2015

Facial Image Analysis for Body Mass Index, Makeup and Identity

Lingyun Wen

Facial Image Analysis for Body Mass Index, Makeup and Identity

by

Lingyun Wen

Dissertation submitted to the Benjamin M. Statler College of
Engineering and Mineral Resources
at West Virginia University
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
in
Computer Science

Approved by

Guodong Guo, Ph.D. Chair
Xin Li, Ph.D. Co-Chair
Donald Adjeroh, Ph.D.
Arun Ross, Ph.D.
CunQuan Zhang, Ph.D.

Lane Department of Computer Science and Electrical Engineering

Morgantown, West Virginia
2014

Keywords: Facial Image Analysis, Body Mass Index Prediction, Makeup Detection, Robust Face Recognition

Abstract

Facial Image Analysis for Body Mass Index, Makeup and Identity

by Lingyun Wen

The principal aim of facial image analysis in computer vision is to extract valuable information (e.g., age, gender, ethnicity, and identity) by interpreting perceived electronic signals from face images. In this dissertation, we develop facial image analysis systems for body mass index (BMI) prediction, makeup detection, as well as facial identity with makeup changes and BMI variations.

BMI is a commonly used measure of body fatness. In the first part of this thesis, we study BMI related topics. At first, we develop a computational method to predict BMI information from face images automatically. We formulate the BMI prediction from facial features as a machine vision problem. Three regression methods, including least square estimation, Gaussian processes for regression, and support vector regression are employed to predict the BMI value. Our preliminary results show that it is feasible to develop a computational system for BMI prediction from face images. Secondly, we address the influence of BMI changes on face identity. Both synthesized and real face images are assembled as the databases to facilitate our study. Empirically, we found that large BMI alterations can significantly reduce the matching accuracy of the face recognition system. Then we study if the influence of BMI changes can be reduced to improve the face recognition performance. The partial least squares (PLS) method is applied for this purpose. Experimental results show the feasibility to develop algorithms to address the influence of facial adiposity variations on face recognition, caused by BMI changes.

Makeup can affect facial appearance obviously. In the second part of this thesis, we deal with makeup influence on face identity. It is principal to perform makeup detection at first to address makeup influence. Four categories of features are proposed to characterize facial makeup cues in our study, including skin color tone, skin smoothness, texture, and highlight. A patch selection scheme and discriminative mapping are presented to enhance the performance of makeup detection. Secondly, we study dual attributes from makeup and non-makeup faces separately to reflect facial appearance changes caused by makeup in a semantic level. Cross-makeup attribute classification and accuracy change analysis is operated to divide dual attributes into four categories according to different makeup effects. To develop a face recognition system that is robust to facial makeup, PLS method is proposed on features extracted from local patches. We also propose a dual-attributes based method for face verification. Shared dual attributes can be used to measure facial similarity, rather than a direct matching with low-level features. Experimental results demonstrate the feasibility to eliminate the influence of makeup on face recognition.

In summary, contributions of this dissertation center in developing facial image analysis systems to deal with newly emerged topics effectively, i.e., BMI prediction, makeup detection, and the recognition of face identity with makeup and BMI changes. In particular,

to the best of our knowledge, BMI related topics, i.e., BMI prediction; the influence of BMI changes on face recognition; and face recognition robust to BMI changes are first explorations to the biometrics society.

Acknowledgements

First of all, I would like to express my sincere thanks to my advisor, Dr. Guodong Guo and co-advisor, Dr. Xin Li for their knowledgeable guidance, ingenious ideas and valuable advice through my Ph.D. study. What I learned from them will benefit me all my life. With their inspirational guidance, I was able to solve many problems in my research projects and achieved successful experience in facial image analysis.

I would like to thank my committee, Dr. Donald Adjeroh, Dr. Arun Ross and Dr. CunQuan Zhang, for their valuable suggestions and guidance during my Ph.D. study. I also want to thank their precious time and help during the completion of my thesis.

I also want to thank Dr. Hany Ammar, Dr. Thirimachos Bourlai, and Dr. Vinod Kulathumani for their helpful suggestions during my Ph.D. study.

I would like to thank Yu Zhu, Biyun Lai, Xiaolong Wang, Chao Zhang, Wenbin Chen, Cesare Ciaccio, Qin Wu and all my colleagues in the computer vision lab for their friendly nature which made working in the lab more enjoyable.

I would like to thank the Lane Department of Computer Science and Electrical Engineering at West Virginia University for providing me with an excellent study environment during my years as a Ph.D. student.

Finally, I want to thank the support of my family, my parents and my husband.

Contents

List of Figures

List of Tables

Abbreviations

ACA	**A**ccuracy **C**hange **A**nalysis
ASM	**A**ctive **S**hape **M**odel
AUC	**A**rea **U**nder the ROC **C**urve
BMI	**B**ody **M**ass **I**ndex
CJWR	**C**heekbone to **J**aw **W**idth **R**atio
DM	**D**irect **M**atching
EER	**E**qual **E**rror**R**ate
ES	**E**ye **S**ize
FPR	**F**alse **P**ositive **R**ate
FW/LFH	**F**ace **W**idth to **L**ower **F**ace **H**eight
GP	**G**aussion **P**rocess
HOG	**H**istogram of **O**riented **G**radients
LBP	**L**ocal **B**inary **P**attern
LF/FH	**L**ower **F**ace to **F**ace **H**eight
LSE	**L**east **S**quares **E**stimation
MAE	**M**ean **A**bsolute **E**rror
MEH	**M**ean of **E**yebrow **H**eight
MFA	**M**arginal **F**isher **A**nalysis
NIPALS	**N**onlinear **I**terative **P**artial **L**east **S**quares
PAR	**P**erimeter to **A**rea **R**atio
PCA	**P**rincipal **C**omponent **A**nalysis
PLS	**P**artial **L**east **S**quares

RBF	**R**adial **B**asis **F**unction
ROC	**R**eceiver **O**perating **C**haracteristic
SIFT	**S**cale **I**nvariant **F**eature **T**ransform
SVR	**S**upport **V**ector **R**egression
TPR	**T**rue **P**ositive **R**ate
WHR	**W**idth to upper facial **Height R**atio

Chapter 1

Introduction

Human faces, as the major outputs and sensory inputs in communication, encode a lot of valuable information, such as demographic characteristics (e.g., age and gender), identity, and psychological conditions (e.g., happy and sad emotions). Facial image analysis provides a no-intrusive, effective, and economical way to perceive and interpret facial signals, thus having a wide range of critical applications such as biometric identification, surveillance and security, human-computer interaction, image/video retrieval, and computer animation [1].

In this thesis, our objective is to study several newly emerged topics in facial image analysis for BMI, makeup, and identity. In this chapter, we at first introduce the background and motivation of our studies. Then we summarize our contributions. Finally, the organization of the dissertation will be illustrated.

1.1 Background and Motivation

In this Section, at first, we introduce the concept of facial identity. Then we illustrate the background and motivation for BMI prediction and facial identity with BMI changes.

Finally, the background and motivation for makeup detection, and facial identity with makeup changes will be given.

1.1.1 Facial Identity

Human face is an important biometrics character, as important as other characters such as fingerprint, palmprint, and iris. A facial recognition system is a computer application for automatically identifying or verifying a person from a digital facial image. It is a general concept including two modes: face verification (or authentication) and face identification. Face verification is to verify whether two facial images come from the same person or not. Face identification is to determine the identity of a person given his/her facial image and a facial image database with known identity. Face recognition is one of the most important applications of image analysis and understanding due to its wide range of government and commercial applications: law enforcement, access control security, entertainment, smart card and so on.

The first batch of face recognition research using computer were in the 1960's. Among those works, Woodrow W. Bledsoe described problems influenced by variations in illumination, head rotation, facial expression, aging during his face recognition study in 1966 [2]. Even after almost 50 years, these problems are still hot topics related to robust face recognition. Many efforts have been made to develop a face recognition system that is invariant or insensitive to above variations, such as pose [3–5], illumination [6, 7], expression [8, 9], ageing [10, 11] and so on. To develop robust face recognition for real applications, plenty of other factors should be dealt with. In this thesis, two new factors: makeup and BMI variations are studied because of the current social demands.

1.1.2 Body Mass Index

BMI is a measure of body fatness for an individual. Gallagher et al. [12] showed that BMI is independent of age, gender, and ethnicity. Given an individual's height and weight, the calculation of body mass index (BMI) is given by Formula 1.1 [13], depending on the units to use, e.g., meter (m) for height and kilogram (kg) for weight, or inch (in) for height and pond (lb) for weight.

$$\textbf{BMI} = \begin{cases} \frac{weight(kg)}{height(m)^2} \\ \\ \frac{weight(lb)\times 703}{height(in)^2} \end{cases} \tag{1.1}$$

According to the values of BMI, people are divided into four categories: under weight, normal, overweight, and obese. The category standard is shown in Table 1.1.

Table 1.1: BMI categories. According to BMI distribution, people are divided into four BMI categories: under weight, normal, overweight and obese.

Category	BMI
under weight	< 18.5
normal	$18.5 - 24.9$
overweight	$25.0 - 29.9$
obese	> 30

Besides a measure of body fatness, BMI is also valuable in many other applications. BMI is even related to personal attractiveness. In most cultures, a person with normal BMI would be more attractive comparing with other BMI categories. While in some cultures, people with larger BMI would be considered more attractive [14]. What's more, BMI is also an important measurement in medicine domain, as important as temperature, blood pressure, and heartbeat. Some research also reveals that BMI is associated with the risk of some common adult cancers [15].

It is easy for a human to judge the fatness of a person by only observing his/her face. This phenomenon indicates that human faces are related to BMI. However, there is no previous work in computer vision to decode BMI from facial signals. Recent studies in psychology and human perception [16–18] have found that facial features are related to human weight or BMI. These studies focus on finding the related, specific features on faces, and measuring the correlations between the facial features and BMI. Motivated by these studies, we at first validate these conclusions on a large database. Then based on these conclusions, we propose a computational approach to BMI prediction from face images by formulating the BMI prediction as a regression problem.

1.1.2.1 Facial Identity with BMI Changes

Nowadays, more than two-thirds of U.S. adults are overweight or obese [19]. Obesity and overweight account for nearly one of every 10 American deaths and can cause many health problems, such as diabetes, hypertension, heart disease and even some common adult cancers [20, 21]. Recently, obesity has been officially recognized as a disease by the American Medical Association to promote its treatment and prevention [22].

Popular weight loss methods include good diet and exercise, and medical weight loss treatments (including drug therapy and surgery treatments). In particular, surgery treatments provide powerful solutions for quick weight loss [23]. Weight changes can change facial appearance obviously. However, no previous research has studied the influence of BMI variations on face recognition. In this thesis, We use BMI changes to describe the variation of fatness caused by weight gain and loss in adult group. Our objective is to study the influence of BMI changes on face recognition. What's more, we also aim to develop face recognition system more robust to BMI changes.

1.1.3 Facial Makeup

Social scientists conducted many surveys about the behavior for women to wear makeup. One survey reveals that over 70% British women do not want to show the makeup-free face in their study or work places. 44% American women feel unattractive when they don't put makeup on. One in three women will never leave home without makeup [24]. The aim of performing facial makeup is to temporarily change the facial appearance to appear more attractive according to the aesthetic standard.

The origin of makeup can even be back up to the year 4,000 BC in ancient society [25]. Archaeological evidence of cosmetics dates at least back to the ancient Egypt and Greece [26, 27]. As the standard of beauty varies in different culture, ethnics over history, makeup products and styles also vary a lot [28]. In the modern society, as a result of globalization, similar makeup products and styles spread all over the world. In this thesis, we focus on modern facial makeup with the effects of popular makeup products, such as eye shadow, foundation, lipstick and so on.

1.1.3.1 Standard of Facial Beauty

Then the question is how human beings define beauty? Is there any universal measurement to measure beauty to guide women to perform makeup? For centuries, artists and scientists tried to find the rules to measure beauty. Even during the 1st century BC, the ancient Roman architect Marcus Vitruvius Pollio in his best known treatise De Architectura [29] described ideal human body proportions with geometry. In 1487, Leonardo da Vinci draw the famous "Vitruvian Man" according to this description. Among these descriptions, some are related to the face geometry [29] as follows: It is one-tenth of the height of a man from the hairline to the bottom of the chin; The ratio from below the chin to the top of the head is one-eighth of the height of a man.

Facial beauty is studied extensively by artists and scientists especially in mordern times [30–35]. [30] and [36] studied the relation between specific adult female facial features and the attraction. Cunningham et al. [31] examined attractiveness ratings across cultural groups. Research in [32, 33] proved that faces representing the average value of the population would be consistently judged as attractive.

In the last decades, computer has an important effect in the study of beauty. Johnston et al. [36] used computers to generate a beauty female face based on personally judgment using genetic algorithm. Etcoff et al. [35] discovered that the positions of the eyes, mouth, and the nose are related to the beauty of the individual. Aarabi et al. [37] developed an automatic facial beauty scoring system based on ratios between facial features. The ratios including: (1)Ratio of horizontal distance between eyes to average vertical distance between eyes and mouth. (2)Ratio of horizontal distance between eyes to the width of the face. (3)Ratio of average vertical distance between eyes and mouth to height of the face. (4) Ratio of distance between mouth and left eye. (5) Ratio of average vertical distance between top of face and eyes to average vertical distance between eyes and bottom of face. (6) Ratio of vertical distance between the mouth and the center of the face to the vertical distance between mouth and the bottom of the face. (7) Ratio of vertical distance between the eyes and the center of the face to the vertical distance between the center of the face and the bottom of the face. (8) Ratio of the horizontal distance between the middle of the eyes and the center of the face to the width of the face.

How makeup helps to improve attractiveness has also attracted much attention. Sharma et al. found that makeup can cover unwanted facial structures, such as wrinkles or pores and also emphasize specific features which are considered attractive [38]. The improved attractiveness using cosmetics has been studied in [39, 40].

When women enjoy their beauty, it brings difficulties to facial image analysis. Because facial makeup can change the perceived appearance of the face temporarily, such as more attractive and younger [41–43]. The accuracy of face recognition will decline when matching faces between makeup and non-makeup groups. Dantcheva et al. showed that facial makeup can impact the performance of face recognition obviously [43]. Besides, other facial image analysis tasks such as age estimation can also give bias results under the influence of makeup. To address the problems brought by makeup, it is important to firstly detect whether a woman wears makeup or not. Then proper approaches can be applied to eliminate the influence of makeup on facial image analysis applications. In this thesis, we develop a facial makeup detection approach based on four facial cues, i.e., skin color tone, skin smoothness, texture, and highlight.

1.1.3.2 Facial Identity with Makeup Changes

In human perception and psychology studies [41, 42], it is revealed that light makeup slightly helps recognition, while heavy makeup significantly decreases human ability to recognize faces. In a computational approach, the impact of facial makeup on face recognition has been presented very recently [43]. It was also shown that facial makeup can significantly change the facial appearance, both locally and globally [43]. In this thesis, we study the makeup effects on face appearance in a semantic level. Facial attributes, such as big eyes, small eyes, etc. are studied from faces with and without makeup, separately. We want to give answers to questions as follows. Which facial attribute changes only a little before and after makeup? Which attribute is enhanced by makeup? Which attribute changes a lot with makeup effect? Existing face matching methods based on contrast and texture information can be impacted by the application of facial makeup [43]. In our study, we also develop face recognition systems to decrease the influence of makeup changes.

1.2 Contributions

We summarize our contributions in this thesis as follows:

1) An automatic facial image analysis approach to BMI prediction is proposed. Particularly, it is the first attempt to predict BMI from facial images automatically. We also study the influence of BMI changes on face recognition. A correlation mapping method is explored to increase the accuracy of face recognition with BMI changes.

2) We develop a facial makeup detection system with high accuracy. Furthermore, facial attributes are explored to describe the makeup effect on facial identities in a semantic level. Two face verification systems, including PLS and dual attribute methods are developed to deal with makeup changes separately.

1.3 Organization

The remaining structure of this thesis is as follows:

Chapter 2 focuses on decoding BMI information from facial images. Firstly, the relations between facial ratios and BMI is validated on a large database. This validation provides the basis to develop a computational BMI prediction system. In the following, we will introduce the BMI prediction system based on a statistical learning formulation.

In Chapter 3, the influence of both small and large BMI changes on face recognition will be illustrated at first. The experimental results show that only large BMI changes can influence face recognition obviously. Then correlation mapping based face recognition will be introduced to decrease the influence of large BMI changes.

In Chapter 4, proposed facial makeup detection will be illustrated. We will show the steps to extract features related to four makeup cues, including skin color tone, skin smoothness, texture, and highlight. Discriminative mapping method and patch selection scheme will be illustrated in the following sections.

In Chapter 5, dual-attributes will be learned from faces with and without makeup, separately. By accuracy change analysis (ACA) for the attribute classification results, we can obtain semantic interpretation of makeup influence on face appearance. Chapter 5 also illustrates approaches to develop face verification robust to makeup changes. PLS approach is used to reduce the influence of makeup changes. Combining makeup detection method as illustrated in Chapter 4, a complete automatic face verification system can be developed. Besides, dual attributes are proposed to be used as feature descriptors to build a robust face verification system.

In the end, Chapter 6 presents the future work and conclusions of the dissertation.

Chapter 2

BMI Prediction based on Facial Features

2.1 Introduction

Facial image analysis offers a convenient and practical way to obtain valuable information without contacting the subject directly. How to decode BMI information from face images is a new problem. To develop a facial signal decoder for BMI, we need to at first know BMI cues on the face. Recent studies in psychology revealed the correlations between seven facial features and BMI. In our study, we firstly validate their conclusions on a large database which includes over 14,500 facial images to provide basis for BMI prediction.

To compute an individual's BMI, the traditional method is to measure both the body weight and height. Thus some tools are required, such as a ruler and a scale. As shown in Figure 2.1, we can give an intuitive prediction about how fat a person is just by observing the face image. This phenomenon gives us the inspiration that face images can offer cues to predict BMI [44].

Recent studies in psychology and human perception [16–18] have found that facial features have relations to BMI. These studies focus on finding the related, specific features on faces,

Figure 2.1: Some face images with BMI values and categories. Increased facial adiposity can be observed obviously as the BMI increasing.

and measuring the correlations between the facial features and BMI. In Chapter 2, we validate these conclusions on a large facial image database using automatic facial image analysis technologies. Based on the psychology studies and our intuitive observation, we believe that it is worth to investigate a computational approach to body mass index prediction from faces. In our computational approach, we will perform BMI prediction and evaluate its accuracy on a large database. To the best of our knowledge, no previous work has been done to predict BMI from face images in computer vision.

One advantage of predicting BMI from face images is that, the approach is non-invasive. There is no need to measure an individual's height and weight in order to compute his or her BMI. This is a nice property for some practical uses, such as in on-line photos or surveillance videos of faces, where it is impossible to use traditional measures of weight and height for BMI calculation. Recently, there are more and more online dating or friend search sites, where possibly only face photos are shown for each individual. The automated prediction of BMI from face photos can be useful to judge bodily fatness in such scenarios.

2.1.1 Related Work

The studies in psychology and human perception [16–18] have found that facial features are correlated to BMI. These conclusions provide basis for BMI prediction based on face images.

Coetzee et al. [16] showed that facial adiposity, or the perception of weight in the face, can predict perceived health and attractiveness. They recruited 84 Caucasian participants (43 female and 41 male) to capture face photos. The weight, height, and other information, e.g., blood pressure, of all participants were recorded. Then they recruited another four groups of people to rate the facial images manually. They showed that the rated facial adiposity is related to BMI.

Coetzee et al. [17] studied three facial features: width-to-height ratio, perimeter to area ratio, and cheek to jaw width ratio. They captured face photos of 95 Caucasian and 99 African participants. Each facial image was manually delineated by defining 179 feature points and aligned according to interpupillary distance using a computer software. Pearson's correlations were computed. They showed that the three facial features were significantly related to BMI in males, while only width-to-height ratio and cheek to jaw ratio were significantly related to BMI in females.

Pham et al. [18] studied correlation between 7 facial features and BMI, in young and elderly people in Korean. They recruited 911 participants in two age groups: people in twenties and sixties. A well-trained operator was asked to manually label the facial features. Then the Pearson's correlation coefficient was calculated to assess the association of facial features to BMI.

However, there are some limitations in psychology studies or human perception: (1) usually a small number of face images are used; (2) facial ratios have to be manually labeled; (3) a limited number of age groups are captured; and (4) the study results have not been validated on a large scale database.

Motivated by the psychology and human perception studies in [16–18], we want to validate the psychology study results on a large database with automatical computer vision technique to provide solid basis for automatic BMI prediction.

In computer vision, the most related work is to predict body weight based on human metrology using a copula model in Bayes theorem by Cao et al. [45]. In their work, they showed the effect to predict body weight using metrology, e.g. the length of the arm. Some of the characters are from the head, such as bizygomatic breadth, face length, head breadth, head circumference, head length. Their goal is to predict the body weight, while we focus on predicting the measure of fatness, i.e. BMI. Besides, we only use facial cues while excluding cues from other body parts.

In the remaining of the Chapter, we firstly give some introduction about seven facial features used in our work in Section 2.2. Then we introduce the calculation of the Pearson's Correlation Coefficient between facial features and BMI in Section 2.3. In Section 2.4, we introduce the framework that we develop to calculate facial features and their correlations with BMI. In Section 2.7.3, we show that the correlations in [16–18] also exist on our large scale database. BMI prediction is modeled as a regression problem. We present our computational approaches for BMI prediction in Section 2.5. The experimental results on a large scale database for BMI prediction are described in Section 2.7.3. And finally, we give the summary in Section 2.8.

2.2 Facial Ratios

Seven facial ratios are used in our approach. They are cheekbone to jaw width ratio (CJWR), width to upper facial height ratio (WHR), perimeter to area ratio (PAR), eye size (ES), lower face to face height ratio (LF/FH), face width to lower face height ratio (FW/LFH), and mean of eyebrow height (MEH). These features are similar to those in [18], but are detected and computed automatically. Details of these facial features are shown in Figure 2.2.

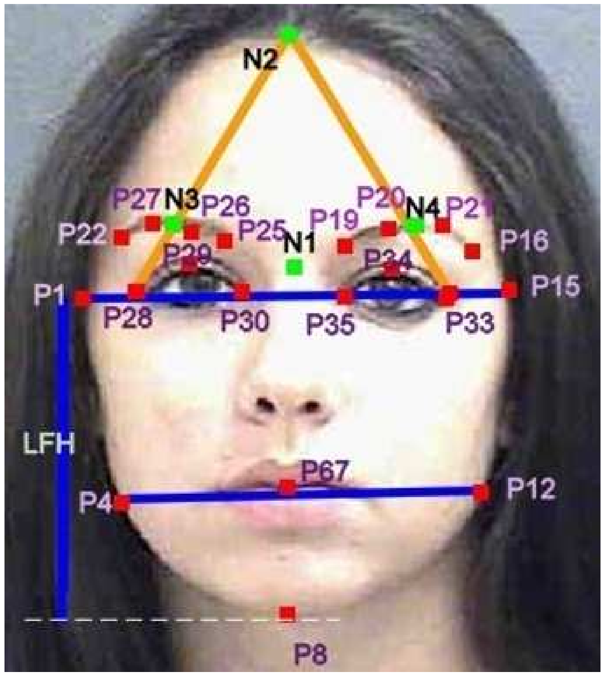


FIGURE 2.2: Illustration of the facial features for BMI prediction.

All the facial ratio and geometry features are described below in detail, and related points are shown in Figure 2.2:

(1) CJWR is the ratio of the cheekbone width to jaw width. The cheekbone width is the distance between P_1 and P_{15}, and jaw width is the distance between P_4 and P_{12};

(2) WHR is the ratio of the cheekbone width to upper facial height which is the distance between P_{67} and N_1. The formula of WHR is $\frac{|P_1P_{15}|}{|P_{67}N_1|}$;

(3) PAR is the ratio of the perimeter to area of polygon running through $P_1P_4P_8P_{12}P_{15}P_1$;

(4) ES is the average size of eyes, which is the average distance between P_{28} and P_{33} minus distance between P_{30} and P_{35}. The formula is $\frac{1}{2}(|P_{28}P_{33}| - |P_{30}P_{35}|)$;

(5) LF/FH is the lower face to face height ratio. The lower face is the part of the face include and below the cheekbone (line P_1 P_{15}). The lower face height is the distance between the cheekbone and the lowest point in the jaw which is shown in Figure 2.2 as the distance LFH. The face height is the distance between the highest point N_2 and the lowest point P_8 on the face. The formula is $\frac{|LFH|}{|N_2P_8|}$;

(6) FW/LFH is the face width to the lower face height ratio. The face width is the cheekbone width.

(7) MEH is the average distance between eyebrows and the upper edge of eyes, which is the mean of distances between P_{22} to P_{28}, N_3 to P_{29}, P_{30} to P_{25}, P_{19} to P_{35}, N_4 to P_{34}, P_{16} to P_{33}. The formula is $\frac{1}{6}|P_{22}P_{28}| + |N_3P_{29}| + P_{30}P_{25}| + |P_{19}P_{35}| + |N_4P_{34}| + |P_{16}P_{33}|$.

2.3 Pearson's Correlations and Hypothesis Testing

As mentioned above, Coetzee et al. [17] and Pham et al. [18] both used Pearson's Correlation Coefficient to reveal relationship between facial features and BMI in their work. Although Pearson's Correlation Coefficient is not used often in computer vision, it is the most widely used measure of relationship. For example, Carroll (1961), in his presidential address to the Psychometric Society, called the correlation coefficient as "one of the most frequently used tolls of psychometricians" [46]. We also calculate Pearson's Correlation Coefficient on the face images database: Morph-II, which is a very large database, to validate whether the relationship between facial features and BMI is still existing on a large scale database. In this section, we will first introduce the calculation of Pearson's Correlation Coefficient. Then we introduce the hypothesis testing, which is used to test the significance of the obtained relationship between facial features and BMI. That's because the Pearson's correlation coefficient is obtained from a sample of observations. By the hypothesis testing, we can draw a conclusion whether the facial features and BMI are correlated in the population.

2.3.1 Pearson's Correlation Coefficient

Pearson's Correlation Coefficient is also called Pearson's r. Pearson developed the mathematical formula in 1895, and it is still most commonly used to measure correlation.

Pearson focused on the correlation coefficient as a computational index used to measure bivariate association. The correlation coefficient constitutes the principal statistical methodology for observational experiments in many disciplines. [46]

To calculate the correlation between Facial features and BMI, Pearson's correlation coefficient is used. It is calculated by:

$$r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}}$$

where r is Pearson's correlation coefficient, which is between -1 and 1. If $r < 0$, it indicates a negative correlation between X and Y. If $r > 0$, it indicates a positive correlation. If $r = 0$, it indicates there is no correlation between X and Y [47]. If Pearson's r is close to -1 and +1, it indicates a very strong correlation. A value of 0 means that no correlation exists. Pearson's r measures the strength of a linear relationship between two variables. Closer the value of r is to 0, weaker the linear correlation becomes.

2.3.2 Hypothesis Testing

If facial features have strong associations with BMI according to the Pearson's correlation coefficient, regression process can be performed to predict BMI using facial features. However, these are the sample correlation coefficients obtained from observations in our training set. We are interested to know whether the facial features and BMI are correlated in the population. Only the associations which is also strong in the population can be used in practice. In this case we can use Hypothesis Testing for these correlation coefficients.

In statistics, p-value describes how certain a hypothesis is true. There are two possible interpretations of one observation. The first one is called the null hypothesis and the second one is called alternate hypothesis. In most situations, the null hypothesis says that the effect indicated by the sample is due only to random variation between the sample and the population. The alternate hypothesis says that the effect indicated by the sample is real, in that it accurately represents the whole population [48]. The p-value is used to indicate whether a significant correlation exists between facial features and BMI.

In performing a hypothesis test, null hypothesis is assumed to be true at the beginning. The hypothesis test involves measuring the strength of the disagreement between the sample and null hypothesis to produce p-value which is a number between 0 and 1. The p-value measures the plausibility of null hypothesis. The smaller the p-value, the stronger the evidence is against null hypothesis. If the p-value is sufficiently small, null hypothesis would be abandoned and alternate hypothesis is believed [48].

Null hypothesis is defined as there is no correlation between facial features and BMI. And the alternate hypothesis is defined as there is correlation between facial features and BMI.

As we mentioned above, p-value indicates the correlation between facial features and BMI. Smaller p-value, stronger correlation exists. Usually, we can set a threshold such as 0.001, 0.01 or 0.05. If the p-value is equal or smaller than the threshold, it indicates a significant correlation between facial features and BMI.

The Pearson correlation coefficient and corresponding p-value between facial features and BMI in Morph-II will be shown in Section 2.7.3.

2.4 The Framework of Validation

An automatic framework of validation the correlations between facial ratios and BMI is proposed. The whole framework is shown in Figure 2.3. In the following, we describe all related procedures.

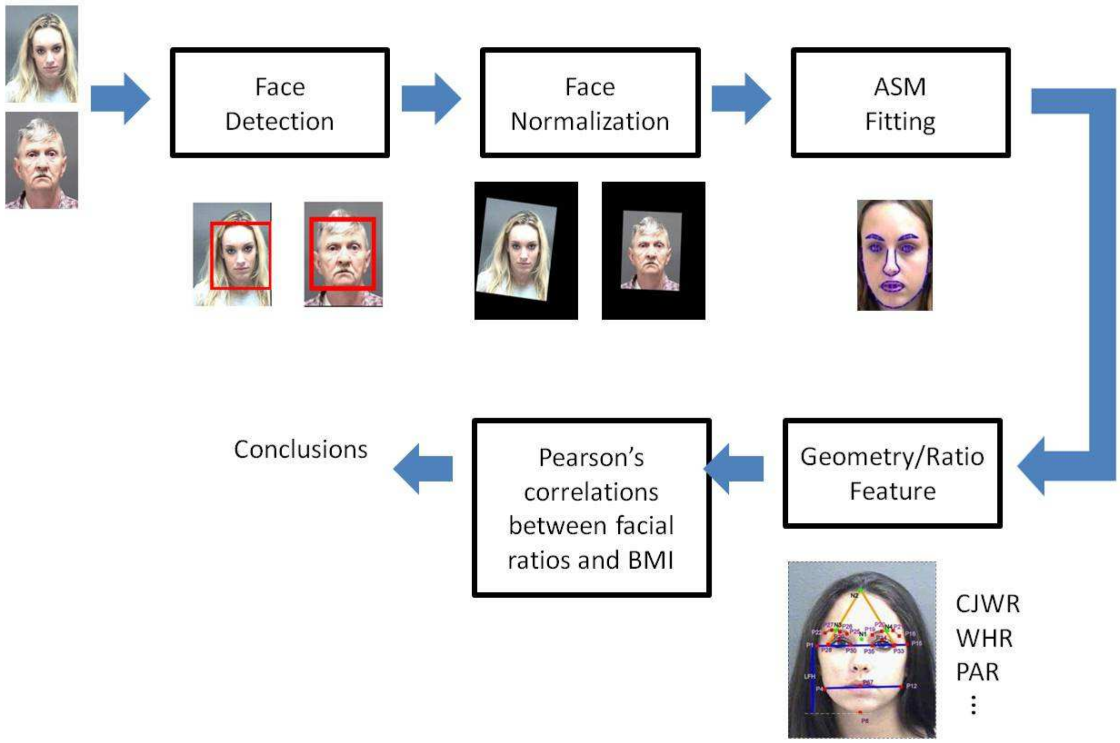


FIGURE 2.3: The proposed framework for validation of correlations between facial ratios and BMI.

2.4.1 Face Detection, Alignment, and Fitting

For a given face image, the first step is to execute face detection, and the detection of two eyes using a technique similar to face detection [49]. The difference is what templates are used for training the AdaBoost classifier. For face detection, the face templates are used, while for eye detection, the templates are replaced with eyes. Then each detected face is normalized based on the detected eye coordinates. The normalization is basically to perform translation, rotation, and scaling of the faces so as to align all face images into the common eye coordinates. These processing steps are important in developing a robust computational system.

Next, the active shape model (ASM) is used to detect a number of fiducial points in each face image. The ASM method was originally proposed by Cootes et al. [50]. In the ASM model, the principal component analysis (PCA) is applied on the locations of facial components (e.g., eyes, nose, lips, facial contours, etc.), then ASM is presented as connected point distributions from a variety of manually labeled images with variations such as pose, illumination, and expression changes.

Milborrow [51] extends the ASM with a so-called Constrained Local Model to improve its performance. The ASM can perform well in facial feature extraction [52], structure locating in medical images [53], and many other applications. An example of the ASM fitting on a test face image is shown in Figure 2.4. Please note that we apply the ASM model to the normalized face images, rather than the original images that may contain various sizes of faces and head rotations. Letting the ASM fitting work on the normalized face images is helpful to make the fiducial point detection more robust and accurate. The detected facial points by the ASM are used as the input for the next step to compute facial geometry or ratio features.

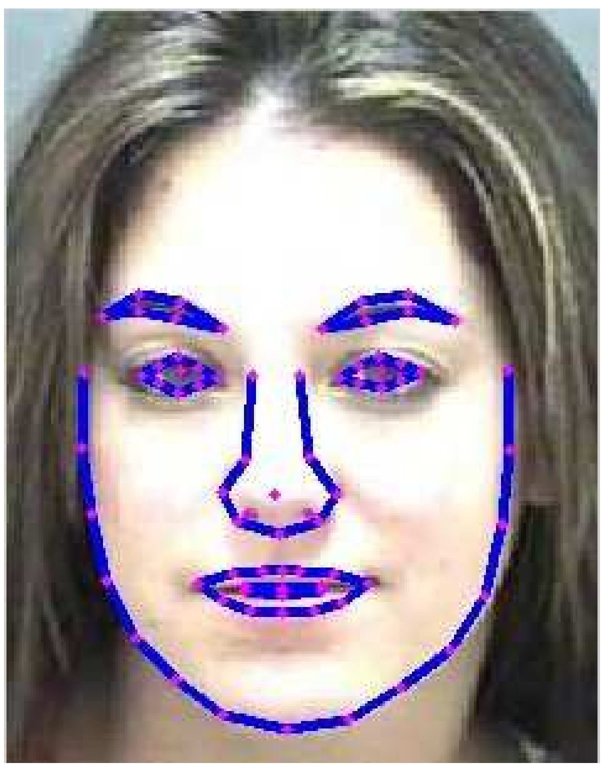


FIGURE 2.4: The result of the ASM fitting on a face image.

The ASM model can produce 76 points on a face image [54]. We use 20 out of 76 points to produce the facial features that are related to BMI. Those points are marked as P_* in

Figure 2.2 (* indicates the *th points from the total 76). Because 76 points still cannot generate all the features needed, four additional points are added to generate all the features. The added additional points are marked as $N_1 - N_4$ in Figure 2.2.

2.4.2 Automatic Feature Computation

Note that even with the big number of 76 points detected by the ASM, as shown in Figure 2.4, there are still some other points that are needed to compute the features. Actually, another four points are needed but cannot be detected by the ASM method, denoted by $N_1 - N_4$ in Figure 2.2. We estimate the positions of these four points based on the detected fiducial points by the ASM method.

N_1 is the midpoint of P_{29} and P_{34}. N_2 is the intersection point of line P_{28} N_3 and line P_{33} N_4. N_3 is the midpoint of P_{26} and P_{27}. N_4 is the midpoint of P_{20} and P_{21}. We use N_2 as an estimate of the highest point on a face. This estimate is based on our observations. To evaluate this estimation method, we randomly selected 100 images from our experiment database: MORPH-II, and manually labeled all needed points to produce N_2 and also the real highest point of the face. We use N_2 as an estimate of the highest point on a face. This estimate is based on our observations and a quantitative evaluation that will be presented in the next.

Given the fitted fiducial points by the ASM method, and the estimated points described above, i.e., $N_1 - N_4$, all of the facial geometry and ratio features, introduced in Section 2.2, can be computed automatically. These features are supposed to be correlated to BMI, motivated by the psychology studies [17, 18]. We believe that it is a significant progress to show that the BMI related facial features can be computed automatically in each face image.

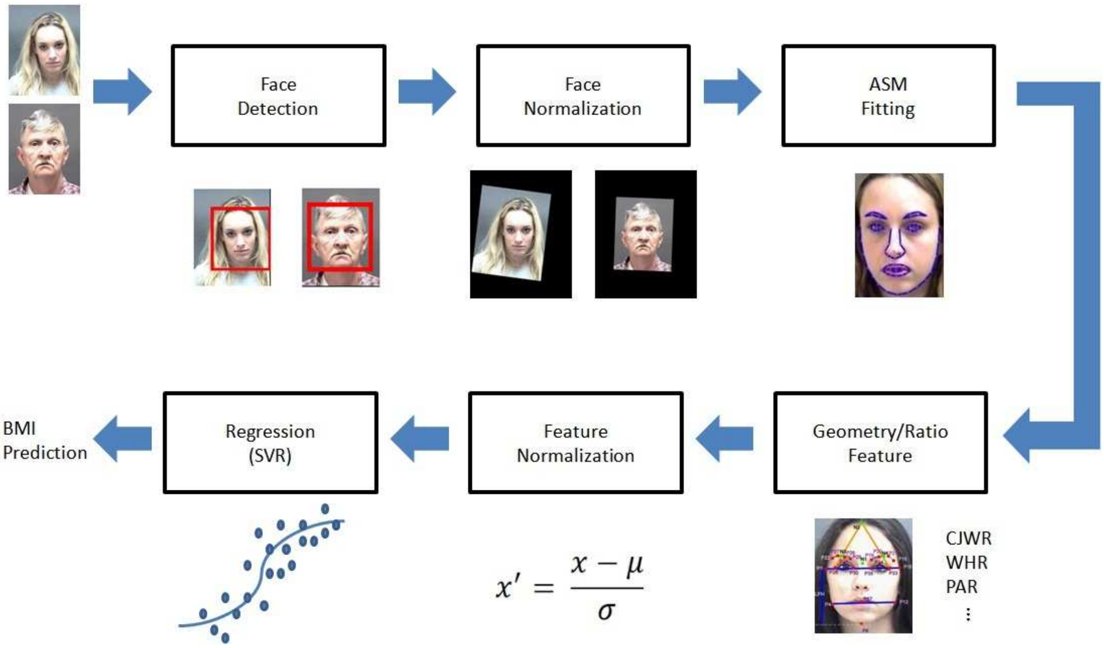


FIGURE 2.5: The proposed framework for body mass index prediction from a face image.

2.5 Prediction Framework

The computational framework for BMI prediction are shown in Figure 2.5. The pre-processing steps are the same as used in the Chapter 2 for correlation study, including face detection, face normalization and generating facial Ratio and Geometry features. After that, a normalization is executed to get normalized features by

$$x' = \frac{x-\mu}{\sigma}, \tag{2.1}$$

where μ is the average value and σ is the standard deviation, computed from the training data along each feature dimension. After feature normalization, we perform a statistical learning to build the relation between facial features and BMI. Specifically, we consider the BMI prediction as a regression problem. In the end, the regression model can predict the BMI when a face image is provided.

In Section 2.6, we will describe the statistical learning methods that can be used for BMI prediction in details.

2.6 Statistical Learning Methods

We define the facial features based BMI prediction as a statistical learning problem. Three regression methods, support vector regression (SVR) , the least squares estimation (LSE), and Gaussian process regression (GP) are compared for learning the BMI prediction.

2.6.1 Support Vector Regression

Given a set of training data $(x_1, y_1), ..., (x_l, y_l) \subset X \times R$, where X denotes the space of the input patterns (e.g. $X = R_d$). SVR is to learn a function $f(x)$ that has at most ε deviation from the actually obtained targets y_i for all the training data, and at the same time is as flat as possible. Taking the form

$$f(x) \quad = \quad < w, x > + b \tag{2.2}$$

where $< ., . >$ denotes the dot product in X. Flatness in this form means that one seeks a small w. One way to ensure this is to minimize the norm. So this problem can be written as a convex optimization problem. [55]

$$\begin{array}{ll} minimize & \quad \frac{1}{2} \parallel w \parallel^2 \\ subject\ to & \begin{cases} y_i - < w, x_i > - b \ \leq \varepsilon \\ < w, x_i > + b \ - \ y_i \leq \varepsilon \end{cases} \end{array} \tag{2.3}$$

A quadratic programming is usually used to solve the optimization problem [55].

2.6.2 Least Squares Regression

Least squares estimation is a traditional method to estimate the relation between two sets of variables. A basic assumption underlying the application of the least squares method is that the error terms in the regression model are independent [56]. Suppose observations are $\{X, y\}$. The model is

$$y = X\beta + \epsilon \tag{2.4}$$

where β is a vector of parameters given by

$$\beta = (X'X^{-1}X)^{-1}X'y \tag{2.5}$$

with X' as the transpose of X.

2.6.3 Gaussian Process Regression

Given observations $\mathcal{D} = (x_1, y_1), ..., (x_l, y_l) = (X, y)$. We will review the Bayesian analysis of the standard linear regression model with Gaussian noise [57]

$$\begin{aligned} f(X) &= X^T W, \\ y &= f(X) + \varepsilon \end{aligned} \tag{2.6}$$

, where X is the input vector, W is a vector of weights of the linear model, f is the function value and y is the observed target value. ε is a bias weight or offset. Assuming that the observed values y differ from the function values $f(X)$ by additive noise, and further assuming that this noise follows an independent, identically distributed Gaussian distribution with zero mean and variance σ_n^2, we get $\varepsilon \sim \mathcal{N}(0, \sigma_n^2)$.

In the Bayesian formalism, we have a zero mean Gaussian prior with covariance matrix Σ_p on the weights $W \sim \mathcal{N}(0, \Sigma_p)$. To make predictions for a test case x_*, the predictive distribution for $f_* \triangleq f(x_*)$ at x_* is given by averaging the output of all possible linear models with respect to the Gaussian posterior,

$$\begin{aligned} p(f_*|x_*, X, y) &= \int p(f_*|x_*, W)p(W|X, y)dW \\ &= \mathcal{N}(\tfrac{1}{\sigma_n^2} x_*^T A^{-1} X y, x_*^T A^{-1} x_*) \end{aligned} \tag{2.7}$$

,where $A = \frac{1}{\sigma_n^2} X X^T + \Sigma_p^{-1}$

2.7 Experiments

Now we perform experimental validations on a large database Morph-II as introduced in section 2.7.1, and then report the experimental results. Our experiments are operated automatically. The experiment results show that facial features which are relevant with BMI on small databases in psychology studies are still relevant with BMI on a large scale database. Next, we report the correlations between seven facial ratios and BMI.

The number of samples used in our experiments is much larger than those used in psychology studies [17, 18]. We have the ground truth of weight and height for each individual, thus the true BMI values are known for our evaluation. The BMI values are mainly in the range of 15 to 35, which contains all BMI categories: underweight, normal, overweight and obese.

2.7.1 Database

The MORPH-II database is used in our research for BMI study [58]. There are about 55,000 face images in MORPH-II. Because of the imbalance of the age, gender and ethnicity distribution on MORPH-II database, (e.g., 96% of the images are Black and White faces, but only 4% are Hispanic, Asian, Indian, and other), we only used the White and Black faces in our study. Also considering the gender and age group distributions, we selected part of face images, and divided into two sets: Set1 and Set2. The division into two sets is motivated by the age estimation study in [59]. The details about the distribution can be seen from Table 2.1. Since some BMI information is not available for some subjects, the distribution of Set1 and Set2 are not strictly equal.

Table 2.1: The selected data of Set1 and Set2. There are 4 gender and ethnicity groups in both Set1 and Set2: Black female, Black male, White female, and White male. They are also corresponding to six age groups: age10s (< 20), age20s (20-29), age30s (30-39), age40s (40-49), age50s (50-59) and age60s (≥ 60)..

	Set1	Set2
Black female	908	902
Black male	2642	2692
White female	1006	1022
White male	2717	2707
Age10s	992	993
Age20s	2687	2692
Age30s	1901	1901
Age40s	1243	1251
Age50s	413	455
Age60s	24	20

2.7.2 On Facial Points

Our goal is to perform BMI prediction from face images automatically. To achieve the goal, one major step is to detect facial points automatically for feature extraction. In our current approach, most facial points can be detected automatically by the ASM method, however, there are some points that cannot be obtained by the ASM. As stated in Section 2.4.1, some facial points are predicted or estimated by the other ASM detected points, rather than by manual labeling. It is not practical or convenient to ask the users to manually label some facial points, especially in test face photos. To have a brief evaluation of our estimation method for the facial points, especially the highest point of a face, N2, we randomly selected 100 images from our experimental database: MORPH-II, and manually labeled all the needed points to produce N2 and also the real highest point of each face. Then we found the average deviation of our estimate w.r.t. the manually labeled positions of the highest points is about 20.19 pixels. While the average face size of the 100 images is about 182×230 pixels, this deviation is about 8.8% of the face height. We consider this deviation is small for our automated feature extraction. Moreover, the BMI prediction errors are not large (see results in this section), which indirectly show the feasibility of estimating the point location for feature extraction. Next, we present the experimental results, including both correlation measures and the BMI prediction performance analysis.

2.7.3 Correlation Results

In the next step, we want to investigate the influences of age, gender, and ethnicity to the correlations between facial features and BMI, as operated in [17] and [18]. Because of the similar population distribution, also Pearson's correlation coefficient and the p-value of Set1 and Set2 in Table 2.2 are nearly the same. So we just need to conduct our experiments in Set1 or Set2. We choose Set1 for our experiments. The result is shown in Table 2.3 and Table 2.4. In these tables, $p - value = 0.0000$ indicates a very small p-value, which is less than 0.00001.

TABLE 2.2: Pearson's correlation coefficient r between facial features and the BMI on Set1 and Set2, and the corresponding p-values.

	Set1		Set2	
	r	p-value	r	p-value
CJWR	-0.20	0.0000	-0.20	0.0000
WHR	0.25	0.0000	0.28	0.0000
PAR	-0.07	0.0000	-0.04	0.0006
ES	-0.10	0.0000	-0.11	0.0000
LF/FH	-0.07	0.0000	-0.09	0.0000
FW/LFH	0.15	0.0000	0.19	0.0000
MEH	0.04	0.0002	0.05	0.0001

In the following discussion, we will consider correlations with $p - value \leq 0.05$ as the significant correlation. And we also consider that there is no correlation exist if Pearson's r is less than or equal to 0.01. Smaller CJWR and larger WHR indicates a wider and

TABLE 2.3: Pearson's r between facial features and the BMI in age groups (in Set1).

	age10s		age20s		age30s		age40s		age50s		age60s	
	r	p-value	r	p-value	r	p-value	r	p-value	r	p-value	r	p-value
CJWR	-0.16	0.0000	-0.21	0.0000	-0.21	0.0000	-0.21	0.0000	-0.11	0.0213	-0.22	0.3046
WHR	0.27	0.0000	0.24	0.0000	0.27	0.0000	0.26	0.0000	0.23	0.0000	0.16	0.4622
PAR	-0.10	0.0013	-0.10	0.0000	-0.08	0.0002	-0.07	0.0163	0.07	0.1631	-0.09	0.6870
ES	-0.09	0.0050	-0.07	0.0001	-0.09	0.0000	-0.10	0.0003	-0.08	0.1249	0.06	0.9412
LF/FH	-0.05	0.0956	-0.10	0.0000	-0.08	0.0007	-0.10	0.0005	-0.04	0.4727	-0.20	0.3456
FW/LFH	0.19	0.0000	0.15	0.0000	0.16	0.0000	0.11	0.0000	0.24	0.0000	0.06	0.7845
MEH	0.05	0.0981	0.09	0.0000	0.04	0.1012	0.07	0.0139	-0.10	0.0598	0.15	0.4853

TABLE 2.4: Pearson's correlation coefficient r between facial features and the BMI in gender-ethnicity groups (in Set1).

	White Female		White Male		Black Female		Black Male	
	r	p-value	r	p-value	r	p-value	r	p-value
CJWR	-0.25	0.0000	-0.19	0.0000	-0.26	0.0000	-0.14	0.0000
WHR	0.21	0.0000	0.25	0.0000	0.30	0.0000	0.20	0.0000
PAR	-0.19	0.0000	-0.07	0.0006	-0.22	0.0000	-0.06	0.0016
ES	0.01	0.7500	-0.10	0.0000	0.04	0.1900	-0.05	0.0117
LF/FH	-0.03	0.4200	-0.02	0.2000	-0.06	0.0600	-0.03	0.1400
FW/LFH	0.13	0.0001	0.22	0.0000	0.14	0.0000	0.15	0.0000
MEH	-0.00	1.0000	-0.03	0.1200	0.07	0.0400	-0.01	0.7600

squarer face[18]. Smaller PAR indicates more round low face which also means a wider and squarer face. Smaller ES indicates smaller eyes. Smaller LF/FH indicates smaller rate between lower face to whole face length. Larger FW/LFH also indicates a wider face. Since usually, the height of people's face is larger than the width, we consider a wider face as a squarer face.

Table 2.3 shows the correlations between facial features and BMI in six age groups. We can see the tendency between facial features and BMI in different groups. In age10s group, a person with a squarer face (CJWR, WHR, PAR, and FW/LFH) and smaller eyes (ES) has a larger BMI. In age20s group, a person with a squarer face (CJWR, WHR, PAR, and FW/LFH), smaller eyes (ES), smaller rate between lower face to whole face length (LF/FH), larger eyebrow height (MEH) has larger BMI. In age30s group, a person with a squarer face (CJWR, WHR, PAR, and FW/LFH), smaller rate between low face to whole face length (LF/FH), and smaller eyes (ES) have larger BMI. In age40s group, a person with a squarer face (CJWR, WHR, PAR, and FW/LFH), smaller rate between lower face to whole face length (LF/FH), smaller eyes (ES), and larger eyebrow height (MEH) has bigger BMI. In age50s group, a person with a squarer face (CJWR, WHR, and FW/LFH) has larger BMI. There is no feature indicating significant correlations with BMI in age60s group. This may be caused by the insufficient number of samples used in this age group. So a hypothesis testing is necessary to claim an correlation.

Table 2.4 shows correlations between facial features and BMI in four gender-ethnicity groups. In white female group, a person with a squarer face (CJWR, WHR, PAR, and FW/LFH) has larger BMI. In white male group, a person with a squarer face (CJWR, WHR, PAR, and FW/LFH), smaller eye size (ES) have larger BMI. In black female group, a person with a squarer face (CJWR, WHR, PAR, and FW/LFH), larger eyebrow height (MEH) has larger BMI. In black male group, a person with a squarer face (CJWR, WHR, PAR, and FW/LFH), smaller eyes (ES) has larger BMI.

Table 2.5: Correlations between BMI and facial features.

	Coetzz et al. [17]		Pham et al. [18]		Our Work	
	Male	Female	Male	Female	Male	Female
	(n = 187)	(n = 194)	(n = 230)	(n = 229)	(n = 1973)	(n = 714)
PAR	-0.22	-0.12#	-0.30	-0.23	-0.16	-0.21
WHR	0.17	0.36	0.30	0.28	0.29	0.27
CJWR	-0.20	-0.29	-0.40	-0.29	-0.24	-0.26
ES			0.29	0.26	-0.07#	-0.05#
LF/FH			0.12#	0.10#	-0.02#	-0.02#
FW/LF			0.12#	0.05#	0.20	0.15
MEH			0.05#	0.16	-0.04#	0.01#

The correlation between our work and psychology study results are shown in Table 2.5. Our results have similar correlations with those in [18] and [17], except ES. Maybe that's because in [18], participants are Korean. We also notice that in table 2.2, all the correlations are significant in a whole. So we can use all these facial ratios together to predict BMI using regression models regardless of the age, gender, and ethnicity.

2.7.4 BMI Prediction Results

Our experiments are operated automatically on the Morph-II database as introduced in Section 2.7.1. We use Set1 and Set2 as training and test dataset or vice versa. The performance of BMI estimate is measured by the mean absolute error (MAE). The MAE is defined as the average of the absolute errors between the estimated BMIs and the ground truth BMIs. MAE $= \sum_{k=1}^{N} |\hat{b}_k - b_k|$, where b_k is the ground truth BMI for the test image k, $\hat{b}_k$ is the estimated BMI for image k, and N is the total number of test images [60].

Figure 2.6 shows the MAEs in the whole Set1 or Set2. The left shows the MAEs when training in Set2 while testing in Set1. The right shows the MAEs when training in Set1 while testing in Set2. From this figure, we can observe that overall the SVR method performs better than both the GP and LSE methods.

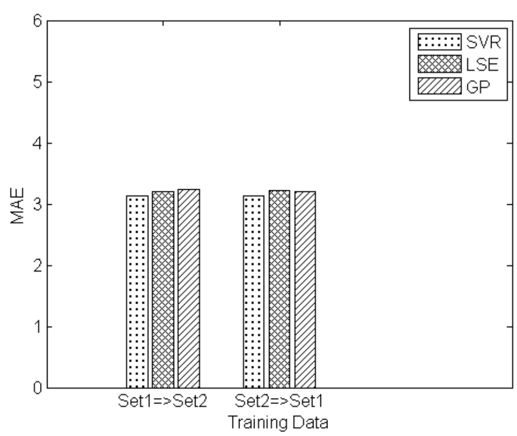


FIGURE 2.6: Visualization of the MAEs of SVR, GP and LSE in total average. Left: Set1 $\Rightarrow$ Set2; Right: Set2 $\Rightarrow$ Set1.

2.7.4.1 MAEs in Different Age Groups

Table 2.6 shows the MAEs of BMI prediction using the regression methods, SVR (with the RBF kernel), GP, and LSE. In Table 2.6, the MAEs are shown in different age groups. The MAEs of BMI is among a range of 3.00 ± 1.00. This indicates that the error of the

TABLE 2.6: The BMI prediction errors measured by the mean absolute error (MAE) on a large database. The results are presented based on age groups and in total average. "Set1⇒ Set2" means the data in Set1 is used for training, while Set2 is used for testing. Three different regression methods, SVR, GP, and LSE, are used to learn the BMI prediction function.

	Set1 ⇒ Set2			Set2 ⇒ Set1		
	SVR	LSE	GP	SVR	LSE	GP
Age10s	2.79	3.01	3.02	2.64	2.90	2.89
Age20s	3.03	3.11	3.15	3.08	3.19	3.18
Age30s	3.47	3.44	3.47	3.42	3.42	3.41
Age40s	3.13	3.17	3.21	3.20	3.21	3.18
Age50s	3.24	3.41	3.26	3.21	3.22	3.19
Age60s	2.52	2.23	2.23	2.66	2.75	2.59
Avg.	3.14	3.21	3.23	3.14	3.22	3.20

BMI prediction is relatively small, comparing to the wide range of BMI from 15 to 35. As shown in Table 2.6, the MAEs are slightly different between different groups. For example, the age10s group has a smaller MAE than other age groups.

TABLE 2.7: This table shows the number of face images in four BMI categories. The normal BMI category contains the most number of faces, which is about 55% of the whole set. Overweight category is the second largest, which contains about 30% of faces

	Set1	Set2
underweight	308	303
normal	4007	4042
overweight	2096	2093
obese	838	862

2.7.4.2 MAEs in Different BMI Categories

Figure 2.7 shows the MAEs of SVR, GP and LSE in different BMI categories: underweight, normal, overweight, and obese. We can see that all the three methods perform well in the normal and overweight BMI categories. In the normal BMI categories, the MAE of SVR is only 1.91. In Table 2.7, the normal BMI category contains about 55%

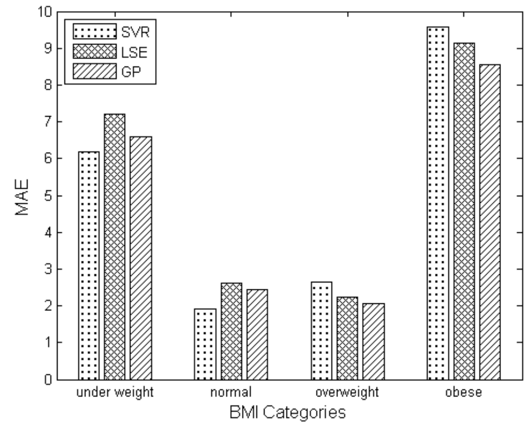


FIGURE 2.7: Visualization of the MAEs of SVR, GP, and LSE in different BMI categories. The SVR has the lowest errors for underweight and normal, while the GP has the smallest errors for overweight and obese categories.

of the faces. The overweight category contains about 30%. These two categories possess about 85% of the whole set. As shown in Figure 2.7, in the normal BMI category, the MAE of SVR is the smallest. In the overweight category, the MAE of GP is the smallest. As a result, we can get a higher accuracy in predicting the BMI according to facial features in these two categories.

The lower accuracies in another two categories, i.e., underweight and obese, are probably because of the smaller number of samples for both training and testing, as shown in Table 2.7. The result suggests that a large number of training examples are needed in order to have a better BMI prediction performance.

Even though the SVR, GP and LSE have similar MAE results, the SVR is better for the underweight and normal BMI categories, while the GP performs better for overweight and obese BMI categories. Overall the SVR has the smallest errors as shown in Figure 2.6.

2.7.4.3 MAEs in Different Gender and Ethnicity Groups

Finally, we study the BMI prediction performance when it is executed in separate gender and ethnicity groups. In this study, the training and test faces come from the same gender and ethnicity groups. The same group in Set1 is used for training, and the data in Set2 is used for testing, and vice versa. Since the SVR method performs the best overall, we use the SVR for this experiment. The results are shown in Table 2.8.

Table 2.8: The BMI prediction errors (measured by the MAE) based on separated training in each of the four groups, using the SVR method.

	Set1 ⇒ Set2	Set2 ⇒ Set1
Black female	4.18	3.91
Black male	3.07	3.16
White female	3.44	3.22
White male	2.69	2.78
Average	3.13	3.12

Table 2.9: The BMI prediction errors (measured by the MAE) based on the SVR method using mixed training (without gender and ethnicity separation). The results are presented for the four groups separately.

	Set1 ⇒ Set2	Set2 ⇒ Set1
Black female	4.29	3.94
Black male	3.05	3.13
White female	3.64	3.44
White male	2.65	2.76
Average	3.14	3.14

For comparison, we also show the BMI prediction results using mixed training data, i.e., without separation of the four groups in training, in Table 2.9. From Tables 2.8 and 2.9, we can observe that the MAE differences are small between the separated training and mixed training, although the MAE can be reduced in some cases, when the separated training is executed. So there is no need to perform a separate training for BMI prediction based on our current experiments. Note that in order to do the BMI computation in separated groups, the gender and ethnicity information has to be recognized first, which is beyond the scope of our topic.

From Tables 2.8 and 2.9, we can also observe that the BMI prediction errors are different for different groups of people. For example, the MAEs for the male are small, while the errors for the female are large. This result suggests that our computational BMI prediction is more accurate for males than for females.

2.7.5 Other Features

In our study, the facial features that we used are motivated by the observations in psychology and human perception studies [16–18]. One may argue that probably some other features could be used too. One possible choice is the ASM shape feature [61] , which is popular for facial shape analysis and face recognition. To get an understanding on this, we use the ASM feature for facial feature characterization for BMI prediction, using the SVR for regression.

Experimentally, we found that the MAE of the ASM shape feature is 3.34, based on the average of Set1 to Set2, and also Set2 to Set1. This MAE is slightly higher than the 3.14 for the seven features that we used and motivated by the human perception studies [16–18]. This comparison shows that the features that we used are capable of characterizing the facial property better than the ASM features for BMI prediction. Our interpretation is that the ASM shape parameters characterize the generic shape features and pertain to the individual's shape, and thus may not deliver the "salient features" pertaining to the BMI prediction.

Based on our exploration, it may inspire some future investigations to develop more discriminative features that can characterize and predict the BMI with a better performance.

2.8 Summary

We develop an automatic and computational system for validation of correlations between facial features and the BMI. It is motivated by the recent studies in psychology and human perception. The validation of the psychology studies is operated on a large database with more than 14,500 face images in a statistically meaningful manner. This conclusion provides the basis to develop a computational BMI prediction system based on a statistical learning formulation.

To decode BMI information from facial images, an automated and computational system is developed for body mass index prediction from face images. Seven facial ratio and geometry features are extracted automatically from the face image. Three classic regression methods, i.e. support vector regression, least squares regression, and Gaussian process regression are explored. Among them, support vector regression gives the best performance on aggregate. Our developed system is probably the first computational approach to predict BMI from face images automatically. Our work shows that a computational approach can be developed for BMI prediction using machine vision and statistical learning techniques.

Chapter 3

Face Recognition with BMI Changes

3.1 Introduction

BMI is the measure of body fatness. The BMI calculation is introduced in Section 1.1.2. It is calculated using body weight and height. In this thesis, we only consider adult BMI. For adults, the body height is relatively steady, so we use the concept of body weight changes in this Chapter to describe the adult BMI changes.

It is a common phenomenon that facial appearance changes along with body weight gain or loss. In this Chapter, we want to study the influence of body weight changes on face recognition. Further, we explore the feasibility to develop a face recognition system robust to body weight changes. PLS method is utilized to develop face recognition system robust to body weight changes [62].

Body weight changes bring the increase and decrease of facial adiposity. As shown in Figure 3.1, the large body weight changes (bottom row) can cause facial shape variation more significantly than small body weight changes. However, no previous work has studied the influence of body weight changes on face recognition computationally. It is for the first time to study this problem, to the best of our knowledge.

Figure 3.1: Real face images (of the same subject) with facial shape variations caused by small (top) or large (bottom) BMI changes.

The most related work is the research about relation between face and body weight or body mass index (BMI). Previous studies in psychology and human perception [16–18] have found that facial geometry and ratios have relations to human weight or BMI. The correlations have been validated on a large database computationally in Chapter 2. These previous works indicate some trends for BMI changes, e.g., facial squareness and facial ratios/geometries, etc. These conclusions help us to synthesize faces with different BMIs.

To facilitate the study of the influence of body weight changes on face recognition, we assembled two face databases. First, synthesized face images are assembled according to trends as we validated in Chapter 2 based on both human perception and computation [16–18]. Synthesized faces can simulate faces with small or large body weight changes, since it is hard to collect real faces with both small and large body weight changes. Further, synthesized faces can be used to focus on weight changes while exclude other factors, e.g., age, pose, illumination and expression variations. Second, a real face image database is collected as well. Each subject has two images, one is overweight or obesity (e.g., before weight loss), while the other is close to normal (e.g., after weight loss).

And also, we develop an approach to reduce the influence of body weight changes on face recognition. The partial least squares (PLS) method [63] is used to learn projection matrices for both overweight/obese and normal faces, so that the features can be transformed into a common subspace. Both synthetic and real face databases are employed to evaluate the proposed approach.

The rest of the Chapter is organized as follows. Section 3.2 introduces two databases that we assembled. In Section 3.3, feature descriptors and the PLS method are introduced. Section 3.4 shows experimental results. Finally, Section 3.5 gives the the summary of this Chapter.

3.2 New Databases

Many face databases have been assembled to study the effect of different variations on face recognition, such as age, pose, illumination, expression and even makeup variations. However, few face databases have been collected with body weight changes. To facilitate the study, we assembled two face databases with body weight changes. The first database consists of synthetic faces with both small and large body weight changes to facilitate comparative studies. The second is a real face database, collected from the Internet with large body weight changes typically.

3.2.1 Synthetic Face Database

As we introduced in Chapter 2 that an adult with a squarer face, smaller eyes, smaller ratio between the lower face to whole face length, and larger eyebrow height tends to have a larger BMI [16–18]. According to this conclusion, synthesis operations are operated to simulate different BMI effects using PhotoShop[1]. First of all, the shape of the face edge is

[1]We would like to thank Prof. Ayres's group at Yale for the help of generating synthetic data for this study.

changed to get desired facial squareness and facial ratios/geometries. Then texture effects are adjusted mainly around eyes, checks, chin and also the neck to get vivid face effects corresponding to different BMI effects. The original face images of these subjects (before weight changes) are selected from some existing face databases captured in a controlled environment. In total, there are 120 subjects in the synthetic face database, including 29 black males, 29 white males, 30 black females and 32 white females.

Synthesis operations are operated mainly around eyes, checks, chin and also the neck to simulate different body weight effects. To compare and measure the influences of different body weight changes on face recognition, faces with both small and large body weight changes are rendered. Some faces with small body weight changes are shown in Figure 3.2(a). While Figure 3.2(b) shows some examples with large body weight changes. In Figure 3.2, two faces of the same subject are put in the same column, from which only body weight variation can be observed (i.e., controlled scenario without age, pose, illumination, or expression changes).

(a) Faces with Small Weight Changes

(b) Faces with Large Weight Changes

FIGURE 3.2: Some examples of synthetic face images with body weight changes. (a). small facial adiposity changes, and (b) large facial adiposity changes. Two faces of the same subject are in each column.

FIGURE 3.3: Some examples of real face images with large body weight changes, collected from the Internet. Two faces of the same subject are next to each other in the same row.

3.2.2 Real Face Database

Even though the synthetic face images can simulate different body weight effects well, we still need to validate the results obtained from the synthetic database with real faces. We collected face images with large body weight changes from the Internet. It is difficult to collect a large database with both small and large body weight changes for each subject. Since weight loss is a relatively new fashion, images with large body weight changes can be easily found from the Internet. Some examples from the real face database are shown in Figure 3.3.

There are totally 242 subjects on the real face database, including 6 black males, 28 black females, 23 white males, 90 white females, 13 Asian males and 82 Asian females. There are two photos for each subject: one is overweight or obese and the other is relatively normal. Typically, the face image pairs are before and after weight loss of the same subjects. Besides BMI changes, other face variations, such as illumination, pose, expression, age and so on can also be observed, but the BMI changes are the main variation in this database.

3.3 Method

To study the influence of body weight changes on face recognition, we chose to use the LBP and SIFT as the feature descriptors. A correlation mapping method - Partial Least Squares is used to eliminate the influence of BMI changes on face recognition.

3.3.1 Local Binary Pattern (LBP)

LBP feature descriptor binarizes a neighbourhood using the center pixel as the threshold. The binary values of the neighbors are encoded into a number as the LBP code of the center pixel [64],

$$f_{LBP}(x) = \sum_{i=0}^{P-1} b_i 2^i,$$

where x is the center pixel, P is the number of neighbors, and b_i is the binary value of the i-th neighbor pixel.

By obtaining LBP codes for each pixel as the center, histogram features can be computed from a region. For face recognition, we usually divide the whole face into patches with overlap and concatenate the LBP histograms together as the face feature descriptor.

3.3.2 Scale Invariant Feature Transform (SIFT)

SIFT is a descriptor proposed by Lowe [65]. With the stable location, scale, and orientation information for each keypoint, the SIFT descriptor characterizes the appearance of the keypoint. A set of orientation histograms are computed with samples in a 16×16

neighborhoods around the keypoint. To get the SIFT descriptor of a region, samples in one of the four subregions are accumulated together with weights according to the gradient norm. Also a Gaussian weight is employed according to the distance to the center. Finally orientations are quantized into 8 bins, and the total size of the SIFT descriptor is 128 histogram bins.

Similar to the LBP feature, SIFT descriptors are also calculated from facial patches and then concatenated together as the feature descriptor of the face image.

The LBP-based face recognition is employed as the baseline in our work. SIFT has been shown relatively robust to expression, accessory and pose variations for face recognition [66]. Here we want to investigate if the SIFT is robust to facial adiposity changes caused by body weight changes.

3.3.3 Partial Least Squares

Suppose we have two data sets $\mathbf{X}$ and $\mathbf{Y}$, containing n data samples each, where $\mathbf{X} \subset \mathbf{R}^N$ and $\mathbf{Y} \subset \mathbf{R}^M$, and $\mathbf{R}^N$ and $\mathbf{R}^M$ stand for N-dimensional and M-dimensional spaces, respectively. PLS models the relations between $\mathbf{X}$ and $\mathbf{Y}$ by means of score vectors such that

$$\begin{aligned} \mathbf{X} &= \mathbf{T}\mathbf{P}^T + \mathbf{E} \\ \mathbf{Y} &= \mathbf{U}\mathbf{Q}^T + \mathbf{F}, \end{aligned} \tag{3.1}$$

where $\mathbf{T}$ and $\mathbf{U}$ are extracted score vectors (latent vectors), the $\mathbf{P}$ and $\mathbf{Q}$ represent matrices of loadings, and the $\mathbf{E}$ and $\mathbf{F}$ are the matrices of residuals [63].

The PLS usually uses a greedy strategy to find multiple basis vectors that project $\mathbf{X}$ and $\mathbf{Y}$ to a latent space. Its classical form is based on the nonlinear iterative partial least squares (NIPALS) algorithm [68] to find weight vectors $\mathbf{w}, \mathbf{c}$ to satisfy

$$\begin{aligned} [cov(\mathbf{t}, \mathbf{u})]^2 &= [cov(\mathbf{Xw}, \mathbf{Yc})]^2 \\ &= max_{|r|=|s|=1} [cov(\mathbf{Xr}, \mathbf{Ys})]^2 , \end{aligned} \tag{3.2}$$

where $cov(\mathbf{t}, \mathbf{u}) = \frac{\mathbf{t}^T\mathbf{u}}{n}$ denotes the sample covariance between the score vectors $\mathbf{t}$ and $\mathbf{u}$ [63]. The linear PLS models can have variants based on the deflation difference [69].

The PLS models were originally derived for regression or classification problems [68–70]. The PLS has been successfully applied to face recognition with variations of pose [5]. Here, the PLS method is investigated to deal with a new variation - facial shape variations caused by body weight changes. The framework for face recognition robust to BMI changes based on PLS is shown in Figure 3.4.

$\mathbf{X}$ and $\mathbf{Y}$ are face images but with BMI changes. Elements in corresponding rows of $\mathbf{X}$ and $\mathbf{Y}$ are from the same subject. The PLS can project features from both $\mathbf{X}$ and $\mathbf{Y}$ into a common subspace $\mathbf{N}$. After feature projections, new features in the sub-space $\mathbf{N}$ can be used for face recognition. We hope the use of correlation mapping bases learned from the PLS can reduce the influence of weight changes on face verification.

3.4 Experiments

One of the main goals in this study is to explore the influence of BMI changes on face recognition. To achieve the goal, faces with both small and large weight changes on the synthetic database are employed. Match scores are compared to observe the effects of different BMI changes on face recognition. Then, the PLS based face recognition is explored to see if the influence of large BMI changes can be reduced. Both synthetic faces with large weight changes and real face databases are used for the PLS based face recognition.

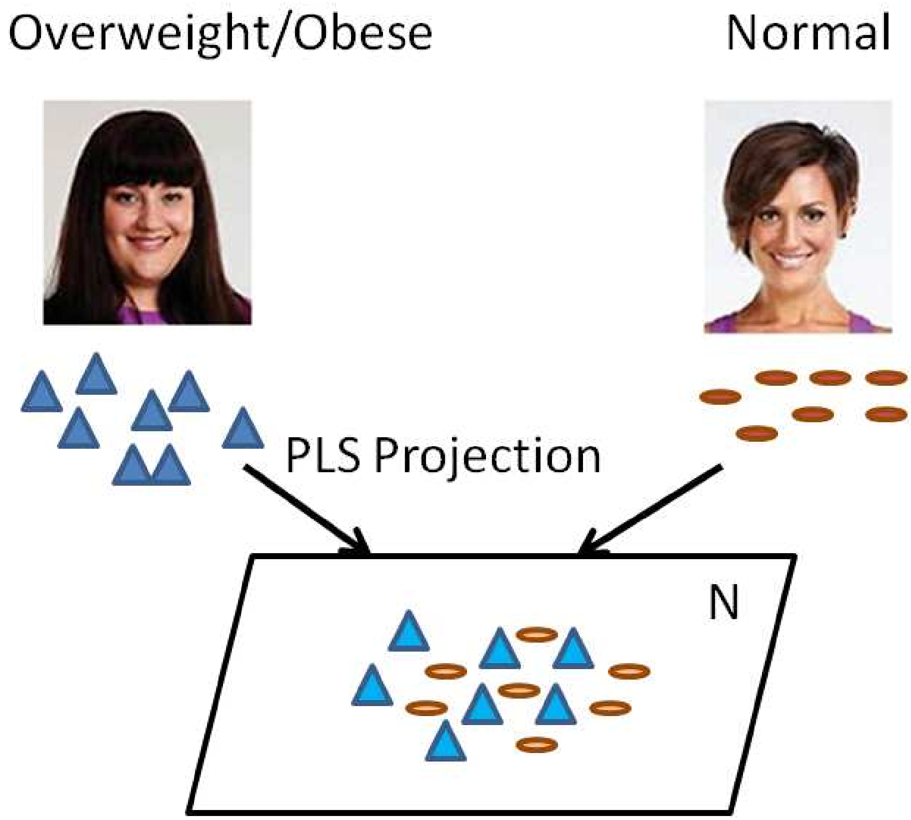


FIGURE 3.4: The basic framework based on the PLS for face recognition with large BMI changes.

3.4.1 Small vs. Large Body Weight Changes

To carefully examine the influence of BMI changes on face recognition, faces with both small and large BMI changes are investigated. Firstly, a direct face matching is executed. Secondly, face verification performance are compared to investigate the influence of BMI changes on face recognition. After face detection and alignment, gray scale faces are used for face recognition.

3.4.1.1 Match Score Differences

In this experiment, we compare match scores of genuine synthetic face pairs of the 120 subjects in cases of small and large BMI changes. Direct face matching is performed.

So all the samples can be employed for face matching. Histogram intersection similarity measure [67] can be used to measure the match score between two histogram features as shown below,

$$P(H_1, H_2) = \frac{\sum_{k=1}^{n} min(H_1(k), H_2(k))}{min(\sum_{k=1}^{n} H_1(k), \sum_{k=1}^{n} H_2(k))},$$

where H_1 and H_2 are two histograms and n is the dimension of the feature descriptor.

If there is no influence of BMI variation on face recognition, the match score between two genuine faces on synthetic faces should be close to 1, while the reduced match scores can reveal the differences brought by BMI variations. Figure 3.5 shows comparisons of match scores between small and large BMI changes. Figure 3.5(a) shows match scores using the LBP feature, and Figure 3.5(b) is based on the SIFT feature. From Figure 3.5, we can observe that large BMI changes cause significant decrease of match scores than small BMI changes, by 9.1% using the LBP feature, and 9.8% using the SIFT feature. Besides, SIFT feature gives better match scores than LBP feature for both small and large weight changes.

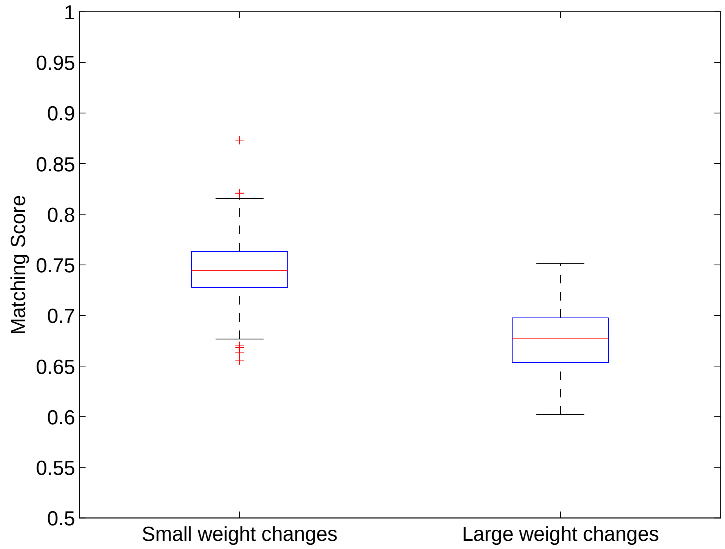
3.4.1.2 Face Verification

For face verification experiment, we used 120 genuine face pairs from the synthetic face database and 120 impostor face pairs selected from faces with the highest match scores of different subjects. The cosine distance between the features is computed for similarity measure. The Receiver Operating Characteristic (ROC) curves are drawn in Figure 3.6, where SWC stands for “small weight changes“ and LWC stands for “large weight changes“.

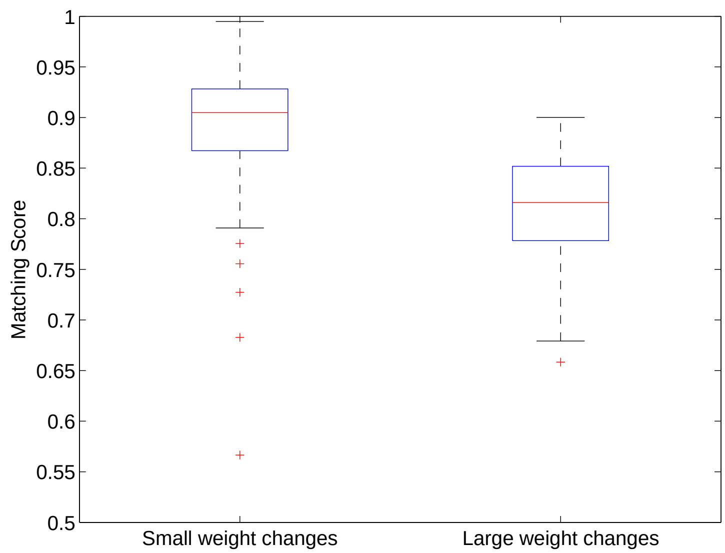
Table 3.1: EER(%) of the face verification results on the synthetic face database.

	Small Weight Changes	Large Weight Changes
LBP	3.9%	26.7 %
SIFT	4.2 %	17.2 %

As shown in Figure 3.6, face pairs with small BMI changes have very high face verification accuracies for both LBP and SIFT features. On the contrary, face pairs with large



(a) LBP



(b) SIFT

FIGURE 3.5: Boxplot of match scores of genuine face pairs on synthetic database: Comparisons between small and large BMI changes.

BMI changes have lower accuracies significantly. Table 3.1 shows the EER results of face verification on the synthetic face database. The EERs of Large weight changes are substantially higher than small weight changes. These comparisons show that large BMI changes can reduce the recognition accuracy significantly.

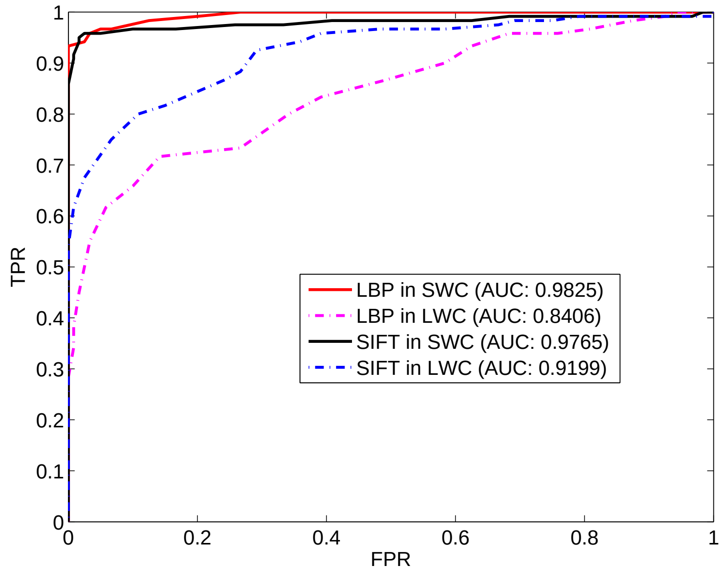


FIGURE 3.6: ROC Curves of face verification results on the synthetic face database.

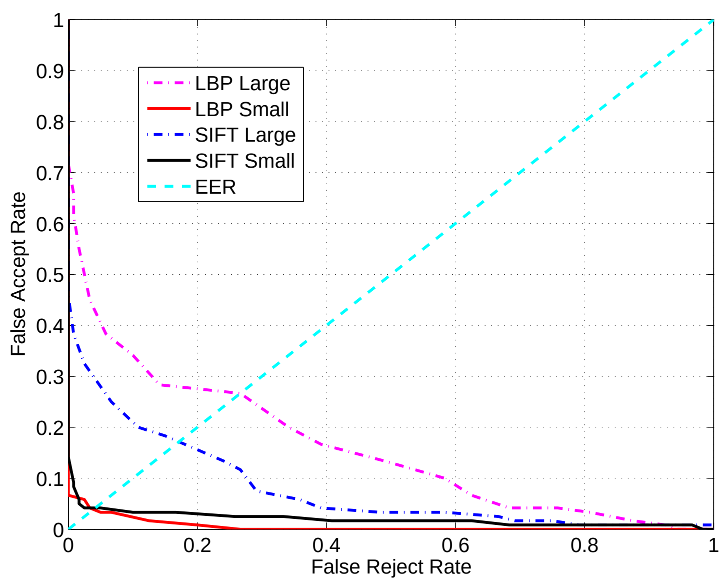


FIGURE 3.7: FAR and FRR curves with EER.

We can also observe that the LBP performs sightly better than the SIFT in small BMI changes, while the SIFT performs much better than the LBP in large BMI changes.

In the following, real faces will be explored to study how to reduce the influence of weight changes on face recognition.

Besides accuracy, the time cost is also under consideration. We use a desktop computer manufactured by Dell Inc. The computer is with windows 7 professional 32-bit operating system. The processor of this computer is Intel(R) Core(TM)2 Duo CPU E7500 @ 2.93 GHZ and memory is 8192MB RAM.

3.4.2 PLS based Face Recognition

In this experiment, we study if the PLS method can be used to reduce the influence of BMI changes on face recognition. Both synthetic and real face databases are employed.

3.4.2.1 PLS on the Synthetic Face Database

As shown in Table 3.1, as well as Figure 3.6, the face verification accuracy for faces with small weight changes on the synthetic database can be very high, indicating that there is no need to address the influence of small weight changes on face recognition. In this experiment, we only use faces with large BMI changes to validate the performance of the PLS. A 5-fold cross validate scheme is used for the face verification experiment. In each trial, 4/5 genuine face pairs are used for training the PLS, while the remaining genuine face pairs and the impostor face pairs are used for testing. The comparisons between Direct matching (DM) and PLS are shown in Table 3.2. The PLS can improve the verification accuracy by 21.1% for the LBP feature, and 12.6% for the SIFT feature.

Table 3.2: Comparison of face verification accuracy between direct matching (DM) and PLS based face recognition using faces with large BMI changes on the synthetic face database.

	LBP	SIFT
DM	65.4 % ± 7.3 %	78.8% ± 5.0 %
PLS	79.2 % ± 3.9 %	88.7% ± 1.9 %

3.4.2.2 PLS on the Real Face Database

The synthetic faces can exclude other variations and focus on the weight changes. The real face database that we assembled contains some other variations, in addition to the major change of body weight. A 5-fold cross validation scheme is used on the real face database. In each trial, 192 genuine face pairs are used for PLS training, while 48 genuine pairs and 48 impostor pairs are used for testing. Table 3.3 shows the comparison of face verification accuracies between the direct matching and the PLS based face recognition. The results show improvement by 16.7% for LBP, and 13.2% for SIFT feature, respectively.

Besides, we show some examples, which are falsely verified by direct matching, but are corrected by the PLS method. Figure 3.8 shows correctly verified face pairs that were false negative results by the direct matching. Figure 3.9 shows correctly verified face pairs that were false positive results by the direct matching. The test face pairs are listed in the same column in Figures 3.8 and 3.9. The difference is that the face pairs are genuine in Figure 3.8, while the face pairs are impostor in Figure 3.9. From the results, we can see that the PLS based method is useful to address body weight changes in face recognition.

From Table 3.3, we can also observe that the SIFT is a little bit more robust than the LBP for BMI changes. Further more, significant improvements are obtained on the real face database when using the PLS method, compared to the direct matching. The face verification accuracy on the real database (shown in Table 3.3) is lower than accuracy on the synthetic database (shown in Table 3.2), which is probably caused by other variations except BMI changes, e.g., age, pose, illumination, or expression changes, etc. Besides, we record the running time of PLS based face verification using LBP and SIFT separately. With LBP feature, the running time is 3.32 seconds, while the time cost is 5.95 seconds using SIFT feature. Even though SIFT feature spends more time, considering the robustness of SIFT feature, SIFT feature is still a good choice for face recognition with BMI changes.

Table 3.3: Comparison of face verification accuracy between direct matching (DM) and PLS on the real face database.

	LBP	SIFT
DM	58.8 % ± 3.3 %	61.5% ± 2.1 %
PLS	68.6 % ± 6.0 %	69.6% ± 0.9 %

Figure 3.8: Some false negative examples from real database by the direct matching, but corrected by the PLS method. Two face images in each row next to each other (left two columns and right two columns, separately) are the genuine face pairs.

3.5 Summary

We study the influence of BMI changes on face recognition. To the best of our knowledge, this is for the first time to explore the BMI variation issue in face recognition. To facilitate the study, two face databases, named synthetic and real face databases are employed. Experimental results have shown that large BMI changes can reduce the accuracies of face recognition significantly. Then we explored the PLS method to reduce the influence of BMI changes on face recognition. Preliminary results have shown significant improvements using the PLS method on both the synthetic and real face databases.

Figure 3.9: Some false positive examples from real database by the direct matching, but corrected by the PLS method. Two faces in each row next to each other (left two columns and right two columns, separately) are the impostor face pairs.

Chapter 4

Facial Makeup Detection

4.1 Introduction

Makeup detection problem is to classify a face image into the makeup or non-makeup class. There are some recent works related to this problem. The appearance of cosmetic foundation on face images was studied using a multi-band camera system and the "oily-shine" regions in makeup face images were detected using a clustering method [71]. There is no result on classification of makeup and non-makeup faces in [71]. A system was developed in [72] to detect eye-shadow, lipstick, and liquid foundation, separately, in makeup faces of only 21 subjects. Both the HSV color space and texture features were used. However, there is no classification between makeup and non-makeup faces in [72]. Chen et al. [73] explored the binary classification of makeup and non-makeup faces, using the shape, texture and color features.

In our study, we consider more makeup cues, and use 12 facial patches (only 3 patches in [73]) for facial makeup detection. Meanwhile, patch selection and discriminative mapping are applied in our makeup detection system. Those steps are essential to obtain a high accuracy for makeup detection, as shown in our experiments. Finally, our makeup detection experiments are conducted on a much larger database of 500 subjects [74].

Based on how humans apply facial makeup [75], we present four categories of features to characterize the facial makeup in Section 4.2. Different schemes, e.g., patch selection as illustrated in Section 4.3 and discriminative mapping in Section 4.4, are applied to improve the makeup detection performance. In Section 4.5, we show the experimental results. In the end, we summarize this study in Section 4.6.

4.2 Characterization of Facial Makeup

We propose four different facial cues to characterize facial makeup, based on the knowledge about makeup [75], i.e. skin color tone, skin smoothness, skin texture, and highlight effect.

4.2.1 Local Patches

Different facial regions could be applied with different cosmetic products and with different amount or degrees. For example, the eye makeup may use various colors (e.g., gold soft smoky, bronze, etc.) to make the eye more youthful and the lashes extra-long, while red or other coloration may be applied to the mouth to reshape (e.g., reduce, increase or bring balance to the lips) to have a beautiful look. We use local patches on face images to extract different makeup cues separately. The patches we use in each face image are illustrated in Figure 4.1.

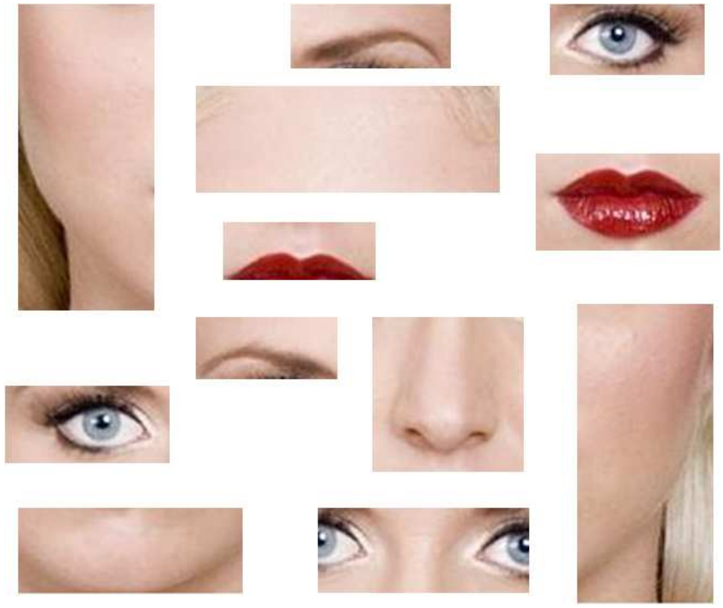


FIGURE 4.1: Facial patches used in our experiments. There are 12 local patches selected for each face image to perform local patch-based facial analysis.

4.2.2 Skin Color Tone

The facial skin color tone may be changed after applying cosmetic products on a face. The changed skin color tone may make a person (especially a women) look beautiful and younger [75]. Thus it is intuitive that the facial skin color tone can be used to characterize facial makeup. In other words, faces with and without makeup can be separated based on measuring the skin color difference.

We compute the mean, standard deviation, and entropy, denoted by μ, σ, and E, respectively, to characterize the skin color tone in each of the three color channels, R, G, and B, for each facial patch. As a result, nine features are extracted for each image patch in a face image.

Suppose the normalized histogram of the pixel values in each patch is denoted by $H(i)$, where i is the index of the pixel values. The number of pixels within the patch is N. Then, the mean, standard deviation, and entropy are computed for each image patch by

Equations 4.1, 4.2, 4.3 in each color channel $c \in \{r, g, b\}$, separately.

$$\mu_c = \sum_i H(i) * i, \tag{4.1}$$

$$\sigma_c = \sqrt{\sum_i (i - \mu)^2 * H(i)}, \tag{4.2}$$

$$E_c = -\sum_i H(i) \log_2 H(i), \tag{4.3}$$

4.2.3 Skin Smoothness

When the faces have makeup, the facial smoothness may be changed. Through applying foundation cream or other cosmetic products to faces, the facial appearance may look smoother. Thus the measure of facial smoothness may be used to determine whether a face has makeup or not.

In characterizing the facial smoothness, only the image intensity values are used within each patch in a face image. We compute the mean, standard deviation, and entropy, respectively, using Equations (4.1), (4.2), (4.3), but in image patch intensities only.

4.2.4 Skin Texture

With the use of cosmetic products on faces, the facial skin texture might look different. The texture change may be coupled with smoothness changes, but we want to characterize the skin texture specifically. To characterize the skin texture patterns, we use the local binary pattern (LBP) [76], which is a simple and popular method to extract texture features.

LBP feature is computed by comparing the intensity value of a center pixel with its surrounding neighbors. The details of LBP are illustrated in Section 3.3.1.

4.2.5 Highlight

The facial highlight may be perceived differently between faces with makeup and without. Characterizing the facial highlight might be useful to discriminate makeup faces from non-makeup faces. The makeup foundation may change the highlight on face images [71]. To explore this, we compute facial highlight in face images, and then extract related features to characterize the highlight component for facial makeup detection.

To compute the facial highlight, we adopt the dichromatic reflection model [77] to characterize the facial reflection, which is a simple linear combination of specular $\mathcal{I}^S$ and diffuse $\mathcal{I}^D$ components for the reflected light color $\mathcal{I}$ at each pixel,

$$\mathcal{I} = \mathcal{I}^D + \mathcal{I}^S, \tag{4.4}$$

where $\mathcal{I} = \{I_r, I_g, I_b\}$ is the image color with three color components. Define chromaticity or *normalized* color as

$$\sigma_c = \frac{I_c}{\sum_{c\in\{r,g,b\}} I_c}, \tag{4.5}$$

where $c \in \{r, g, b\}$. Similarly, we can define diffuse chromaticity Λ_c and illumination chromaticity Γ_c by

$$\Lambda_c = \frac{I_c^D}{\sum_{c\in\{r,g,b\}} I_c^D}, \quad \Gamma_c = \frac{I_c^S}{\sum_{c\in\{r,g,b\}} I_c^S}. \tag{4.6}$$

The maximum chromaticity can be defined [78] by

$$\sigma_{max} = \max\{\sigma_r, \sigma_g, \sigma_b\}, \tag{4.7}$$

and similarly for maximum diffuse chromaticity by

$$\Lambda_{max} = \max\{\Lambda_r, \Lambda_g, \Lambda_b\}. \tag{4.8}$$

Then the diffuse component can be computed by

$$I_c^D(\Lambda_{max}) = I_c - \frac{\max_{c\in\{r,g,b\}} I_c - \Lambda_{max}\sum_{c\in\{r,g,b\}} I_c}{1 - 3\Lambda_{max}}, \tag{4.9}$$

according to [78]. Therefore the highlight detection problem can be formulated as the searching for the maximum diffuse chromaticity Λ_{max} which changes from pixel to pixel [78], and can be improved based on bilateral filtering in Λ_{max} estimation [79]. The method of bilateral filtering based estimation is used for highlight detection in our makeup detection problem.

Given the detected highlight, we compute the mean, standard deviation, and entropy, in each of the designated patch in the highlight image, using Equations. similar to (4.1), (4.2), (4.3).

4.3 Patch Selection

In above, we introduced four different facial cues to characterize facial makeup. Each feature is computed in each facial patch independently. 12 patches are used for makeup detection including left and right eyebrow, eye, cheek regions, as well as forehead, nose, mouth, upper lip, and chin regions. The whole face region are the 12th patch. To explore makeup effects on each face region, we perform patch selection using a greedy method.

Specifically, we measure the accuracy or capability of each patch for facial makeup detection. All of the patches are sorted in a descending order based on their classification accuracy results. Starting from the rank-1 patch, we sequentially add patches, one by one, and measure the accuracy again when each new patch is added. It follows the similar procedure until all patches are added. Then the peak value with the highest accuracy will be selected to obtain the proper number of patches and which patches to use for facial makeup classification. This process is done for each of the four features separately. Based on this greedy search method, we can find all useful patches in each feature representation for makeup detection.

4.4 Discriminative Mapping

After patch selection, we obtain the useful patches for makeup classification. However, there are still two things to consider: (1) if the features on the selected patches can be integrated together; and (2) if the features on the selected patches can be made more discriminative in order to improve the recognition accuracy. To address these, we will combine the selected patches with each corresponding feature, and use a discriminative mapping method to make the combination more efficient and discriminative.

For discriminative mapping, we exploit the Marginal Fisher Analysis (MFA) [80], which is a supervised manifold learning algorithm with Fisher criterion. It constructs the within-class graph $\mathcal{G}_{\mathrm{w}}$ and between-class graph $\mathcal{G}_{\mathrm{b}}$ considering both discriminant and geometrical structure in the data. Define the within-class affinity weight $s_{ij}^{(\mathrm{w})} = 1$ when $\mathbf{x}_i$ and $\mathbf{x}_j$ are k nearest neighbors of each other with the same class label, otherwise $s_{ij}^{(\mathrm{w})} = 0$. Define symmetric matrix $\mathbf{S}_{\mathrm{w}}(i,j) = s_{ij}^{(\mathrm{w})}$, diagonal matrix $\mathbf{D}_{\mathrm{w}}(i,i) = \sum_j s_{ji}^{(\mathrm{w})}$, and Laplacian matrix $\mathbf{L}_{\mathrm{w}} = \mathbf{D}_{\mathrm{w}} - \mathbf{S}_{\mathrm{w}}$. Similarly, define the between-class affinity weight $s_{ij}^{(\mathrm{b})} = 1$ when $\mathbf{x}_i$ and $\mathbf{x}_j$ are k nearest neighbors of each other with different class labels, otherwise $s_{ij}^{(\mathrm{b})} = 0$. Thus $\mathbf{S}_{\mathrm{b}}$, $\mathbf{D}_{\mathrm{b}}$ and $\mathbf{L}_{\mathrm{b}}$ are obtained. The objective of MFA is to obtain the optimal projection vector $\mathbf{p}^*$ such that

$$\mathbf{p}^* = \operatorname*{argmin}_{\mathbf{p}} \frac{\mathbf{p}^T \mathbf{X} \mathbf{L}_{\mathrm{w}} \mathbf{X}^T \mathbf{p}}{\mathbf{p}^T \mathbf{X} \mathbf{L}_{\mathrm{b}} \mathbf{X}^T \mathbf{p}} \tag{4.10}$$

Here we use the optimal $\mathbf{p}^*$ as the basis for discriminative mapping of the facial makeup patterns after patch selection. This exploration of a discriminative mapping method for facial makeup detection is one of our novel contributions.

4.5 Experiments

4.5.1 Makeup Database

A face image database of 1002 face images that is 501 pairs of female individuals. It contains mainly adult women of Asian or Caucasian descent, and is larger than datasets used in [71–73]. Each pair has one makeup and one non-makeup face images of the same individual.

(1) (2) (3) (4)

FIGURE 4.2: Pairs of face images of the same individuals in makeup (columns 1 and 3) or nonmakeup classes (columns 2 and 4), showing significant appearance changes. We study how to perform face authentication that is robust to makeup changes.

4.5.2 Experimental Setting

Based on the knowledge about cosmetics, we propose to study four categories of features, i.e., skin color tone, skin smoothness, texture, and highlight, as stated in Section 4.2. Through experimental validations and comparisons, we can understand what features are useful for makeup detection, how to use these features effectively, and if it is possible to combine these features together for a better performance.

In the following, we study the proposed four categories of features separately. There are about 4/5 faces for training, and the remaining for testing in our five-fold cross validation. Makeup detection problem is considered as a two-class problem: with and without makeup. The support vector machines (SVM) with radial basis function (RBF) kernel [81] are used as the classifier.

4.5.3 Skin Color Tone

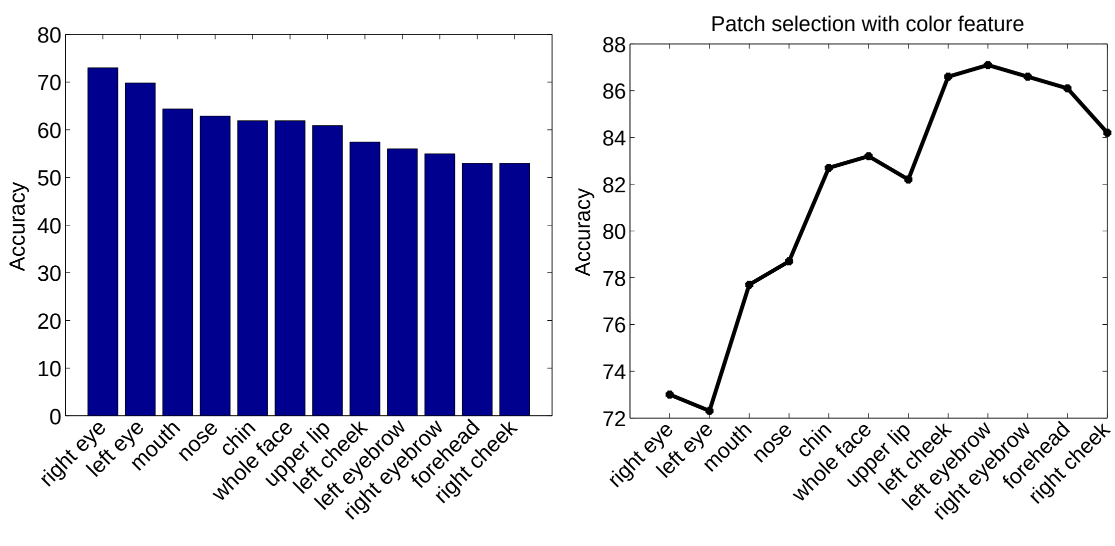
One purpose of using cosmetic products on faces is to change the skin color tone to make the face look younger and more beautiful. So our first feature on makeup detection is about color. From each local patch in a face image, we compute the mean, standard deviation, and entropy in each of the three channels: Red, Green, and Blue. Thus there is a 9-dimension feature vector extracted from each local patch.

To understand the differences among the local patches in makeup detection, we use only one local patch each time for makeup classification. Then the local patches are sorted in a descending order of accuracy. The result is shown in the left of Figure 4.3. We can observe that colors in regions of eyes and mouth are the top patches. This result tells that the eyes and mouth regions have more color changes between makeup and non-makeup faces. There is a slight difference between left eye and right eye, that is because of the existence of other variations in face images, such as pose, illumination, etc. The colors in some other facial regions are also useful but with lower accuracies for makeup detection.

To explore how many local patches can be combined together to enhance the makeup detection performance, we perform patch selection, as stated in Section 4.3. A greedy algorithm is used to select patches by adding one patch each time, starting from the most discriminative local patch first. The process of patch selection for the color feature is shown in the right of Figure 4.3. We can see the accuracy increases by adding more patches, and reaches the peak when the 6 most accurate patches (out of 12) are combined together. This result shows that the patch selection is needed, and our scheme is effective to find the most useful local patches.

To show the results quantitatively, we put the makeup detection accuracies based on color in the first row of Table 4.1. When all patches are used, the makeup detection accuracy is 83.0%, which is increased to 85.0% when our patch selection scheme is executed. We also show that it is not good to use the whole face image. The accuracy reduces significantly to 55.5% when the whole face is used.

Next, we will do similar analysis for the other three features.



(a) (b)

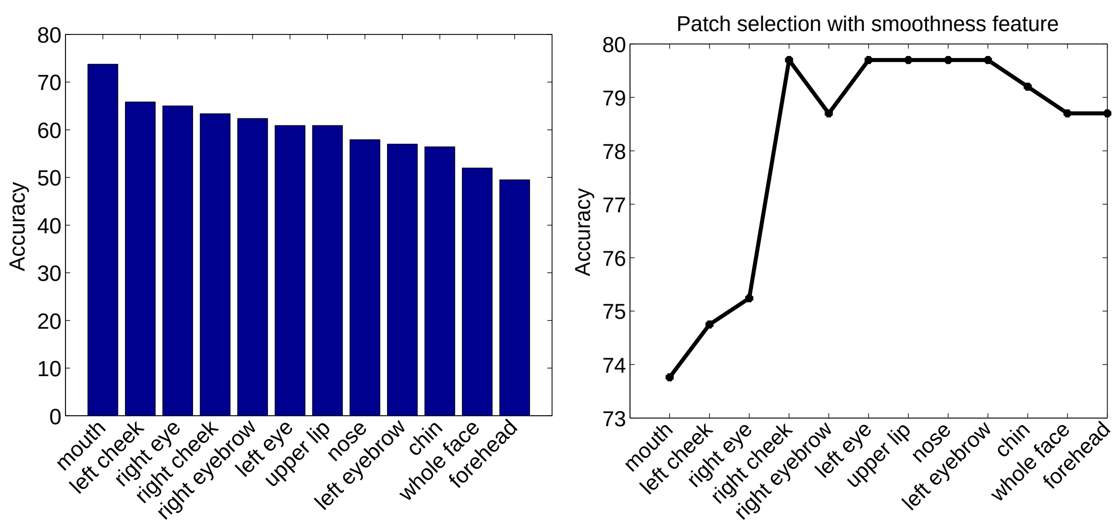
FIGURE 4.3: Facial makeup detection using color feature. The RGB color space is used. (a) Recognition accuracy of the color feature on each local patch is sorted in descending order. (b) Accuracies of sequential adding of local patches based on the sorted patches are shown. The accuracy increases when more patches are added but drops when too many patches are used. The peak is selected based on our patch selection scheme.

4.5.4 Skin Smoothness

The second category of features we explore is the facial skin smoothness. Women like to use makeup products to hide certain facial flaws, such as freckles, pimples, acne scars, wrinkles, and pores. In addition to make the face look younger and more beautiful, the makeup also makes the facial skin smoother. To separate makeup and non-makeup faces, the skin smoothness feature may be useful.

To compute the skin smoothness, we use the intensity values from each local patch in a face image. As stated in Section 4.2.3, we compute the mean, standard deviation, and entropy of the intensity values in each local patch.

The results of patch sorting and patch selection based on the smoothness feature are shown in Figure 4.4. We can observe that the top patches for smoothness measure are different from those using color features. The number of the selected local patches is also different from the color features.



(a) (b)

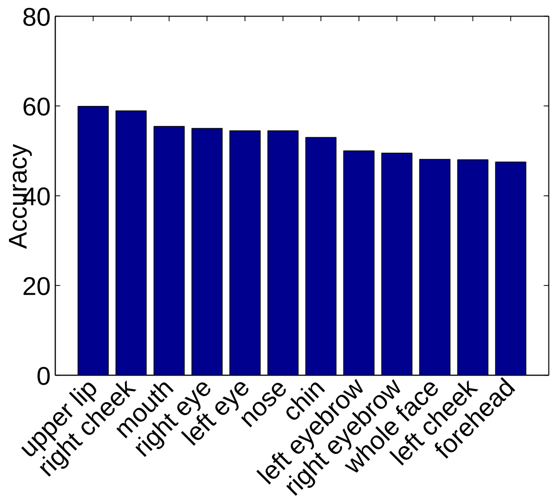
FIGURE 4.4: Facial makeup detection using facial smoothness feature. (a) Recognition accuracy of the smoothness feature on each local patch is sorted in a descending order. (b) Accuracies of sequential adding of local patches based on the sorted patches. The accuracy increases when more patches are added. The peak is selected based on our patch selection scheme.

The quantitative results of the smoothness feature are shown in the second row of Table 4.1. When all patches are used, the makeup detection accuracy is 80.0%, which stays the same of 80.0% when our patch selection scheme is executed. Again, it is not good to use the whole face image. The accuracy reduces to 52.0% when the whole face is used for smoothness measure.

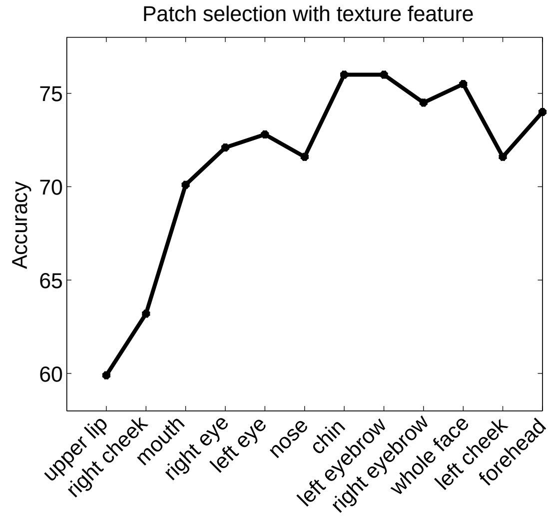
4.5.5 Skin Texture

In addition to skin smoothness change, cosmetic products may also change the facial skin texture. To explore this, we extract the texture feature for the separation of makeup and non-makeup faces. Since the LBP feature is usually used for facial texture characterization for face recognition, we use the LBP texture feature for makeup detection, as stated in Section 4.2.4. The LBP feature is computed in each local patch of image intensities.

The results of patch sorting and patch selection based on the texture feature are shown in Figure 4.5. We can observe that the top patches for texture measures are different from those using color or smoothness features. Two most accurate local patches are selected to achieve the highest accuracy in the patch selection process.



(a)



(b)

FIGURE 4.5: Facial makeup detection using facial texture feature. The LBP is used to characterize the texture. (a) Recognition accuracy of the texture feature on each local patch is sorted in a descending order. (b) Accuracies of sequential adding of local patches based on the sorted patches. The accuracy increases when more patches are added but drops when too many patches are used. The peak is selected based on our patch selection scheme.

The quantitative results of the texture feature are shown in the third row of Table 4.1. When all patches are used, the makeup detection accuracy is 81.0%, which is increased to 85.0% when the patch selection scheme is executed. When the whole face image is used for texture measure, the accuracy reduces to a lower value of 72.5%.

4.5.6 Highlight

The last category of feature that we explore is the facial highlight. Intuitively, the diffuse reflection and highlight distributions maybe different in makeup and non-makeup faces. Based on this, we compute the highlight in each face image based on the dichromatic reflection model, and extract features to characterize the highlight from each patch in a face image, as stated in Section 4.2.5.

The results of patch sorting and patch selection based on the highlight feature are shown in Figure 4.6. We can see the top patches for highlight measure and the accuracy variation during the patch selection process. To reach a high accuracy, six patches are selected as the most accurate, which are different from the color, smoothness, or texture features, individually.

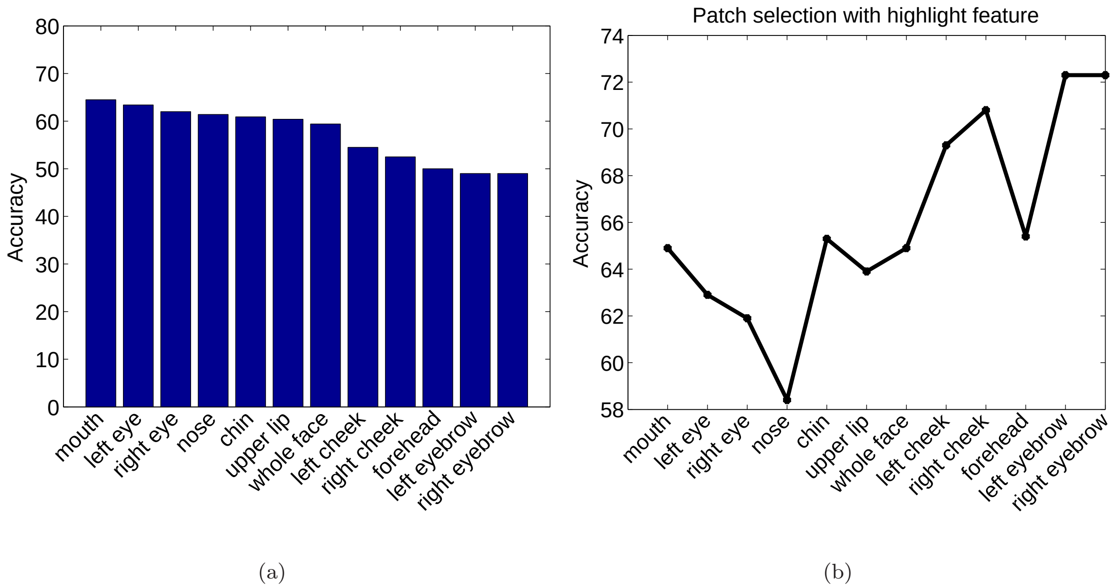


FIGURE 4.6: Facial makeup detection using facial highlight feature. The highlight is extracted in each face image. (a) Recognition accuracy of the highlight feature on each local patch is sorted in a descending order. (b) Accuracies of sequential adding of local patches based on the sorted patches.

The quantitative results of the highlight feature are shown in the last row of Table 4.1. When all patches are used, the makeup detection accuracy is 64.5%, which is increased to 68.5% based on patch selection. It is still not good to use the whole face for highlight characterization, since the accuracy reduces to 52.5% in that case.

Overall, we have shown that several kinds of features can be extracted for classification of makeup and non-makeup faces. We found that the performance will be much lower when these features are computed in the whole face image. The reason might be that the cosmetic products are used differently on different facial parts, and some parts may see the cosmetic "effect" more significantly than others. So we emphasize the use of local patches for facial makeup characterization. Further, the designated facial patches might not be equally useful for facial makeup detection. Further, our patch selection scheme is effective to improve the recognition accuracies. The next questions are: (1) can we make these features more discriminative? and (2) can we combine the different features together? We will address them in the following.

Table 4.1: Facial makeup detection accuracies using four different categories of features in three different cases: (1) whole face, (2) all local patches, and (3) selected patches.

	Whole face	All patches	Patch selection
Color	55.5%	83.0%	85.0%
Smoothness	52.0%	80.0%	80.0%
Texture	72.5%	81.0%	85.0%
Highlight	52.5%	64.5%	68.5%

4.5.7 Discriminative Mapping

We explore a discriminative mapping approach to combine the selected patches with different features together, and make them more discriminative. To verify the idea, we apply the MFA method introduced in Section 4.4.

The results are shown in Table 4.2. We found that the discriminative mapping by the MFA can improve the color feature for makeup detection from 85.0% to 88.0%. Combining color and smoothness features can further improve the makeup detection accuracy to 96.0%, which is the highest accuracy we obtain. On the other hand, when all four features are concatenated together and then mapped discriminatively by the MFA method, the accuracy is 92.5%, which is higher than each single feature, i.e., color, smoothness, texture, and highlight, but is still lower than the color and smoothness feature combination. We also tried to use other feature combinations, as those shown in Table 4.2, but there is no accuracy that can be higher than the 96.0%. Based on this result, we observe that the color and smoothness features can complement each other, and their combination with discriminative mapping performs significantly better than other combinations for makeup detection. The texture feature itself can get a high accuracy, but its combination with other features cannot reach the highest accuracy.

In summary, the combination of color and smoothness features is the best for makeup detection in our current experiment. A discriminative mapping via the MFA method can improve it to a high accuracy of 96.0%. Other features are useful for makeup detection, however, their performance is not so good as the color and smoothness combination. The texture and highlight cues may not be sufficiently complementary to the color and smoothness cues, and thus the combination of all of these cues gives a lower accuracy than using the color and smoothness cues.

TABLE 4.2: Facial makeup detection accuracies using discriminative mapping by the MFA method. Each feature is based on patch selection.

Discriminative Mapping on Various Feature Combinations	Accuracy
Color	88.0%
Color+Smoothness	**96.0%**
Color+Texture	91.0%
Color+Smoothness+Texture	94.0%
Color+Smoothness+Texture+Highlight	92.5%

We show the results of wrong classification face images in Figure 4.7. There are six non-makeup face images incorrectly classified as makeup (left), while two makeup faces incorrectly classified as non-makeup (right). The reasons may include the low image quality, and expression and pose changes. Sometimes, it is even challenging for human perception to separate makeup and non-makeup faces.

4.6 Summary

In this Chapter, we develop an automatic makeup detection system. Based on the knowledge of how to apply cosmetics, we have proposed to investigate four categories of features,

Figure 4.7: The faces of wrong classification results. Left: six non-makeup faces incorrectly classified as makeup; Right: two makeup faces incorrectly classified as non-makeup.

including facial skin tone, skin smoothness, texture, and highlight. These features are computed in local patches of face images. We have presented a patch selection scheme and discriminative mapping to improve the makeup detection accuracies effectively. We have found that the color feature combined with smoothness performs better than others for makeup characterization. A high accuracy of 96.0% has been achieved.

Chapter 5

Face Recognition with Makeup Changes

5.1 Introduction

According to the standard of beauty, women use makeup to hide facial flaws, as well as to enhance some traits to appear more attractive. Different facial regions will be applied with different cosmetic products. For example, because big eyes are usually considered attractive, to visually create an effect of big eyes, colorful eye shadows may be applied around the eyes. Some studies have been done to explore the influence of makeup changes on face recognition [41–43]. However, no previous study shows how facial makeup can effect facial attributes. In this Chapter, we explore makeup effects on facial attributes. Our key idea is that the dual-attributes can be learned from faces with and without makeup, separately [82]. The analysis of accuracy changes can be performed between within-group and cross-group attribute classification. The study of facial attributes variation with makeup changes can give a semantic explanation of the makeup effects on facial appearance.

How to develop face recognition systems robust to makeup changes is a relatively new problem. There are two schemes concerned in our work. One is based on PLS method [74]. The other one is based on dual attributes [82].

In the rest of this Chapter, at first, we introduce dual attributes of female faces with makeup changes in Section 5.2. A dual-attributes based method is also concerned to utilize dual attributes as the feature descriptors for face matching to reduce the makeup influence on face recognition. In the following, we explore the PLS method for makeup-invariant face authentication in Section 5.2.5. In Section 5.3, experiment results are analyzed. Finally, we summarize this study in Section 5.4.

5.2 Dual Attributes

5.2.1 Facial Attributes

Attributes are a semantic level description of visual traits [83, 84]. For example, a horse can be described as four-legged, with fur, a long tail, etc. Facial attributes, such as big eyes, or a pointed chin, are semantic level descriptions of visual traits in faces [85]. Facial attributes can also be used for face image ranking and search [86, 87]. The query can be given by semantic facial attributes, e.g., a man with moustache, rather than just a query image.

In this thesis, we only focus on attributes in female group. 28 attributes are defined to describe female faces with and without cosmetics. So there is no attribute related to children or males. The list of the 28 attributes is given in Table 5.1. Some examples of these attributes are shown in Figure 5.1. The attributes used in our approach are learned from face images with and without cosmetics, separately. As shown in Figure 5.1, we can visually check the facial appearance differences between faces with and without cosmetics.

To better describe the female faces with and without cosmetics, we used some detailed attribute features, compared to the attributes used in [85–87]. There is no specific consideration of facial cosmetics for face recognition in [85–87]. For example, in [87] and [86], only the attribute of eye glasses is used to describe the characteristics of eyes, while only the attributes of eye width, eyes open, and glasses are used in [85]. In our approach, more detailed attributes are used, such as double-fold eyelids, slanty eyes, and eye crows feet, which include more details related to female beauty and facial cosmetics.

From Figure 5.1, one can get an intuitive observation of the appearance variations of the same attributes between non-makeup and makeup faces. To quantify the differences, it might be necessary to learn the attribute classifiers separately, for the makeup or non-makeup face images. Therefore, we propose the dual-attributes classifiers, which will be presented in the next.

Table 5.1: List of the 28 Attributes used in our approach.

Asian	Eyebrow thickness	Pointed chin
Attractive	Eyebrow wave	Red cheek
Big eyes	Full bang	Small eye distance
Big nose	Half bang	Slanty eyes
Black eyes	High nose	Square jaw
Blue eyes	Lip line	Thick mouth
Double-fold eyelids	Nose middle heave	Wide cheek
Eye bage	No bang	White
Eye crows feet	Open nose	
Eyebrow long	Philtrum	

5.2.2 Dual-attributes Classifiers

To address the visual appearance differences between faces with and without makeup in attribute learning, two classifiers are learned for each attribute, one is on faces with makeup, and the other is on faces with non-makeup. They are called dual attributes classifiers.

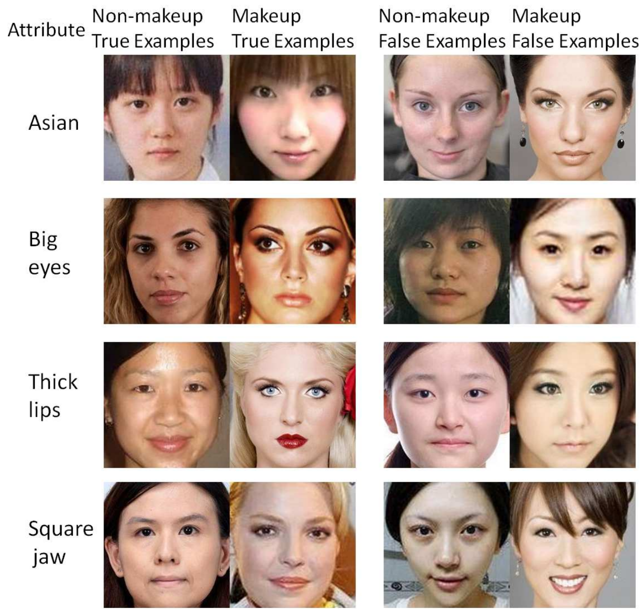


Figure 5.1: Example face images for some attributes. Faces in the same row are the positive and negative examples for a given attribute label. From left to right: positive non-makeup face, positive makeup face, negative non-makeup face, and negative makeup face.

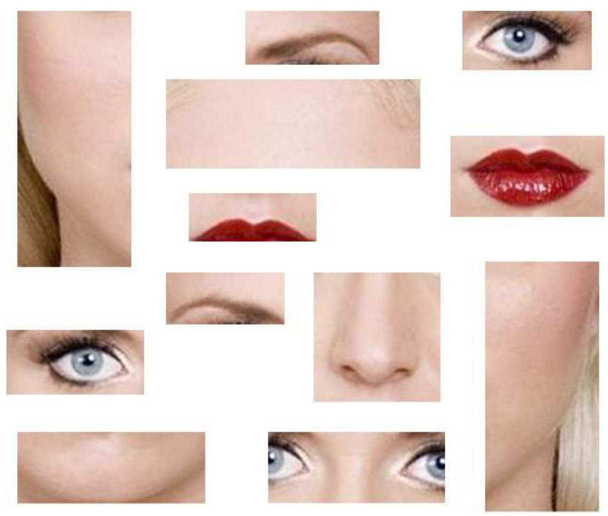


Figure 5.2: Local patches in faces for dual-attributes learning.

In Section 5.2.1, we introduced 28 attributes. Three attributes among the 28 are related to color: blue eyes, black eyes, and red cheek. The pixel value of R, G, and B channels of an image (using local patches) are combined together to form the feature. For other attributes, Histogram of Oriented Gradients (HOG) feature descriptor is used to represent the facial patches as shown in Figure 5.2.

5.2.2.1 Histogram of Oriented Gradients

HOG is a widely used feature descriptor recently. The basic idea about HOG is that local object appearance and shape within an image can be described by the distribution of intensity gradients or edge directions [88]. Suppose a selected region R in a image, given a grid cell C with center $p_c = (x_c, y_c)$ from region R. For each point p in the cell, both gradient magnitudes $G_x(x, y)$ and $G_y(x, y)$ and gradient orientation $\Theta(x, y)$ can be calculated. Usually, orientations from $0 - 180°$ are equally divided into 9 orientation bins F_C. Then each pixel can contribute a weighted vote for one orientation bin which its gradient orientation value $\Theta(x, y)$ belongs to. The weight value is related to its gradient magnitudes. The votes are accumulated into orientation bins F_C for each cell. The HOG descriptor of the whole region R can be calculated by $F_R = \sum_{C \in R} F_C$. In our study, we use L2-norm to normalize the F_R into a normalized histogram feature descriptor for each region.

5.2.2.2 Within-Group and Cross-Group Attribute Classifications

To verify the influence of facial cosmetics on facial attributes, we compare the performance of attribute classifications within the same group of faces or across groups (i.e., from makeup to non-makeup, or vice versa), which is an accuracy change analysis (ACA). This further proves the influence of facial makeup on face analysis, and the necessity to deal with the influence of facial makeup on facial image analysis.

5.2.3 Accuracy Change Analysis

We evaluate the dual-attributes classification results and conduct ACA to examine the facial attributes variation under the influence of makeup. ACA is to perform a cross-group attributes classification to investigate the influence of facial cosmetics on attribute classification. Here we have two groups, one contains makeup faces, and the other contains non-makeup faces. In other words, each individual has two face images in our database, one is in the makeup group, and the other is in the non-makeup group. When both learning and testing are in the same group for attribute classifications, it is called within-group classification, while the training and testing are in two different groups, e.g., from makeup to non-makeup, or vice versa, it is called cross-group or cross-makeup attribute classification.

5.2.4 Dual-attributes based Face Verification

The above learned dual-attributes in Chapter 5 can be combined together for face matching. Note that this matching is on the semantic level represented by the attributes, rather than using low-level facial features for direct matching. In our dual-attributes approach, the attributes for a makeup face will be detected using the attribute classifiers learned from makeup faces, while the attributes for a non-makeup face will be obtained by using the classifiers learned from non-makeup faces.

For face verification, all available attributes extracted from a pair of faces form a vector of values, denoted by $F(I) = \langle MC(I), NMC(I) \rangle$ for individual I, where $\langle MC(I)$ is an attribute vector for a makeup face, while $\langle NMC(I)$ is an attribute vector from a non-makeup face. $\langle MC(I) =< MC_1, MC_2, ..., MC_n >$, while the attribute vector for a non-makeup face is $\langle NMC(I) =< NMC_1, NMC_2, ..., NMC_n >$, where n is the number of attributes, I means the I-th person.. $\langle NMC_i$ and $\langle MC_i$ are used to name the i-th non-makeup and makeup attribute classifier separately. A SVM face verification classifier will be trained to distinguish whether two face images are the same woman or not. In [85], the attribute vector is binary with values either 0 and 1, where 0 means the face does not have the attribute, while 1 means the face has the attribute. In our experiments, we use distances from the test sample to the SVM hyperplane instead of the binary values, and found the face verification accuracy can be improved by about two percent over the binary values.

Although attributes have been used for face verification in previous works, e.g., [85], there are several differences between our work and that in [85]: (1) We study facial makeup on face verification while there is no specific study on facial makeup in [85]; (2) The main focus in [85] is about face verification with pose, illumination, and expression (PIE) changes, which also appears in our database, but we emphasize the makeup influence and study face verification across facial makeup specifically; (3) In our database, there are faces with large head pose variations, while the face images in [85] were filtered by a frontal-view face detector and thus the pose variations are not very large; (4) The population contains women adults only in our database, while there are both men and women, young and adult in [85], where many attributes such as the age group, gender, beards, etc., can be used in [85], but those attributes cannot be used to help face authentication in our study. More importantly, we propose *dual-attributes* based face verification as shown in Figure 5.3, which has not been presented in any previous attribute-base approaches, to the best of our knowledge.

5.2.5 PLS-based Face Verification

In Chapter 3.4.2, the PLS method has been successfully applied to reduce the influence of BMI changes on face recognition. To develop a robust face recognizer insensitive to facial makeup changes, we study the effect of PLS method to reduce the influence of makeup changes on face recognition. The basic framework of our proposed approach can be illustrated in Fig. 5.4.

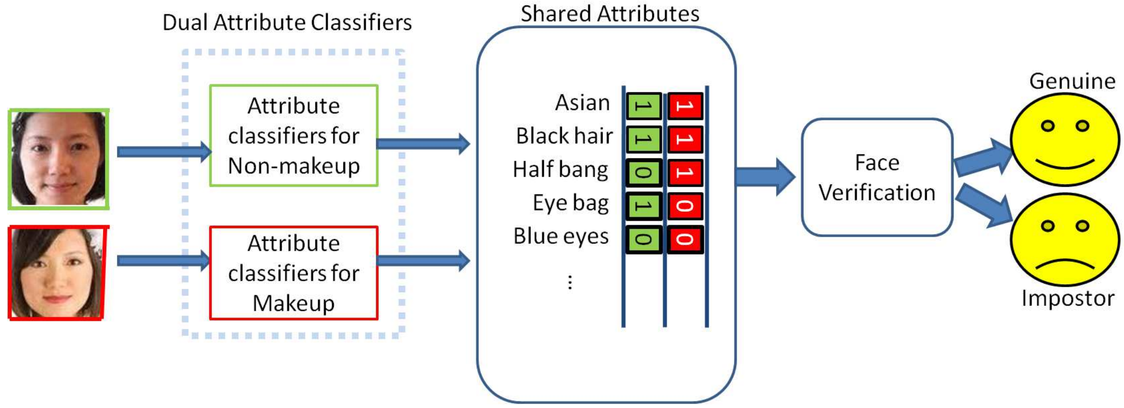


FIGURE 5.3: Illustration of our *dual-attributes* approach to face verification. Two sets of attributes are learned in face images with and without cosmetics, separately. In testing, the pair of query faces undergo two different groups of attribute classifiers, and the shared attributes are used to measure the similarity between the pair of faces in a semantic level, rather than a direct match with low-level features.

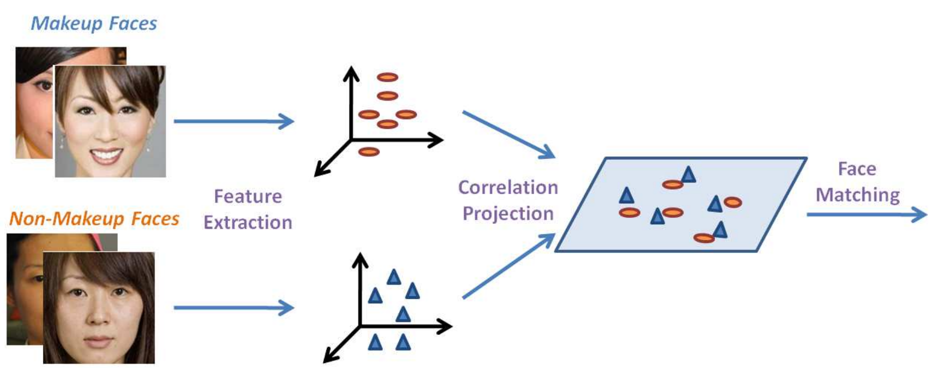


FIGURE 5.4: The basic framework that we propose for face recognition robust to cosmetic changes based on a correlation mapping.

To learn the correlation between makeup and non-makeup face images, we explore PLS method. We investigate if PLS is useful for face recognition invariant to facial cosmetics.

5.2.5.1 Facial Feature Representation

The correlation method PLS will be applied to the extracted features in face images, rather than raw pixel values. The reason is that facial makeup may change the facial appearance significantly in terms of single pixel values, and the changes of pixel values may be different for different facial parts and for different individuals. Building the correlations on extracted features rather than raw pixel values may relieve the problem of pixel value changes caused by facial makeup to some degree.

In extracting facial features, we use some standard methods for facial feature representation, such as the local binary patterns (LBP) [76], and the histogram of oriented gradient (HOG) [88]. The principal component analysis (PCA) is usually used for dimensionality reduction, it can compute the Eigenfaces [96] for facial feature characterization. The projection coefficients are used as features for face representation in Eigenfaces [96], which are used frequently in face recognition. So we also investigate the Eigenfaces representation (called PCA) in our correlation analysis. Our emphasis is that the correlations need to be performed on those features rather than on raw pixels directly.

In some previous approaches, e.g., [94, 97], correlation is applied to raw face images. We found that the performance will be deteriorated when the correlation methods are applied to raw images, which can be demonstrated in our experiments. This fact tells that our problem of face authentication with facial makeup changes has some special properties itself, and is different from the problems studied in [94, 97, 98]. In addition, our problem might be more challenging than those in [94, 97], since the direct correlation on raw pixels values has a significantly lower accuracy in our problem. Finally, we use local patches rather than the whole face image, which is essential for our problem, and will be presented next.

5.2.5.2 Local Patches

In addition to learning correlations on extracted features rather than raw pixels, we also found that local patches have to be used in computing the correlations and extracting features. This can be demonstrated in our experiments. The reason is that different facial regions could be applied with different cosmetic products and with different amount or degrees. For example, the eye makeup may use various colors (e.g., gold soft smoky, bronze, etc.) to make the eye look youthful and the lashes extra-long, while red or other coloration may be applied to the mouth to reshape (e.g., reduce, increase or bring balance to the lips) to have a beautiful look. It might be difficult to learn a correlation of global faces between makeup and non-makeup. To deal with this problem, we use local patches on face images. Another advantage of using local patches is that the local patches can also make our approach robust to head pose changes, facial expressions, or illumination variations, in addition to cosmetic changes. As a result, both the feature extraction and correlation learning are based on local patches, rather than a global face image. Again, this is different from the use of whole images in [94, 97]. The patches we use in each face image are the same with used in Chapter 5 for dual attributes study.

5.2.5.3 Using Makeup Detection for Automated Face Verification

The makeup detection result can help to make it automatic for our correlation mapping based face authentication. The idea is that for a given pair of face images, we first perform makeup detection. If a face is detected to have makeup, it will be projected by using the correlation bases corresponding to the makeup class, otherwise, the bases corresponding to the non-makeup class will be used for projection or feature transform. This way, the correlation bases selection can be performed automatically, and the process of face recognition can be executed automatically, without asking the user to provide the makeup or non-makeup information. As a result, we can develop a complete framework for face authentication with an automated makeup detection, as shown in Fig. 5.5.

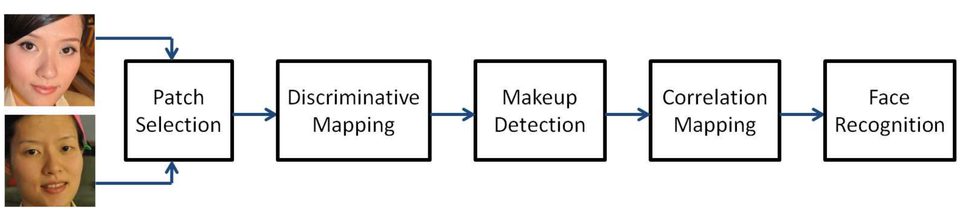


FIGURE 5.5: A complete framework of integrated makeup detection and face verification. Given a pair of face images with extracted features, the system performs patch selection, discriminative mapping, and makeup detection. Then the correlation mapping can be executed based on the recognized makeup or non-makeup. Finally, face verification is accomplished. The system can do both face matching and makeup detection.

It is useful in practice to make the face recognition process automatic without user interaction. However, the accuracy of facial makeup detection will have a direct impact on correlation bases selection and the final face recognition accuracy as well. We need to investigate if the final accuracy of face recognition is influenced significantly or not, using the makeup detection result. We will validate this automated approach to face verification in our experiments.

5.3 Experiments

5.3.1 Database

The same makeup database introduced in 4.5.1 are used to study the facial attributes variations under the influence of makeup. The 28 attributes are manually labeled by five human subjects. The five participants discussed with each other for each attribute label to get a majority voting if there is an inconsistency. Then the labeled attributes are used as the ground truth for our study.

5.3.2 Dual-attributes Classification

The attribute classification results are shown in Table 5.2. The first column is the attribute classification accuracies using non-makeup faces for both training and testing. Please note there is no overlap of individuals between the training and testing sets. The second column shows the classification results using the same non-makeup faces for testing, but with different attribute classifiers trained by using makeup faces. The third column 'decrease1' shows the decrease of classification accuracy when the second column is compared to the first column in each row. If the decrease is negative, it means the accuracy is increased. The remaining three columns have the similar meanings.

Table 5.2: Attribute classification results for both within-group and cross-group cases. We use "$A \rightarrow B$" to indicate that 'A' is for training, while 'B' is for testing. The two groups are 'makeup' and 'non-makeup' faces. The column 'decrease1' shows accuracy decrease when the training group is changed to makeup. The column 'decrease2' shows accuracy decrease when the training group is changed to non-makeup faces.

	non-makeup → non-makeup	makeup → non-makeup	decrease1	makeup → makeup	non-makeup → makeup	decrease2
Big nose	74 %	74 %	0 %	64 %	64 %	0%
Nose middle heave	73 %	73 %	0 %	65 %	65 %	0 %
Philtrum	61 %	60 %	1 %	71 %	70 %	1 %
Wide cheek	85 %	87 %	-2 %	84 %	82 %	2 %
White	88 %	88 %	0 %	84 %	85 %	-1 %
Half bang	85 %	84 %	1 %	81 %	80 %	1 %
High nose	70 %	73 %	-3 %	78 %	73 %	4 %
Lip line	63 %	66 %	-3 %	69 %	65 %	4 %
Asian	84 %	87 %	-3 %	88 %	84 %	4 %
Double-fold eyelids	66 %	62 %	4 %	62 %	68 %	-4 %
Small eye distance	74 %	69 %	5 %	67 %	67 %	0 %
Open nose	72 %	68 %	4 %	74 %	67 %	7 %
Pointed chin	90 %	83 %	7 %	85 %	75 %	10 %
Thick lips	63 %	61 %	2 %	71 %	68 %	3 %
Eyebrow long	68 %	58 %	10 %	64 %	58 %	6 %
Eyebrow thickness	64 %	32 %	32 %	84 %	48 %	36 %
Eyebrow wave	68 %	68 %	0 %	62 %	53 %	9 %
Blue eyes	75 %	74 %	1%	73 %	70 %	3 %
Black eyes	85 %	81 %	4 %	78 %	76 %	2 %
Eye bag	63 %	65 %	-2 %	71 %	66 %	5 %
Slanty eyes	67 %	54 %	13 %	59 %	58 %	1 %
Eye crows feet	71 %	65 %	6 %	69 %	67 %	2 %
Big eye	64 %	59 %	5 %	67 %	60 %	7 %
Square jaw	64 %	62 %	2 %	71 %	60 %	11 %
Red cheek	67 %	61 %	6 %	71 %	67 %	4 %
Attractive	68 %	64 %	4 %	78 %	78 %	0 %
Full bang	92 %	92 %	0 %	96 %	89 %	7 %
No bang	95 %	95 %	0 %	85 %	80 %	5 %

5.3.2.1 Dual-attributes Categories

Specifically, we categorize the accuracy changes into four cases: "Small change", "M-preferred", "NM-preferred", and "Big change", where "M" is for makeup, and "NM" is for non-makeup. When the makeup group is used for training, while the cross-group accuracy is higher than the within-group, it is categorized as "M-preferred". Similarly, we may have "NM-preferred".

From Table 5.2, one can observe the four cases. Firstly, the attribute classification results belonging to the case of "Small change" include: 'Big nose', 'Nose middle heave', 'Philtrum', 'Wide Cheek', and 'White'. Our interpretation is that the facial cosmetics does not change those attributes (mainly on shape) too much. In this case, the attribute classification accuracy changes are small when the cross-group accuracies are compared to the within-group. To tolerate the small estimation errors caused by other variations (pose, expression, illumination, etc.) and consider the fact that the training database is not big, we use threshold values between -2% and 2% to determine this case. In other words, if the accuracy changes are within the range of -2% to 2%, it is termed as a small change.

The second case is called "M-preferred", which means the cross-group accuracy is increased when the makeup group is used for training attribute classifiers. In this case, the attribute class includes: 'High nose', 'Lip line', and 'Asian'. Our interpretation is that the facial makeup will enhance the classification accuracy for those attribute classes.

Similarly, the third case is called "NM-preferred", which means the cross-group accuracy is increased when the non-makeup group is used for training attribute classifiers. In this case, the attribute class includes only the "Double-fold eyelids". Our interpretation is that this attribute may be easier to be learned from non-makeup face images.

The last case is the "Big change", which means that cross-group classifications result in significant accuracy decreases. There are 18 attributes belonging to this case, as can be seen in Table 5.2.

Our decision of the four cases are based on the rules as below:

$$\textbf{ACA} = \begin{cases} \textit{Small change}, & d1 \textit{ and } d2 \in [-2\%, 2\%], \\ \textit{M-preferred}, & d1 < -2\% \textit{ and } d2 \geq 0, \\ \textit{NM-preferred}, & d1 \geq 0 \textit{ and } d2 < -2\%, \\ \textit{Big change}, & \quad \textit{otherwise} \end{cases}$$

where ACA is for 'Accuracy change analysis', d1 is for 'decrease1,' d2 is for 'decrease2.' M-preferred means 'Makeup preferred', and NM-preferred means 'Non-makeup preferred.' The summary of the four cases is shown in Table 5.3 for a clearer view.

TABLE 5.3: Attribute classification ACA. The change analysis results in four cases: Small change, Makeup preferred, Non-makeup preferred, and Big change.

Attribute Classification	Attributes
Small change	Big nose, Nose middle heave, Philtrum, Wide Cheek, White, Half bang,
Makeup preferred	High nose, Lip line, Asian,
Non-Makeup preferred	Double-fold eyelids,
Big change	Small eye distance, Open nose, Pointed chin, Thick lips, Eyebrow thickness, Eyebrow long, Eyebrow wave, Blue eyes, Black eyes, Slanty eyes, Eye crows feet, Big eyes, Square jaw, Red cheek, Attractive, Full bang, Forehead visible,

From Table 5.3, we can understand that most of these attribute changes or no changes are close to our common knowledge. Women like to wear makeup near eyes to appear more attractive. Usually there is less makeup designed to change the appearance of nose. That is why 2 out of 4 attributes related to nose have small changes, while most eye-related attributes (8 out of 9) have big changes. That is the makeup effect. One interesting attribute is 'double-fold eyelids'. Usually we can observe whether a female is double-fold eyelids or not when there is no makeup around the eyes. But with makeup, it may bring some difficulty to determine if the woman has double-fold eyelids or not, especially using faked eyelashes for makeup. As a result, we categorized the 'double-fold eyelids' as non-makeup preferred attribute.

5.3.2.2 Dual-attributes Classification Interpretation

Our dual-attributes-based analysis makes it easier to understand and interpret the computational approaches in terms of the effect of facial cosmetics on facial image analysis. This is a nice property using the attributes, and is different from previous approaches, either purely based on human perception [41, 42] or based on low-level features [43].

On the other hand, we can see that the accuracies of many attribute classifiers are decreased, some may be decreased by more than 30%. This phenomenon also indicates the necessity of developing face recognition robust to makeup changes.

We study dual attributes and also correlation projection based methods to deal with makeup changes on face verification. All the experiments are using the same training and testing data.

5.3.3 Face Verification across Makeup

For face verification, we have about 100 pairs of positive faces for testing, while 400 pairs for training in each round. To have some negative pairs in testing, we randomly selected one face image from each of the remaining individuals in the test set. The random selection keeps the makeup and non-makeup group information, i.e., if the face is non-makeup under consideration, the randomly selected negative face will come from the makeup group. So our face matching is always between a pair of makeup and non-makeup faces, no matter whether they are positive or negative pairs. The random pairing process was repeated for each individual. As a result, we have about 100 negative pairs of face images. In total, we have 200 pairs of faces for testing each time. This process is repeated five times to compute the average in our five-fold cross validation scheme.

All faces are aligned based on the eye coordinates. The eyes can be detected by some existing methods, e.g., [99]. 12 patches the same as used in Chapter 5 for dual attributes study are selected based on the relative locations. Then various features will be extracted from the designated local patches in each face image. To show the importance of extracting features from local patches rather than the whole face, we compare the recognition accuracies in the two cases with a variety of features. The cosine distance (angular) measure is used for feature matching.

5.3.3.1 Experimental Setting

We randomly divide the individuals into training and testing sets for attribute classification. Each pair of faces for an individual are either in the training or testing set. There are about 4/5 individuals or face pairs are in the training set, and the remaining 1/5 are used for testing the attribute classification. The testing set is also used for the face verification experiments. For each attribute, two SVM classifiers are trained, using two groups of faces, makeup and non-makeup, separately.

5.3.3.2 Whole Face or Local Patches?

First, we want to verify which is better for face verification: whole face or local patches? Using the same training and test image sets, we measure the face verification accuracies based on different methods, such as the PCA, HOG, LBP for feature extraction in the whole face images and local patches, respectively. We also used the PLS methods on raw images and local patches to compare the accuracies. The results are shown in Table 5.4. Both the recognition accuracy and the area under the ROC (receiver operating curve), denoted by AUC, are computed. Note that the accuracy and AUC are two different measures, which can be used to represent the authentication results from two aspects. For accuracy computation, the threshold value is learned and adjusted in the tuning set, which is part of the training set (about 20% of the training data). The ROC curve characterizes the performance by computing the true positive rate (TPR) and false positive rate (FPR) using many different threshold values.

TABLE 5.4: Face verification with different methods: PCA, HOG, and LBP, for direct face matching; or PLS for correlation-based matching, using the whole face or local patches, respectively.

	Whole face		Patch based	
	Accuracy	AUC	Accuracy	AUC
PCA	64.0%	0.68	72.5%	0.75
HOG	65.0%	0.66	71.5%	0.74
LBP	67.5%	0.68	69.5%	0.72
PLS	64.5%	0.65	71.5%	0.71

TABLE 5.5: The improvement of face verification accuracies when the PLS method is used on various features: PCA, HOG, and LBP, respectively. The use of PLS correlation can improve the accuracies significantly for various features.

	Accuracy without PLS	Accuracy with PLS
PCA	72.5%	81.5%
HOG	71.5%	80.0%
LBP	69.5%	76.0%

From Table 5.4, we can observe several things: (1) the local patch based approaches are usually better than the whole face images, especially for the PCA, HOG, LBP, and the PLS methods. For example, the accuracy for PCA is 64.0% when the whole face image is used, while it increases to 72.5% when the local patches are used. This result tells us that using local patches can reduce the influence of facial makeup on face recognition to some extent. Usually the local patches are used to deal with head pose variation, but we show that local patches are also important to cope with facial makeup changes in face recognition; (2) the correlation methods, such as the PLS, cannot work well on raw images (whole images). For example, the PLS has an accuracy of 64.5% when the global raw images are used, which is lower than the PCA, HOG, and LBP. The three correlation methods have similar accuracies when the whole images are used; (3) using local patches rather than the whole face image, can improve the recognition accuracies for PLS;

Table 5.6: Face verification accuracies of the PLS method on various features: PCA, HOG, and LBP, respectively, based on the automated facial makeup detection result.

	Face Verification Accuracy with Makeup Detection
PCA+PLS	80.5%
HOG+PLS	80.0%
LBP+PLS	72.0%

5.3.3.3 Combining with Makeup Detection

In our face verification experiments in Section 5.3.3, we assume that the makeup or non-makeup information is known for each face image for PLS method. Based on our makeup detection experiments presented in Chapter 4, we can use the computer recognized makeup or non-makeup information for correlation mapping. If a face is classified as a non-makeup face, it will use the projection matrix corresponding to non-makeup set. Otherwise, the projection matrix for makeup set is used.

Based on the 96.0% accuracy of makeup detection, the face verification accuracies are shown in Table 5.6. The correlation with PLS is applied to different features in local patches, including the PCA, HOG, and LBP features. By comparing Tables 5.6 and 5.5, we found that the face verification accuracies do not drop significantly for the PCA and HOG features, but it does change from 76.0% to 72.0% for the LBP feature. This also indicates that the makeup changes the facial texture which reduces the LBP performance. In contrast, even with the automated makeup detection, our system can still get a good performance, e.g., an 80.5% accuracy using the "PCA+PLS" method, and an 80.0% accuracy based on the "HOG+PLS" method.

As a result, we have shown that our makeup detection approach is effective, and the accuracy of face verification with cosmetic changes can be as high as 80.5%, when proper methods are developed.

5.3.3.4 PLS vs. Dual-attributes

In Section 5.2.4, we propose to use dual-attributes for face verification with makeup changes. In this part, we want to show the comparison between PLS and dual-attributes.

We use ground truth dual-attributes as facial features as introduced in Section 5.2.4. The face verification is 79.5%. This experiment proves the feasibility to use dual-attributes as facial features for face verification to reduce the influence of makeup changes.

In Chapter 5, we study dual-attributes classifiers for each dual-attribute. When we use the predicted dual-attributes information as facial features, the accuracy of face verification is 72%, much lower than using ground truth information. More accuracy dual-attributes classifiers are needed to bring dual-attributes based face verification in real applications.

Besides accuracy, we also record the running time. The same computer as illustrated in Section 3.4.1.2 is used here. The running time includes time of all the training and test steps. The running time of PLS based method is 21.16 seconds, while dual-attributes based method needs 4488.96 seconds (including time of all the training processes for dual-attributes classifiers).

PLS based face verification gives 81.5% face verification accuracy with ground truth makeup information. When using automatic predicted makeup information, the accuracy is 80.5%, much higher than dual-attribute approach. Besides, PLS method costs much less time than dual-attributes approach. So PLS based method is more suitable for real applications.

5.4 Summary

In this Chapter, we develop a dual-attributes based method to explore makeup effects on facial attributes. It provides an easy way to understand and interpret the influences of facial cosmetics on facial identities in a semantic level, which is difficult for previous approaches that are purely based on low-level features. Our study shows that makeup can influence most of facial attributes largely.

To deal with the face recognition with makeup changes, we use PLS method to perform correlation mapping to match makeup and non-makeup faces. Besides, dual-attributes based face recognition method is also proposed to eliminate makeup influence. Finally, we present a complete system that can perform automated face verification utilizing the makeup detection result. A high accuracy of about 80.0% can be obtained for automated face authentication in conjunction with a makeup detection scheme.

Chapter 6

Future Work and Summary

This dissertation deals with facial image analysis for BMI prediction, facial makeup detection, and robust face recognition with makeup and adult BMI changes. The objective to study those relatively new or original topics is to extend the capability of facial image analysis to extract more information from face images. In this Chapter, we discuss future extensions of our work, our contributions and also a summary.

6.1 Future Work

In this thesis, we study some relatively new topics in facial image analysis. In the following, we give some concerns about future work related to our topics.

Our BMI prediction work is the first work to automatically decode BMI information from face images. The insufficiency of our study is that the accuracy of BMI prediction declines when predicting underweight or obese people. The reason is probably due to the lack of enough underweight and obese samples in our database. To apply BMI prediction for real application, it is important to develop BMI prediction robust to all BMI categories.

Dual-attribute is an interesting concept and has been proved effective in decreasing the influence of makeup on face verification. It is also interesting to discover its effect to deal with adult BMI variations. Further, how to develop dual-attributes classifiers with high accuracy is also an important task to develop a robust face recognition system.

Correlation mapping method and dual-attribute based method are used to decrease the influence of makeup and BMI changes on face recognition. Other methods can also be investigated to make the face recognition system more robust to deal with makeup and BMI changes.

Currently, only adults are concerned in face recognition with BMI changes. Body weight changes can be significant during human growing. In addition to normal weight changes along with aging, the overweight and obesity is quite common for adolescents as well. It is also interesting to develop face recognition systems robust to weight changes for adolescents.

There are also other topics can be studied related to current study. e.g., BMI is related to attractiveness and makeup is also used to improve attractiveness. It would be interesting to study the influence of BMI and makeup changes on facial attractiveness.

6.2 Contributions

Overall, this dissertation has contributions as follows,

-construction of automated decoders to interpret BMI, makeup information from face images. Particularly, to the best of our knowledge, it is the first attempt to predict BMI from face images in computer vision.

-exploration of the influence of makeup and BMI variations on face recognition. Dual-attributes are proposed for giving a semantic explanation of the influence of makeup on facial appearance. Besides, experimental results show that only large BMI variations can influence face recognition obviously.

-development of robust face recognition for both makeup and BMI changes separately. Correlation mapping based and dual attributes based approaches are explored to deal with makeup changes, while PLS method is utilized to eliminate influence of BMI changes on face recognition.

6.3 Closing Summary

In this dissertation, we study some newly emerged topics to extend the capability of facial image analysis, including developing facial signal decoders for BMI, makeup information, as well as robust face recognition with facial makeup and adult BMI changes.

At first, we explore BMI prediction from faces. Seven facial ratio and geometry features are extracted to capture BMI cues on the face. After validating correlations between facial features and BMI automatically on a large database, three classic regression methods, i.e. least square estimation, Gaussian processes for regression, support vector regression are utilized to predict the BMI, separately. Experimental results show the feasibility to predict BMI from faces.

Secondly, we conduct facial makeup detection. Four makeup cues, including skin color tone, skin smoothness, texture, and highlight are considered. Makeup detection is considered as a binary classification problem. A discriminative mapping method and patch selection scheme are employed to improve the performance of makeup detection. A makeup detection has been obtained with accuracy as high as 96% which provides us the confidence to use the proposed makeup detection method for real applications.

Thirdly, we study the effect of dual-attributes from faces with and without makeup, separately, to reflect facial appearance changes caused by makeup. Dual-attributes classifiers are developed for each dual-attributes. ACA is operated to divided dual-attributes into four categories: small change, makeup preferred, non-makeup preferred, and big change. We also develop a face recognition system robust to facial makeup changes. We propose two solutions. One is to use dual-attributes as facial features to match faces with and without makeup. The other is to perform correlation mapping between makeup and non-makeup faces on features extracted from local patches. What's more, a complete system is developed for face verification utilizing the makeup detection result. Experimental results show the effect of both dual-attributes and PLS methods on face recognition with makeup changes. PLS based face recognition is feasible for real applications.

Fourthly, we develop a robust face recognition system with adult BMI changes. At first, we study the influence of adult BMI changes on face recognition. The effect to match faces with both small and large BMI changes are evaluated in adult group. PLS method is utilized to improve face recognition with large adult BMI changes. Experimental results demonstrate the feasibility to address the influence of adult BMI variations on face recognition.

Appendix A

Publications

The following list shows published documents of the author.

Lingyun Wen, Xin Li, Guodong Guo and Yu Zhu, “Automated Depression Diagnosis based on Facial Dynamic Analysis and Sparse Coding,” *IEEE Trans. Information Forensics and Security*, pp. 1-10, Mar. 2015.

Lingyun Wen and Guodong Guo, “Dual attributes for face verification robust to facial cosmetics,” *Journal of Computer Vision and Image Processing*, Vol. 3, no. 1, pp. 63-73, Jan. 2013.

Lingyun Wen and Guodong Guo, “A computational approach to body mass index prediction from face images,” *Image and Vision Computing*, Vol. 31, no. 5, pp. 392-400, Mar. 2013.

Guodong Guo, **Lingyun Wen** and Shuicheng Yan, “Face authentication with makeup changes,” *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 24, no. 5, pp. 814-825, May 2014.

Lingyun Wen, Guodong Guo and Xin Li, “A Study on the Influence of Body Weight Changes on Face Recognition,” *International Joint Conference on Biometrics (IJCB)*, Clearwater, Florida, Sep. 2014.

Cesare Ciaccio, **Lingyun Wen**, and Guodong Guo, "Face recognition robust to head pose changes based on the RGB-D sensor," *IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems*, Washington DC, Sep. 29, 2013.

Bibliography

[1] D. Huang, C. Shan, M. Ardebilian, and L. Chen. Facial image analysis based on local binary patterns: A survey. *IEEE Trans. Systems, Man, and Cybernetics, Part C*, 41(6):765–781, 2011.

[2] W. W. Bledsoe. The model method in facial recognition. *Panoramic Research Inc., Palo Alto, CA, Rep. PR1*, 15:47, 1966.

[3] Y. Liu, K. L. Schmidt, J. F. Cohn, and S. Mitra. Facial asymmetry quantification for expression invariant human identification. *Computer Vision and Image Understanding*, 91(1):138–159, 2003.

[4] C. Ciaccio, L. Wen, and G-D. Guo. Face recognition robust to head pose changes based on the rgb-d sensor. In *IEEE Conf. on Biometrics: Theory, Applications and Systems*, pages 1–6, 2013.

[5] A. Sharma and D. W. Jacobs. Bypassing synthesis: Pls for face recognition with pose, low-resolution and sketch. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 593–600, 2011.

[6] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 23(6):643–660, 2001.

[7] S. Z. Li, R. Chu, S. Liao, and L. Zhang. Illumination invariant face recognition using near-infrared images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 29 (4):627–639, 2007.

[8] Y. Liu, K. L. Schmidt, J. F. Cohn, and S. Mitra. Facial asymmetry quantification for expression invariant human identification. *Computer Vision and Image Understanding*, 91(1):138–159, 2003.

[9] A. Bronstein, M. Bronstein, and R. Kimmel. Expression invariant 3d face recognition. *Audio- and Video-Based Biometric Person Authentication*, pages 62–70, 2003.

[10] G-D. Guo, G. Mu, and K. Ricanek. Cross-age face recognition on a very large database: the performance versus age intervals and improvement using soft biometric traits. In *International Conf. on Pattern Recognition*, pages 3392–3395, 2010.

[11] Z. Li, U. Park, and A. K. Jain. A discriminative model for age invariant face recognition. *IEEE Trans. Information Forensics and Security*, 6(3):1028–1037, 2011.

[12] D. Gallagher, M. Visser, D. Sepulveda, Pierson R. N., Harris T., and S. B. Heymsfield. How useful is body mass index for comparison of body fatness across age, sex, and ethnic groups? *Evidence-Based Complementary and Alternative Medicine*, 2011 (749209):228–239, 2011.

[13] A. Keys, F. Fidanza, M. J. Karvonen, N. Kimura, and H. L. Taylor. Indices of relative weight and obesity. *Journal of Chronic Diseases*, 25(6-7):329–343, 1972.

[14] Does size matter in africa? `http://news.bbc.co.uk/2/hi/africa/4566870.stm`, 2006. Accessed Set. 10, 2014.

[15] A. G. Renehan, M. Tyson, M. Egger, R. F. Heller, and M. Zwahlen. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet*, 2008(371):536–546, 2008.

[16] V. Coetzee, D. I. Perrett, and L. D. Stephen. Facial adiposity: A cue to health? *Perception*, 38:1700–1711, 2009.

[17] V. Coetzee, D. I. Perrett, and L. D. Stephen. Deciphering faces: quantifiable visual cues to weight. *Perception*, 39:51–61, 2010.

[18] D. D. Pham, J. H. Do, and B. Ku. Body mass index and facial cues in sasang typology for young and elderly persons. *Evidence-Based Complementary and Alternative Medicine*, 2011(749209):1–9, 2011.

[19] E. A. Yates, A. K. Macpherson, and J. L. Kuk. Secular trends in the diagnosis and treatment of obesity among us adults in the primary care setting. *Obesity*, 20(9): 1909–1914, 2012.

[20] SIRS Issues Researcher. Obesity in america: What's driving the epidemic? *Harvard Men's Health Watch*, pages 1–1, February 2012.

[21] A. G. Renehan, M. Tyson, M. Egger, R. F. Heller, and M. Zwahlen. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *The Lancet*, 371(9612):569–578, 2008.

[22] Andrew Pollack. A.M.A. recognizes obesity as a disease. *The New York Times*, June (19):B1, 2013.

[23] R. Hodara and A. Kacharava. Obesity and cardiovascular disease: Steps to reduce the risk. *Consultant*, 53(2):67–72, 2013.

[24] New survey results indicate there's more to makeup use than meets the eye. `http://www.prnewswire.com/news-releases/new-survey-results-indicate-theres-more-to-` 2012. Accessed Oct. 10, 2014.

[25] History of cosmetics. `http://en.wikipedia.org/wiki/Makeup`. Accessed Oct. 10, 2014.

[26] L. Adkins and R. A. Adkins. *Handbook to Life in Ancient Greece*. Oxford University Press, 1998.

[27] B. Burlando, L. Verotta, L. Cornara, and E. Bottini-Massa. *Herbal Principles in Cosmetics: Properties and Mechanisms of Action*, volume 8. CRC Press LLC, 2010.

[28] R. Corson. Fashions in makeup. *London: Peter Owen*, pages 1–1, 1972.

[29] P. Vitruvius, V. Rose, and H. Müller-Strübing. *De architectura libri decem*. Teubneri, 1867.

[30] M. R Cunningham. Measuring the physical in physical attractiveness: quasi-experiments on the sociobiology of female facial beauty. *Journal of personality and social psychology*, 50(5):925, 1986.

[31] M. R. Cunningham, A. R. Roberts, A. P. Barbee, P. B. Druen, and C-H Wu. " their ideas of beauty are, on the whole, the same as ours": Consistency and variability in the cross-cultural perception of female physical attractiveness. *Journal of Personality and Social Psychology*, 68(2):261, 1995.

[32] J. H. Langlois and L. A Roggman. Attractive faces are only average. *Psychological Science*, 1(2):115–121, 1990.

[33] J. H. Langlois, L. A. Roggman, and L. Musselman. What is average and what is not average about attractive faces? *Psychological Science*, 5(4):214–220, 1994.

[34] C. Darwin. *The Descent of Man*. D. Appleton and Company, 1871.

[35] Nancy Etcoff. *Survival of the Prettiest: The Science of Beauty*. Random House LLC, 2011.

[36] V. S. Johnston and M. Franklin. Is beauty in the eye of the beholder? *Ethology and Sociobiology*, 14(3):183–199, 1993.

[37] P. Aarabi, D. Hughes, K. Mohajer, and M. Emami. The automatic measurement of facial beauty. In *IEEE International Conf. on Systems, Man, and Cybernetics*, volume 4, pages 2644–2647, 2001.

[38] K. Scherbaum, T. Ritschel, M. Hullin, T. Thormählen, V. Blanz, and H-P. Seidel. Computer-suggested facial makeup. *Computer Graphics Forum*, 30(2):485–492, 2011.

[39] N. Guéguen. The effects of women's cosmetics on men's courtship behavior. *North American Journal of Psychology*, 10(1):221–228, 2008.

[40] A. Dantcheva and J. L. Dugelay. Female facial aesthetics based on soft biometrics and photo-quality. In *IEEE Conf. on Institute for Computational and Mathematical Engineering*, pages 1–6, 2011.

[41] G. Rhodes, A. Sumich, and G. Byatt. Are average facial configurations attractive only because of their symmetry? *Psychological Science*, 10(1):52–58, 1999.

[42] S. Ueda and T. Koyama. Influence of make-up on facial recognition. *Perception*, 39 (2):260, 2010.

[43] A. Dantcheva, C. Chen, and A. Ross. Can facial cosmetics affect the matching accuracy of face recognition systems? In *IEEE Conf. on Biometrics: Theory, Applications, and Systems*, pages 391–398, 2012.

[44] L. Wen and G-D. Guo. A computational approach to body mass index prediction from face images. *Image and Vision Computing*, 31(5):392–400, 2013.

[45] D. Cao, C. Chen, D. Adjeroh, and A. Ross. Predicting gender and weight from human metrology using a copula model. In *IEEE Conf. on Biometrics: Theory, Applications, and Systems*, pages 162–169, 2012.

[46] J. L. Rodgers and W. A. Nicewander. Thirteen ways to look at the correlation coefficient. *The American Statistician*, 42(1):59–66, 1988.

[47] R. R. Wilcox. *Fundamentals of Modern Statistical Methods substantially Improving Power and Accuracy*. Springer, 2010.

[48] William. N. *Statistics for Engineers and Scientists 2nd ed.* McGraw-Hill, 2008.

[49] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages I–511–I–518, 2001.

[50] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models-their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[51] S. Milborrow and F. Nicolls. Locating facial features with an extended active shape model. *European Conf. on Computer Vision*, pages 504–513, 2008.

[52] M. H. Mahoor and M. Abdel-Mottaleb. Facial features extraction in color images using enhanced active shape model. In *IEEE International Conf. on Automatic Face and Gesture Recognition*, page 5, 2006.

[53] T. F. Cootes, A. Hill, C. J. Taylor, and J. Haslam. Use of active shape models for locating structures in medical images. *Image and Vision Computing*, 12(6):355–365, 1994.

[54] Stephen M. *Locating Facial Features with Active Shape Models.* University of Cape Town for the Degree of Master of Science, 2007.

[55] A. J. SMOLA and B. SCHÖLKOPF. A tutorial on support vector regression. *Statistics and Computing*, 14:199–222, 2004.

[56] J. Durbin and G. S. Watson. Testing for serial correlation in least square: I. *Biometrika*, 37(3/4):409–428, 1950.

[57] C. Rasmussen and C. Williams. *Gaussian Processes for Machine Learning.* MIT Press, Cambridge, MA, 2006.

[58] K. Ricanek and T. Tesafaye. Morph: A longitudinal image database of normal adult age-progression. In *IEEE International Conf. on Automatic Face and Gesture Recognition*, pages 341–345, 2006.

[59] G-D. Guo and G. Mu. Human age estimation: what is the influence across race and gender? In *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, pages 71–78, 2010.

[60] G-D. Guo, G. Mu, Y. Fu, and T. S. Huang. Human age estimation using bio-inspired features. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 112–119, 2009.

[61] L. A. Zebrowitz. Finally, faces find favor. *Social Cognition*, 24(5):657–701, 2006.

[62] L. Wen, G-D. Guo, and X. Li. A study on the influence of body weight changes on face recognition. In *IEEE International Joint Conf. on Biometrics*, pages 1–6, 2014.

[63] R. Rosipal. Nonlinear partial least squares: An overview. In H. Lodhi and Y. Yamanishi, editors, *Chemoinformatics and Advanced Machine Learning Perspectives: Complex Computational Methods and Collaborative Techniques*, pages 169–189. ACCM, IGI Global, Hershey, PA, USA, 2011.

[64] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.

[65] D. G. Lowe. Object recognition from local scale-invariant features. In *IEEE Cof. on Computer vision*, volume 2, pages 1150–1157, 1999.

[66] J. Luo, Y. Ma, E. Takikawa, S. Lao, M. Kawade, and B-L. Lu. Person-specific sift features for face recognition. In *IEEE Cof. on Acoustics, Speech and Signal Processing*, volume 2, pages II–593, 2007.

[67] M. J. Swain and D. H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.

[68] H. Wold. Path models with latent variables: The nipals approach. In H. M. Blalock, A. Aganbegian, F. M. Borodkin, R. Boudon, and V. Capecchi, editors, *Quantitative Sociology: Internnational perspectives on mathematical and statistical model building*, pages 307–357. Academic Press, 1975.

[69] R. Rosipal and N. Kramer. Overview and recent advances in partial least squares. In C. Saunders, M. Grobelnik, S. Gunn, and J. Shawe-Taylor, editors, *Subspace, Latent Structure and Feature Selection*, pages 34–51. Springer, 2006.

[70] H. Wold. Partial least squares. In S. Kotz and N. Johnson, editors, *Encyclopedia of Statistical Sciences*, volume 6, pages 581–591. Wiley, New York, 1985.

[71] R. Ohtsuki, S. Tominaga, and R. Hikima. Appearance analysis of human skin with cosmetic foundation. In *IS&T/SPIE Electronic Imaging*, pages 82920Q–82920Q, CA, USA, January 2012.

[72] S. Varshovi. Facial makeup detection using hsv color space and texture analysis. Master's thesis, Concordia University, 2012.

[73] C. Chen, A. Dantcheva, and A. Ross. Automatic facial makeup detection with application in face recognition. In *International Conf. on Biometrics*, pages 1–8, 2013.

[74] G-D. Guo, L. Wen, and S. Yan. Face authentication with makeup changes. *IEEE Trans. Circuits and Systems for Video Technology*, 24(5):814–825, 2014.

[75] G. Goldberg. How to apply makeup. `http://www.wikihow.com/Apply-Makeup`. Accessed Oct. 10, 2014.

[76] T. Ahonen, A. Hadid, and M. Pietikainen. Face recognition with local binary patterns. In *European Conf. on Computer Vision*, pages 469–481, 2004.

[77] S. Shafer. Using color to separate reflection components. *Color Research & Application*, 10(4):210–218, 1985.

[78] R. Tan and K. Ikeuchi. Separating reflection components of textured surfaces using a single image. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 27(2):178–193, 2005.

[79] Q. Yang, S. Wang, and N. Ahuja. Real-time specular highlight removal using bilateral filtering. In *European Conf. on Computer Vision*, pages 87–100, Crete, Greece, 2010.

[80] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, and S. Lin. Graph embedding and extensions: A general framework for dimensionality reduction. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 29:40–51, 2007.

[81] V. N. Vapnik. *Statistical Learning Theory*. John Wiley, New York, 1998.

[82] L. Wen and G-D. Guo. Dual attributes for face verification robust to facial cosmetics. *Journal of Computer Vision and Image Processing*, 3(1):63–73, 2013.

[83] C. H. Lampert, H. Nickisch, and S. Harmeling. Learning to detect unseen object classes by between-class attribute transfer. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 951–958, 2009.

[84] A. Farhadi, I. Endres, D. Hoiem, and D. Forsyth. Describing objects by their attributes. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 1778–1785, 2009.

[85] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and simile classifiers for face verification. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 365–372, 2009.

[86] N. Kumar, P. Belhumeur, and S. Nayar. Facetracer: A search engine for large collections of images with faces. *European Conf. on Computer Vision*, pages 340–353, 2008.

[87] B. Siddiquie, R. S. Feris, and L. S. Davis. Image ranking and retrieval based on multi-attribute queries. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 801–808, 2011.

[88] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 886–893, 2005.

[89] H. Hotelling. Relations between two sets of variates. *Biometrika*, 28(3/4):312–377, 1936.

[90] D. Hardoon, S. Szedmak, and J. Shawe-Taylor. Canonical correlation analysis: An overview with application to learning methods. *Neural Comput.*, 16(12):2639–2664, 2004.

[91] D. Hardoon, C. Saunders, S. Szedmak, and J. Shawe-Taylor. A correlation approach for automatic image annotation. In *Advanced Data Mining and Applications*, pages 681–692, 2006.

[92] Kim T-K., Wong S-F., and Cipolla R. Tensor canonical correlation analysis for action classification. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 1–8, 2007.

[93] A. Li, S. Shan, X. Chen, and W. Gao. Maximizing intra-individual correlations for face recognition across pose differences. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 605–611, 2009.

[94] A. Sharma and D. Jacobs. Bypassing synthesis: PLS for face recognition with pose, low-resolution and sketch. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 593–600, 2011.

[95] L. Sun, S. Ji, and J. Ye. Canonical correlation analysis for multilabel classification: A least-squares formulation, extensions, and analysis. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 33:194–200, 2011.

[96] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[97] W. Schwartz, H. Guo, and L. Davis. A robust and scalable approach to face identification. In *European Conf. on Computer Vision*, pages 476–489, Crete, Greece, September 2010.

[98] A. Sharma, M. A. Haj, J. Choi, L. S. Davis, and D. W. Jacobs. Robust pose invariant face recognition using coupled latent space discriminant analysis. *Computer Vision and Image Understanding*, 116(11):1195–1110, 2012.

[99] M. Everingham, J. Sivic, and A. Zisserman. "hello! my name is... buffy" – automatic naming of characters in TV video. In *Proceedings of the British Machine Vision Conference*, 2006.