state-of-the-art methods: (a). the 3D aging model [160]; (b). the discriminative aging model [39]; (c). the hidden factor analysis model (2013) [40]; (d). the maximum entropy model (2015) [145]; (e). the cross-age reference coding (CARC) model [170], and (f). local pattern selection with the hidden factor analysis (LPS+HFA) model [38], which is an extension of [145] and has achieved the best recognition accuracy on MORPH, to date. The comparison results of the different methods are compared and shown in Table 7-5.

Table 7-5. Rank-1 recognition accuracies on the MORPH dataset.

Algorithms	**Recognition Accuracies**
3D aging model (2010) [160]	79.8%
Discriminative aging model (2011) [39]	83.9%
Hidden factor analysis model (2013) [40]	91.14%
Maximum entropy model (2015) [145]	92.26%
CARC model (2015) [145]	92.8%
LPS+HFA model (2016) [38]	94.87%
Proposed AG-IIM trained on FGNET	**93.12%**
Proposed AG-IIM trained on MORPH	**95.62%**

As some existing methods have already achieved impressive performance results on the MORPH dataset, the improvement of our method in terms of the recognition accuracy

is marginal, but our method is still comparable to the state-of-the-art methods. More importantly, we found that, even if we replace the training data with the 1,002 images from FGNET and with their identity labels, a similar recognition performance can still be obtained. This finding is exciting in the sense that, from Fig. 7-6, most of the images from the FGNET and the MORPH datasets have different races. However, because the subjects in the FGNET dataset have a wide range of age difference, our algorithm can capture more substantial identity information, which compensates for the difficulties of recognition across races.

7.3.3 Face verification on the CACD dataset

In order to fully examine the generalization power of our proposed framework for age-invariant face recognition, we further conducted face verification on the CACD verification subset (CACD-VS). It comes as a part of the CACD dataset, and contains 2,000 positive pairs (images of the same person across ages) and 2,000 negative pairs. They are carefully selected and annotated to make sure that each of the images has a correct identity tag. To perform verification, we use the marginalized likelihood as the metric learning for recognition. For each pair of face images, we compute the likelihood $p_r(\boldsymbol{x}_p, \boldsymbol{x}_g)$ that they belong to the same identity, and the likelihood $p_r(\boldsymbol{x}_p)p_r(\boldsymbol{x}_g)$ that they are from different identities. Then, we compute the likelihood ratio to form a threshold for face verification as in [165]:

$$R(\boldsymbol{x}_p, \boldsymbol{x}_g) = \frac{\textbf{\textit{likelihood}}\textit{(same)}}{\textbf{\textit{likelihood}}\textit{(diff)}} = \frac{p_r(\boldsymbol{x}_p, \boldsymbol{x}_g)}{p_r(\boldsymbol{x}_p)p_r(\boldsymbol{x}_g)}. \tag{7.14}$$

Following the same experiment set-up in [170], we partition the whole subset into ten folds, with 400 image pairs (200 positive and 200 negative) in each fold. In order to test

the generalization power of the proposed framework, we still use all the FGNET images for training to obtain the identity subspace. What's more, we notice that, as all the subjects in CACD are celebrities retrieved from online, most of the women's faces are wearing make-up, which makes the recognition across ages become more difficult. To further assist the model in adapting different variations, we also added 798 face images of 100 identities (which were chosen from one of the folders and then excluded from testing) from the CACD dataset. In this way, the training set size is increased to 1800 images, in total. The threshold for (7.14) is learnt by using 3,200 pairs of images from eight of the folds, while the remaining single fold serves as the testing set. The experiments are repeated nine times, and the average verification results are shown in Table 7-6. The performance of our method is compared to several state-of-the-art methods, as well as the human voting reported in [170], in terms of the receiver-operating characteristic (ROC) curves, as shown in Fig. 7-8.

Table 7-6. Verification accuracies on the CACD-VS dataset.

Algorithms	**Verification Accuracies**
High Dimensional LBP (2013) [127]	81.6%
Hidden factor analysis model (2013) [40]	84.4%
CARC-NT model (2015) [170]	85.6%
CARC model (2015) [170]	87.6%
Proposed AG-IIM trained on FGNET	**89.8%**
Human, Average	85.7%
Human, Voting	94.2%

From the above results, the proposed method outperforms other methods and is better than the average performance of humans, although the number of training samples used

in our method is fewer than the methods reported in [170]. However, the verification performance of our method is relatively low on the CACD-VS dataset, as the images have more variations besides age. It should also be noted that the voting-based human performance is still better than by machine, especially when the faces have not only large age variations, but also variations of poses, expressions, illuminations, etc. A possible future work for our proposed method can investigate age-invariant face recognition under real-life challenges.

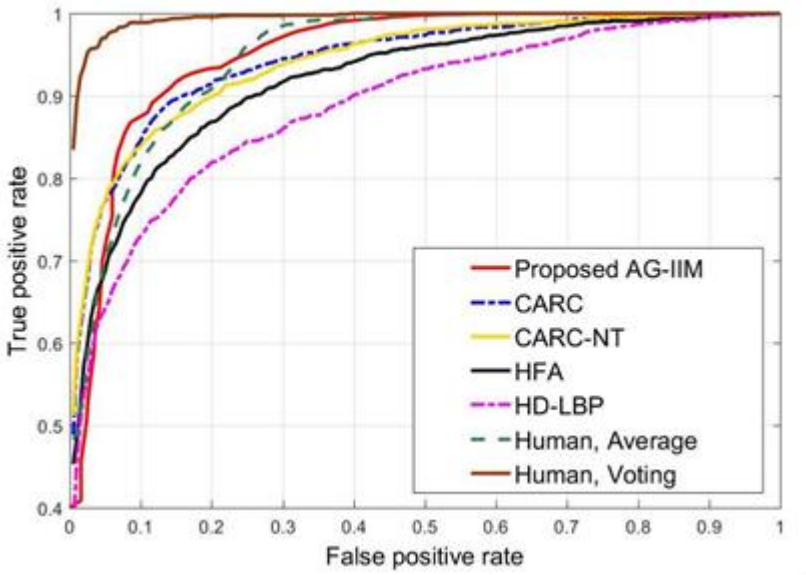


Fig. 7-8: ROC curves for face verification on the CACD-VS dataset.

7.4 Conclusions

In this chapter, we have proposed an aging-guided identity inference model (AG-IIM), based on independent aging subspace learning for age-invariant face recognition. Following the idea that a face can be separated into an identity representation, which is invariant to aging influence, and an aging representation, which changes with age progression, we have applied the PLDA model to obtain two distinctive subspaces. Unlike all the previous methods, which utilize real-age labels, we have proposed using

appearance aging labels to learn the aging subspace independently. It has been shown that, by obtaining the appearance aging information so that computers can understand it better, performance on age-invariant face recognition has been improved. What's more, learning the aging subspace independently enables experiments to rely only on the identity labels given by the datasets, which makes constructing a face-aging dataset much easier. Another contribution of this chapter lies in the projection of two efficient feature representations onto a more correlated subspace using CCA, where their correlation is maximized. Experimental results on different datasets have shown superior performances of our method in terms of recognition accuracy, especially on the most challenging aging dataset FGNET, where face images are under the largest age range.

Chapter 8. Conclusions and future work

In this thesis, we first introduce the concept and development of face recognition, as well as facial-feature analysis techniques. Some of the existing challenges, which serve as the motivation of this research work, are discussed in detail. From Chapter 3 to Chapter 7, we have presented different facial-feature analysis techniques, including an efficient facial-feature localization scheme for pre-processing in Chapter 3; a face super-resolution framework to improve recognition on low-resolution faces in Chapter 4; a high-resolution face verification scheme, based on pore-scale facial features in Chapter 5, and two approaches for age-invariant face recognition in Chapters 6 and 7, respectively.

In this final chapter, we will summarize the main contributions of this research, and discuss some possible directions for future work.

8.1 Summary and conclusions

In our research, we focus on face recognition on low-resolution faces, high-resolution faces, and aging faces, along with facial-feature localization.

In order to improve the alignment performance for most of the face recognition approaches, a shape-appearance-correlated AAM was proposed in Chapter 2. The algorithm includes an efficient initialization scheme, a better representation of shape and appearance information about faces where their correlation is maximized, and a fast simultaneous inverse compositional algorithm for optimization. By conducting experiments on different face datasets and comparing our proposed method with state-of-the-art methods, experimental results show a great improvement in fitting accuracy, especially when the probe face is a face unseen in a gallery, as shown in Fig. 3-10 and Fig. 3-12.

When dealing with the low-resolution face recognition problem in Chapter 3, we apply face super-resolution (face hallucination) to recover the lost information and reconstruct a high-resolution version for better recognition performance. The main contribution of our proposed method is that we can achieve finer global hallucinated faces, based on orthogonal CCA, which has been proved to be more efficient for data reconstruction. Besides, a detail compensation scheme based on linear-mapping further improves the hallucination performance by considering both the inter- and intra-information about face manifolds. With a similar idea to explore face information as much as possible, we have also devised a high-resolution face verification framework in Chapter 4, based on pore-scale facial features, where more discriminative and unique facial characteristics can be found. By establishing the correspondences between facial-pore keypoints of two faces, the proposed method can achieve invariance in pose, expression, and time span. Besides, it has also been proved to be alignment-error-insensitive.

We have also presented two different frameworks for age-invariant face recognition. In the first approach, which is based on feature aging, facial features at different ages are predicted with the assumption that facial features of a person at two different ages are correlated. Results shown in Tables 6-1 and 6-2 prove that the assumption is correct. What is more interesting is that both the results also show that predicting features at an older age group from a younger one is more accurate than the other way around. This is easy to understand as faces are much more similar to each other when they are young, and gradually become more and more distinctive as people grow older. On the other hand, we also notice that the aging process is quite personalized and a person may look younger or older than another person with the same age. Thus, we explore another direction for age-invariant face recognition, which uses appearance age labels instead of real ages. We

model human-identity and aging variables simultaneously using Probabilistic LDA, and obtain the underlying identity information, based on aging subspace learning with appearance ages. Experiment results have shown significant improvement on the FGNET aging dataset, which is known as one of the most challenging datasets with the largest age range. Another contribution of this framework comes from the possibility of making the collection of aging photos easier. As shown in Tables 7-5 and 7-6, the recognition results stay almost the same, even when we trained the identity-inference model with a different dataset. This enables us to collect aging datasets more efficiently as only identity labels are required for all the face data.

8.2 Future work

This thesis, which has presented a number of new ideas and methods, is just a snapshot of our ongoing research, undertaken in the field of robust facial-feature analysis and face recognition. In this section, some directions of possible future research will be discussed. Future research may be carried out in the following fields:

(a). The facial-feature localization problem: With the access to photos taken under different environments, we are facing the facial-feature localization challenges under a combination of numerous variations. Most of the current methods can perform well, only under controlled settings. Besides, localizing features on a face whose identity is unseen in a gallery set, adds even more difficulties to this task. To continue our work on facial-feature analysis, we will focus on more reliable representations of human faces in addition to those simple visual features, like LBP or HOG. More sophisticated models will be investigated to represent face information and explore the correlations between landmarks or between face shape and texture.

(b). The low-resolution face recognition problem: As introduced before, most of the current work can be categorized into super-resolution-based methods and resolution-robust feature representation methods. Super-resolved faces not only can provide more information for recognition, but also help with other applications. However, extracting features from reconstructed high-resolution faces may lose more information than super-resolving the features directly from low resolution to high resolution. Thus, we aim to further explore efficient features from low-resolution faces and predict its corresponding high-resolution version for robust face recognition in the future.

(c). The high-resolution face recognition problem: With the proposed pore-scale facial-feature matching scheme, we aim to investigate the fusion of pore-scale facial features from high-resolution images with larger-scale facial features from low-resolution images to improve the current face recognition systems. Besides, we will also study how to further improve the efficiency of the proposed framework, and apply it to other important areas like 3D face reconstruction, where dense correspondences between two faces with pose variations are highly desired.

(d). The age-invariant face recognition problem: In order to further improve face recognition accuracy on aging faces, we aim to combine aging-feature mapping with appearance-age modeling. Besides, more training samples at different ages will be collected so the learning of the mapping functions and the projection matrices can be more accurate. We will further generalize our identity-inference model and make it robust against different datasets. It should be noted that all the frameworks proposed in this thesis do not work alone. Instead, they can be integrated to become a facial-feature analysis system for recognition tasks under all kinds of variations.

Reference

[1] W. W. Zou and P. C. Yuen, "Very low resolution face recognition problem," *IEEE Transactions on Image Processing,* vol. 21, pp. 327-340, 2012.

[2] S. Escalera, M. Torres, B. Martinez, X. Baró, and H. J. Escalante, "Chalearn looking at people and faces of the world: Face analysis workshop and challenge 2016," in *Proceedings of IEEE conference on Computer Vision and Pattern Recognition Workshops*, 2016.

[3] J. Shi, X. Liu, and C. Qi, "Global consistency, local sparsity and pixel correlation: A unified framework for face hallucination," *Pattern Recognition,* vol. 47, pp. 3520-3534, 2014.

[4] H. Huang, H. He, X. Fan, and J. Zhang, "Super-resolution of human face image using canonical correlation analysis," *Pattern Recognition,* vol. 43, pp. 2532-2543, 2010.

[5] L. An and B. Bhanu, "Face image super-resolution using 2D CCA," *Signal Processing,* vol. 103, pp. 184-194, 2014.

[6] X. Ma, J. Zhang, and C. Qi, "Hallucinating face by position-patch," *Pattern Recognition,* vol. 43, pp. 2224-2236, 2010.

[7] Y. Zhuang, J. Zhang, and F. Wu, "Hallucinating faces: LPH super-resolution and neighbor reconstruction for residue compensation," *Pattern Recognition,* vol. 40, pp. 3178-3194, 2007.

[8] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE transactions on image processing,* vol. 19, pp. 2861-2873, 2010.

[9] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, 2004, pp. I-I.

[10] A. M. Burton, S. Wilson, M. Cowan, and V. Bruce, "Face recognition in poor-quality video: Evidence from security surveillance," *Psychological Science,* vol. 10, pp. 243-248, 1999.

[11] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE transactions on pattern analysis and machine intelligence,* vol. 28, pp. 2037-2041, 2006.

[12] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*, 1991, pp. 586-591.

[13] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 815-823.

[14] A. K. Jain and S. Z. Li, *Handbook of face recognition*: Springer, 2011.

[15] C. Ding, C. Xu, and D. Tao, "Multi-task pose-invariant face recognition," *IEEE Transactions on Image Processing,* vol. 24, pp. 980-993, 2015.

[16] S. Liao, A. K. Jain, and S. Z. Li, "Partial face recognition: Alignment-free approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 35, pp. 1193-1205, 2013.

[17] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE transactions on pattern analysis and machine intelligence,* vol. 31, pp. 210-227, 2009.

[18] B. Fasel and J. Luettin, "Automatic facial expression analysis: a survey," *Pattern recognition,* vol. 36, pp. 259-275, 2003.

[19] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and Vision Computing,* vol. 27, pp. 803-816, 2009.

[20] W. Gu, C. Xiang, Y. Venkatesh, D. Huang, and H. Lin, "Facial expression recognition using radial encoding of local Gabor features and classifier synthesis," *Pattern Recognition,* vol. 45, pp. 80-91, 2012.

[21] T. M. Chaplin and A. Aldao, "Gender differences in emotion expression in children: a meta-analytic review," *Psychological Bulletin,* vol. 139, p. 735, 2013.

[22] L. Zhong, Q. Liu, P. Yang, B. Liu, J. Huang, and D. N. Metaxas, "Learning active facial patches for expression analysis," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012, pp. 2562-2569.

[23] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino, "2D and 3D face recognition: A survey," *Pattern Recognition Letters,* vol. 28, pp. 1885-1906, 2007.

[24] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," in *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, 1999, pp. 187-194.

[25] G. Fanelli, M. Dantone, J. Gall, A. Fossati, and L. Van Gool, "Random forests for real time 3D face analysis," *International Journal of Computer Vision,* vol. 101, pp. 437-458, 2013.

[26] Y. Lei, M. Bennamoun, M. Hayat, and Y. Guo, "An efficient 3D face recognition approach using local geometrical signatures," *Pattern Recognition,* vol. 47, pp. 509-524, 2014.

[27] H. Tang, B. Yin, Y. Sun, and Y. Hu, "3D face recognition using local binary patterns," *Signal Processing,* vol. 93, pp. 2190-2198, 2013.

[28] N. Magnenat-Thalmann, E. Primeau, and D. Thalmann, "Abstract muscle action procedures for human face animation," *The Visual Computer,* vol. 3, pp. 290-297, 1988.

[29] J. Zhang, J. Yu, J. You, D. Tao, N. Li, and J. Cheng, "Data-driven facial animation via semi-supervised local patch alignment," *Pattern Recognition,* vol. 57, pp. 1-20, 2016.

[30] C. Cao, Q. Hou, and K. Zhou, "Displaced dynamic expression regression for real-time facial tracking and animation," *ACM Transactions on Graphics (TOG),* vol. 33, p. 43, 2014.

[31] I. S. Pandzic and R. Forchheimer, "MPEG-4 facial animation," *The standard, implementation and applications. Chichester, England: John Wiley&Sons,* 2002.

[32] S. Bouaziz, Y. Wang, and M. Pauly, "Online modeling for realtime facial animation," *ACM Transactions on Graphics (TOG),* vol. 32, p. 40, 2013.

[33] Y. Sun, D. Liang, X. Wang, and X. Tang, "Deepid3: Face recognition with very deep neural networks," *arXiv preprint arXiv:1502.00873,* 2015.

[34] C. Ding and D. Tao, "Robust face recognition via multimodal deep face representation," *IEEE Transactions on Multimedia,* vol. 17, pp. 2049-2058, 2015.

[35] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Technical Report 07-49, University of Massachusetts, Amherst2007.

[36] B. Li, H. Chang, S. Shan, and X. Chen, "Low-resolution face recognition via coupled locality preserving mappings," *IEEE Signal Processing Letters,* vol. 17, pp. 20-23, 2010.

[37] Z. Wang, Z. Miao, Q. J. Wu, Y. Wan, and Z. Tang, "Low-resolution face recognition: a review," *The Visual Computer,* vol. 30, pp. 359-386, 2014.

[38] Z. Li, D. Gong, X. Li, and D. Tao, "Aging Face Recognition: A Hierarchical Learning Model Based on Local Patterns Selection," *IEEE Transactions on Image Processing,* vol. 25, pp. 2146-2154, 2016.

[39] Z. Li, U. Park, and A. K. Jain, "A discriminative model for age invariant face recognition," *IEEE transactions on information forensics and security,* vol. 6, pp. 1028-1037, 2011.

[40] D. Gong, Z. Li, D. Lin, J. Liu, and X. Tang, "Hidden factor analysis for age invariant face recognition," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 2872-2879.

[41] S. Milborrow and F. Nicolls, "Locating facial features with an extended active shape model," in *European conference on computer vision*, 2008, pp. 504-513.

[42] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012, pp. 2879-2886.

[43] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 23, pp. 681-685, 2001.

[44] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Computer vision and image understanding,* vol. 61, pp. 38-59, 1995.

[45] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision,* vol. 57, pp. 137-154, 2004.

[46] C. Liu, "A Bayesian discriminating features method for face detection," *IEEE transactions on pattern analysis and machine intelligence,* vol. 25, pp. 725-740, 2003.

[47] C. Erdem, S. Ulukaya, A. Karaali, and A. T. Erdem, "Combining Haar feature and skin color based classifiers for face detection," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 1497-1500.

[48] Y. Sun, X. Wang, and X. Tang, "Deep convolutional network cascade for facial point detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3476-3483.

[49] J. S. Bruner and R. Tagiuri, "The perception of people," DTIC Document1954.

[50] W. W. Bledsoe, "The Model Method in Facial Recognition," Panoramic Research Inc., Palo Alto, CA1964.

[51] M. D. Kelly, "Visual identification of people by computer," DTIC Document1970.

[52] T. Kanade, *Computer recognition of human faces* vol. 47: Birkhäuser, 1977.

[53] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of cognitive neuroscience,* vol. 3, pp. 71-86, 1991.

[54] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," *IEEE Transactions on Pattern analysis and Machine intelligence,* vol. 12, pp. 103-108, 1990.

[55] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 19, pp. 711-720, 1997.

[56] A. M. Martínez and A. C. Kak, "Pca versus lda," *IEEE transactions on pattern analysis and machine intelligence,* vol. 23, pp. 228-233, 2001.

[57] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face recognition by independent component analysis," *IEEE Transactions on neural networks,* vol. 13, pp. 1450-1464, 2002.

[58] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using Laplacianfaces," *IEEE transactions on pattern analysis and machine intelligence,* vol. 27, pp. 328-340, 2005.

[59] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *NIPS*, 2001, pp. 585-591.

[60] J. Yang, D. Zhang, A. F. Frangi, and J.-y. Yang, "Two-dimensional PCA: a new approach to appearance-based face representation and recognition," *IEEE transactions on pattern analysis and machine intelligence,* vol. 26, pp. 131-137, 2004.

[61] J. Yang, D. Zhang, X. Yong, and J.-y. Yang, "Two-dimensional discriminant transform for face recognition," *Pattern recognition,* vol. 38, pp. 1125-1129, 2005.

[62] S. Chen, H. Zhao, M. Kong, and B. Luo, "2D-LPP: a two-dimensional extension of locality preserving projections," *Neurocomputing,* vol. 70, pp. 912-921, 2007.

[63] C. Liu, "Gabor-based kernel PCA with fractional power polynomial models for face recognition," *IEEE transactions on pattern analysis and machine intelligence,* vol. 26, pp. 572-581, 2004.

[64] B. Moghaddam, "Principal manifolds and probabilistic subspaces for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 24, pp. 780-788, 2002.

[65] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural computation,* vol. 10, pp. 1299-1319, 1998.

[66] J. Yang, A. F. Frangi, J.-y. Yang, D. Zhang, and Z. Jin, "KPCA plus LDA: a complete kernel Fisher discriminant framework for feature extraction and recognition," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 27, pp. 230-244, 2005.

[67] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Face recognition using kernel direct discriminant analysis algorithms," *IEEE Transactions on Neural Networks,* vol. 14, pp. 117-126, 2003.

[68] B. Scholkopft and K.-R. Mullert, "Fisher discriminant analysis with kernels," *Neural networks for signal processing IX,* vol. 1, p. 1, 1999.

[69] J.-B. Li, J.-S. Pan, and S.-C. Chu, "Kernel class-wise locality preserving projection," *Information Sciences,* vol. 178, pp. 1825-1835, 2008.

[70] X. Niyogi, "Locality preserving projections," in *Neural information processing systems*, 2004, p. 153.

[71] J. Cheng, Q. Liu, H. Lu, and Y.-W. Chen, "Supervised kernel locality preserving projections for face recognition," *Neurocomputing,* vol. 67, pp. 443-449, 2005.

[72] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *science,* vol. 290, pp. 2319-2323, 2000.

[73] L. K. Saul and S. T. Roweis, "Think globally, fit locally: unsupervised learning of low dimensional manifolds," *Journal of Machine Learning Research,* vol. 4, pp. 119-155, 2003.

[74] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science,* vol. 290, pp. 2323-2326, 2000.

[75] A. L. Yuille, P. W. Hallinan, and D. S. Cohen, "Feature extraction from faces using deformable templates," *International journal of computer vision,* vol. 8, pp. 99-111, 1992.

[76] R. Brunelli and T. Poggio, "Face recognition: Features versus templates," *IEEE transactions on pattern analysis and machine intelligence,* vol. 15, pp. 1042-1052, 1993.

[77] T. Kanade, "Picture processing system by computer complex and recognition of human faces," *Doctoral dissertation, Kyoto University},* vol. 3952, pp. 83-97, 1973.

[78] G. Tzimiropoulos and M. Pantic, "Optimization problems for fast aam fitting in-the-wild," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 593-600.

[79] J. Sung, T. Kanade, and D. Kim, "A unified gradient-based approach for combining ASM into AAM," *International Journal of Computer Vision,* vol. 75, pp. 297-309, 2007.

[80] H. Zhou, K.-M. Lam, and X. He, "Shape-appearance-correlated active appearance model," *Pattern Recognition,* vol. 56, pp. 88-99, 2016.

[81] G. Tzimiropoulos, J. Alabort-i-Medina, S. Zafeiriou, and M. Pantic, "Generic active appearance models revisited," in *Asian Conference on Computer Vision*, 2012, pp. 650-663.

[82] H.-S. Lee and D. Kim, "Tensor-based AAM with continuous variation estimation: Application to variation-robust face recognition," *IEEE transactions on pattern analysis and machine intelligence,* vol. 31, pp. 1102-1116, 2009.

[83] D. Cristinacce, T. F. Cootes, and I. M. Scott, "A Multi-Stage Approach to Facial Feature Detection," in *BMVC*, 2004, pp. 1-10.

[84] D. Cristinacce and T. F. Cootes, "Feature Detection and Tracking with Constrained Local Models," in *BMVC*, 2006, p. 3.

[85] C. Kotropoulos, A. Tefas, and I. Pitas, "Frontal face authentication using morphological elastic graph matching," *IEEE Transactions on Image Processing,* vol. 9, pp. 555-560, 2000.

[86] U. Prabhu, J. Heo, and M. Savvides, "Unconstrained pose-invariant face recognition using 3D generic elastic models," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 33, pp. 1952-1961, 2011.

[87] H. Shin, S.-D. Kim, and H.-C. Choi, "Generalized elastic graph matching for face recognition," *Pattern Recognition Letters,* vol. 28, pp. 1077-1082, 2007.

[88] A. Tefas, C. Kotropoulos, and I. Pitas, "Using support vector machines to enhance the performance of elastic graph matching for frontal face authentication," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 23, pp. 735-746, 2001.

[89] L. Wiskott, J.-M. Fellous, N. Kuiger, and C. Von Der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 19, pp. 775-779, 1997.

[90] J. Zhang, Y. Yan, and M. Lades, "Face recognition: eigenface, elastic matching, and neural nets," *Proceedings of the IEEE,* vol. 85, pp. 1423-1435, 1997.

[91] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Wurtz, *et al.*, "Distortion invariant object recognition in the dynamic link architecture," *IEEE Transactions on computers,* vol. 42, pp. 300-311, 1993.

[92] S.-H. Lin, S.-Y. Kung, and L.-J. Lin, "Face recognition/detection by probabilistic decision-based neural network," *IEEE transactions on neural networks,* vol. 8, pp. 114-132, 1997.

[93] S.-Y. Kung and J.-S. Taur, "Decision-based neural networks with signal/image classification applications," *IEEE Transactions on Neural Networks,* vol. 6, pp. 170-181, 1995.

[94] G. Bebis and M. Georgiopoulos, "Feed-forward neural networks," *IEEE Potentials,* vol. 13, pp. 27-31, 1994.

[95] D. Svozil, V. Kvasnicka, and J. Pospichal, "Introduction to multi-layer feed-forward neural networks," *Chemometrics and intelligent laboratory systems,* vol. 39, pp. 43-62, 1997.

[96] J. Ilonen, J.-K. Kamarainen, and J. Lampinen, "Differential evolution training algorithm for feed-forward neural networks," *Neural Processing Letters,* vol. 17, pp. 93-105, 2003.

[97] K. Khan and A. Sahai, "A comparison of BA, GA, PSO, BP and LM for training feed forward neural networks in e-learning context," *International Journal of Intelligent Systems and Applications,* vol. 4, p. 23, 2012.

[98] A. Graves, A.-r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *2013 IEEE international conference on acoustics, speech and signal processing*, 2013, pp. 6645-6649.

[99] Y. Liu, Z. Wang, and X. Liu, "Global exponential stability of generalized recurrent neural networks with discrete and distributed delays," *Neural Networks,* vol. 19, pp. 667-675, 2006.

[100] J. Cao and J. Wang, "Global asymptotic stability of a general class of recurrent neural networks with time-varying delays," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications,* vol. 50, pp. 34-44, 2003.

[101] P. J. Angeline, G. M. Saunders, and J. B. Pollack, "An evolutionary algorithm that constructs recurrent neural networks," *IEEE transactions on Neural Networks,* vol. 5, pp. 54-65, 1994.

[102] Y. Sun, Y. Chen, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification," in *Advances in Neural Information Processing Systems*, 2014, pp. 1988-1996.

[103] N. Jindal and V. Kumar, "Enhanced face recognition algorithm using pca with artificial neural networks," *International Journal of Advanced Research in Computer Science and Software Engineering,* vol. 3, pp. 864-872, 2013.

[104] H. A. Rowley, S. Baluja, and T. Kanade, "Rotation invariant neural network-based face detection," in *Computer Vision and Pattern Recognition, 1998. Proceedings. 1998 IEEE Computer Society Conference on*, 1998, pp. 38-44.

[105] H. M. El-Bakry, "Fast face detection using neural networks and image decomposition," in *International Computer Science Conference on Active Media Technology*, 2001, pp. 205-215.

[106] Y. Lu, N. Zeng, Y. Liu, and N. Zhang, "A hybrid Wavelet Neural Network and Switching Particle Swarm Optimization algorithm for face direction recognition," *Neurocomputing,* vol. 155, pp. 219-224, 2015.

[107] G. B. Huang, H. Lee, and E. Learned-Miller, "Learning hierarchical representations for face verification with convolutional deep belief networks," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012, pp. 2518-2525.

[108] Y. Sun, X. Wang, and X. Tang, "Hybrid deep learning for face verification," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1489-1496.

[109] Y. Taigman, M. Yang, M. A. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1701-1708.

[110] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural processing letters*, vol. 9, pp. 293-300, 1999.

[111] E. Osuna, R. Freund, and F. Girosit, "Training support vector machines: an application to face detection," in *Computer vision and pattern recognition, 1997. Proceedings., 1997 IEEE computer society conference on*, 1997, pp. 130-136.

[112] B. Heisele, P. Ho, and T. Poggio, "Face recognition with support vector machines: Global versus component-based approach," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, 2001, pp. 688-694.

[113] G. Guo, S. Z. Li, and K. Chan, "Face recognition by support vector machines," in *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, 2000, pp. 196-201.

[114] R. J. Baron, "Mechanisms of human facial recognition," *International Journal of Man-Machine Studies*, vol. 15, pp. 137-178, 1981.

[115] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang, "Recognizing partially occluded, expression variant faces from single training image per person with SOM and soft k-NN ensemble," *IEEE Transactions on Neural Networks*, vol. 16, pp. 875-886, 2005.

[116] S. Chen, J. Liu, and Z.-H. Zhou, "Making FLDA applicable to face recognition with one sample per person," *Pattern recognition*, vol. 37, pp. 1553-1555, 2004.

[117] B. Kepenekci, F. B. Tek, and G. B. Akar, "Occluded face recognition based on Gabor wavelets," in *Image Processing. 2002. Proceedings. 2002 International Conference on*, 2002, pp. I-293-I-296 vol. 1.

[118] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, pp. 971-987, 2002.

[119] O. Déniz, G. Bueno, J. Salido, and F. De la Torre, "Face recognition using histograms of oriented gradients," *Pattern Recognition Letters*, vol. 32, pp. 1598-1603, 2011.

[120] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005, pp. 886-893.

[121] X. Liu and T. Cheng, "Video-based face recognition using adaptive hidden markov models," in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, 2003, pp. I-340-I-345 vol. 1.

[122] H.-S. Le and H. Li, "Recognizing frontal face images using hidden Markov models with one training image per person," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, 2004, pp. 318-321.

[123] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *2011 International conference on computer vision*, 2011, pp. 2564-2571.

[124] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Computer vision and image understanding*, vol. 110, pp. 346-359, 2008.

[125] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, 2004, pp. II-506-II-513 Vol. 2.

[126] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, pp. 91-110, 2004.

[127] D. Chen, X. Cao, F. Wen, and J. Sun, "Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3025-3032.

[128] H. Zhou and K.-M. Lam, "Face hallucination using orthogonal canonical correlation analysis," *Journal of Electronic Imaging*, vol. 25, pp. 033005-033005, 2016.

[129] H. Zhou, J. Hu, and K.-M. Lam, "Global face reconstruction for face hallucination using orthogonal canonical correlation analysis," in *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, 2015, pp. 537-542.

[130] X. Wang and X. Tang, "Hallucinating face by eigentransformation," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 35, pp. 425-434, 2005.

[131] S. Biswas, K. W. Bowyer, and P. J. Flynn, "Multidimensional scaling for matching low-resolution face images," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, pp. 2019-2030, 2012.

[132] Z. Lei, T. Ahonen, M. Pietikäinen, and S. Z. Li, "Local frequency descriptor for low-resolution face recognition," in *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, 2011, pp. 161-166.

[133] J. Y. Choi, Y. M. Ro, and K. N. Plataniotis, "Color face recognition for degraded face images," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, pp. 1217-1230, 2009.

[134] X. Zhang, K.-M. Lam, and L. Shen, "Image magnification based on a blockwise adaptive Markov random field model," *Image and Vision Computing,* vol. 26, pp. 1277-1284, 2008.

[135] H. He and L. P. Kondi, "An image super-resolution algorithm for different error levels per frame," *IEEE Transactions on Image Processing,* vol. 15, pp. 592-603, 2006.

[136] M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images," *IEEE transactions on image processing,* vol. 6, pp. 1646-1658, 1997.

[137] X. Li, K. M. Lam, G. Qiu, L. Shen, and S. Wang, "Example-based image super-resolution with class-specific predictors," *Journal of Visual Communication and Image Representation,* vol. 20, pp. 312-322, 2009.

[138] Y. Hu, K. M. Lam, T. Shen, and W. Wang, "A novel kernel-based framework for facial-image hallucination," *Image and Vision Computing,* vol. 29, pp. 219-229, 2011.

[139] C. Liu, H.-Y. Shum, and W. T. Freeman, "Face hallucination: Theory and practice," *International Journal of Computer Vision,* vol. 75, pp. 115-134, 2007.

[140] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 24, pp. 1167-1183, 2002.

[141] D. Lin and X. Tang, "Recognize high resolution faces: From macrocosm to microcosm," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, 2006, pp. 1355-1362.

[142] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *Workshop on statistical learning in computer vision, ECCV*, 2004, pp. 1-2.

[143] U. Park and A. K. Jain, "Face matching and retrieval using soft biometrics," *IEEE Transactions on Information Forensics and Security,* vol. 5, pp. 406-415, 2010.

[144] N. Srinivas, G. Aggarwal, P. J. Flynn, and R. W. V. Bruegge, "Facial marks as biometric signatures to distinguish between identical twins," in *CVPR 2011 WORKSHOPS*, 2011, pp. 106-113.

[145] D. Gong, Z. Li, D. Tao, J. Liu, and X. Li, "A maximum entropy feature descriptor for age invariant face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5289-5297.

[146] T. Cootes and A. Lanitis, "The FG-NET aging database," ed: ed, 2008.

[147] Y. Fu, G. Guo, and T. S. Huang, "Age synthesis and estimation via faces: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, pp. 1955-1976, 2010.

[148] A. Lanitis, "A survey of the effects of aging on biometric identity verification," *International Journal of Biometrics*, vol. 2, pp. 34-52, 2009.

[149] H. Ling, S. Soatto, N. Ramanathan, and D. W. Jacobs, "A study of face recognition as people age," in *2007 IEEE 11th International Conference on Computer Vision*, 2007, pp. 1-8.

[150] K. Zhu, D. Gong, Z. Li, and X. Tang, "Orthogonal gaussian process for automatic age estimation," in *Proceedings of the 22nd ACM international conference on Multimedia*, 2014, pp. 857-860.

[151] A. Montillo and H. Ling, "Age regression from faces using random forests," in *2009 16th IEEE International Conference on Image Processing (ICIP)*, 2009, pp. 2465-2468.

[152] G. Mu, G. Guo, Y. Fu, and T. S. Huang, "Human age estimation using bio-inspired features," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 112-119.

[153] Y. Fu and T. S. Huang, "Human age estimation with regression on discriminative aging manifold," *IEEE Transactions on Multimedia*, vol. 10, pp. 578-584, 2008.

[154] X. Geng, Z.-H. Zhou, and K. Smith-Miles, "Automatic age estimation based on facial aging patterns," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 29, pp. 2234-2240, 2007.

[155] J. Suo, X. Chen, S. Shan, and W. Gao, "Learning long term face aging patterns from partially dense aging databases," in *2009 IEEE 12th International Conference on Computer Vision*, 2009, pp. 622-629.

[156] X. Shu, J. Tang, H. Lai, L. Liu, and S. Yan, "Personalized Age Progression with Aging Dictionary," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 3970-3978.

[157] J. Suo, S.-C. Zhu, S. Shan, and X. Chen, "A compositional and dynamic model for face aging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 385-401, 2010.

[158] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Toward automatic simulation of aging effects on face images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 442-455, 2002.

[159] B.-C. Chen, C.-S. Chen, and W. H. Hsu, "Cross-age reference coding for age-invariant face recognition and retrieval," in *European Conference on Computer Vision*, 2014, pp. 768-783.

[160] U. Park, Y. Tong, and A. K. Jain, "Age-invariant face recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, pp. 947-954, 2010.

[161] L. Du and H. Ling, "Cross-age face verification by coordinating with cross-face age verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2329-2338.

[162] C. Otto, H. Han, and A. Jain, "How does aging affect facial components?," in *European Conference on Computer Vision*, 2012, pp. 189-198.

[163] H. Ling, S. Soatto, N. Ramanathan, and D. W. Jacobs, "Face verification across age progression using discriminative methods," *IEEE Transactions on Information Forensics and security*, vol. 5, pp. 82-91, 2010.

[164] S. J. Prince and J. H. Elder, "Probabilistic linear discriminant analysis for inferences about identity," in *2007 IEEE 11th International Conference on Computer Vision*, 2007, pp. 1-8.

[165] S. Prince, P. Li, Y. Fu, U. Mohammed, and J. Elder, "Probabilistic models for inference about identity," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, pp. 144-157, 2012.

[166] S. Ioffe, "Probabilistic linear discriminant analysis," in *European Conference on Computer Vision*, 2006, pp. 531-542.

[167] Y.-J. Zhang, *Advances in Face Image Analysis: Techniques and Technologies: Techniques and Technologies*: IGI Global, 2010.

[168] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the royal statistical society. Series B (methodological)*, pp. 1-38, 1977.

[169] K. Ricanek and T. Tesafaye, "Morph: A longitudinal image database of normal adult age-progression," in *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, 2006, pp. 341-345.

[170] B.-C. Chen, C.-S. Chen, and W. H. Hsu, "Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset," *IEEE Transactions on Multimedia*, vol. 17, pp. 804-815, 2015.

[171] O. Çeliktutan, S. Ulukaya, and B. Sankur, "A comparative study of face landmarking techniques," *EURASIP Journal on Image and Video Processing*, vol. 2013, p. 1, 2013.

[172] R. T. Chin and C. R. Dyer, "Model-based recognition in robot vision," *ACM Computing Surveys (CSUR)*, vol. 18, pp. 67-108, 1986.

[173] W.-P. Choi, K.-M. Lam, and W.-C. Siu, "An adaptive active contour model for highly irregular boundaries," *Pattern Recognition*, vol. 34, pp. 323-331, 2001.

[174] S. Yan, C. Liu, S. Z. Li, H. Zhang, H.-Y. Shum, and Q. Cheng, "Face alignment using texture-constrained active shape models," *Image and Vision Computing*, vol. 21, pp. 69-75, 2003.

[175] S. Milborrow and F. Nicolls, "Active Shape Models with SIFT Descriptors and MARS," in *VISAPP (2)*, 2014, pp. 380-387.

[176] K. Sun, H. Zhou, and K. M. Lam, "An adaptive-profile active shape model for facial-feature detection," in *Pattern Recognition (ICPR), 2014 22nd International Conference on*, 2014, pp. 2849-2854.

[177] X. Shen, Z. Lin, J. Brandt, and Y. Wu, "Detecting and aligning faces by image retrieval," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3460-3467.

[178] B. M. Smith, J. Brandt, Z. Lin, and L. Zhang, "Nonparametric context modeling of local appearance for pose-and expression-robust facial landmark localization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1741-1748.

[179] X. Zhao, S. Shan, X. Chai, and X. Chen, "Locality-constrained active appearance model," in *Asian Conference on Computer Vision*, 2012, pp. 636-647.

[180] K. Yu, T. Zhang, and Y. Gong, "Nonlinear learning using local coordinate coding," in *Advances in neural information processing systems*, 2009, pp. 2223-2231.

[181] S. Milborrow, T. Bishop, and F. Nicolls, "Multiview active shape models with SIFT descriptors for the 300-W face landmark challenge," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2013, pp. 378-385.

[182] X. P. Burgos-Artizzu, P. Perona, and P. Dollár, "Robust face landmark estimation under occlusion," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1513-1520.

[183] N. Brunet, F. Perez, and F. De la Torre, "Learning good features for active shape models," in *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*, 2009, pp. 206-211.

[184] R. Gross, I. Matthews, and S. Baker, "Generic vs. person specific active appearance models," *Image and Vision Computing*, vol. 23, pp. 1080-1093, 2005.

[185] J. Saragih and R. Göcke, "Learning AAM fitting through simulation," *Pattern Recognition*, vol. 42, pp. 2628-2636, 2009.

[186] X. Liu, "Generic face alignment using boosted appearance model," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1-8.

[187] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, and N. Kumar, "Localizing parts of faces using a consensus of exemplars," *IEEE transactions on pattern analysis and machine intelligence,* vol. 35, pp. 2930-2940, 2013.

[188] J. M. Saragih, S. Lucey, and J. F. Cohn, "Face alignment through subspace constrained mean-shifts," in *2009 IEEE 12th International Conference on Computer Vision*, 2009, pp. 1034-1041.

[189] D. Cristinacce and T. Cootes, "Automatic feature localisation with constrained local models," *Pattern Recognition,* vol. 41, pp. 3054-3067, 2008.

[190] X.-B. Shen, Q.-S. Sun, and Y.-H. Yuan, "Orthogonal canonical correlation analysis and its application in feature fusion," in *Information Fusion (FUSION), 2013 16th International Conference on*, 2013, pp. 151-157.

[191] M. Bevilacqua, A. Roumy, C. Guillemot, and M.-L. A. Morel, "Single-image super-resolution via linear mapping of interpolated self-examples," *IEEE Transactions on image processing,* vol. 23, pp. 5334-5347, 2014.

[192] K.-H. Pong and K.-M. Lam, "Multi-resolution feature fusion for face recognition," *Pattern Recognition,* vol. 47, pp. 556-567, 2014.

[193] G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang, "Image-based human age estimation by manifold learning and locally adjusted robust regression," *IEEE Transactions on Image Processing,* vol. 17, pp. 1178-1188, 2008.

[194] D. Cai, X. He, J. Han, and H.-J. Zhang, "Orthogonal laplacianfaces for face recognition," *IEEE transactions on image processing,* vol. 15, pp. 3608-3614, 2006.

[195] N. Wang, X. Gao, D. Tao, and X. Li, "Facial feature point detection: A comprehensive survey," *arXiv preprint arXiv:1410.1037,* 2014.

[196] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Comparing Active Shape Models with Active Appearance Models," in *BMVC*, 1999, pp. 173-182.

[197] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor, "Canonical correlation analysis: An overview with application to learning methods," *Neural computation,* vol. 16, pp. 2639-2664, 2004.

[198] C. Goodall, "Procrustes methods in the statistical analysis of shape," *Journal of the Royal Statistical Society. Series B (Methodological),* pp. 285-339, 1991.

[199] I. Matthews and S. Baker, "Active appearance models revisited," *International Journal of Computer Vision,* vol. 60, pp. 135-164, 2004.

[200] S. Baker and I. Matthews, "Lucas-kanade 20 years on: A unifying framework," *International journal of computer vision,* vol. 56, pp. 221-255, 2004.

[201] M. M. Nordstrøm, M. Larsen, J. Sierakowski, and M. B. Stegmann, "The IMM face database-an annotated dataset of 240 face images," Technical University of Denmark, DTU Informatics, Building 3212004.

[202] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, *et al.*, "Bosphorus database for 3D face analysis," in *European Workshop on Biometrics and Identity Management*, 2008, pp. 47-56.

[203] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and simile classifiers for face verification," in *2009 IEEE 12th International Conference on Computer Vision*, 2009, pp. 365-372.

[204] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "300 faces in-the-wild challenge: The first facial landmark localization challenge," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2013, pp. 397-403.

[205] J. Yang, Z. Lin, and S. Cohen, "Fast image super-resolution based on in-place example regression," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1059-1066.

[206] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, *et al.*, "The CAS-PEAL large-scale Chinese face database and baseline evaluations," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 38, pp. 149-161, 2008.

[207] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, pp. 600-612, 2004.

[208] X. Tao, X. Chen, X. Yang, and J. Tian, "Fingerprint recognition with identical twin fingerprints," *PloS one*, vol. 7, p. e35704, 2012.

[209] D. Li and K.-M. Lam, "Design and learn distinctive features from pore-scale facial keypoints," *Pattern Recognition*, vol. 48, pp. 732-745, 2015.

[210] T. H. Cormen, *Introduction to algorithms*: MIT press, 2009.

[211] W. H. Press, *Numerical recipes 3rd edition: The art of scientific computing*: Cambridge university press, 2007.

[212] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, pp. 381-395, 1981.

[213] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-pie," *Image and Vision Computing*, vol. 28, pp. 807-813, 2010.

[214] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, *et al.*, "Overview of the face recognition grand challenge," in *2005 IEEE Computer*

Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 2005, pp. 947-954.

[215] T. S. Lee, "Image representation using 2D Gabor wavelets," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 18, pp. 959-971, 1996.

[216] J. R. Kettenring, "Canonical analysis of several sets of variables," *Biometrika,* vol. 58, pp. 433-451, 1971.

[217] J. G. Daugman, "Two-dimensional spectral analysis of cortical receptive field profiles," *Vision research,* vol. 20, pp. 847-856, 1980.

[218] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *JOSA A,* vol. 2, pp. 1160-1169, 1985.

[219] D.-H. Liu, K.-M. Lam, and L.-S. Shen, "Optimal sampling of Gabor features for face recognition," *Pattern Recognition Letters,* vol. 25, pp. 267-276, 2004.

[220] Q.-S. Sun, Z.-d. Liu, P.-A. Heng, and D.-S. Xia, "A theorem on the generalized canonical projective vectors," *Pattern Recognition,* vol. 38, pp. 449-452, 2005.

[221] K. A. Deffenbacher, T. Vetter, J. Johanson, and A. J. O'Toole, "Facial aging, attractiveness, and distinctiveness," *Perception,* vol. 27, pp. 1233-1243, 1998.

[222] K.-M. Lam and H. Yan, "Locating and extracting the eye in human face images," *Pattern recognition,* vol. 29, pp. 771-779, 1996.

[223] H. Zhou, K.-W. Wong, and K.-M. Lam, "Feature-aging for age-invariant face recognition," in *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, 2015, pp. 1161-1165.