2011

Gender Classification from Facial Images

Cunjian Chen
West Virginia University

Gender Classification from Facial Images

by

Cunjian Chen

Thesis submitted to the
College of Engineering and Mineral Resources
at West Virginia University
in partial fulfillment of the requirements
for the degree of

Master of Science
in
Computer Science

Don Adjeroh , Ph.D.
Xin Li , Ph.D.
Arun Ross, Ph.D., Chair

Lane Department of Computer Science and Electrical Engineering

Morgantown, West Virginia
2011

Keywords: Gender Classification, Face Images, Cross-Spectrum, Facial Attributes

Abstract

Gender Classification from Facial Images

by

Cunjian Chen

Gender classification based on facial images has received increased attention in the computer vision community. In this work, a comprehensive evaluation of state-of-the-art gender classification methods is carried out on publicly available databases and extended to real-life face images, where face detection and face normalization are essential for the success of the system. Next, the possibility of predicting gender from face images acquired in the near-infrared spectrum (NIR) is explored. In this regard, the following two questions are addressed: (a) Can gender be predicted from NIR face images; and (b) Can a gender predictor learned using visible (VIS) images operate successfully on NIR images and vice-versa? The experimental results suggest that NIR face images do have some discriminatory information pertaining to gender, although the degree of discrimination is noticeably lower than that of VIS images. Further, the use of an illumination normalization routine may be essential for facilitating cross-spectral gender prediction. By formulating the problem of gender classification in the framework of both visible and near-infrared images, the guidelines for performing gender classification in a real-world scenario is provided, along with the strengths and weaknesses of each methodology. Finally, the general problem of attribute classification is addressed, where features such as expression, age and ethnicity are derived from a face image.

Acknowledgements

I would first like to thank my committee chair and advisor, Dr. Arun Ross, for giving me the opportunity to work with him and his students. This thesis would not be possible without his constant guidance and steadfast support. His passion and hard-work for the research is always an inspiration source and influence to me. He is also an outstanding professor in teaching, from which I have benefited a lot in his classes.

I would also like to thank Dr. Donald Adjeroh and Dr. Xin Li for being on my committee. Dr. Adjeroh is always very kind to students and tries his best to help them when they are in need. His rigorous attitude toward research always touches me. Dr. Li has also provided lots of very useful advice for my research work. I have taken two classes from him, whom I shall remember for his passion in the class and valuable insights in the research work.

My sincere thanks also goes to Dr. Guodong Guo, Dr. Tim McGraw, for offering me the discussion opportunities either in the projects or classes. Special thanks goes to Dr. Matthew Valenti who has kindly provided the thesis template.

In my daily work I have been blessed with a friendly and cheerful group of fellow students. Rui Guo, as well as Deng Cao have suggested useful ideas related to my research. I also appreciate my colleagues in the lab, such as Brian DeCann and Raghunandan Pasula, and Manisha SamSunder who provided useful suggestions on the thesis work or the discussion in relevant work.

Last but not the least, I wish to thank my parents, my brother, and all other relatives. They raised me, supported me, taught me, and loved me. To them I dedicate this thesis. I also want to thank other people who are part of my life and help me transit in USA. They make it a wonderful time for me to study and live here.

Contents

List of Figures

List of Tables

Notation

We use the following notation and symbols throughout this thesis.

x_i	:	Vector representation of an image i
$\hat{x}$	:	Mean vector representation
$(\cdot)^T$	:	Matrix transpose
$\sum_g$	:	Covariance Matrix
$\langle \cdot \rangle$	:	Dot product of two vectors
$sign(\cdot)$	:	Sign function
$\| \cdot \|$	:	Euclidian norm
$\Re\{\cdot\}$	:	Real part of the argument
$\Im\{\cdot\}$	:	Imaginary part of the argument
$\mathcal{N}$	:	Gaussian Distribution

Bold upper case letters denote matrices and bold lower case letters denote vectors.

Chapter 1

Introduction

1.1 Background

Gender classification plays an important role in Human-Computer Interaction (HCI), upon which more complex visual systems are built [10]. Recognizing a person's gender will enhance the HCI's ability to respond in a user-friendly and socially acceptable manner. In the realm of biometrics, gender is viewed as a soft biometric trait that can be used to index databases or enhance the recognition accuracy of primary traits such as face [11]. Moreover, gender classification is also an essential part in automatically labeling images with demographic attributes such as ethnicity, gender, and others. Apart from the research work done in computer vision, psychologists are particularly interested in how humans perceive the gender from face images [12].

Gender classification is a fundamental task for human beings, as many social activities depend on the precise identification of gender. In this work, gender identification is considered as a binary classification problem: male or female. Often times, the way humans perceive gender does not only rely upon the perception of the face region, but also on the surrounding context, such as hair, dress and skin tone [13]. The problem of predicting the gender from face images is the scope of our study. It is possible that the hair information might also be included in the face region, but the majority of the information presented are the facial features, such as eyes, nose, mouth and cheeks. A recent work in [14] discusses an interesting topic of changing genders, but preserving the biometric identity of the individu-

als. Such gender conversion suggests that the perception of gender is a mixture of factors related to the skin, hairstyle, facial components and facial hair.

1.2 Relevant Work

The study of automatic gender classification from face images dates back to the early 1990s and is one of the recent hot topics in studying facial attributes. Most techniques for gender classification approach the problem from the perspective of machine learning, as it is essentially a two-class classification problem.

Golomb et al. [15] trained a back-propagation neural network (BPNN) to identify gender from human face images at a resolution of 30×30 pixels. An average classification rate of 91.9% on 90 exemplars was obtained compared to a human performance of 88.4%.

Gutta et al. [16] used hybrid classifiers consisting of an ensemble of radial basis function (RBF) networks and decision trees. The experiments were conducted on a collection of 3006 face images corresponding to 1009 subjects from the FERET database. The cross-validation results yielded an average accuracy of 96% on the gender classification task.

Later on, Moghaddam et al. [17] utilized a support vector machine (SVM) for gender classification, based on low-resolution thumbnail face images of resolution 21×12. The average rate for five-fold cross-validation on 1755 FERET face images was 96.62% with the use of the Gaussian RBF kernel. They also compared their technique against other classifiers such as RBF neutral network, Fisher Linear Discriminant (FLD) and Bayesian classifier. Among all the methods, the SVM classifier gave the best performance. Their work also pointed out that the SVM classification of low-resolution face images was very effective, compared to other methods.

Baluja et al. [18] presented the use of Adaboost classifier to identify the gender of a person from a low-resolution face image. The proposed system was extremely fast and yet comparable to the SVM-based classifier. They reported an accuracy over 93% on a dataset of 2409 FERET faces images with resolution 20×20 pixels.

Although the use of low-resolution face images in [15, 17, 18] results in very good performance for facial gender classification, it is a simple feature representation that may not be

robust enough in some complex scenarios involving pose and illumination changes. Therefore, other types of feature extraction methods have also been proposed.

BenAbdelkader et al. [19] presented an appearance-based method for gender classification based on features extracted from local regions. The matching was performed on local regions with the commercial FaceIt software. 94.2% was the best performance achieved on a database of approximately 13,000 near-frontal images using the SVM classification methods.

Recently, with the popularity of Local Binary Patterns (LBP) for face recognition [20], Yang et al [21] used the LBP histogram features for gender feature representation, and the real adaboost algorithm to learn the best local features for classification. Experiments were performed to predict the age, gender and ethnicity information from face images. A similar work was presented in [22], where LBP features and Adaboost classifier were combined to achieve better performance.

Local based descriptors have also been adopted in the work of gender classification. For example, Guo et al. [23] evaluated gender classification results based on LBP, histograms of oriented gradients (HOG), and Biologically-Inspired Features (BIF) with SVM as the classifier. It was demonstrated that gender prediction was affected by age variations on a large database. Wang et al [24] proposed a novel gender recognition method in terms of the Scale Invariant Feature Transform (SIFT) descriptor and shape contexts. Again, Adaboost was used to select features from face images to form the strong classifier. Other approaches utilized gender-specific information, such as hair, to enhance gender prediction [13], or genetic algorithms to select features that encoded gender information [25].

All the aforementioned work mainly focus on datasets that were collected under well-constrained environments. Recently, gender classification on unconstrained real-world face images has been attempted. Chen et al. [26] built a gender classification system on real-world face images where the decision is based on the surrounding regions from face detection and the associated context-regions. Shan [27] proposed to use the boosted LBP features to represent face images and applied SVM to determine the gender on the Labeled Faces in the Wild (LFW) dataset. They obtained a performance of 94.44% on a dataset of 7,443 face images. Gallagher et.al [1] used social context information in real-world group images to accomplish gender classificaiton. They argued that the struture information within the group provides

meaningful context for individuals. For example, men were more likely to stand at the corner of an image than women. Gao et.al [28] targeted face gender classification on consumer images in a multiethnic enviroment. To overcome the non-uniformity of pose, expression, and illumination changes, they proposed a robust Active Shape Model (ASM) to normalize the face texture. The consideration of ethnic factors can help improve gender classification accuracy in a multiethnic enviroment. Recently, Toews [29] extended gender classificaiton to arbitrary viewpoints and under occlusions. A viewpoint-invariant appearance model was learned for the object class and a bayesian classifier was trained to identify the model features that indicate gender. In the work of [30], images with unconstrained pose, expression and light conditions were considered for gender classification based on additive logistic models.

In principle, a gender classification method can be divided into two components: (a) a feature extractor that extracts features from the face and (b) a feature classifier that assigns the extracted features into one of two classes - male or female. Feature extraction methods include the use of low resolution face images [17, 18], Principle Component Analysis (PCA) [31], Linear Discriminant Analysis (LDA) [32], Independent Component Analysis (ICA) [33] and LBP [21, 27, 22]. Some feature selection algorithms [34] have also been used to select gender specific features. Most gender classifiers are based on Neural Network [15, 34], Adaboost [18, 21, 22, 27], Gaussian Process (GP) classifier [35] and SVM [17, 19, 23]. A systematic overview of methods for gender classification from face images in the visible spectrum can be found in [10].

The overview of gender classification methods and their accuracies are summarized in Table 1.1. It gives a brief summary of different algorithms used in the past. Only those algorithms that predicted the genders from facial images are selected, and not those based on body [36, 37] or gait [38]. Furthermore, gender prediction from speech [39] has also not been included for comparison. The list of datasets used varies from one work to the other. The authors may have presented numerous results on multiple datasets in a single paper, but only one of them is listed. Based on the information presented, it is easy to trace the trend of gender features and classifiers used in the literature.

Table 1.1: Overview of recent studies on gender classification.

Study	**Features**	**Classifier**	**Name, Size**	**Perf.**
1990 [15]	Raw pixels	Neural Network	Private, 90	91.9%
1998 [16]	Raw pixels	Hybrid Classifier	FERET, 3006	96%
2002 [17]	Raw pixels	SVM	FERET, 1755	96.62%
2004 [33]	ICA	LDA	FERET, 500	99.3%
2005 [19]	Local features	SVM	Identix, 13,000	94.2%
2005 [34]	PCA	Neural Network	Private, 400	88.7%
2006 [22]	LBP	Adaboost	FERET, 2000	95.75%
2006 [35]	Raw pixels	GPC	AR, 515	97%
2007 [18]	Raw pixels	Adaboost	FERET, 2409	93%
2007 [21]	LBP	Adaboost	Private, 3540	96.32%
2008 [40]	LBP,Gabor	SVM	CAS-PEAL, 10,784	93.74%
2009 [28]	ASM	Adaboost	Private, 1300	92.89%
2009 [29]	SIFT	Bayesian	FERET, 994	83.7%
2010 [27]	LBP	Adaboost	LFW, 7,443	94.44%
2010 [23]	LBP,HOG, BIF	SVM	YGA, 8,000	89.28%
2010 [24]	SIFT,Context	Adaboost	FERET, 2409	95%
2011 [32]	PCA	LDA	FERET, 994	93.33%

1.3 Thesis Outline

Before we delve into gender classification, some preprocessing steps such as face detection and face normalization are necessary. These are important components of a gender classification system deployed in practical applications. Without reliable output from face detection, results from gender prediction would become meaningless. For more detailed discussions concerning face detection, the reader is encouraged to refer to [41].

In Chapter 2, the key concepts in gender classification are reviewed, including face detection, face normalization, feature extraction and classification. The gender feature representation methods mainly considered in this work include PCA, LDA, GaborFace and LBP. PCA is a very useful dimension reduction tool, which can be applied to generate a compact feature descriptor. Such a dimensionality reduction technique retains only the essential information that is useful for gender recognition while reducing the computation cost. Due to less discrimination power of PCA used for classification, LDA is introduced to overcome the limitations of PCA. Then, Gabor and LBP descriptors are introduced to represent the gender features. LBP was previously investigated in the work of [21, 27, 22], where LBP histogram (LBPH) features derived from local regions of face images were concatenated to form the holistic feature vector. Among the variants of LBP-based descriptors, Multi-block based LBP (MBLBP) is chosen to extract the gender features. It has been shown that such a LBP descriptor gives very good performance in discriminating gender information. Therefore, it is particularly interesting to determine how to use the LBP features in order to improve gender classification performance. Apart from LBP, Gabor descriptor is another local feature descriptor used in our work. Therefore, both global descriptors (PCA and LDA) and local descriptors (LBP and Gabor) are tested in the framework of gender feature extraction.

Recalling the gender classification work from [17], it has been shown that the SVM-based gender classifier can achieve very good results. Therefore, we adopt SVM classifier in most of our tasks, with the emphasis on the selection of kernel and parameters for optimization. We also introduce other type of classifiers such as Adaboost and FLD to perform comparison. The strength and weakness of each methodology is evaluated through numerous experiments on various databases. Such a comparison is necessary to understand the performance of

gender classification under different scenarios thereby providing guidelines for future work.

In Chapter 3, we extend the gender classification work from visible (VIS) images to near-infrared (NIR) and thermal (THM) images. It is demonstrated that gender classification can be successful in NIR and THM spectra. Furthermore, we study the possibility of cross-spectrum gender classification, where the trained classifier is in one domain (e.g., VIS) and the test samples are from another (e.g., NIR). Due to the fact that appearance information presented in multiple spectra are contrastingly different, an illumination normalization approach is adopted to reduce the difference between those two spectra. Similar to gender classification in the visible domain, we prefer the SVM classifier [17] to predict gender from NIR spectrum or THM spectrum images. The performance is evaluated on public databases and the experiments demonstrate promising results.

As has been mentioned, gender feature is one of several visual traits that can be observed from face images. Other attributes such as ethnicity, age and expression can also be perceived from the human face to a certain degree. The current work is extended from gender classification to attributes-based classification in Chapter 4.

Finally in Chapter 5, the main contributions of the thesis are reviewed, as we conclude this thesis work. Future work for potential improvement of system's performance is also discussed.

Chapter 2

Gender Classification

Gender classification is a fundamental task for human beings, as many social activities depend on the successful perception of gender information. In demographic data collection applications, information such as gender requires accurate gender identification. Automatic gender classification is also a useful preprocessing step for face recognition since it is possible to reduce the number of potential face candidates. Many studies have been proposed to address the gender identification problem. But it is still unclear what kind of gender features are useful for discrimination. Besides, there are few benchmark datasets that have been used to compare different approaches. In this chapter, we systematically compare various gender feature extraction methods after using different classifiers on numerous publicly available datasets. The structure of the whole chapter can be viewed in the system flowchart (Figure 2.1). In our research we have investigated several different feature extraction methods and discriminant classifiers. Developed framework consists of two parts: automatic face detection on images or video and applying of a gender classification algorithm to detected faces.

In this chapter, automatic gender classification is divided into different parts in order

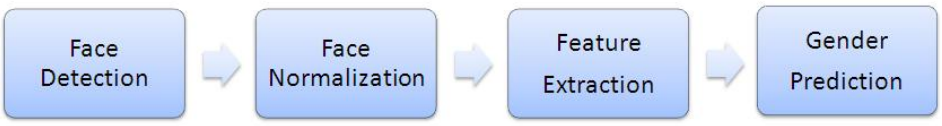


Figure 2.1: Overview of the system flowchart.

to provide detail analysis for each component. In Section 2.1 different approaches for face detection and facial feature localization are studied. After briefly discussing the face detector, the importance of face normalization based on the detected feature points is emphasized. This essentially reduce the intra-class variations between samples and potentially improve the classification accuracy. In Section 2.2, the feature extraction and classification methods are introduced. During the feature extraction procedure, a compact feature descriptor is derived from the normalized face image and then fed into the classifier to predict the gender. Finally a summary of the chapter will be given in section 2.5.

2.1 Face Detection

Face detection has a wide range of applications such as automatic face recognition, face tracking, and surveillance. It is also a very critical pre-step for automatic gender classification. Most of the current face detection algorithms treat this task as a two-class (face/non-face) classification problem and employ neural-network based methods [42], support vector machines [43], and adaboost [44]. The final accuracy of gender classification depends heavily on the output from the face detector. This means that we need to locate all the faces in the images while reducing false positives.

However, face detection is still a challenging problem due to the large variations in the face images. The variations associated with face are due to pose, expression and lighting changes. Among these factors, pose and lighting account for most of the failures in detection. Another issue is partial occlusion, caused either by another face or by other objects presented in the image. The image conditions can also vary. Some images are taken in an indoor environment (Figure 2.2), while others might be captured in an outdoor environment (Figure 2.4). The quality and resolution of the images all contribute to the complexity of face detection (Figure 2.3). The searching of potentially large candidate regions makes the algorithm time consuming and unsuitable for real-time application.

Thanks to the cascaded face detector proposed by Viola and Jones [44], face detection can be done accurately in real-time. Cascaded face detector searches for faces using a sub-window approach and each sub-image is passed to the layers of classifiers to determine whether the

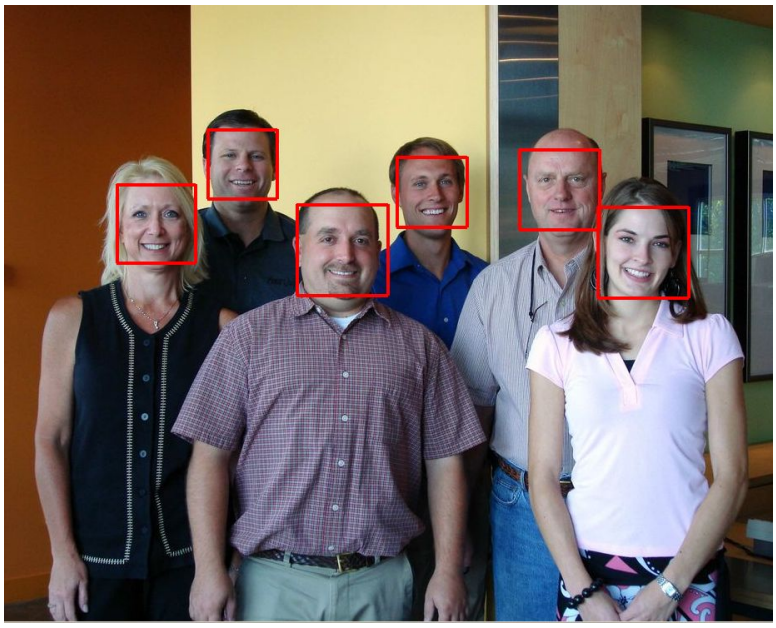


Figure 2.2: An example of face detection using the Adaboost method from OpenCV; all of the faces are correctly detected. Image is from a dataset collected by [1].

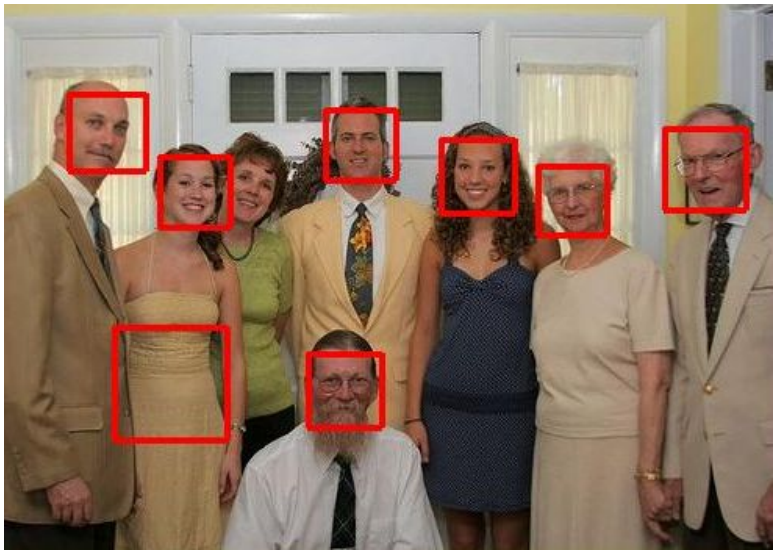


Figure 2.3: Face detection using Adaboost: one false positive and one false negative. Image is from a dataset collected by [1].

region contains a face or not. If the sub-image is successfully classified as a face by all the classifiers then the face detector will claim that this sub-image contains the face [10]. The survey of face detection methods can be found in [45]. The Adaboost classifier based on Haar-like features generally gives very good performance and is adopted in the automatic

Figure 2.4: Face detection in a complex environment. Image is from a dataset collected by [1].

gender classification system.

2.1.1 Feature Localization

The output of face detector only gives the coarse location of the face image. It does not specify the feature locations of eyes, mouth and nose. To locate the features present in the face region, a simple way is to train individual classifiers for locating the eye, mouth and nose region inside the detected face. But this still has the same limitation as the face detector. Other approaches utilize the relationship or structure between facial features to build a generative model [46]. The probability distribution of the joint feature positions is modeled as a mixture of Gaussians. The appearance of each facial feature is assumed to be independent of the other and a discriminant classifier is trained to determine each feature position using Haar-like features.

Another recent approach is to use Active Appearance Model (AAM) [47, 48] or Active Shape Models (ASM) [3] to localize the facial features. The shape of a face image is represented by v vertices that define a mesh,

$$s = [(x_1, y_1), (x_2, y_2), \ldots, (x_v, y_v)]^T. \tag{2.1}$$

AAM imposes linear constrains on shape variation, and so an input shape can be represented

as the linear combination of N base shapes,

$$s = s_0 + \sum_{i=1}^{N} p_i s_i. \tag{2.2}$$

Here, s_0 is the mean shape, s_i is the i^{th} base shape, and p_i is the corresponding weight vector for this shape. The texture is defined as the pixel intensities that are within the shape boundary. It can be defined as a vector of intensities $A(x)$:

$$A(x) = A_0(x) + \sum_{i=1}^{M} \lambda_i A_i(x), \tag{2.3}$$

where $A_0(x)$ is the mean texture and $A_i(x)$ is the i^{th} texture vector. Unlike AAM, ASM seeks to only match the positions of the feature points, although some models may incorporate the texture information. Such a model is usually referred to as constrained AAM. The AAM fitting problem is usually defined by a cost function, which tries to minimize the following:

$$r(p) = (A_i(x) - A_m(x))^T (A_i(x) - A_m(x)). \tag{2.4}$$

This classical optimization problem can be solved in an iterative way. Matthews and Bakers [48] proposed a popular AAM fitting method within the framework of the Lucas-Kanade algorithm. But it cannot generalize well to unseen subjects for locating feature points. Sometimes, the notations of AAM and ASM are exchangeable as they might be used in different scenarios. An example of ASM fitting for localization of facial landmarks is shown in Figure 2.5. We use the Stasm library [3] for the ASM fitting. Usually, the face detector is invoked first to provide the coarse location for the initialization of AAM, then the model would fit onto the face images until convergence condition is satisfied. Currently, the search for feature points using AAM is not accurate for face images with large pose changes. As our study is mainly constrained to near-frontal face images, the AAM can localize the features with very accurate results.

2.1.2 Face Normalization

Once the localization of face region and facial features are completed, it is necessary to normalize the face images based on both geometry and appearance. In case of face images

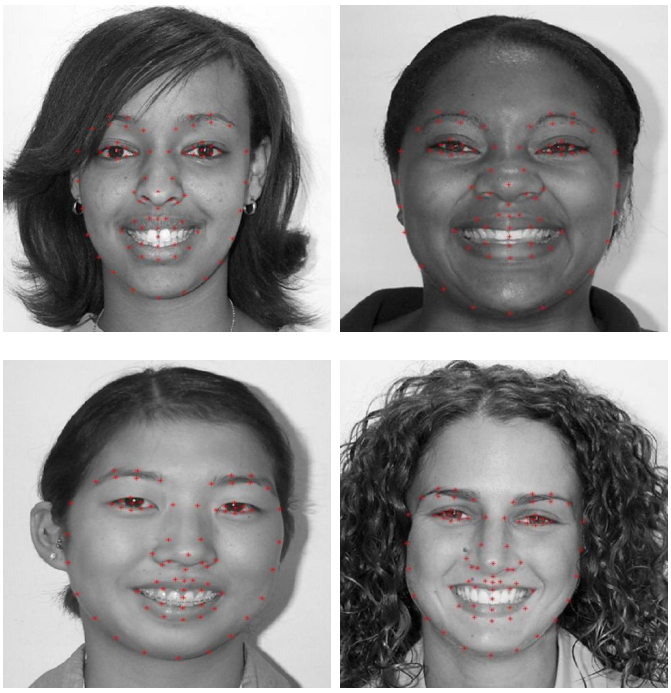


Figure 2.5: ASM used to locate various features in face images. Images are from lifespan database [2]. The output is based on the algorithms described in [3].

with certain pose and illumination changes, such a preprocessing step has been proved to be effective for the succeeding tasks [10], viz., face recognition and gender classification. The geometry of the face images are normalized by registering the images according to the detected features such as eye coordinates. Some researchers also use other features such as mouth. The appearance normalization can be achieved by applying histogram equalization or other illumination normalization approaches.

We follow the normalization approach proposed by Bolme [4]:

- Geometric normalization: the image is scaled so as to make the distance between the eyes constant or fixed. The standard FERET normalization approach crops the image to the size of 150x130 pixels with 70 pixels between the centers of the two eyes.

- Masking: a mask is applied in order to zero out pixels not in an oval that contains the typical face region. Thus, hair, shirt collars, etc. are usually removed. It can be implemented through ellipse fitting.

- Histogram equalization: equalization is used to smooth the distribution of grey scale values for the non-masked pixels. That can help alleviate the impact of illumination changes.

- Pixel Normalization: the image is normalized so that the non-masked pixels have mean zero and standard deviation one.

As shown, the normalized samples (Figure 2.6(b)) possess far less variations across the database than the original unprocessed samples (Figure 2.6(a)). The application of mask on the face images eliminates some unnecessary background information. But in some cases, part of the hair information is still retained.

(a) (b)

Figure 2.6: (a). Original samples from FERET dataset; (b). Corresponding normalized images from FERET dataset. The output is based on the algorithms described in [4].

Another much simpler face alignment method is to normalize the face images based on the eye coordinates, such that the two eye centers of all the face images are at fixed positions after translation, rotation and scaling operations. This normalization approach is described in [10] and briefly introduced below:

- Locate the center positions of two eyes, either manually or automatically.
- Rotate the image so that the eyes are vertically aligned. The angle of rotation is calculated according to the eye positions.
- Calculate the Euclidean distance d_0 between the eyes in the rotated image.
- Calculate the ratio $r = d_0/d_t$, where d_t is the distance of the eyes in the resized image.
- Compute the width w_0 and height h_0 around the areas of eyes as $w_0 = r * w_t$ and $h_0 = r * h_t$, where w_t and h_t are the width and height of the resized image.
- Compute the coordinates for the corners of the face area in the rotated image.

The resized image is fixed to a specific dimension such as 128×128 or 64×64. After the alignment of images based on eye coordinates, all the resultant samples will have the same image size. The distance between the eye coordinates are fixed and placed in the same position. Such a geometric-based normalization does not account for illumination changes and may include some background information. The choice of normalization depends on the specific application and the feature extraction method and classifier used. For instance, in the task of age estimation, where the texture information is affected by histogram equalization, a simple alignment approach is preferred.

One limitation about the two aforementioned methods is that they do not account for pose changes. In other words, the pose correction procedure is not included. The AAM method [1] described in section 2.1.1 can be used to normalize the pose, apart from feature localization [49]. We can perform Delaunay triangulation and piece-wise affine warping to bring the arbitrary pose to a neutral pose. An example is shown in Figure 2.7. This normalization is particularly useful when the input face images have large pose variations.

2.2 Proposed Methods

The previous sections mainly focus on the detection and normalization of face images. In this section, different feature extraction and classification methods are discussed for gender

[1] AAM-API: http://www2.imm.dtu.dk/ aam/aamapi/

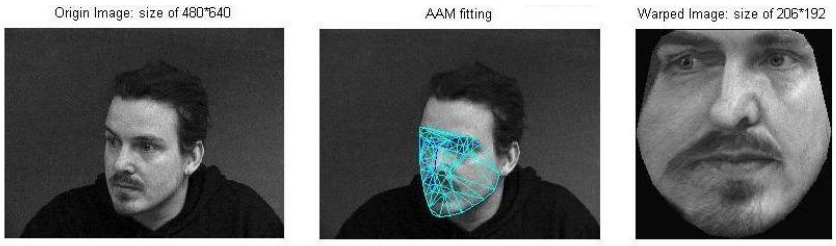


Figure 2.7: Pose normalization by AAM. Image is from IMM face database [5].

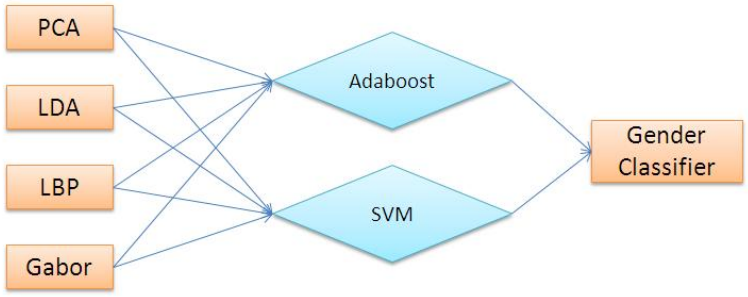


Figure 2.8: Different possible combinations of gender classification methods.

recognition on normalized images. The well-known Eigenface, Fisherface, and Gaborface methods for gender feature representation are introduced, in addition to Local Binary Pattern (LBP) descriptor. The classifiers that are used include Adaboost and SVM. That could provides us with many different combinations (Figure 2.8). Some other classifiers would also be introduced in the later chapter. Since we approach gender classification from a machine learning perspective, the appearance-based methods are expected to be suitable for this task. The learning procedure aims to automatically learn discriminative features for gender representation based on a pool of features, and seeks to compute the decision boundary which can separate the male and female class.

2.2.1 Principle Component Analysis

Previous work on gender classification in the visible domain utilized features extracted via PCA [17, 50, 25] or Haar-like features [18]. In this work, we use the PCA features since it has been successfully used in previous literature. Consider a labeled set of N training samples $\{(x_i, y_i)\}_{i=1}^{N}$, where x_i is the facial image and y_i is the associated class label. Here,

$y_i \in \{-1, 1\}$, where a -1 (+1) indicates a female (male). The PCA is performed on the covariance matrix of vectorized images.

$$\Sigma_g = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(x_i - \bar{x})^T \tag{2.5}$$

where x_i is the sample image after vectorization and $\bar{x}$ is the mean vector of the training set. The eigenvectors, known as eigen faces, can be obtained through the decomposition,

$$\Sigma_g \Phi_g = \Phi_g \Lambda_g \tag{2.6}$$

where Φ_g are the eigenvectors and Λ_g are the corresponding eigenvalues. The gender features can be extracted by projecting the sample image onto the subspace expanded by eigenvectors:

$$s_i = \Phi_g^T(x_i - \bar{x}) \tag{2.7}$$

where s_i is the feature vector to represent the gender information of sample x_i. The feature vectors corresponding to the training set and their label information $\{s_i, y_i\}$ are stored in the database. In the testing stage, when an unknown facial image is presented, the same feature extractor is invoked to obtain the feature set, which is fed into a classifier G to predict the gender. Examples of eigen faces are shown in Figure 2.9. One of the most important parameters in the PCA analysis is the number of eigvectors. Generally, we select the top K eigenvectors to construct the subspace that includes both male and female information based on the ordering of eigenvalues (Figure 2.10). The value K can be set to 60, for instance.

2.2.2 Fisherface and Gaborface

Fisherface takes advantage of the fact that the within class variation lies in a linear subspace that is convex and separable [51]. It models the between-class scatter matrix as,

$$S_B = \sum_{i=1}^{c} N_i(\mu_i - \mu)(\mu_i - \mu)^T \tag{2.8}$$

and the within-class scatter matrix as,

$$S_W = \sum_{i=1}^{c}\sum_{x_k \in X_i}(x_k - \mu_i)(x_k - \mu_i)^T \tag{2.9}$$

Figure 2.9: Eigenface computed from a training set from the FERET database.

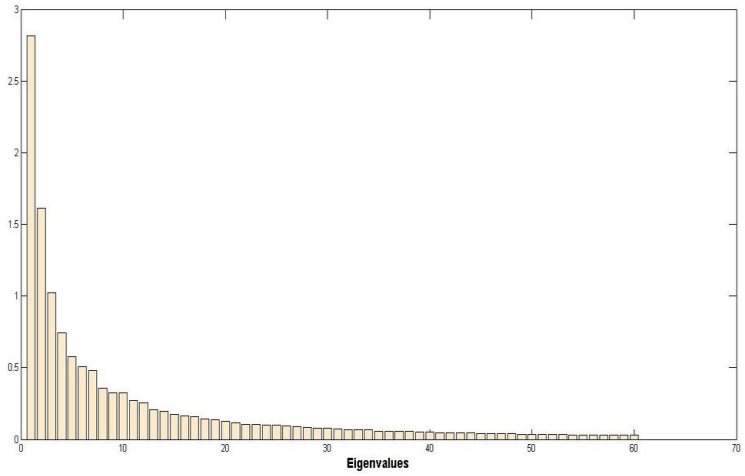


Figure 2.10: Distribution of EigenValues, from largest to smallest.

where $\mu_i = \frac{1}{N_i}\sum_{i=1}^{N_i} x_i$ is the mean sample of class X_i and N_i is the number of samples in class X_i. The summation of $N = \sum_{i=1}^{c} N_i$ is the total number of samples for the dataset. The objective function is to maximize the between-class scatter matrix S_B while minimizing the within-class scatter matrix S_W, i.e.,

$$W_{sub} = \arg\max_{W} \frac{|W^T S_B W|}{|W^T S_W W|}, \tag{2.10}$$

where the resulting subspace W_{sub} is obtained by solving the generalized eigenvectors problem,

$$S_B w_i = \lambda_i S_W w_i \tag{2.11}$$

Therefore, $W_{sub} = \{w_i | i = 1, 2, \cdots, m\}$ and m is the number of eigenvalues. An upper bound on m is $c-1$, where c is the number of classes. For the binary classification problem ($c = 2$), the number of nonzero eigenvalues is 1. However, for the Eigenface methods, there is no such restriction on the selection of available eigenvectors. It is very common to have fewer sample vectors than features (pixels). Therefore, the within-class scatter matrix S_W can be singular and the LDA projection matrix W_{sub} cannot be computed directly. Instead, PCA is often applied first to retain eigenvectors with nonzero eigenvalues and then LDA is applied to the reduced transformation space. Such a combination of PCA and LDA is often termed as Fisherface. The merits of each method and their potential applications can be found in [51].

Apart from the work on Eigenface and Fisherface, Gaborface [52] provides another way for encoding facial features via texture information. Gabor wavelets [2] have been extensively investigated in face recognition [52] and expression recognition [53] due to its optimal localization properties in both spatial and frequency domain, similar to the 2D receptive field profiles of the mammalian cortical simple cells [52]. The basic idea is to decompose image into multiple scales and orientations to capture texture information. The gabor wavelets are defined as follows [52],

$$\varphi_{\mu,\upsilon}(z) = \frac{||k_{\mu,\upsilon}||}{\sigma^2} e^{-\frac{||k_{\mu,\upsilon}||^2}{||z||^2}} e^{ik_{\mu,\upsilon}z - e^{-\frac{\sigma^2}{2}}} \tag{2.12}$$

where μ and υ denote the orientation and scale of the Gabor kernels. The wave vector $k_{\mu,\upsilon}$ is given by,

$$k_{\mu,\upsilon} = k_\upsilon e^{i\phi_\mu} \tag{2.13}$$

where $k_\upsilon = k_{max}/f^\upsilon$ and $\phi_\mu = \pi\mu/8$. Here, k_{max} is the maximum frequency and f is the spacing factor between kernels in the frequency domain. In real applications, the parameters

[2]Gabor wavelet: http://www2.it.lut.fi/project/simplegabor/

of μ and υ are chosen as 8 orientations and 5 scales, resulting in total of 40 images of Gabor response. To encompass all the features produced by Gabor kernels, one can choose to apply feature extraction methods to augment the Gabor feature vector. Common feature extraction methods like Eigenface and Fisherface can be applied. The output of the Gabor response is complex. Usually, the magnitude representation of the Gabor response is selected (Figure 2.11).

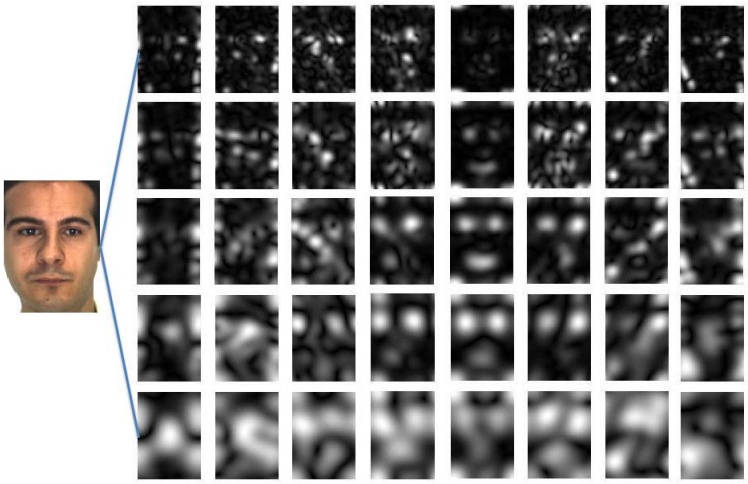


Figure 2.11: Gabor magnitude representation of a face image. Image is from AR face database [6].

2.2.3 Local Binary Pattern

Another promising feature extraction method is the Local Binary Pattern [3] texture descriptor which shows good discrimination power for many face-related applications. LBP [54] are features calculated from the pixel intensities within a pixel neighborhood. Ahonen et al. [55] first extended the research work to face recognition and demonstrated that this local feature descriptor was very efficient in face representation. After that, numerous algorithms based on LBP have been proposed. Shan [56] adopted LBP to solve the facial expression recognition problem. Furthermore, it has also been demonstrated that illumina-

[3]LBP implementation: http://www.cse.oulu.fi/MVG/Downloads/LBPMatlab

tion invariant feature face representation can be obtained by integrating the techniques of LBP in [57, 58]. In addition, LBP in [59] is applied to estimate the head pose. Due to the fact that multi-resolution technique is a very useful tool to analyze images at different scales and orientations, the face image is modeled as a concatenation of the histograms of all the local regions derived from local Gabor magnitude [60].

The LBP operator was first introduced as a texture descriptor that computes patterns in an image by thresholding 3×3 neighborhoods based on the value of the center pixel, and then converting the resulting binary pattern into a decimal value. Later, it was extended to include neighborhoods of different sizes to account for textures at different scales.

The local neighborhood is defined as a set of sampling points evenly spaced on a circle. The LBP operator is described as $LBP^{u^2}_{P,R}$, where P refers to the number of sampling points placed on a circle with radius R. The symbol u^2 represents the uniform pattern which, in our case, refers to those binary patterns that have at most two bitwise transitions from 0 to 1 or 1 to 0. For instance, 10011111 is a uniform binary pattern while 10100111 is not. Uniformity is an important concept as it characterizes micro-features (structural information) such as lines, edges and corners in the image. Although only 58 out of the 256 8-bit patterns are uniform, nearly 90% of all observed image neighbourhoods are uniform [61]. We chose to use $LBP^{u^2}_{8,1}$ in all our experiments based on empirical evidence. The binary pattern for pixels lying in a circle (f_p, $p = 0, 1, \ldots, P-1$) with the center pixel f_c, is computed as follows:

$$S(f_p - f_c) = \begin{cases} 1 & \text{if } f_p - f_c \geq 0; \\ 0 & \text{if } f_p - f_c < 0. \end{cases} \tag{2.14}$$

Then a binomial weight 2^P is assigned to each sign $S(f_p - f_c)$ to compute the LBP code,

$$LBP_{P,R} = \sum_{p=0}^{P-1} S(f_p - f_c)2^P. \tag{2.15}$$

Our approach using LBP is described as in Figure 2.12. The original image is first divided into non-overlapping small blocks, and then the LBP histogram is computed for each block. After deriving the histogram sequence for each block, the final global representation is obtained by concatenating the individual sequences. This is not the only way to extract the histogram features from image. It is possible to change the assumption of non-overlapping blocks to dense sampling of blocks and applying Adaboost method to select LBPH features [21].

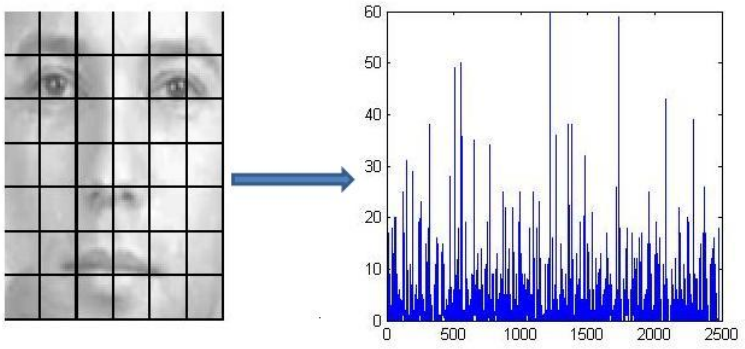


Figure 2.12: Multiblock-based LBP representation for a face image. Image is from FERET database [7].

2.2.4 Support Vector Machine

A Support Vector Machine (SVM) [4] is a machine learning technique used for pattern classification and regression analysis. It is based on the concept of searching linear boundary between two classes of patterns as follows,

$$y(x) = w^T \phi(x) + b \tag{2.16}$$

where $\phi(x)$ denotes the transformation of the original feature-space and b is the bias. The training set comprises of a set of N training samples $\{x_1, \cdots, x_N\}$, with corresponding label values $\{t_1, \cdots, t_N\}$ where $t_i \in \{-1, 1\}$. The incoming new data point x is classified based on the sign of $y(x)$. Currently, we assume that the training dataset is linearly separable in the transformed feature space, which indicates that there would be a linear boundary defined by parameters of $\{w, b\}$ that satisfies the conditions:

$$y(x_i) = \begin{cases} 1 & \text{if } t_i = +1; \\ 0 & \text{if } t_i = -1. \end{cases} \tag{2.17}$$

The final criteria is to make sure that $t_i \cdot y(x_i) > 0$ for all the training data points. There are many possible ways to search for the linear boundary, such as using a perceptron [62] or employing Fisher Linear Discriminant [63]. The SVM addresses this problem from the perspective of maximizing the margin, which is defined to be the smallest distance between

[4]LIBSVM: http://www.csie.ntu.edu.tw/ cjlin/libsvm/

the decision boundary and any of the samples among the patterns. The subset of points that lie on the boundary are called support vectors. The maximum margin solution for SVM is obtained by solving

$$\arg\max_{w,b} \frac{1}{\|w\|} \min_{i} \ [t_i(w^T\phi(x_i) + b)] \tag{2.18}$$

An equivalent solution can be obtained by minimizing the following function which is much easier to solve:

$$\arg\min_{w,b} \ \frac{1}{2}\|w\|^2 \tag{2.19}$$

subject to the constraint,

$$t_i(w^T\phi(x_i) + b) \geq 1, i = 1, \cdots, N. \tag{2.20}$$

This solution is to minimize the constrained quadratic function. To classify the new data points using the trained model, SVM uses the sign of the output to determine the membership.

$$f(s) = \sum_{i=1}^{M} y_i\alpha_i \cdot k(s, s_i) + b, \tag{2.21}$$

where $k(s, s_i)$ represents the kernel function and the sign of $f(s)$ determines the class label of s (gender). The linear kernel is the simplest function, and it is computed by the dot product $\langle s, s_i\rangle$ plus an optional constant c. Any vector s_i that refers to a non-zero α_i is called a support vector (SV) of the optimal hyperplane that separates the two classes. The common kernels used are the radial basis function (RBF) kernel and the linear kernel. However, in some cases where the features are derived from histogram representation, such as LBP and HOG, the histogram intersection kernel might be more effective.

$$k(x, y) = \sum_{i=1}^{n} min(x_i, y_i) \tag{2.22}$$

where x_i and y_i are the i^{th} histogram bin for the feature vectors of x and y. If we know the dataset is linearly separable, it is preferable to use a linear kernel, which often gives much faster solutions.

However, the training patterns may not always be linearly separable due to variation in data samples or noise. Thus the goal of SVM is to maximize the margin and penalize the outliers simultaneously [64]. Hence,

$$\min_{w,b,\epsilon} \ \frac{1}{2}w^T w + C\sum_{i=1}^{l}\epsilon_i \tag{2.23}$$

subject to the constrain,

$$y_i(w^T\phi(x_i)+b) \geq 1-\epsilon_i, \epsilon_i \geq 0. \tag{2.24}$$

The variable ϵ is used to measure the degree of misclassification. The RBF kernel function ϕ is used to map the data into high dimension so that the dataset may become linearly separable in the high dimension space. It is defined as $K(x_i, x_j) = exp(-r||x_i - x_j||^2), r > 0$. There are two parameters for the RBF kernel: the cost C (penalty) and the gamma r. The goal is to identify a good pair of parameters (C, r) so that the classifier can accurately predict unknown data. A common strategy is to split the training set into v subsets of equal size. Only one subset is used for testing based on the trained classifier on the other datasets. Such v-fold cross validation can prevent the overfitting problem. Usually, a grid search approach is used to find the parameters C and r. The idea is to try various pairs of (C, r) and select the one with the best performance accuracy. Due to the large searching space for the pair (C, r), an easy way to reduce the searching cost is to constrain the parameter space to exponentially distributed patterns, such as $C = 2^{-5}, 2^{-3}, \cdots, 2^{15}$ and $r = 2^{-15}, 2^{-13}, \cdots, 2^3$.

2.2.5 Adaboost Classifier

In the Adaboost algorithm [5], specific features are selected among a large pool of features based on their discrimination capability. It has been proved to be effective in [22, 21, 27] to select LBP features. It is an algorithm for constructing a strong classifier as the cascaded linear combination of simple weak classifiers. In the cascaded arrangement, the subsequent classifiers are built in the sense that they are tweaked in favor of those training patterns

[5]AdaBoost: http://cmp.felk.cvut.cz/cmp/software/stprtool/

misclassified by previous classifiers.

$$f(x) = \sum_{t=1}^{T} \alpha_t h_t(x) \tag{2.25}$$

where $h_t(x)$ refers to the weak classifier operating on the input feature set x. The $sign(f(x))$ is the final strong classifier. α_t is the corresponding weight for each weak classifier. For example, in the Viola Jones method for face detection [44], rectangular Haar-like features are used as weak classifiers. For gender classification with boosted LBP, the histogram bin features are considered to be the weak classifiers. One criteria for constructing the weak classifiers is that they should be able to separate the positive/negative classes with certain accuracy. Besides, they should also be very easy and fast to compute. Adaboost calls a weak classifier $h_t(x)$ repeatedly in a series of rounds $t = 1, \cdots, T$. For each round, a distribution of weights D_t is updated accordingly based on the importance of examples in the training dataset for classification. Initially, all the samples are assumed to have equal weights. Then, the weights of misclassified samples are increased so that the classifiers can focus more on those misclassified samples. The algorithm can be described as follows [65]:

Consider a labeled two-class dataset: $(x_1, y_1), \cdots, (x_m, y_m)$, where $x_i \in X$, $y_i \in Y = \{-1, 1\}$. Here, m denotes the total number of samples in the dataset, X refers to the set of training samples and Y is the label information.

Initialize $D_1(i) = \frac{1}{m}, i = \{1, \cdots, m\}$.

For $t = 1, \cdots, T$:

- Find the classifier $h_t : X \rightarrow \{-1, +1\}$ that minimizes the error with respect to the distribution D_t
- If $\epsilon_t \geq 0.5$, where $\epsilon_t = \sum_{i=1}^{m} D_t(i)(y_i \neq h_t(x_i))$ then stop
- Choose $\alpha_t \in R$, usually $\alpha_t = \frac{1}{2} ln \frac{1-\epsilon_t}{\epsilon_t}$
- Update
  $D_{t+1}(i) = \frac{D_t(i) exp(-\alpha_t y_i h_t(x_i))}{Z_t}$ where Z_t is a normalization factor such that D_{t+1} is a distribution.

Output the final classifier:

$$H(x) = sign(\sum_{t=1}^{T} \alpha_t h_t(x)) \tag{2.26}$$

One of the main ideas of the algorithm is to maintain a distribution of weights over the training set. The weights on each training example i at round t is denoted as $D_t(i)$. Initially, all the weights are equal. After each round, the weights of misclassified samples are updated so that the weak learners can focus on those hard samples. There are many variants of Adaboost, such as Discrete Adaboost, Real Adaboost and Gentle Adaboost [66]. We choose the Real Adaboost method because it is more resilient to noise and outliers. It uses a slightly different rule to update the weights.

2.2.6 Other Gender Classifiers

LDA: In the work of [32], the authors argued that the use of linear classification techniques is preferred in the context of limited computational resources. LDA classifier tries to maximize the separation of male and female class based on the Fisher's criterion:

$$J(w) = \frac{w^T S_B w}{w^T S_W w}, \tag{2.27}$$

where S_B is the between class scatter matrix and S_W is the within-class scatter matrix. $J(w)$ is the objective function that we are trying to maximized with respect to w. And the maximum value (projection matrix) is obtained by solving the generalized eigenvalue problem of $S_W^{-1} S_B$.

Random Forest: Random Forest (RF) [36] is an ensemble classifier that consists of many decision trees. Each decision tree is trained independently and successively based on a boot-strapped sampling of the training dataset [36]. The individual learners are combined through bootstrap aggregation. Given an input feature vector, it successively moves through the individual trees in the forest. The final classification is based on a majority voting over all the trees. Recent work [67] on the task of gender classification from infants to seniors has also shown the superiority of using Random Forest for feature selection.

GMM: In the GMM classifier, the probability density function for each class (i.e., male and female) is modeled as a multivariate Gaussian distribution:

$$p(x) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} exp\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\}. \tag{2.28}$$

Here, D is the dimension of the feature vector that encodes gender information and μ is the D-dimensional mean vector. Σ is the $D \times D$ covariance matrix and its determinant is denoted by $|\Sigma|$. The classification is done by maximizing the posterior function: $p(C_i|x_t) = p(x_t|C_i \cdot p(C_i)), i \in \{1, 2\}$.

MLP: A MLP neutral network is composed of an input layer, a hidden layer and an output layer. It utilizes a supervised learning technique (backpropagation) to train the network. The number of input nodes is equal to the dimension of the feature vector. We select 20 hidden nodes and the number of training cycles is set to be 40. The output value is based on the classification threshold, for instance, 0.5. An output value above the threshold is classified as male and a value below is classified as female.

2.3 Experiments

In this section, we evaluate the performance and compare the above mentioned approaches with different datasets. We used three image datasets in this experiment: FERET database [7], AR face database [6] and an image dataset collected from world wide web (WWW) [1]. The FERET dataset has been well studied in the work of [66]. The authors use this benchmark dataset for the evaluation of various gender classification methods. Due to the limited size of the FERET dataset, we use the AR face database to confirm our results. The WWW dataset is used to measure the gender classification performance in unconstrained environments. Apart from single database being tested, the trained classifier from FERET is evaluated on the real-world dataset to show the generalization capability of the proposed methods. The characteristic of each database is summarized as follows,

- **FERET:** A benchmark dataset, used to compare different gender classification algorithms.

- **AR Face:** A dataset with occlusions, used to test the resilience of gender classification algorithms to occlusion.

- **WWW:** A real-world dataset, used to characterize difficult scenario where gender classification needs to improve.

2.3.1 FERET Database

The FERET database [7] contains good quality gray scale face images. A small subset of the FERET is provided by Makinen [66]. It is publicly available on the author's website [6]. One of the benefits of using such a dataset is to compare the results with the work of other researchers and also make the results reproducible. The dataset contains 304 training images and 106 testing images. The number of male and female subjects are equal in the training dataset. There are 59 males and 47 females in the testing set. Samples of the images are shown in Figure 2.13. Here all of the images are aligned and cropped according to the eye coordinates based on the method introduced in Section 2.1.2. The size of each image after alignment is 128×128. No illumination normalization is applied here.

Figure 2.13: Sample images from the FERET dataset.

[6] http://www.cs.uta.fi/hci/mmig/vision/datasets/

All the experimental results (Figure 2.14) were obtained by using the same training set and testing set described above. One sample per subject is used and there is no overlapping of subjects between training and test set. In order to make the work comparable to [10], we do not apply cross-validation or random splitting methods. The reason for using the same number of males and females for training is to make the dataset well-balanced. The number of males and females in the test set has no such restriction.

The PCA with Nearest Neighbor classifier (PCA+KNN) is used as a baseline to compare against. The length of feature vector after dimension reduction is 60 if not specified. In other words, only the top 60 eigenvectors are kept. The SVM implementation has employed both linear and RBF kernels. The local LBP descriptor of face image is also included here to compare against holistic-based PCA presentation. Three discriminant classifiers, viz., Fisher Linear Discriminant (FLD), Quadratic (Quad) and Adaboost (Boost) classifiers were also included. Overall, we have three types of face representations and five classifiers resulting in 15 different approaches. Here, we only select eight representative methods to illustrate the results for gender classification.

With the same linear SVM classifier, the LBP representation (LBP+SVM) is less effective than PCA (PCA+SVM). The PCA captures the global information while the LBP descriptor characterizes the local information. This does not necessarily imply that global-based approaches are better than local approaches for gender classification. The current evaluation is conducted on a single dataset and cannot be used to infer the results on other datasets. One of the main objectives in gender classification is to seek discriminant gender features. Those gender features are derived from appearance information, either locally or globally. An interesting phenomena is that the raw pixels of images also provide sufficient information to discriminate between the genders (Raw+SVM). This has already been verified in the work of [17], which shows that gender classification of thumbnail images can achieve very high classification rate. Experimental evaluation has established that SVM provides superior performance among all classifiers. The linear FLD classifier [68] also achieves very good performance with PCA features at 86.79%. However, the Quadratic classifier [68] does not generalize well to unseen images due to the over-fitting problem. It produces the largest classification error. The pixel-wise Adaboost classifier is a bit sensitive to the noise in this

case and the resulting performance is not as good as the SVM classifier. Another explanation is that the Adaboost classifier is not well-suited to the small size of the training dataset. The best performance achieved on this dataset is 91.51% with the method of PCA+SVM. The RBF kernel is used for the SVM and the gamma paramter is set to be 2. The kernel-based PCA with SVM (KPCA+SVM) [69] has the next best performance at 90.57%.

The image has been resized to a resolution of 22×16 for all methods, except the LBP and Adaboost approaches. The LBP method uses an image size of 64×64, while the Adaboost approach uses an image size of 16×16. Here, the number of eigenvector is kept as 150 for PCA-based methods. The individual classification results for male and female, along with some other parameters are shown in Table 2.1. The male classification result is computed by counting the correctly classified samples within the male group. The same scenario applies to the female classification result. As shown, the role of gender affect the final output of prediction. Even humans perceive males and females differently.

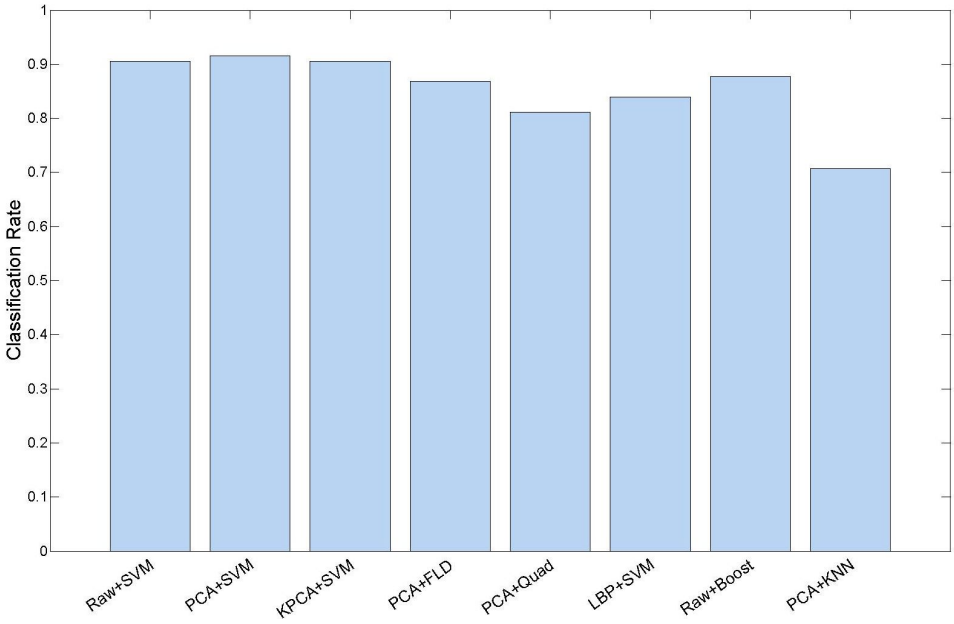


Figure 2.14: Comparison of gender classification methods on FERET dataset.

Table 2.1: Gender classification accuracies on FERET dataset.

Algorithms	**Rank-1 Rate**	**Male**	**Female**
RawPixels+SVM(RBF)	0.9057	0.8983	0.9149
PCA+SVM(RBF)	0.9151	0.8983	0.9362
KPCA+SVM(Linear)	0.9057	0.8983	0.9149
PCA+FLD	0.8679	0.8475	0.8936
PCA+Quadratic	0.8113	0.7797	0.8511
LBP+SVM(Linear)	0.8396	0.8475	0.8298
RawPixels+Adaboost	0.8774	0.8814	0.8723
PCA+KNN(L1)	0.7050	0.8136	0.5745

2.3.2 AR Face Database

The AR face dataset [6] contains 50 male subjects and 50 female subjects. Images feature frontal view faces with different facial expressions, illumination conditions, and occlusions (sun glasses and scarf). Each person participated in two sessions, separated by two weeks (14 days) time. The same subject was taken in both sessions (Figure 2.15). 25 males and 25 females were included in the training set, which consists of 1300 samples. The remaining 25 males and 25 females are used for testing, resulting in a total of 1300 samples. Notice that the subjects are not overlapping in the training and test set. But we use multiple samples per subject. The size of each image is 165×120.

The reason that gender classification is performed on this database is to show how the classifier behaves in the presence of occlusion and changes in illumination. The human perception of gender can encounter difficulties when subjects wear scarfs or glasses, leading to the loss of important facial features (Figure 2.15). The experimental design and algorithms used in this experiment are the same as in Section 2.3.1. Compared to the FERET dataset, AR face database is well aligned and the quality of images are also much better, except that some samples are captured under occlusion conditions. The relatively high performance of gender classification on this dataset (Figure 2.16) shows that the occlusions will not affect the final results significantly as long as enough gender discriminant information is retained.

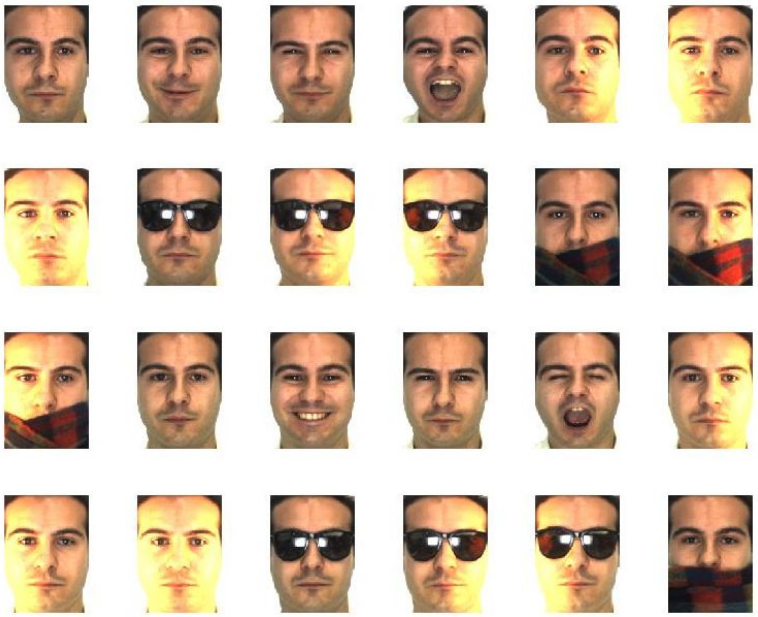


Figure 2.15: Samples of images from AR face dataset.

Similar to the work of [70] where the authors deliberately expand the training data with misaligned examples of face images so that the gender classifier can be more robust to face misalignments, the same idea is applied to include occluded samples in the training set to make the classifier resilient to occlusion. This assumption has been verified with experiments on this dataset.

Another important observation is the high performance obtained from the LBP descriptor, which is better than the PCA representation and the raw pixels-based methods on this dataset. The classification accuracy reaches 90.62% on a test dataset of 1300 images. The next best performance is the combination of KPCA and SVM classifier (90.31%). As in this dataets, samples are captured under various illumination conditions. The LBP representation of face images can account for such changes, whereas the raw pixels might be sensitive to illuminations. Among all the classifiers used, the SVM still gives the best performance (averaging 90.02% for all the four methods), while the performance of quadratic classifier degrades significantly to 50%. The over-fitting problem for Quadratic clasisifer is more evident in this large database. It simply fails as it classifies all the test examples as male. The Adaboost-based classifier performed well on this dataset with the raw pixels representation

presenting an accuracy of 89.46%. All of the parameters are kept the same as in section 2.3.1. More results are shown in Table 2.2. The male and female classification performances are almost the same for this dataset. The image size is chosen as 42×36. For LBP method, the image size remains to be 64×64. To account for the illumination changes, we apply the histogram equalization methods.

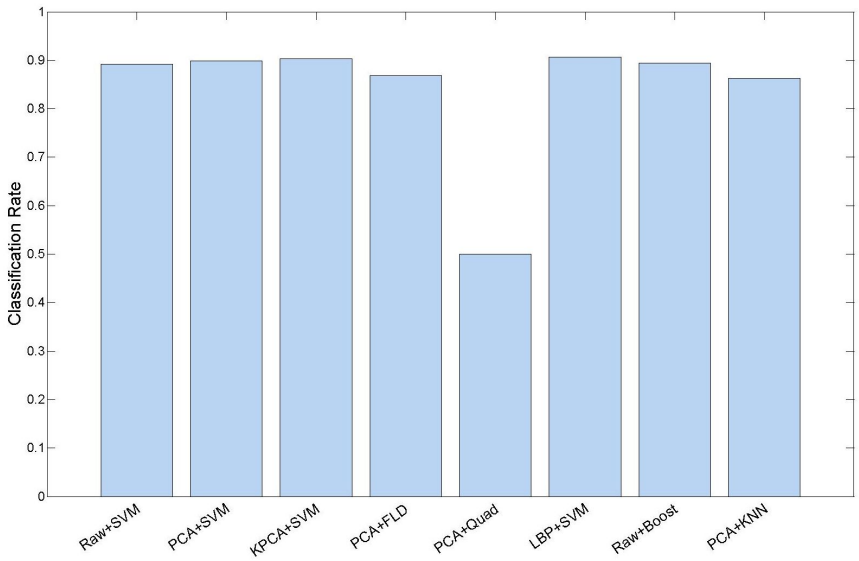


Figure 2.16: Comparison of gender classification methods on AR face database.

Table 2.2: Gender classification accuracies on AR dataset.

Algorithms	**Rank-1 Rate**	**Male**	**Female**
RawPixels+SVM(RBF)	0.8923	0.8892	0.8954
PCA+SVM(RBF)	0.8992	0.8969	0.9015
KPCA+SVM(Linear)	0.9031	0.9000	0.9062
PCA+FLD	0.8685	0.8785	0.8585
PCA+Quadratic	0.5000	N/A	N/A
LBP+SVM(Linear)	0.9062	0.9015	0.9108
RawPixels+Adaboost	0.8946	0.9077	0.8815
PCA+KNN(L2)	0.8623	0.8600	0.8646

2.3.3 Real-World Dataset

The previous experiments were restricted to images from controlled datasets, i.e., FERET and AR. Further, all the training and test images come from the same dataset. But it is important to train and test the algorithms on two different datasets. Therefore, experiments were conducted where the gender classifier (PCA+SVM) was trained using the FERET database and tested on real-world images. It aims to show how well the gender classifier generalizes to unseen images. The output of the face detector is fed into the classifier to predict gender. The face images are normalized according to the detected eye landmarks. All the test images were selected from the group image database [1].

Figure 2.17: Gender prediction: the image has two persons-one male and one female.

From the results in Figure 2.17, the gender classifier successfully predicts the gender label for every face in the image. Surprisingly, it can also be used to predict gender from children faces (Figure 2.18). The previous trained dataset did not include any images from children. The results from Figure 2.19 is another good example to show the effectiveness of the trained gender classifier. As stated before, one of the limitations in the automatic gender classification system is that the final prediction results depend on the successful detection of human faces. In the example of Figure 2.20, the complex background causes the failure of the face detector. Some non-face regions are also mis-classified as faces. Therefore, the

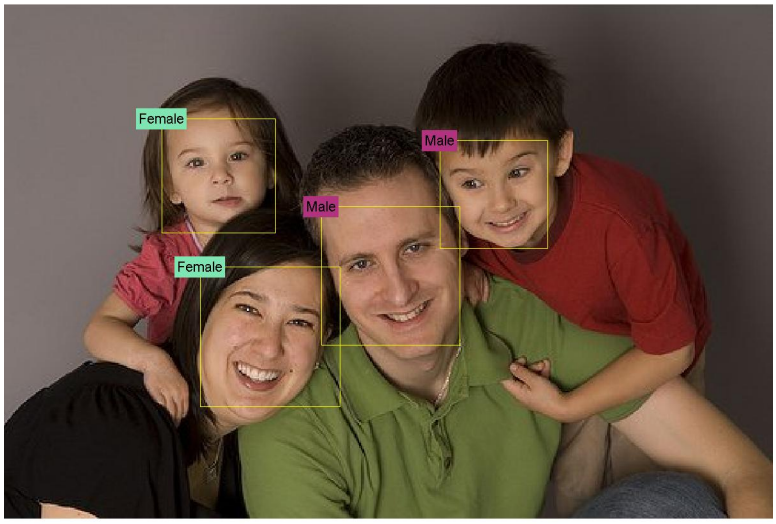


Figure 2.18: Gender prediction: the image has four persons-two males and two females.

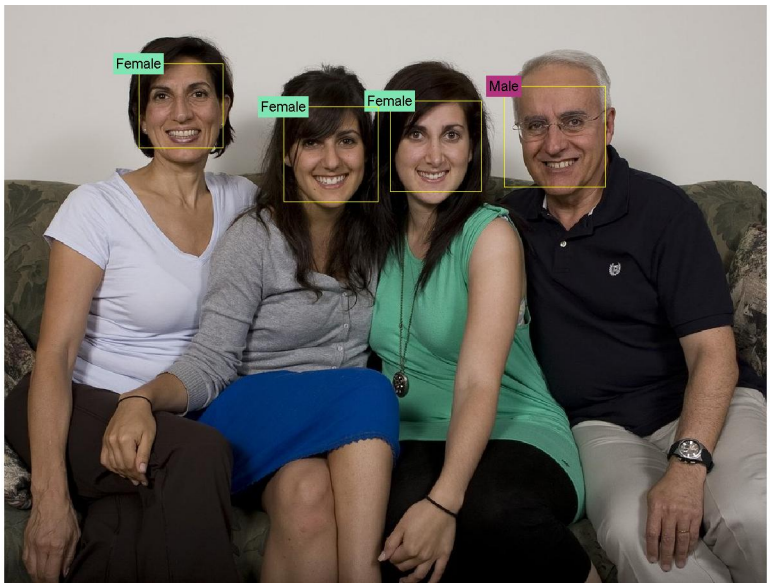


Figure 2.19: Gender prediction: the image has four persons-three females and one male.

corresponding output from the gender classifier is not reliable. There are many possible ways to improve the gender classification accuracy. One is to make the face detector more reliable. Another is to improve the design of the classifier with more options, for example, a rejection scheme. The gender classifier should be able to reject the input regions if the possibility of faces in that region is low. Since the gender classification involves inference with uncertainty,

Figure 2.20: Gender prediction: the image has six persons-three males and three females.

it is possible to incorporate the Bayesian framework into the classification framework [8]. Such a classifier could potentially make the prediction more reliable. Although the decision value from SVM contains the probability about whether the output is male or female, its objective is to maximize the margin between male and female classes, which is different from the Bayesian approach. Some examples of the prediction results are shown in Figure 2.21 and Figure 2.22. The low probability indicates that the decisions are made without enough confidence.

The above cross-database tests on group face images show the generalization capability of the proposed gender classifier, but the results are only based on a few sample images selected from database. The remaining task is to perform a single database test on real-world face images. Here, the images are collected from the Internet [1]. There are 500 male and 500 female subjects, respectively. 200 male and 200 female subjects are randomly selected to train the classifier and the rest are reserved for testing. The size of each image is 61×49. These images possess all kinds of variations such as occlusion, age, pose, expression, and illumination. The quality of these images are also not ideal. The mixture of these impacting factors simulates the real-world problem and poses a big challenge for current gender classification methods. Some samples from the database are shown in Figure 2.23.

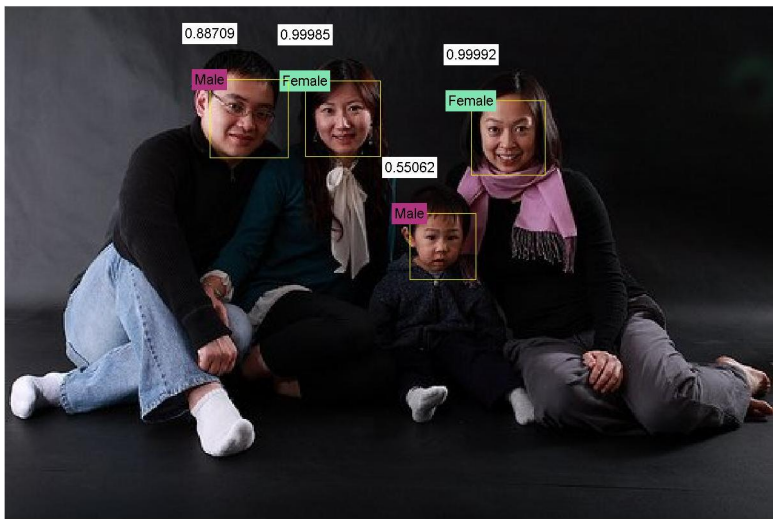


Figure 2.21: Gender prediction with probability estimation. The output is based on the implementation from [8].

Figure 2.22: Gender prediction with probability estimation. The output is based on the implementation from [8].

The gender classification results are listed in Table 2.3. The feature extraction method of PCA (computed via SVD) is selected, resulting in a total of 30 dimensions. Here we restrict the choice of feature extraction and test the performance of various discriminant classifiers.

The discrepancy of gender classification accuracies between males and females is very evident in this case. The result of male classification is much better than female classification.

The best performance (80%) obtained on male classification is with the LDA classifier. The corresponding female classification accuracy is 71%. The SVM classifier obtains 79.50% accuracy on males and 65.50% accuracy on females. The KNN classifier gives the worst overall performance at 61.67%. Currently, the average performance of all the classifiers on the male group is 75.50% and on the female group is 63.52%. It is possible to use other type of features, rather than PCA, to improve the performance. Besides, the database used here is much more difficult than other real-world images with good quality used in the literature. Some previously reported results assumed that faces can be accurately localized (no misalignment) [27] or that pose normalization was reliable [49]. Such conditions unavoidably add some bias to the test results. The main objective here is to provide the comparison results for different type of classifiers [7].

(a) (b)

Figure 2.23: Samples from WWW dataset: (a) male subjects; (b) female subjects.

[7]Some Classifiers are from: http://www.informedia.cs.cmu.edu/yanrong/MATLABArsenal/

Table 2.3: Gender classification accuracies on Real-world dataset.

Algorithms	**Male**	**Female**	**Overall**
KNN	0.7350	0.5900	0.6167
SVM	0.7950	0.6550	0.6617
GMM	0.7850	0.5950	0.6233
LDA	0.8000	0.7100	0.7200
Neural Network	0.6650	0.6550	0.6900
Naive Bayes	0.7850	0.5950	0.6217
Adaboost	0.7200	0.6467	0.6833

2.4 Discussion

The above experiments demonstrate the gender classification performance on three public datasets: FERET, AR and WWW. The trained classifier from FERET is also tested on real-world images, which shows very promising results in Section 2.3.3. This simulates the case where a trained classifier is often obtained from a well-constrained dataset and tested on unseen images from different scenarios. However, there are still many issues with respect to the design of gender classification experiments that are not clearly resolved. Mainly, they are related to the choice of different face normalization approaches and datasets to perform the experiments.

2.4.1 Impact of Face Normalization

Recall that in section 2.1.2, we have enumerated three type of face normalization approaches:

1. Geometry alignment: face image is aligned and cropped based on eye coordinates.

2. Photometry processing: the cropped image is processed with histogram equalization and mask fitting operation after geometry alignment.

3. Pose correction: the image is further processed to bring the pose to frontal.

In real-world applications, the first two steps are usually sufficient for the preprocessing of face images. The third step can be applied when there is a large pose change. But this step often requires the construction of a 3D model or AAM. Generally, it is expected that pose correction will improve the final classification accuracy. Some pose correction results can be observed in Section 4.1. Here, we make extensive comparisons between the first two normalization methods on the same FERET dataset and show the merits of each method. Most of the tested methods have been used in the literature before. 'FERET_NORM' refers to the database that goes through both geometry and photometry processing. Meanwhile, 'FERET' only goes through geometry alignment.

First, the impact of different face normalization approaches on gender classification accuracy is tested. The difference between those two approaches is to determine whether it is necessary to eliminate the background information and apply the histogram equalization to reduce the illumination variation. Previous research work in [10] suggested that the hair information might be useful for gender classification. However, other researchers preferred facial images without much hair information [17, 18].

In order to test the performance on the impact of those factors, the same FERET dataset with 410 samples is used.
We define the protocols for this experiment as follows [71]:

- Both image sets are resized so that the largest side is 32.
- The training-test random split is based on the 40%-60% rule.
- The experiments are repeated 10 times to reduce the bias.
- The final classification accuracy is the average over all the trials.

As can be seen in Figure 2.24, except for the PCA and ICA [33] methods, face normalization with both photometric and geometric transformations results in better accuracy than face normalization with geometric normalization only. In other words, face images going through both step 1 (geometry) and step 2 (photometry) are better than step 1. The main classifier used here is SVM. For the methods of PCA and ICA [72][8], the Nearest Neighbor classifier

[8]http://mplab.ucsd.edu/ marni/code.html

is adopted. The best performance 89% is achieved with the low-resolution face images (RawPixels+SVM). PCA representation of face image (PCA+SVM) or the Gabor-based SVM (Gabor+SVM, GaborFisher+SVM) resulted on performances of 84.5% and 88.6%, respectively. However, the Gaborface representation with SVM does not perform better than the enhanced GaborFisher Face representation for this particular task. Based on the experiments, PCA, LBP and Gabor features are good choices to represent gender features. All of the methods, except the GaborFisher based approach, have been tested before in the gender classification task. The original use of GaborFisher method was for face recognition [73]. We adapted this approach for the task of gender classification.

In general, the normalization approach that can exclude as much background information as possible is preferred. In case of dataset with illumination changes, histogram equalization would also be necessary. Since the performance only degrades slightly when geometry normalization is used, it can also be utilized in many applications. For example, when the LBP and SVM methods are used, an accuracy of 87.8% is achieved on geometry normalized data as opposed to 86.5% on photometric normalized data. The LBP method is robust to uniform illumination changes in itself, and therefore photometric normalization might not be necessary.

2.4.2 Impact of Datasets

Another factor that is often neglected by researchers is the choice of datasets to perform the experiments. From the previous work in Section 1.2, there are many different datasets used for the evaluation of gender classifiers. Some datasets are subsets of larger datasets, leading to variations across trials [17]. Some are not currently publicly available for other researchers to compare [27]. Therefore, gender classification results across different literature are sometimes not comparable. This also leads us to the dilemma of choosing the most suitable algorithms for gender classification. Due to differences across datasets, good performance achieved in one dataset may not be duplicable in another dataset. In order to make the experimental results reproducible, publicly available datasets are chosen so that other researchers could easily compare the results against our implementation.

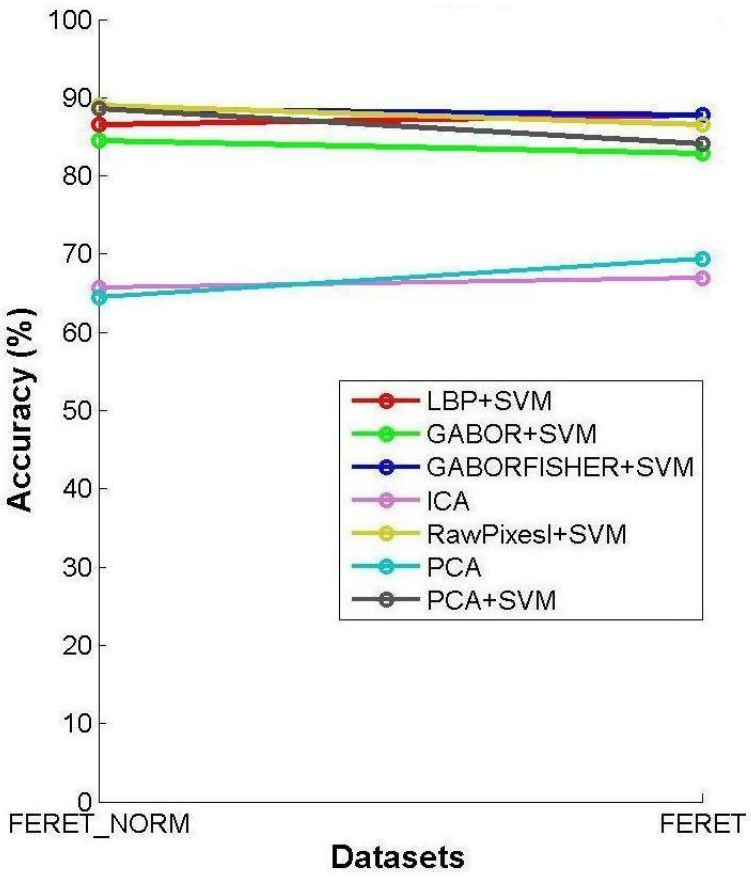


Figure 2.24: Impact of normalization on gender prediction.

A comprehensive performance evaluation on different gender datasets [9] is provided. The criteria of choosing these datasets is that they are publicly accessible to the researchers and the gender annotation information is also available.

- MUCT [74]: The database was created to provide more diversity of lighting, age, and ethnicity of 2D landmarked databases. We use the 276 subjects from category-a, with 131 males and 145 females. The first encountered sample of each subject is selected.

- The FERET [10] : The database contains gray scale face images with varying facial expression, lighting and pose changes. 410 subjects are selected: 211 males and 199 females.

[9]The Medical dataset is available at: http://scien.stanford.edu/pages/labsite/2001/ee368/projects2001/dropbox/project16/med_students.tar.gz

Table 2.4: Overview of datasets used for gender classification.

Datasets	**Males**	**Females**	**Normalization**	**AVG. Rate**
MUCT	131	145	Complex	81.5250%
F.NORM	211	199	Complex	83.9875%
FERET	211	199	Aligned	83.6875%
MEDICAL	200	200	NONE	88.5625%
Combination	342	344	Complex	81.1875%

- MEDICAL: The medical dataset is a collection of students from Standford school. There are 200 males and 200 females. The face occupies most of the image.
- The combination of FERET and MUCT dataset: The database consists of 686 subjects, with 342 males and 344 females.

Notice that not all face images from the entire database are used since we want to separate the identity and gender information by selecting only one sample per subject. The total number of available samples in our test depends on the number of subjects in the database. That is why the size of each database may not be large. But such separation of identity and gender could reduce the bias of gender classification. The majority of face images are near-frontal. But they also exhibit a certain degree of change in pose. The profile face images are not used here. The detailed description of the datasets, along with their individual performances are presented in Table 2.4.

From Figure 2.25, various gender classification methods on five publicly available databases are presented. The experiments are carried out by training the gender classifier on 40% of the database and testing on the remaining 60%. Because the number of male subjects and female subjects in the whole database are almost the same, this train-test split can still ensure that the training dataset is well-balanced. The three databases (MUCT, FERET_NORM, Combination) are normalized based on the approaches in 2.1.2. "Combination" refers to the merging of both MUCT and FERET. The FERET database is normalized based on geometry only. The Medical database is not processed with any normalization method. The reason for using three types of datasets is as follows:

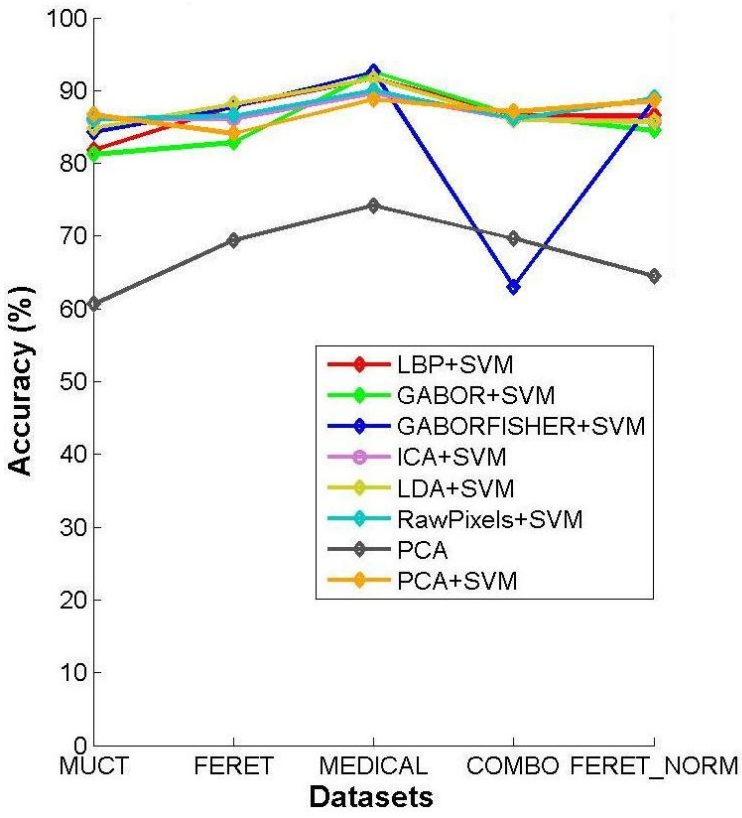


Figure 2.25: Gender classification on five different datasets.

- To test the performance difference between normalized database and unnormalized database
- To show the average performance on each dataset

Currently, the MUCT dataset has the lowest average performance. The MEDICAL face dataset provides the best classification performance as high as 88.56%. Among all the methods, the PCA with the KNN classifier performs the worst in every dataset. That is why the average performance on each dataset is low. Details of each experiment are also available in Table 2.5. Compared to the methods used in Section 2.4.1, we replace the ICA method [72] with ICA+SVM and also add another algorithm (LDA+SVM).

Table 2.5: Overview of each algorithm used in different datasets.

Algorithm	**MUCT**	**FERET**	**MED.**	**COMBO**	**F.NORM**	**AVG**
LBP+SVM	81.8	87.8	91.7	86.6	86.5	86.88
PCA	60.6	69.4	74.2	69.6	64.5	67.66
LDA+SVM	84.8	88.2	91.7	85.9	85.7	87.26
Gabor+SVM	81.2	82.9	92.5	86.6	84.5	85.54
ICA+SVM	86.1	86.1	89.6	86.1	85.7	86.72
GaborFisher	84.2	87.8	92.5	63.0	88.6	83.22
Raw+SVM	86.1	86.5	90.0	86.1	89.0	87.54
PCA+SVM	86.7	84.1	88.8	87.1	88.6	87.06

2.5 Chapter Summary

In this chapter, the complete process of automatic gender classification was discussed. The system consists of face detection, face normalization, feature extraction and gender classification. Face detection and face normalization are vital to the success of gender classification system when applied to real-world images. Without accurate face detection, the output of gender prediction would be meaningless. For experiments conducted on controlled dataset, face normalization is also a necessary step in alleviating within-class variations.

The choice of gender feature extraction methods varies across different datasets and applications. We introduce commonly used gender feature representation methods, such as Eigenface, Fisherface, Gaborface and LBP. These are the feature extraction methods that have been used previously by the researchers and have demonstrated good performance. Realizing the design of gender classifier is also critical: two types of classifiers are brought into our framework, SVM and Adaboost.

The experiments are divided into single database test and cross-database test. The single database test include experiments on the public FERET, AR and WWW datasets. The FERET dataset can be considered as a benchmark gender classification testbed. The AR face database is used to test gender classification algorithms against occlusion, which commonly occurs in face acquisition. To address the occlusion challenges, we can deliberately

include samples with occlusions in the training stage. The WWW dataset is used to test gender classification in real-world datasets. The cross-database test is to illustrate the idea of generalization property with gender classifier. Interestingly, we found that gender classifier can generalize well to young faces even the trained system has not encountered them before. Such challenging cross-database experiments show the promising results of gender classifiers, especially on real-world group images.

Next, we tackle some related issues regarding to the design of gender classifier: the choice of face normalization and datasets. These two factors can affect the output of the gender classifier significantly.

The goal of this chapter is to bring currently popular used gender classification methods into a unified framework. All the aspects related to gender classification have been studied systematically. The results are verified on public available datasets and can be easily reproduced. While a much more detailed discussion of the system is possible, this is not the intension of this thesis.

Chapter 3

Cross-Spectrum Gender Prediction

Though gender classification has received much attention from the research community, previous work has focused on face images obtained in the visible spectrum (VIS) (Chapter 2). The aim of this chapter is to explore gender classification in the near-infrared spectrum (NIR) using learning-based algorithms. The use of NIR images for face recognition has become necessary especially in the context of a night-time environment where VIS face images cannot be easily discerned [9]. Further, NIR images are less susceptible to changes in ambient illumination. Thus, cross-spectral matching has become an important topic of research [75, 76, 77]. To the best of our knowledge, this is the first work that explores gender recognition in NIR face images. In this regard, we address the following questions:

- **Q1.** Can gender be predicted from NIR face images?
- **Q2.** Can a gender predictor learned using VIS images operate successfully on NIR images, and vice-versa?

To answer **Q1**, we use an existing gender prediction mechanism based on SVM [17]. In order to address **Q2**, we hypothesize that an illumination normalization scheme may be necessary prior to invoking the gender classifier.

In this chapter, we describe the design of the gender classifier from NIR spectrum in Section 3.1, with special emphasis on illumination normalization approaches in Section 3.2 for cross-spectral gender prediction. Then, we report experimental results that demonstrate the possibility of assessing gender from NIR face images in Section 3.3. Finally, we discuss

the difficulties in cross-spectral gender classification and indicate future directions in Section 3.4.

3.1 System Design

In order to address the questions raised above, we utilize a gender prediction scheme based on SVMs. Such a scheme has been shown to be efficient in the VIS domain [17, 50]. The SVM-based classification scheme is described below. Given a facial image x_i, in either the VIS or NIR domains, the feature extractor is applied to obtain a discriminant feature set s_i. The gender classifier, G, is then invoked to predict the gender. In case of images that are coming from different spectrum, we utilize the illumination normalization approaches to reduce the spectral difference. Our whole system is depicted in Figure 3.1. The design of

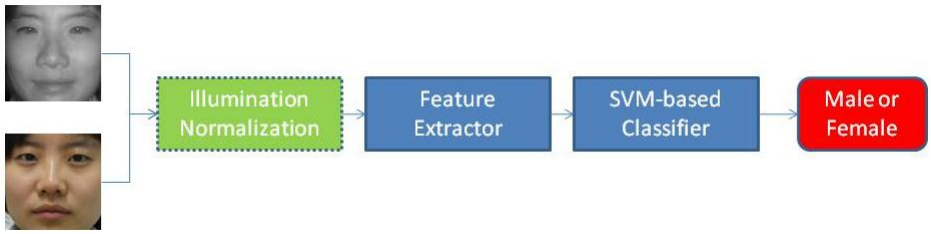


Figure 3.1: Cross-spectral gender classifier.

the system is based on the following observations:

- A gender classifier should be robust to unseen scenarios in the applications.
- The system should be fast and accurate enough to provide useful information regarding to the gender visual attributes.

Before delving into this chapter, we have investigated different feature extraction methods, either globally and locally. Various classification methods have also been studied. Therefore, this chapter will focus on one specific method to explain the cross-spectrum gender prediction problem. Among all the tested algorithms, such as Raw pixels with SVM or Adaboost, PCA with FLD classifier and LBP with SVM, the PCA method with SVM gives the best accuracy.

Therefore, we choose PCA for gender feature representation and SVM for classification in the later sections.

Gender classification using SVM is shown in Figure 3.2. Here, the dimension of the extracted feature vector is reduced to two by PCA in order to visualize the feature distribution. The images are trained on the VIS spectrum and tested on the NIR spectrum. The samples are not linearly separable in this two-dimension space. The SVM maps the feature samples into high dimension and search for the linear boundary by maximizing the width between support vectors.

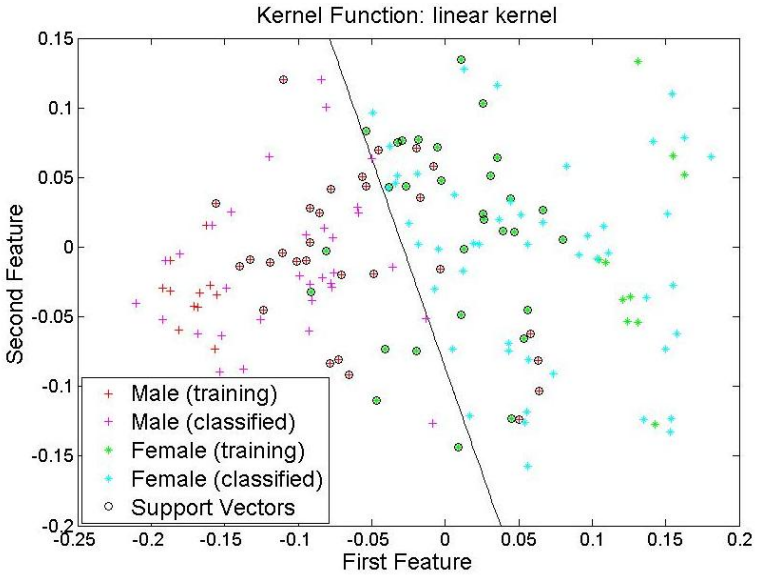


Figure 3.2: Illustration of a SVM-based gender classifier with linear kernel on the HFB database.

3.2 Illumination Normalization

As stated earlier, we hypothesize that the use of an illumination normalization scheme [1] may be necessary to accommodate cross-spectral gender prediction where the training and test sets have images pertaining to different spectral bands.

[1]The INFace toolbox: http://luks.fe.uni-lj.si/en/staff/vitomir/index.html. Only the SQI, Retinex and DCT methods are coming from this toolbox.

Self Quotient Image (SQI): According to the Lambertian model, the image formation process is described as follows:

$$I(x,y) = \rho_w(x,y)n(x,y)s, \tag{3.1}$$

where $\rho_w(x,y)$ is the albedo of the facial surface, n is the surface normal and s is the lighting reflection. To reduce the impact of illumination, we need to separate out the extrinsic factor s from ρ and n. The self-quotient image, Q, of I is defined as [78],

$$Q = \frac{I(x,y)}{\hat{I(x,y)}} = \frac{\rho_w(x,y)n(x,y)s}{G * [\rho_w(x,y)n(x,y)s]}, \tag{3.2}$$

where $\hat{I}$ is the smoothed version of I and G is the smoothing kernel.

Retinex Model: The retinex approach is based on the reflectance illumination model instead of the Lambertian model. It is an image enhancement algorithm [79] proposed to account for the lightness and color constancy of the dynamic range compression properties of the human vision system. It tries to compute the invariant property of reflectance ratio under varying illumination conditions [80, 78]. The retinex model is described as follows:

$$I(x,y) = R(x,y)L(x,y), \tag{3.3}$$

where $I(x,y)$ is the image, $R(x,y)$ is the reflectance of the scene and $L(x,y)$ is the lighting. The lighting is considered to be the low-frequency component of the image $I(x,y)$, and is thus approximated as,

$$L(x,y) = G(x,y) * I(x,y), \tag{3.4}$$

where $G(x,y)$ is a Gaussian filter and $*$ denotes the convolution operator. The output of the retinex approach is the image $R(x,y)$ that is computed as,

$$R(x,y) = \frac{I(x,y)}{L(x,y)} = \frac{I(x,y)}{G(x,y) * I(x,y)}. \tag{3.5}$$

Discrete Cosine Transform (DCT) Model: Since illumination variations typically manifest in the low-frequency domain, it is reasonable to normalize the illumination by removing the low-frequency components of an image. DCT can be first applied to transform an image from the spatial domain to the frequency domain, and then estimate the illumination

of the image via low-frequency DCT coefficients which appear in the upper-left corner of the DCT [80]. By setting the low-frequency components to zero and reconstructing the image, variations due to illumination can be reduced. The 2D $M \times N$ DCT can be computed as,

$$C(u,v) = \alpha(u)\alpha(v) \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} I(x,y) \times cos\left[\frac{\pi(2x+1)u}{2M}\right] cos\left[\frac{\pi(2y+1)v}{2N}\right], \tag{3.6}$$

where $\alpha(u)$ and $\alpha(v)$ are the normalization factors.

CLAHE Normalization: The CLAHE (Contrast Limited Adaptive Histogram Equalization) [81] method applies contrast normalization to local blocks in the image such that the histogram of pixel intensities in each block approximately matches a pre-specified histogram distribution. This scheme is applied to blocks of size 16×16. CLAHE is effective at improving local contrast without inducing much noise. It utilizes the normalized cumulative distribution of each gray level, x, in the block [77]:

$$f(x) = \frac{N-1}{M} \times \sum_{k=0}^{x} h(k). \tag{3.7}$$

Here, M is the total number of pixels in the block, N is the number of gray levels in the block, and h is the histogram of the block. To improve the contrast, the CLAHE technique transforms the histogram of the block such that the histogram height falls below a pre-specified threshold. Gray level counts beyond the threshold are uniformly redistributed among the gray levels below the threshold. The blocks are then blended across their boundaries using bilinear interpolation.

Difference-of-Gaussian (DoG) Filtering: Another type of normalization is proposed in [61], where the local image structures are enhanced. One of the key components in [61] is the Difference-of-Gaussian (DoG) filtering, which can be computed as,

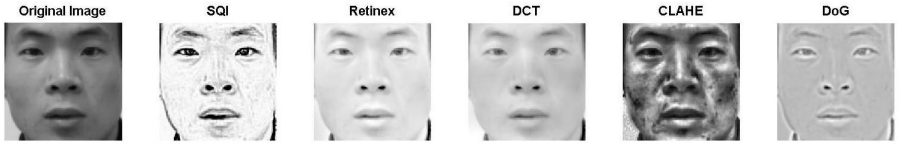
$$D(x,y|\sigma_0,\sigma_1) = [G(x,y,\sigma_0) - G(x,y,\sigma_1)] * I(x,y). \tag{3.8}$$

The symbol * is the convolution operator, and the gaussian kernel function based on σ is,

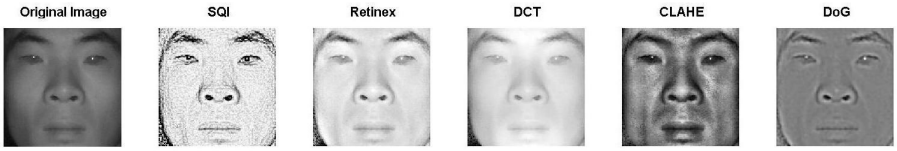
$$G(x,y,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x^2+y^2)/2\sigma^2}. \tag{3.9}$$

This simple filtering scheme has the effect of subtracting two Gaussian filters.

The output of the various illumination normalization schemes are presented in Figure 3.3. The goal of illumination normalization is to facilitate cross-spectral gender classification by mitigating the effect of spectral specific features.



(a)



(b)

Figure 3.3: (a) A VIS image and its corresponding normalized images; (b) A NIR image and its corresponding normalized images. Images are from HFB database [9].

3.3 Experiments

In the previous Section 3.2, we have introduced different illumination normalization approaches in the hope of finding an effective way to reduce the spectrum difference between VIS and NIR. To validate the effectiveness of each method, we perform the gender classificaiton experiments on the HFB database [9].

3.3.1 HFB Database

The HFB face database [2] consists of 100 subjects, including 57 males and 43 females. There are 4 VIS and 4 NIR face images per subject. One of the subjects from the database is displayed in Figure 3.4. The HFB database has also provided the eye coordinates for each sample. We apply the same methodology to crop and align all the images in the database. The cropped version of one subject is shown in Figure 3.5. The variations between corresponding VIS and NIR samples are evident, as they are not captured simultaneously.

[2]HFB database: http://www.cbsr.ia.ac.cn/english/HFB 20Databases.asp

Figure 3.4: Top Row: Samples from VIS spectrum; Bottom Row: Samples from NIR spectrum. Images are from HFB database [9].

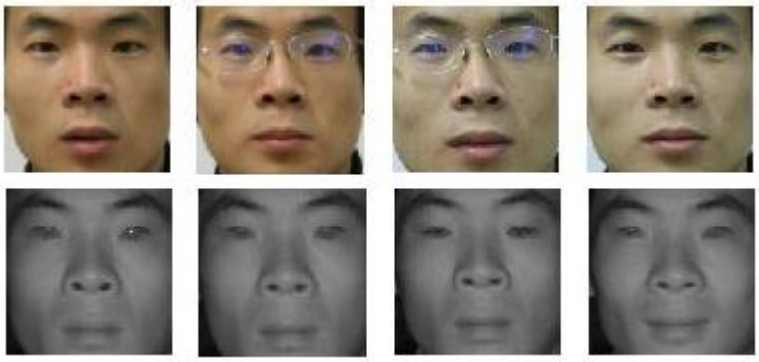


Figure 3.5: Top Row: Cropped Samples from VIS spectrum; Bottom Row: Cropped Samples from NIR spectrum.

3.3.2 Evaluation

The following experiments were conducted on this database: (a) Training and testing using VIS images (VIS-VIS); (b) Training and testing using NIR images (NIR-NIR); (c) Training using VIS images and testing using NIR images (VIS-NIR); (d) Training using NIR images and testing using VIS images (NIR-VIS). In all cases, the subjects used in the training and test sets were mutually exclusive. 20 male and 20 female subjects were randomly selected for training, with 4 samples for each subject. The remaining subjects were reserved for testing. This random partitioning to generate the training and test sets was repeated 10 times

for each experiment in order to understand the variance in classification accuracy. The image size used in our work was 128×128 of the cropped version. For the VIS-VIS experiments, the

Table 3.1: Gender classification results on the HFB database when illumination normalization is **not** used for cross-spectral prediction.

Scenario	**Classification Rate**	**Best**	**Worst**
VIS-VIS	0.9067 ± 0.0397	0.9708	0.8458
NIR-NIR	0.8442 ± 0.0264	0.9000	0.8042
VIS-NIR	0.5625 ± 0.1289	0.7083	0.3833
NIR-VIS	0.6021 ± 0.0769	0.6667	0.3875

Table 3.2: Results for cross-spectral gender classification after applying different normalization schemes.

	VIS-NIR(N)	NIR-VIS(N)
CLAHE	$\mathbf{0.6617 \pm 0.0724}$	$\mathbf{0.6642 \pm 0.0806}$
DoG	0.6446 ± 0.0331	0.6100 ± 0.0354
SQI	0.4512 ± 0.0693	0.4692 ± 0.0611
Retinex	0.5525 ± 0.0537	0.5921 ± 0.0674
DCT	0.5967 ± 0.0840	0.6392 ± 0.0666

average classification rate was 90.67%, with the best performance being 97.08% (Table 3.1). The performance is comparable to the results reported in previous literature on other datasets [17, 18]. This suggests that gender classification can be performed with high accuracy in the VIS domain. For the NIR-NIR experiment, the average performance declined by around 6% compared to VIS-VIS classification resulting in an average accuracy rate of 84.42%. For the VIS-NIR and NIR-VIS experiments, the average classification rates were 56.25% and 60.21%, respectively, suggesting the difficulty in performing cross-spectral gender classification.

However, upon applying certain illumination normalization schemes (to both the training and test images), we observed an improvement in classification accuracy (Table 3.2). Two of the most effective normalization schemes were CLAHE and DoG. In our experiment, the CLAHE gave slightly better performance than DoG. Specifically, the CLAHE normalization scheme improved cross-spectral gender classification for the VIS-NIR(N) and NIR-VIS(N)

experiments to 66.17% and 66.42%, respectively - this represents an improvement of 18% and 10%, respectively. The SQI scheme decreased the performance after normalization, while the retinex model did not impact the accuracy. The DCT algorithm gave slightly better performance, but not as significant as that of CLAHE and DoG.

3.4 Experimental Analysis

Our experimental results indicate the possibility of performing gender classification using NIR face images although the performance is slightly inferior to that of VIS images. This suggests that the gender information observed in the NIR domain *may* not be as discriminative as in the VIS domain. Cross-spectral gender prediction was observed to be difficult - this suggests that the gender related information available in the NIR and VIS face images are significantly different as assessed by the classifier. The key, therefore, is to reduce the variability between these two type of images by applying an illumination normalization routine. Experiments suggest that certain normalization schemes are better than the others. In particular, the CLAHE scheme proved superior than the other models considered in this work.

Next, we consider the reasons for the inferior performance of the other normalization models. The Lambertain model usually assumes that the term $\rho_w(x, y)$ is constant across different lighting sources. However, since the lighting conditions under NIR and VIS spectra are not homogeneous, estimating an illumination invariant albedo $\rho_w(x, y)$ under the Lambertian model for those two type of images is not possible [76]. Therefore, approaches based on the Lambertian model, such as self-quotient image and its variants, are not useful in our application. Since the reflectance is not a stable characteristic of facial features for images captured under the NIR and VIS spectra, the retinex model also does not result in good performance. The DCT method fails since the illumination in NIR images cannot be simply estimated by the low-frequency coefficients of the image. Only those normalization methods based on local appearance-based features (i.e., CLAHE and DoG) result in better accuracy. This could partly be due to the use of PCA-based features in our experiments. The use of other sophisticated features (such as LBP) for gender classification may be useful when the

SQI and retinex models are used for normalization.

When the images (128×128) are downsampled by a factor of 4, the average accuracy of VIS-NIR improved from 66.17% to 71.79% (Table 3.3). Similarly, the average accuracy of NIR-VIS improved from 66.42% to 69.17%. Another observation has to do with the difference in gender classification of males and females. We ran the VIS-NIR experiments on the HFB database 100 times and observed that the female classification rate was 68% while the male classification rate was 77%.

Table 3.3: Impact of image size on gender classification for the VIS-NIR and NIR-VIS scenarios when the CLAHE normalization method is used.

Image Size	VIS-NIR	NIR-VIS
128×128	0.6617 ± 0.0724	0.6642 ± 0.0806
64×64	0.6958 ± 0.0241	0.6596 ± 0.0856
32×32	$\mathbf{0.7179 \pm 0.0208}$	$\mathbf{0.6917 \pm 0.0292}$
16×16	0.6638 ± 0.0362	0.6617 ± 0.0668

Next, we take a look at the histogram distributions of pixel intensities for a VIS image and a NIR image (Figure 3.6). The VIS image has a dense histogram, while the NIR image has a more sparse histogram distribution. This suggests that the VIS image has more intensity values captured than its counterpart. Such a difference indicates the loss in information when forming NIR images. The hypothesis is that histogram normalization can mitigate some of these differences thereby improving cross-spectral gender prediction. We find that by applying the CLAHE normalization approach, it is possible to reduce the difference between the two histograms (Figure 3.6).

3.4.1 Automatic Gender Prediction

One of the motivations in our work is to demonstrate the possibility of using trained VIS classifier to predict genders from NIR face images in real-world applications. In scenarios where the system has already been trained on VIS images, it would be costly to retrain the system with new domain of knowledge, for example NIR images. Therefore, it is a requirement for the designed system to be robust and adaptive to different conditions.

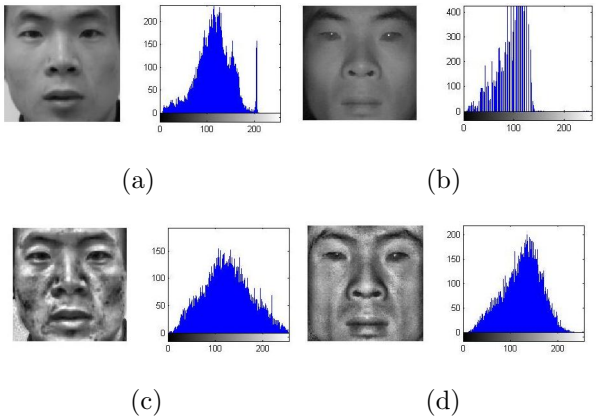
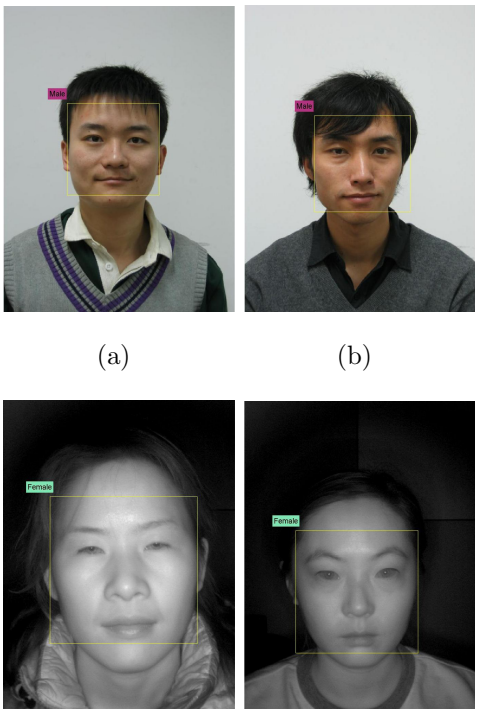


Figure 3.6: (a) VIS image before normalization; (b) NIR image before normalization; (c) VIS image after normalization; (d) NIR image after normalization. Images are from HFB database [9].

Based on the same experiment design and principle in previous sections, we show some real-time gender classification performance (Figure 3.7) to demonstrate the effectiveness of the proposed cross-spectrum gender classifier. Noted that the trained classifier is completely based on visible face images and is generalized to test on unseen NIR face images. The subjects presented in the training dataset and test set are not overlapped. The cross-modality gender prediction is thought to be harder than single-modality test. In the Figure 3.7, top shows the single-modality prediction and the bottowm row shows the cross-modality prediction.

Due to large spectra difference caused by VIS and NIR, the trained classifier are also expected to fail in many cases, demonstrated in Figure 3.8. Sometimes, this is unavoidable if the trained classifier is solely depending on VIS images and the test images are captured from NIR spectrum. It is possible to cope with such spectra difference by introducing NIR images in the training stage. For example, an image classification system is employed to determine whether the test image is from VIS spectrum or NIR spectrum. After that, the corresponding trained VIS classifier or NIR classifier is invoked to predict gender from the test image. Another way is to build a common subspace, instead of individual subspace to address those challenges.



(a) (b)

(c) (d)

Figure 3.7: Automatic prediction of gender from both VIS and NIR images. Images are from HFB database [9].

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

Figure 3.8: Failure cases of gender prediction from both VIS and NIR images. Images are from HFB database [9].

3.4.2 Fusion of VIS and NIR

Since there exists difference between individually trained VIS and NIR classifier, it is possible to utilize the information from both spectra to enhance the final classification. For

Table 3.4: Comparison of different classifiers and the fused results.

Classifier	**Rate**	**Male**	**Female**
Visible	0.9042	0.9324	0.8587
Near-infrared	0.7875	0.8041	0.7609
Fused(OR)	0.8542	0.9730	0.6630
Fused(AND)	0.8375	0.7635	0.9565

instance, some misclassified images in VIS domain would be corrected identified in NIR domain for the same set of subjects, and vice versa. This phenomena is verified from the statistical analysis of the following experiments,

- Both NIR-NIR and VIS-VIS are correct: 0.7375
- VIS-VIS is correct and NIR-NIR is incorrect: 0.1333
- VIS-VIS is incorrect and NIR-NIR is correct: 0.0708
- Both NIR-NIR and VIS-VIS are incorrect: 0.0583

For the same set of subjects, the gender of 73.75% of the samples are correctly identified by both VIS and NIR classifiers. This demonstrates the similarities between those two type of classifiers. On the other hand, 13.33% of the samples are misclassified by NIR classifier while the corresponding VIS classifier can successfully predict the gender. Such difference is also observed for the case where NIR classifier is accurate, but VIS classifier is not. Naturally, we want to fuse the results based on decision-level to enhance the performance. And the following performance is observed 3.4, The result is opposite to what we expect: the fused classifier is not better than the best individual classifier. One of the possible reasons is the use of the feature extraction method or SVM classifier. It is possible to see the improvement with other classifiers on different datasets. But this experiment provides us the instinct as how the individual classifier performs in VIS and NIR spectra, what they have in common and what is the difference.

3.4.3 Gender Prediction on NIR Dataset

In this section, gender classification methods in the NIR spectrum are further evaluated. There is limited work conducted on gender classification exclusively in the NIR dataset. Although gender prediction from NIR spectrum have been tested in HFB database, the size of that dataset is relatively small. With the extension of the performance evaluation on a larger and more challenging dataset, we can better understand gender classification algorithms in terms of spectrum difference.

The CBSR NIR Face Dataset [3] contains 3,940 NIR face images of 197 persons. The image size is 480×640 pixels. Each subject has 20 samples. We manually label the gender information and exclude those subjects that are not easily recognizable. In the end, 135 male subjects and 55 female subjects are identified. Samples of one subject are shown in Figure 3.9(a) and the corresponding normalized versions are shown in Figure 3.9(b). The normalization approach is based on manually located eye coordinates and is described in the work of Bolme et al. [4].

(a) (b)

Figure 3.9: (a). Original NIR face samples of one subject from CBSR dataset. (b). Normalized NIR face samples.

To test gender classification algorithms in the NIR dataset, 15 male subjects and 15 female

[3]CBSR: http://www.cse.ohio-state.edu/otcbvs-bench/

subjects are used for training. The total number of samples is $15 \times 20 + 15 \times 20 = 600$. The remaining 120 male and 40 female subjects are used for testing, resulting in a total of 3200 images. The subjects in the training and test sets are mutually exclusive in this experiment also.

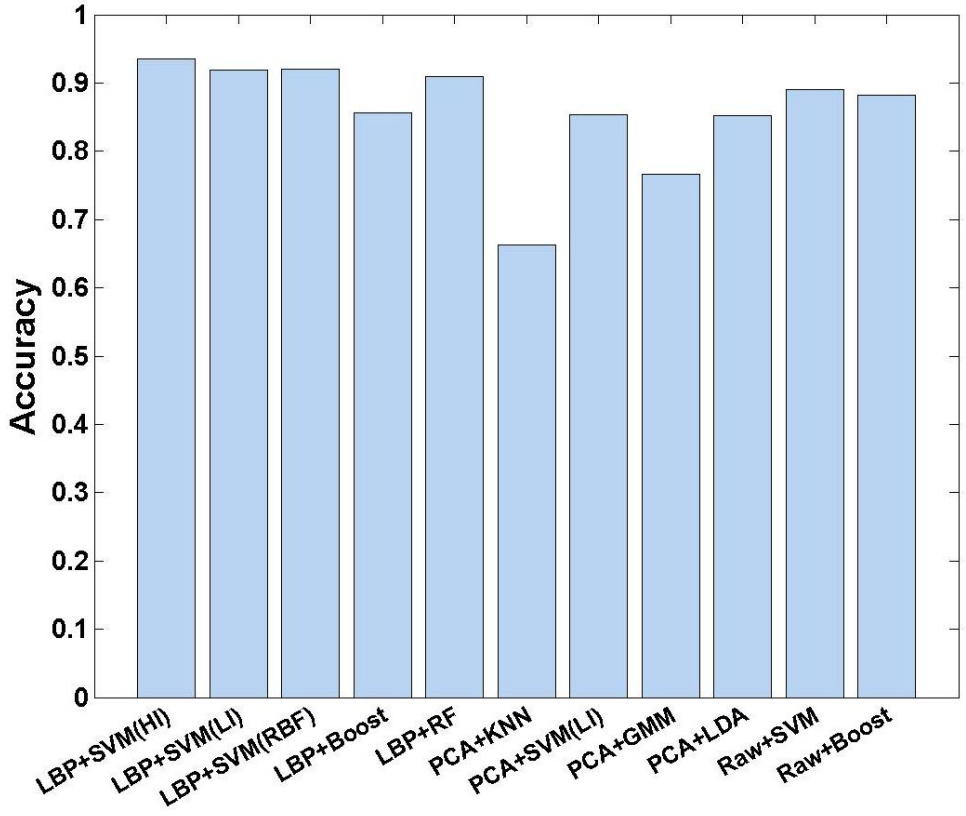


Figure 3.10: Comparison of different gender classification algorithms on CBSR dataset.

From the results shown in Figure 3.10, the LBPH descriptor of facial features gives best classification accuracy with LBP+SVM (HI) resulting in 93.59%, LBP+SVM (LI) resulting in 91.91% and LBP+SVM (RBF, $C = 32, \gamma = 0.001$) resulting in 91.97%. The results indicate that the HI kernel is superior to both the LI and RBF kernels in this case. The combination of LBP and Adaboost resulted in a classification rate of 84.58% for males and 88.5% for females. For the Random Forest classifier with LBP features (LBP+RandomForest), an overall accuracy of 91% is obtained. The individual accuracy of this method for male and female classes were 92.08% and 87.75%, respectively.

We also tested other type of facial features such as PCA and low-resolution image pixels

Algorithm	Overall	Male	Female
LBP+SVM (HI)	0.9359	0.9417	0.9187
LBP+SVM (LI)	0.9191	0.9208	0.9137
LBP+SVM (RBF)	0.9197	0.9375	0.8662
LBP+Adaboost	0.8556	0.8458	0.8850
LBP+RandomForest	0.9100	0.9208	0.8775
PCA+KNN	0.6634	0.6358	0.7462
PCA+SVM (LI)	0.8531	0.8721	0.7963
PCA+GMM	0.7659	0.8013	0.6600
PCA+LDA	0.8516	0.8533	0.8462
RawPixels+SVM [17]	0.8897	0.9083	0.8337
RawPixels+Adaboost [18]	0.8819	0.9033	0.8175

Table 3.5: Gender classification accuracies on near-infrared images using different feature extractors and classifiers.

(See Figure 3.10). For the PCA method, SVD is used to reduce the feature dimensionality to 60. The image was also resized to 32×32 pixels. Among all the PCA-based methods, the SVM (PCA+SVM(LI)) and LDA (PCA+LDA) classifiers result in the best performance, although this is still lower than the LBP-based methods. The remaining classifiers, such as GMM classifier (PCA+GMM) and KNN do not result in good performance.

The low-resolution feature representation has been observed to perform well in the visible spectrum [17, 18]. We apply the same methods used in the work of [17] and [18] on the NIR dataset. Each image was resized to 20×20. The number of weak classifiers used by the Adaboost classifier was 625 (Table 3.5). The SVM classifier used a Gaussian RBF kernel with $\gamma = 0.001$ and $C = 1$, based on a grid-search of the parameter space. Individual classification rates of 88.97% and 88.19% were obtained for SVM and Adaboost, respectively.

The above experiments show the effectiveness of employing LBP features for gender classification in NIR spectrum, compared to PCA or low resolution (RawPixels) features [17, 18].

3.5 Gender Prediction on Thermal Dataset

The thermal database contains one thermal face image each of 1003 subjects. In addition, this database contains two visible-light (VIS) images for each of these subjects. The image size is 480×640 pixels. The size of each image after alignment and cropping is 130×150 pixels. For the LBP methods, the image is resized to 126×90. There are 229 female subjects and 774 male subjects. The subjects have variations in age and ethnicity. Most of the samples are captured in the near-frontal pose (Figure 3.11).

Figure 3.11: Samples images from the thermal database. Top row shows male subjects and the bottom row shows female subjects.

To test gender classification algorithms on thermal images, the first 100 male and 100 female subjects are selected for training and the remaining is used for testing. This ensures that there is no overlapping of subjects between the training and test sets. Thus, there are 674 male subjects and 129 female subjects in the test set. The male (female) classification rate is defined as the percentage of males (females) that are correctly recognized within the male (female) groups. This distinction is to determine if there is any bias of individual groups towards overall classification performance due to imbalanced test data sets. The ratio of males and females in the test set is close to 5:1.

As shown in Table 3.6, the use of histogram intersection (HI) kernel in SVM results in better performance than the linear kernel (LI) in terms of the female classification rate. The LBP+SVM (HI) method achieves 84.50% accuracy for female classification, compared to

79.07% using the LBP+SVM (LI) method. The RBF kernel ($C = 8, \gamma = 0.002$) achieves better results than both HI and LI kernels, except for the female classification. However, it is much slower during the training stage as it searches a large parameter space to seek the optimum values. The performance is further enhanced with the use of PCA to derive a more compact feature descriptor, compared to LBP+SVM (LI). The reduced dimension feature vector is 60 in this experiment based on sigular value decomposition (SVD). Among all tested classifiers, SVM (LBP+PCA+SVM) results in the most balanced accuracy in terms of overall male and female accuracies. Here the linear SVM kernel is preferred, since the derived feature is no longer directly obtained from histogram bin features.

In order to show that LBPH descriptor is much more effective than PCA or low-resolution features extracted from thermal images, we perform experiments on the same dataset with the latter set of features (Figure 3.12). The best overall performance of 81.32% is achieved using the PCA+LDA methods. The individual performances for male and female are 80.71% and 84.50%, respectively.

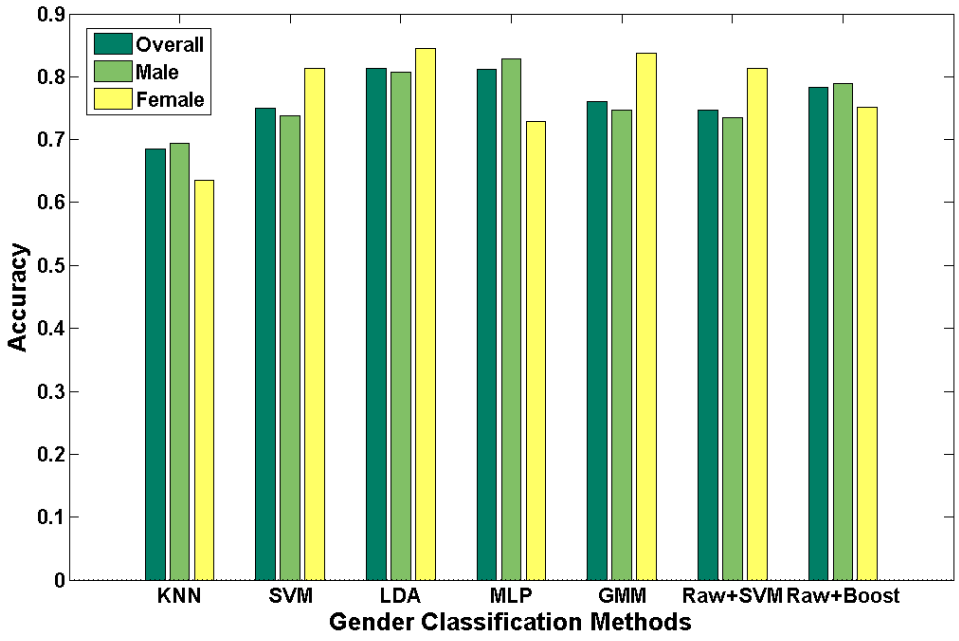


Figure 3.12: Performance evaluation of different gender classifiers based on PCA and low resolution features on the thermal dataset.

Since the thermal database includes visible-light images for each subject, gender classifica-

Algorithm	Overall	Male	Female
LBP+SVM (HI)	0.8792	0.8858	0.8450
LBP+SVM (LI)	0.8705	0.8858	0.7907
LBP+SVM (RBF)	0.9041	0.9184	0.8295
LBP+RandomForest	0.8655	0.8665	0.8605
LBP+PCA+SVM	0.9016	0.9110	0.8527
LBP+PCA+LDA	0.8667	0.8650	0.8760
LBP+PCA+MLP	0.8804	0.8828	0.8682
LBP+PCA+GMM	0.8742	0.8783	0.8527

Table 3.6: Gender classification accuracies on thermal images.

Algorithm	Overall	Male	Female
LBP+SVM (HI)	0.8842	0.8739	0.9380
LBP+SVM (LI)	0.8493	0.8398	0.8992
LBP+SVM (RBF)	0.8941	0.8828	0.9535
LBP+RandomForest	0.8667	0.8546	0.9302
LBP+PCA+SVM	0.9066	0.9036	0.9225
LBP+PCA+LDA	0.8804	0.8724	0.9225
LBP+PCA+MLP	0.8643	0.8501	0.9380
LBP+PCA+GMM	0.8269	0.8205	0.8605

Table 3.7: Gender classification accuracies on visible images in the thermal database.

tion is conducted on the visible spectra as well. It must be noted that the pairs of visible and thermal images are not co-registered. From the results in Table 3.7, the LBP+PCA+SVM method has the best overall performance at 90.66%. The RBF kernel ($C = 2, \gamma = 0.002$) is slightly better than the HI kernel. Meanwhile, the HI kernel is much better than the LI kernel. Compared to the LBP+SVM (LI) method, the enhanced feature reduction and classification methods such as LBP+PCA+SVM, LBP+PCA+LDA and LBP+PCA+MLP achieve better results. The average overall performance of all the seven algorithms on the thermal dataset is 88.03%, compared to 87.16% on the visible dataset. This suggests the feasibility of assessing gender from thermal images as well. This is the first work in the literature that establishes this possibility.

The visual appearance of facial thermal image is remarkably different from that of visible-light and near infrared face images. This poses a big challenge for humans to recognize gender information from face images (cropped) if the entire upper-body image is not available (See Figure 3.11). However, machine learning approaches treat a face image as a bunch of pixels and automatically select the most relevant features from an image to perform gender classification.

In order to compare the performance of the proposed approach against that of humans, we asked four human subjects (observers) to perform gender prediction independently. Only the cropped version of thermal images were provided to human subjects and they were asked to assign one of two labels to each image: male or female. Human observers tend to classify males or females based on presence or absence of mustache, beard and eyelids, which are still observable in the thermal spectrum. Ambiguities arise when such distinctive features are not available in the face images. As seen in Table 3.8, more females are misclassified as males while fewer males are misclassified as females. This is expected since females are usually much more difficult to classify than males when important facial features are missing. Since this database was collected by law enforcement department, the presented male subjects exposed features such as facial hair, mustache, and so on. On the other hand, the machine learning approach automatically selects relevant facial features from thermal images to make this distinction. This experiment demonstrates the advantage of using machine learning approaches for gender classification in complex scenarios involving non-

Observers	Overall	Male	Female
Subject A	0.9142	0.9585	0.6822
Subject B	0.8968	0.9748	0.4884
Subject C	0.8918	0.9496	0.5891
Subject D	0.9080	0.9896	0.4806
Machine	0.9016	0.9110	0.8527

Table 3.8: Gender classification accuracies reported on thermal images based on human perception. Subject A and B are male observers and subject C and D are female observers. Note that the subjects were not very good at classifying female face images.

traditional spectrum. The machine algorithm performance reported in Table 3.8 is based on the LBP+PCA+SVM method in Table 3.6.

3.6 Chapter Summary

This Chapter presents initial experimental results on gender classification using NIR images. A classification accuracy of 84.42% was obtained in the NIR-NIR scenario. The work reported in this chapter represents the first step toward cross-spectral gender recognition where training images and test images originate from different spectral bands. The preprocessing operation involving illumination normalization was observed to improve cross-spectral classification accuracy by up to 18%. But this is still lower than the performance obtained for intra-spectral classification (i.e., VIS-VIS and NIR-NIR scenarios). The trained VIS classifier is extended to unseen NIR images to test the performance in real-time applications.

Currently, we are examining the use of fundamental image formation models to better understand the gender-specific details present in NIR and VIS images. Further, we anticipate that the use of other gender classification approaches (based on LBP and Haar features) may be necessary to improve cross-spectral gender prediction.

In the end, we revisit the NIR-based gender classification on a large-scale database and show that LBP representation with intersection kernel gives very good performance. It

achieves 93.59% accuracy on the test size of 3200 NIR images. Gender classification results with thermal images are also enumerated to show the effectiveness of LBP features.

Chapter 4

Facial Attributes based Classification

4.1 Study of Facial Attributes

In the previous chapters, gender classification has been the main focus of our work, which is one of many facial-based attributes. According to the definition of [82], an attribute classifier is a binary classifier trained to recognize the presence or absence of describable aspects of visual appearance (e.g., gender, ethnicity, and age). Some attributes such as ethnicity and age can be divided into more than two classes. For example, an age classifier can have meta-groups such as "kids", "youth" and "seniors". To mitigate the difference, attribute classification can be posed as a binary classification or a multi-classification problem. In the scenario of multi-class classification, the attribute classifier is used to categorize each attribute to one of several groups.

The ability of current search engines to find face images are based on semantic annotations. Thus, one has to manually label the relevant attributes in an image and build the relationship between the actual content and the associated textual annotations [83]. Such a procedure is time consuming and becomes challenging as the Internet grows rapidly with a host of pictures and videos. Researchers have been working on content-based image retrieval (CBIR) for a long time. Unlike text-based searching, the content-based approach analyzes the contents of the images based on colors, shapes and textures cues.

Recently, there has been a focus on attributes-based searching and classification. For example, users might ask questions as "Males with mustaches" or "Young blonde lady" [83].

The search engine should be able to return results for such specific queries. Such queries include attributes such as gender (Male, Lady), age (Young) and others (Mustache). The gender classification methods discussed earlier can be directly incorporated into the system. In this chapter, other attributes such as age, ethnicity, and expressions are investigated. These attributes are very common and constitute search queries. Besides, the attributes can be directly inferred from the visual appearance of facial regions.

There are few known works that combine all the main attributes together due to complexity of each individual attribute-based classification. Lyons et al. [84] proposed a system that was trained from face image exemplars to classify faces on the basis of high level attributes, such as gender, "race", and expression. The faces were represented by elastic graphs labelled with 2D Gabor wavelet features. Wilhelm et al. [85] compared different feature extraction approaches and classification methods for the task of gender, age, facial expression and identity classification. Gao et al. [28] used attributes of gender, ethnicity and age for photo album management and visual surveillance monitoring. In Kumar et al. [82], the attribute classifier was efficiently incorporated into a face verification system in order to improve the performance. They also built an attributes-based search engine for large collections of images with faces [83].

In the following work, we will describe expression, age and ethnicity based attribute classifiers. The remaining of this chapter will introduce attributes-based classification, including expression classification (Section 4.2), age estimation and classification (Section 4.3), and ethnicity classification (Section 4.4). Each attribute classifier is briefly described and illustrated with experiments.

4.2 Expression Classification

Facial expression recognition has found potential applications in many areas such as human-computer interaction, image retrieval, animation and human motion analysis [86]. Most existing work on facial expression recognition utilize appearance information to classify different expressions [87]. These facial feature extraction methods can be categorized into two groups: Image-based and Model-based. Image-based methods encode facial expression

features using Gabor wavelets [53], LBP [88], Facial Action Units [89] and so on. Model-based method is an alternative to image-based feature extraction scheme. Typical approaches include AAM [90, 91] and 3D deformable models [92]. The advantage of using image-based method to extract expression features is that it does not require extensive knowledge about the object of interests (e.g., face). Further, it is relatively fast and simple to compute. Hence, we provide an image-model based analysis using appearance information from face images to train and classify expressions. Here the attributes contain different facial expressions.

To test the algorithm of classifying expressions, the Taiwanese Facial Expression Image Database (TFEID)[1] is used. The TFEID consists of 7200 face images captured from 40 models, each with eight facial expressions: neutral, anger, contempt, disgust, fear, happiness, sadness and surprise. A subset of images with frontal pose is selected. The total number of images is 336, with an average of 42 samples per expression. The eight different facial expressions are illustrated in Figure 4.1. From top to bottom, the expressions are anger, contempt, disgust, fear, happy, neutral, sadness and surprise. All the image samples are masked using an eclipse-fitting method to isolate the facial region [7]. The purpose is to eliminate the noise information associated with the background and retain only essential information about facial expression. The subjects presented within each expression category are different.

The above mentioned facial expression classification is a multi-class classification problem on the TFEID database. In some applications, a binary-class classification problem such as discrimination between happy faces and neutral faces is sufficient. To illustrate this problem, a database with only happy faces and neutral faces is used [2]. The face database contains seven facial expressions. But we select only the happy and neutral expressions. Each facial expression has subjects aged from 20 to 90 years old, with both males and females. In the end, there are 209 subjects with both happy and neutral expressions. Each subject has the pair of happy-neutral expressions. To ensure that the subjects are not overlapped between the two classes, we randomly select 100 subjects with happy expression and the remaining 109 subjects with neutral expression. The examples of the images are shown in Figure 4.2. All of the samples have been normalized and appropriately processed with background removal.

[1] http://bml.ym.edu.tw/ download/html/

Figure 4.1: Eight different facial expressions from TFEID database. Each row represents one facial expression. From top to bottom, the expressions are anger, contempt, disgust, fear, happy, neutral, sadness and surprise.

The experiments are conducted by randomly selecting 40% samples of each class as training set, and the remaining 60% as the test set. The methods adopted to evaluate the facial expression classification consists of LBP+SVM, Gabor+SVM, LPP+SVM and RawPixels+SVM. This includes descriptors of LBP, Gabor, LPP and Raw Pixels. LBP and Gabor description of facial expressions have been well studied for expression recognition [53, 88]. The expression classification rates for the two databases are shown in Fig-

Figure 4.2: Samples from Lifespan face database. Top Row: neutral expression; Bottom Row: happy expression.

ure 4.3. DFH_GRAY refers to the gray scale version of TFEID database. From the Fig-

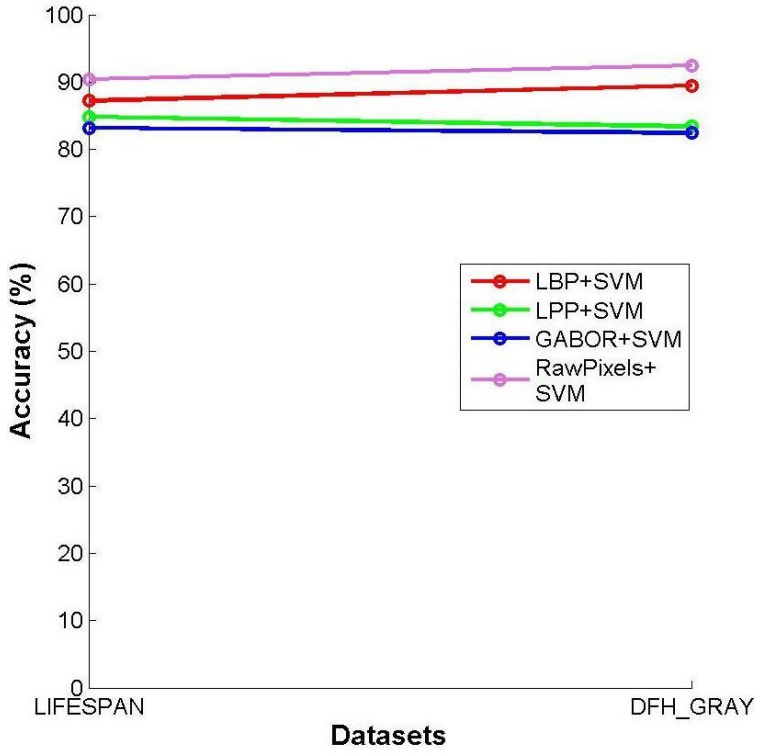


Figure 4.3: Facial expression classification on the two databases.

ure 4.3, the RawPixels+SVM and LBP+SVM methods perform better than LPP+SVM [2] and Gabor+SVM methods. The RawPixels+SVM achieves 90.4% accuracy on the Lifespan

[2]LPP code: www.zjucadcg.cn/dengcai/Data/data.html

database and 92.5% accuracy on the DFH_GRAY dataset. In both datasets, the LBP+SVM exceeds the performance of Gabor+SVM and produces comparable results with that of RawPixels+SVM. Although the descriptor using RawPixels is better than LBP and Gabor descriptors, it is not necessarily true that the Raw Pixel representation is much more superior. Since our datasets have already been normalized to exclude any pose and background changes, it is expected that Raw Pixel descriptor would perform much better here.

Another metric for the performance measurement of facial expression classification is the confusion matrix. We rerun the experiment of applying RawPixels+SVM method on DFH_GRAY dataset to obtain the results (Table 4.2). Surprisingly, there are no misclassifications in the happy and neutral classes. That indicates happy and neutral expressions are easily separable compared to other classes. Those two expressions are also the most common ones encountered [87].

Table 4.1: Expression classification on DFH_GRAY dataset based on confusion matrix.

Classification	Anger	Contempt	Disgust	Fear	Happy	Neutral	Sad	Surprise
Anger	**18**	0	0	0	0	1	1	0
Contempt	0	**37**	1	1	0	1	0	0
Disgust	0	1	**22**	0	0	0	1	0
Fear	0	0	1	**21**	0	1	1	0
Happy	0	0	0	0	**24**	0	0	0
Neutral	0	0	0	0	0	**23**	0	0
Sad	0	0	2	0	0	1	**20**	0
Surprise	2	0	0	0	0	0	0	**19**

4.3 Age Estimation and Classification

Human age is another important personal attribute or trait, which can be directly inferred from facial appearance. Age estimation [93] is the determination of a person's age based on facial features, although other biometric traits can be used. By contrast, age classification (coarse age classification) assigns a person into one of several categories such as kid, young, middle aged and senior. The output from age estimation is a scalar value indicating predicted

age from that person. On the contrary, the result from age classification is a label pointing out which group that person belongs to. Usually, the task of age classification is much easier than age estimation in terms of complexity.

Face aging is a slow, irreversible and complex procedure that is mixed with many uncertainties [94]. During the early stage of growth and development of face, the greatest change is craniofacial growth in the form of shape changes [95]. After that, the skin or texture change is the dominant factor in adult aging. Such changes of patterns from childhood to adult can be observed in Figure 4.4. Although we use normalized shapes to make all the samples shape-free, the relative size or distance between important facial features are different.

The motivation to study age estimation or classification arises from many aspects. Aging effect often decreases the performance of many biometrics systems, such as gender recognition [96], face recognition [97] and expression recognition [98]. Age-related difference in these classification systems reflects the importance of studying age estimation or age classification methods. Classification of age information in advance might improve the system performance. Another reason is that the ability of determining age information from human faces have many other real-world applications, ranging from security control, human-computer interaction to forensic Art [95]. Common appearance-based approaches include Gabor features with fuzzy LDA [99], ICA-based local facial features [100] and LBP [21].

The FGNET database has 82 subjects. Each subject has images collected across different age groups. The minimum age is 0 and the maximum age is 69. The total size of the database is 1002 samples. Samples of the database can be seen from Figure 4.4. All the face images are processed with pose and shape normalization based on previous methods. Normalization stabilizes the shape difference and makes the appearance information most prominent factor in consideration. This is consistent with the previous approaches that we used to extract facial appearance information. Besides, the spatial distribution of facial component is also captured in global-based approaches. Age estimation is posed as a regression problem. Consider a sequence of face images $\{x_i, y_i\} : i = 1 \cdots n$, where x_i is the facial appearance information and y_i is a scalar value indicating the person's age. The objective of regression

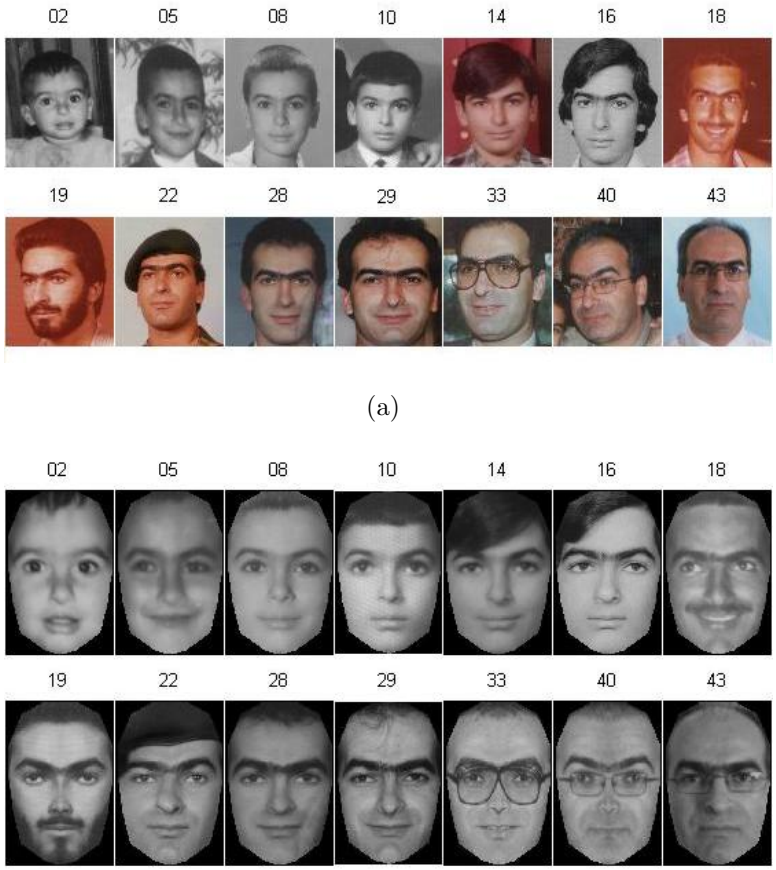


Figure 4.4: One subject across different ages in FGNET database. The number indicates the actual age of that sample. (a) Original samples from FGNET. (b) Samples after pose normalization.

analysis is to minimize the following error,

$$err = \sum_{i=1}^{n}(y_i - f(x_i))^2. \tag{4.1}$$

The simple linear regression model assumes that the relationship between y_i and x_i is linear, and can be modeled as $y = mx + b$. Thus the equation becomes,

$$err(m, b) = \sum_{i=1}^{n}(y_i - mx_i - b)^2. \tag{4.2}$$

Taking the partial derivatives with respect to the parameters m and b will lead to the solution of the regression problem. However, relationship between age (y_i) and facial appearance (x_i) is not linear. Therefore, nonlinear regression model can more accurately characterize the relationship between those two variables. The basic idea of Support Vector Regression (SVR) [64] is to search for a function $f(x_i)$ that has at most ϵ deviation from the target value y_i for the training data $x_i, i = 1, \cdots, n$. At the same time, $f(x_i)$ should be as flat as possible. Such requirements make SVR less sensitive to outliers than linear regression or quadratic regression [101]. The linear SVR is described as,

$$f(x) = \langle w, x \rangle + b \tag{4.3}$$

The objective function is given by,

$$\min_{w,\xi} \frac{1}{2} \parallel w \parallel^2 + C \sum_{i=1}^{n}(\xi_i^+ + \xi_i^-) \tag{4.4}$$

subject to

$$\begin{aligned} y_i - \langle w, x_i \rangle - b &\leq \epsilon + \xi_i^+ \\ \langle w, x_i \rangle + b - y_i &\leq \epsilon + \xi_i^- \\ \xi_i^+, \xi_i^- &\leq 0 \end{aligned} \tag{4.5}$$

where the constant and positive C determines the trade-off between the flatness of function f and the data deviation ϵ. ξ_i^+ and ξ_i^- are the imposed constrains on the optimization problems. Such an optimization problem can be efficiently solved in the dual formation of the former problem [102]. The linear SVR can be transformed into a non-linear SVR via the kernel trick [64].

To demonstrate age estimation using SVR, the FGNET database with 82 subjects is used. Subjects with duplicated age samples are removed, resulting in a total of 988 samples. We select the first 40 subjects for training. The remaining subjects are reserved for testing. First, LBP is applied to extract facial features from each block that encodes gender information, resulting in a feature dimension vector of 59. Then all the extracted features are concatenated to form the holistic feature vector $x_i \in R^{2478}$. Finally, the feature vector x_i and its labeled age information y_i is modeled as a SVR regression problem. The performance of age estimation is given by the mean absolute error (MAE),

$$MAE = \sum_{k=1}^{N} |\hat{s_k} - s_k|/N \tag{4.6}$$

where s_k is the age ground truth and $\hat{s_k}$ is the estimated age. The reported MAE is **7.66** on the 496 samples of FGNET database. That means the average error between predicted age and actual age is around 8. This performance is comparable to the results obtained by others [103, 104]. It is only within 3 years compared to recent work in [101] which reports an MAE of **5.07**. However, the correlation value (one of the outputs from SVM) between the predicted labels and the ground truth is 0.57. It is possible that with the use of more complex features, the performance could be improved further. Some of the prediction results are shown in Figure 4.5. It appears that the ages of young children are more difficult to determine than that of young adults.

Age classification is different in the sense that it is a classification rather than a regression problem. In the Lifespan database, there are no subjects in the age group 0 to 15 (i.e.,"kid"). The minimum age is 18 years old. Only faces with neutral expressions are selected. The database is divided into age groups of 18-29 (223), 30-49 (76), 50-69 (123) and 70-94 (158). The numbers in the parentheses are the number of subjects in each category. Each subject has only one sample available. The reported experimental results are shown in Figure 4.6. All the feature extraction methods use the SVM classifier. The best performance on this database is achieved by LDA-based feature extraction methods with an accuracy of 71%. The LBP descriptor does not perform well in this case. All remaining feature extraction methods have comparable performance. The overall accuracy on this database for age classification is relatively low, compared to other attribute classifiers. One explaination is that there are

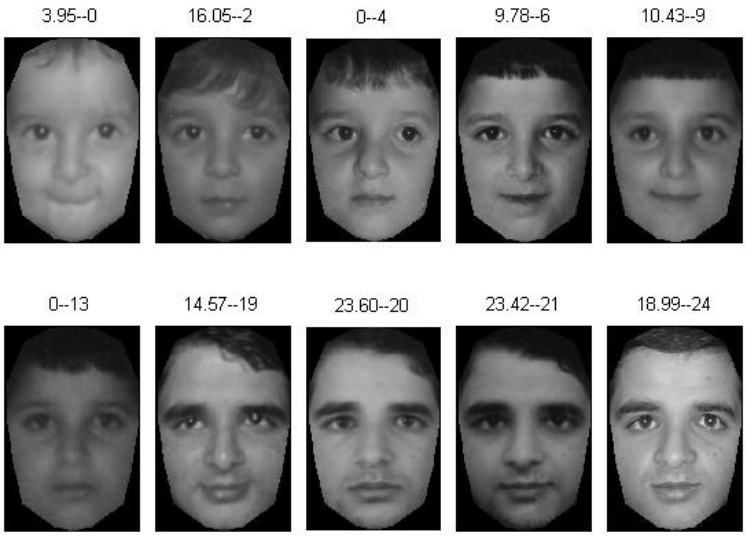


Figure 4.5: Age prediction results for one subject across different ages in FGNET database. The left number is the predicted age and the right number is the actual age.

some overlaps between age groups of 18-29 and 30-49 or age groups of 30-49 and 50-69. Another explanation is that the aging process varies from person to person. Only extracting texture information from the facial appearance might not be disciminative enough to indicate the presence of age information. Some subjects might look older than their actual age.

4.4 Ethnicity Classification

Ethnicity is another demographic attribute that can be deduced from human face images. Lu et al. [105] formulated the study of image-based ethnicity as a machine learning problem. The Linear Discriminant Analysis (LDA) was presented to separate two classes (Asian vs non-Asian).

In the work of [106], it was shown that there exists Anthropometric statistic difference among different racial groups (i.e., Caucasian, African-American, and Asians). For example, the Asian group had the widest faces. Those results were obtained from the measurements of 25 defined facial landmarks. The localization of facial landmarks can be accomplished by applying Active Appearance Model (AAM) or Active Shape Model (ASM). However, current feature localization methods cannot achieve robust results that are essential for the

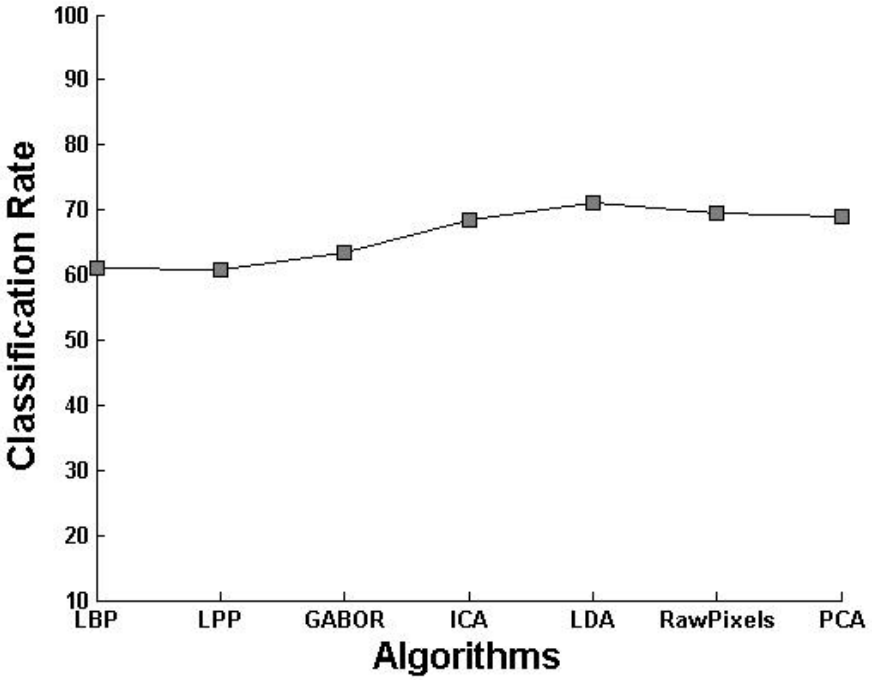


Figure 4.6: Age classification results across different age groups.

measurement of Anthropometric statistics from human face images. Instead we address the problem of ethnicity identification based on gray-scale human face images. The appearance-based scheme has demonstrated its power in gender and expression recognition. To simplify the task, ethnicity recognition is formulated as a two-class classification problem (Asian or Non-Asian). Here, "Asian" refers to people from South Korea, Japan and China. "Non-Asian" refers to Caucasian, African-American and others. Samples of the database are shown in Figure 4.7.

The ethnicity database is a combination of four different face databases, all of which are available in public websites. The database is separated into two groups: Asian and Non-Asian. The Asian group is composed of 188 subjects from CUHK student dataset [107] and 100 subjects from HFB dataset [9]. The non-Asian group is composed of 79 subjects from FERET [7] and 209 subjects from Lifespan [2]. Each subject has only one sample. Therefore, the size of Asian and Non-Asian dataset is the same, each having 288 images. Within the same ethnicity group, the subjects have different expressions, ages and illumination factors.

The experiments are carried out by randomly splitting the training and test set as

(a) (b)

Figure 4.7: Representative faces from the four selected database. (a) Asian; (b) Non-Asian.

40/60(%). The final classification accuracy is the average over 10 random trials (Figure 4.8). As can be seen, the LBP and Gabor representations of ethnicity features produce the best classification rates of 98.0% and 97.4%, respectively. All other descriptors such as PCA, LDA and RawPixels only decline by a small margin. All the feature extraction algorithms adopt the same SVM classification scheme. The average performance of all the algorithms is 97.23%.

Two factors may contribute to the relatively high ethnicity classification performance: one is the use of the SVM classifier and the other could be the use of a small database.

4.5 Chapter Summary

This chapter extends previous work on gender-based attribute classifier to expression, age and ethnicity attributes. Such an extension only requires the adaption of feature extraction methods from one domain to another. Similar to the gender attribute that is categorized as male and female, the expression attribute is categorized as Happy and Neutral, while the ethnicity attribute is classified as Asian and Non-Asian. The age attribute differs slightly since it contains more categories. It would be a little trivial to classify age as young and old. Instead, we divide age into kid, youth, adult and senior.

All the attribute classifiers are evaluated separately on different datasets to show the

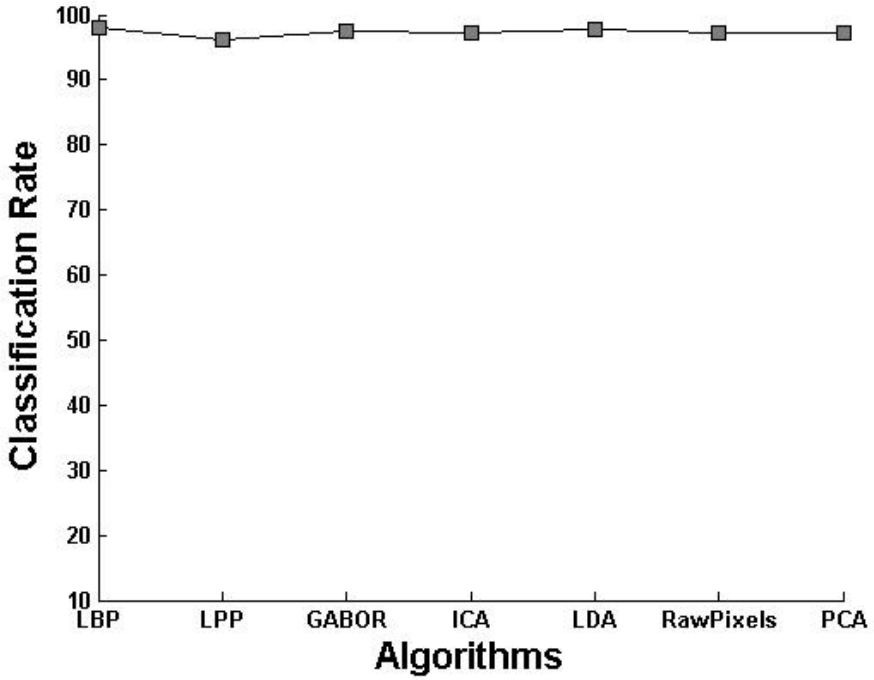


Figure 4.8: Ethnicity classification results with different feature extraction methods.

effectiveness of the proposed scheme. For expression classification, we achieve a 90.4% accuracy on the Lifespan database and 92.5% accuracy on the DFH GRAY dataset. For age estimation, the reported MAE is 7.66 on the 496 test samples of the FGNET database. For age classification using 4 age groups, an accuracy of 71% was obtained on the LifeSpan Database. Ethnicity classification has the highest performance of 98% on a collected database.

Chapter 5

Conclusion

5.1 Summary

Gender classification has been shown in recent literature to be a useful attribute providing many potential applications in computer vision and biometrics. Different feature extraction and classification methods with respect to gender classification were investigated and revisited in order to provide readers with an understanding about ongoing research work in this area (Chapter 1).

First, we introduced gender classification methods from the perspective of feature extraction and classification, along side two pre-processing routines: face detection and normalization. Then the methods were evaluated in various scenarios using public databases in Chapter 2. Since gender classification has not been studied before for near-infrared (NIR) and thermal (THM) images, we extended the research work to the NIR and THM domain where the utilization of NIR and THM information can help cope with illumination changes. To handle cross-spectrum gender classification, where the trained classifiers are based on visible spectrum (VIS) images and the test images are from the NIR domain, we used illumination normalization to alleviate the difference between these two spectra. In addition, we showed that the gender information extracted from NIR images can be as discriminative as VIS images, though the performance degrades a little bit. This provides empirical evidence of the possibility of applying existing gender classifiers to NIR images. Our experimental results performed on different datasets verify the effectiveness of the proposed approaches

(Chapter 3).

Finally, we explored other attribute-based classifiers. In particular, expression, age and ethnicity were explored, as they too can be directly inferred from visual appearance of face images. The objective is to build classifiers that can automatically predict gender, expression, age and ethnicity from facial images. Such attribute classifiers can also be incorporated into content-based retrieval systems to develop an attribute-based face search engine (Chapter 4).

5.2 Contributions

The contributions of this thesis are summarized as follows:

1. Gender recognition techniques were revisited and extensively evaluated using single database as well as cross-database tests. This provides researchers with some guidelines on the selection of effective feature extraction and classification methods. The reported experimental results can also serve as benchmark tests.

2. We extended the problem of gender classification to near-infrared (NIR) and thermal (THM) domains. Limited work has been done in the NIR and THM spectra to predict gender from face images. Our experiments suggest that gender can be predicted from NIR and THM images with reasonable accuracy.

3. We proposed the concept of cross-spectrum gender classification, where the trained classifiers are derived using VIS images and tested on NIR images (and vice-versa). To overcome spectrum difference between those two domains, we used illumination normalization approaches. The effectiveness of these approaches were evaluated in the experiments.

4. Gender classification was extended to include other attributes such as expression, age and ethnicity which were estimated from facial images.

5. A gender classification toolbox was implemented.

Appendix A

Toolboxes Used and Implemented

A.1 Sources

This work uses several existing software. The main toolboxes are listed below:

- Statistical Pattern Recognition Toolbox
- The Matlab Toolbox for Pattern Recognition
- MATLAB Classification toolbox
- Matlab Toolbox for Dimensionality Reduction
- The INface Toolbox for Illumination Invariant Face Recognition

Some of the codes were modified in order to adapt to specific tasks. Thanks to the authors who shared their work for use by others in the research community. To search for these toolboxes, please Google the keywords.

A.2 Gender Classification Toolbox

The gender classification toolbox developed in this thesis includes two parts: (a) single-database test of different gender classification algorithms and (b) cross-database test of these algorithms. Other algorithms can be easily integrated into the system.

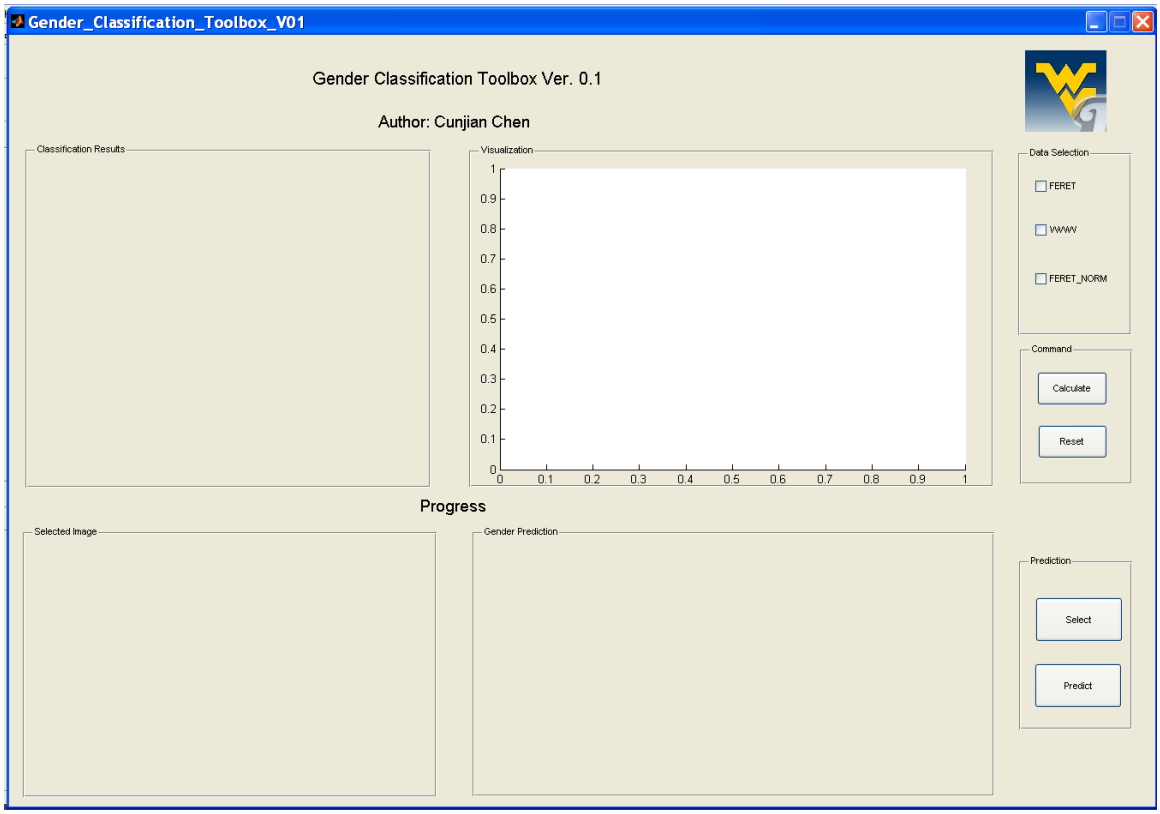


Figure A.1: Initial GUI of the system

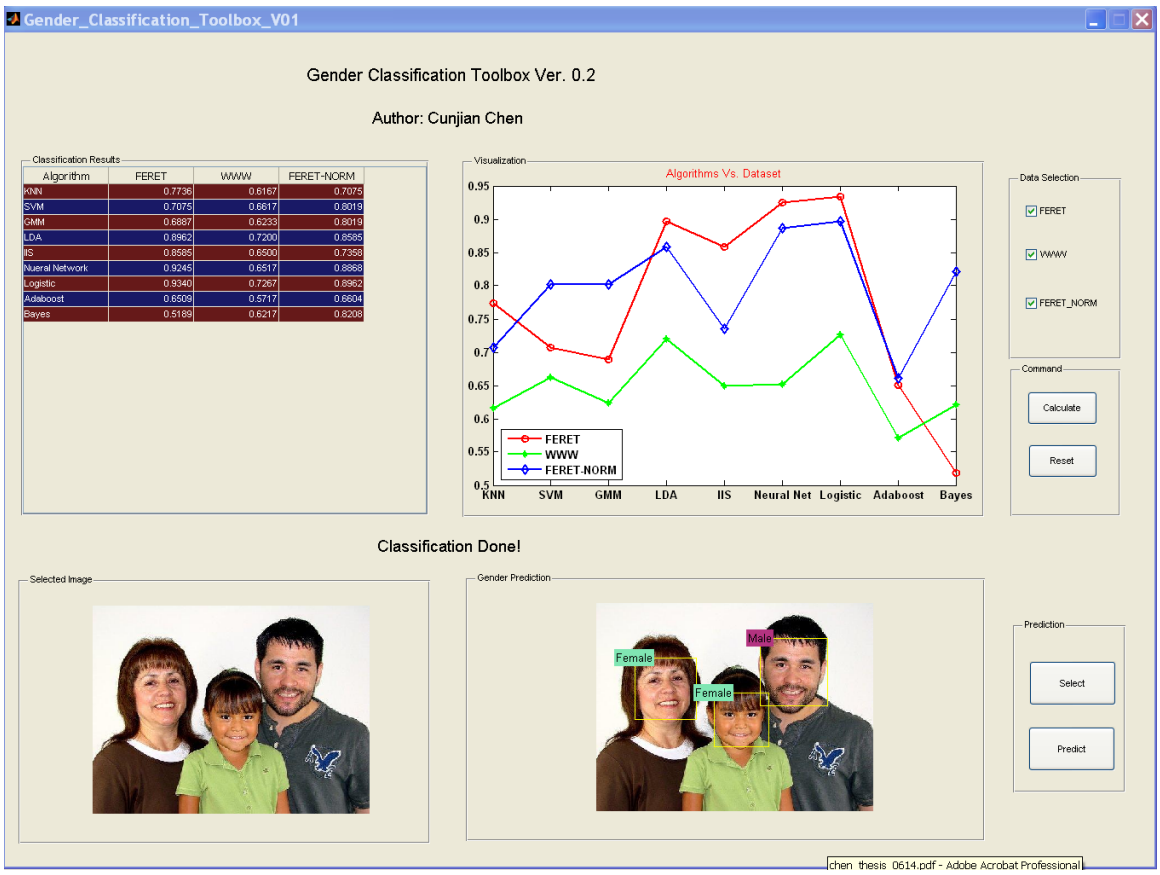


Figure A.2: Gender classification results displayed by the system

References

[1] Andrew C. Gallagher and Tsuhan Chen, "Understanding images of groups of people," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 256–263.

[2] Meredith Minear and Denise C Park, "A lifespan database of adult facial stimuli.," *Behavior Research Methods, Instruments, and Computers*, vol. 36, no. 4, pp. 630–3, 2004.

[3] Stephen Milborrow and Fred Nicolls, "Locating facial features with an extended active shape model," *European Conference on Computer Vision*, pp. 504–513, 2008.

[4] David S. Bolme, J. Ross Beveridge, Marcio Teixeira, and Bruce A. Draper, "The CSU face identification evaluation system: Its purpose, features and structure," in *In International Conference on Vision Systems*, 2003, pp. 304–311.

[5] M. M. Nordstrøm, M. Larsen, J. Sierakowski, and M. B. Stegmann, "The IMM face database - an annotated dataset of 240 face images (technical report)," May 2004.

[6] A.M. Martnez and R. Benavente, "The AR-face database," in *CVC Technical Report 24*, 1998.

[7] P. Jonathon Phillips, Hyeonjoon Moon, Syed A. Rizvi, and Patrick J. Rauss, "The FERET evaluation methodology for face recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.

[8] J. Aghajanian, J. Warrell, S.J.D. Prince, P. Li, J.L. Rohn, and B. Baum, "Patch-based within-object classification," in *International Conference on Computer Vision*, 2009, pp. 1125–1132.

[9] Stan Z. Li, Zhen Lei, and Meng Ao, "The HFB face database for heterogeneous face biometrics research," *International Conference on Computer Vision and Pattern Recognition Workshop*, pp. 1–8, 2009.

[10] Erno Makinen and Roope Raisamo, "Evaluation of gender classification methods with automatically detected and aligned faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 3, pp. 541–547, 2008.

[11] Anil K. Jain, Sarat C. Dass, and Karthik Nandakumar, "Can soft biometric traits assist user recognition?," in *SPIE Conference on Biometric Technology for Human Identification*, 2004, pp. 561–572.

[12] Abigail J. Stewart and Christa McDermott, "Gender in Psychology," *Annual Review of Psychology*, vol. 55, no. 1, pp. 519–544, 2004.

[13] Xiao-Chen Lian and Bao-Liang Lu, "Gender classification by combining facial and hair information," in *Advances in Neuro-Information Processing*. 2008, pp. 647–654, Springer-Verlag.

[14] Jin-Li Suo, Liang Lin, Shiguang Shan, Xilin Chen, and Wen Gao, "High-resolution face fusion for gender conversion," *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, vol. 41, no. 2, pp. 226–237, 2011.

[15] B. A. Golomb, D. T. Lawrence, and T. J. Sejnowski, "Sexnet: A neural network identifies sex from human faces," in *Advances in Neural Information Processing Systems*, 1990, pp. 572–577.

[16] Srinivas Gutta, Harry Wechsler, and P. Jonathon Phillips, "Gender and ethnic classification of face images," in *International Conference on Face and Gesture Recognition*, 1998, pp. 194–199.

[17] Baback Moghaddam and Ming-Hsuan Yang, "Learning gender with support faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 707–711, 2002.

[18] Shumeet Baluja and Henry A. Rowley, "Boosting sex identification performance," *International Journal of Computer Vision*, vol. 71, no. 1, pp. 111–119, 2007.

[19] C. BenAbdelkader and Paul Griffin, "A local region-based approach to gender classification from face images," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2005, vol. 3, p. 52.

[20] Timo Ahonen, Abdenour Hadid, and Matti Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 2037–2041, December 2006.

[21] Zhiguang Yang and Haizhou Ai, "Demographic classification with local binary patterns," in *International Conference on Biometrics*, 2007, pp. 464–473.

[22] Ning Sun, Wenming Zheng, Changyin Sun, Cairong Zou, and Li Zhao, "Gender classification based on boosting local binary pattern," in *Advances in Neural Networks*, 2006, pp. 194–201.

[23] Guodong Guo, C.R. Dyer, Yun Fu, and T.S. Huang, "Is gender recognition affected by age?," in *International Conference on Computer Vision Workshops*, 2009, pp. 2032 –2039.

[24] J.G. Wang, J. Li, W.Y. Yau, and E. Sung, "Boosting dense SIFT descriptors and shape contexts of face images for gender recognition," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2010, pp. 96–102.

[25] Zehang Sun, George Bebis, Xiaojing Yuan, and Sushil J. Louis, "Genetic feature subset selection for gender classification: A comparison study," in *Proceeds of Applications of Computer Vision*, 2002, pp. 165–170.

[26] Duan-Yu Chen and Kuan-Yi Lin, "Real-time gender recognition for uncontrolled environment of real-life images," in *International Conference on Computer Vision Theory and Applications*, 2010, pp. 357–362.

[27] Caifeng Shan, "Gender classification on real-life faces," in *Advanced Concepts for Intelligent Vision Systems*, 2010, pp. 323–331.

[28] Wei Gao and Haizhou Ai, "Face gender classification on consumer images in a multiethnic environment," in *International Conference on Advances in Biometrics*, 2009, pp. 169–178.

[29] Matthew Toews and Tal Arbel, "Detection, localization, and sex classification of faces from arbitrary viewpoints and under occlusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 9, pp. 1567–1581, 2009.

[30] Simon J. D. Prince and Jania Aghajanian, "Gender classification in uncontrolled settings using additive logistic models," in *International Conference on Image Processing*, 2009, pp. 2557–2560.

[31] K. Balci and V. Atalay, "PCA for gender estimation: Which eigenvectors contribute?," in *International Conference on Pattern Recognition*, 2002, pp. 363–366.

[32] J. Bekios Calfa, J.M. Buenaposada, and L. Baumela, "Revisiting linear discriminant techniques in gender recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 4, pp. 858–864, April 2011.

[33] Amit Jain and Jeffrey Huang, "Integrating independent components and linear discriminant analysis for gender classification," in *IEEE International Conference on Automatic Face and Gesture Recognition*, 2004, pp. 159–163.

[34] Asifullah Khan, Abdul Majid, and Anwar M. Mirza, "Combination and optimization of classifiers in gender classification using genetic programming," *International Journal of Knowledge-based and Intelligent Engineering Systems*, vol. 9, pp. 1–11, January 2005.

[35] Hyun-Chul Kim, Daijin Kim, Zoubin Ghahramani, and Sung Yang Bang, "Appearance-based gender classification with gaussian processes," *Pattern Recognition Letters*, vol. 27, pp. 618–626, April 2006.

[36] Liangliang Cao, Mert Dikmen, Yun Fu, and Thomas S. Huang, "Gender recognition from body," in *ACM Multimedia*, 2008, pp. 725–728.

[37] Guodong Guo, Guowang Mu, and Yun Fu, "Gender from body: A biologically-inspired approach with manifold learning," in *Asian Conference on Computer Vision*, 2009, pp. 236–245.

[38] Shiqi Yu, Tieniu Tan, Kaiqi Huang, Kui Jia, and Xinyu Wu, "A study on gait-based gender classification," *IEEE Transactions on Image Processing*, vol. 18, pp. 1905–1910, 2009.

[39] Pawan Kumar, Nitika Jakhanwal, Anirban Bhowmick, and Mahesh Chandra, "Gender classification using pitch and formants," in *International Conference on Communication, Computing and Security*, 2011, pp. 319–324.

[40] Bin Xia, He Sun, and Bao-Liang Lu, "Multi-view gender classification based on local gabor binary mapping pattern and support vector machines," in *IEEE International Joint Conference on Neural Networks*, 2008, pp. 3388–3395.

[41] Ming-Hsuan Yang, David J. Kriegman, and Narendra Ahuja, "Detecting faces in images: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34–58, 2002.

[42] Henry A. Rowley, Shumeet Baluja, and Takeo Kanade, "Neural network-based face detection," *IEEE Transactions On Pattern Analysis and Machine intelligence*, vol. 20, pp. 23–38, 1998.

[43] Edgar Osuna, Robert Freund, and Federico Girosi, "Training support vector machines: an application to face detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, 1997, pp. 130 – 136.

[44] Paul Viola and Michael Jones, "Robust real-time object detection," *International Journal of Computer Vision*, vol. 57, pp. 137–154, May 2004.

[45] Ming-Hsuan Yang, David J. Kriegman, and Narendra Ahuja, "Detecting faces in images: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 34–58, January 2002.

[46] Xiaoyang Tan, Fengyi Song, Zhi-Hua Zhou, and Songcan Chen, "Enhanced pictorial structures for precise eye localization under incontrolled conditions," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 1621–1628.

[47] M. B. Stegmann, "Active appearance models: Theory, extensions and cases-technical report," M.S. thesis, Aug 2000.

[48] Iain Matthews and Simon Baker, "Active appearance models revisited," *International Journal of Computer Vision*, vol. 60, pp. 135–164, 2003.

[49] Zhiguang Yang, Ming Li, and Haizhou Ai, "An experimental study on automatic face gender classification," in *International Conference on Pattern Recognition*, 2006, pp. 1099–1102.

[50] Arnulf B. A. Graf and Felix A. Wichmann, "Gender classification of human faces," in *Biologically Motivated Computer Vision*, 2002, pp. 491–500.

[51] Peter N. Belhumeur, Joao P. Hespanha, and David Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 711–720, July 1997.

[52] C.J. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Transactions on Image Processing*, vol. 11, no. 4, pp. 467–476, 2002.

[53] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with gabor wavelets," in *International Conference on Face and Gesture Recognition*, 1998, pp. 200–205.

[54] Timo Ojala, Matti Pietikinen, and Topi Menp, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[55] Timo Ahonen, Abdenour Hadid, and Matti Pietikäinen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.

[56] Caifeng Shan, Shaogang Gong, and Peter W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and Vision Computing*, vol. 27, pp. 803–816, May 2009.

[57] Stan Z. Li, RuFeng Chu, ShengCai Liao, and Lun Zhang, "Illumination invariant face recognition using near-infrared images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 627–639, April 2007.

[58] Guillaume Heusch, Yann Rodriguez, and Sebastien Marcel, "Local binary patterns as an image preprocessing for face authentication," in *International Conference on Automatic Face and Gesture Recognition*, 2006, pp. 9–14.

[59] Bingpeng Ma, Wenchao Zhang, Shiguang Shan, Xilin Chen, and Wen Gao, "Robust head pose estimation using LGBP," in *International Conference on Pattern Recognition*, 2006, pp. 512–515.

[60] Wenchao Zhang, Shiguang Shan, Wen Gao, Xilin Chen, and Hongming Zhang, "Local gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition," in *IEEE International Conference on Computer Vision*, 2005, pp. 786–791.

[61] Xiaoyang Tan and Bill Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," in *International Conference on Analysis and modeling of Faces and Gestures*, 2007, pp. 168–182.

[62] Michael Collins, "Discriminative training methods for hidden markov models: Theory and experiments with perceptron algorithms," in *Proceedings of Empirical Methods in Natural Language Processing*, 2002, pp. 1–8.

[63] C.J. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Transactions on Image Processing*, vol. 11, no. 4, pp. 467–476, April 2002.

[64] Chih-Chung Chang and Chih-Jen Lin, *LIBSVM: A Library for Support Vector Machines*, 2001, Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[65] Yoav Freund and Robert E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," in *European Conference on Computational Learning Theory*, pp. 23–37.

[66] E. Makinen and R. Raisamo, "An experimental comparison of gender classification methods," *Pattern Recognition Letters*, vol. 29, no. 10, pp. 1544–1556, July 2008.

[67] Yishi Wang, Ricanek K., Cuixian Chen, and Yaw Chang, "Gender classification from infants to seniors," in *IEEE International Conference on Biometrics: Theory Applications and Systems*, 2010, pp. 1 – 6.

[68] V. Franc and V. Hlavac, "Statistical Pattern Recognition Toolbox for Matlab," *Center for Machine Perception*, 2004.

[69] L.J.P. van der Maaten, E. O. Postma, and H. J. van den Herik, "Dimensionality reduction: A comparative review (technical report)," 2008.

[70] M. Mayo and E. Zhang, "Improving face gender classification by adding deliberately misaligned faces to the training data," in *International Conference on Image and Vision Computing*, 2008, pp. 1–5.

[71] Brian Christopher Becker and Enrique G. Ortiz, "Evaluation of face recognition techniques for application to facebook," in *IEEE International Conference on Automatic Face and Gesture Recognition*, 2008, pp. 1–6.

[72] Marian Stewart Bartlett, Javier R. Movellan, and Terrence J. Sejnowski, "Face recognition by independent component analysis," *IEEE Transactions on Neural Networks*, pp. 1450–1464, 2002.

[73] Chengjun Liu and Harry Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Transactions on Image Processing*, vol. 11, no. 4, pp. 467–476, 2002.

[74] S. Milborrow, J. Morkel, and F. Nicolls, "The MUCT Landmarked Face Database," *Pattern Recognition Association of South Africa*, 2010.

[75] Brendan Klare and Anil K. Jain, "Heterogeneous face recognition: Matching NIR to visible light images," *International Conference on Pattern Recognition*, pp. 1513–1516, 2010.

[76] Shengcai Liao, Dong Yi, Zhen Lei, Rui Qin, and Stan Z. Li, "Heterogeneous face recognition from local structures of normalized appearance," in *International Conference on Biometrics*, 2009, pp. 209–218.

[77] Thirimachos Bourlai, Nathan D. Kalka, Arun Ross, Bojan Cukic, and Lawrence Hornak, "Cross-spectral face verification in the short wave infrared (SWIR) band," in *International Conference on Pattern Recognition*, 2010, pp. 1343–1347.

[78] Haitao Wang, Stan Z. Li, Yangsheng Wang, and Jianjun Zhang, "Self quotient image for face recognition," in *International Conference on Image Processing*, 2004, pp. 1397–1400.

[79] Daniel J. Jobson, Ziaur Rahman, and Glenn A. Woodell, "Properties and performance of a center/surround retinex," *IEEE Transactions on Image Processing*, vol. 6, no. 3, pp. 451–462, 1997.

[80] Weilong Chen, Meng Joo Er, and Shiqian Wu, "Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain," *IEEE Transactions on Systems, Man and Cybernetics, Part B*, vol. 36, no. 2, pp. 458–466, 2006.

[81] Karel Zuiderveld, *Contrast limited adaptive histogram equalization*, pp. 474–485, Academic Press Professional, Inc., 1994.

[82] Neeraj Kumar, Alexander C. Berg, Peter N. Belhumeur, and Shree K. Nayar, "Attribute and simile classifiers for face verification," in *IEEE International Conference on Computer Vision*, 2009.

[83] Neeraj Kumar, Peter Belhumeur, and Shree Nayar, "Facetracer: A search engine for large collections of images with faces," in *European Conference on Computer Vision*, 2008, pp. 340–353.

[84] Michael Lyons, Julien Budynek, Andre Plante, and Shigeru Akamatsu, "Classifying facial attributes using a 2-d gabor wavelet representation and discriminant analysis," in *IEEE International Conference on Automatic Face and Gesture Recognition*, 1999, pp. 202–207.

[85] Torsten Wilhelm, Hans-Joachim Böhme, and Horst-Michael Gross, "Classification of face images for gender, age, facial expression, and identity," in *International Conference on Artificial Neural Networks*, 2005, pp. 569–574.

[86] Maja Pantic and Leon J. M. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1424–1445, December 2000.

[87] Beat Fasel and Juergen Luettin, "Automatic facial expression analysis: a survey," *Pattern Recognition*, vol. 36, no. 1, pp. 259 – 275, 2003.

[88] Guoying Zhao and Matti Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 915–928, June 2007.

[89] Ying li Tian, Takeo Kanade, and Jeffrey F. Cohn, "Recognizing action units for facial expression analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 97–115, 2001.

[90] Ahmed Bilal Ashraf, Simon Lucey, Jeffrey F. Cohn, Tsuhan Chen, Zara Ambadar, Ken Prkachin, Patty Solomon, and Barry J. Theobald, "The painful face: pain expression recognition using active appearance models," in *International Conference on Multimodal Interfaces*, 2007, pp. 9–14.

[91] Dragoş Datcu and Léon Rothkrantz, "Facial expression recognition in still pictures and videos using active appearance models: A comparison approach," in *International Conference on Computer systems and technologies*, 2007, pp. 1–6.

[92] Christoph Mayer, Matthias Wimmer, Martin Eggers, and Bernd Radig, "Facial expression recognition with 3d deformable models," in *International Conferences on Advances in Computer-Human Interactions*, 2009, pp. 26–31.

[93] Yun Fu, Guodong Guo, and Thomas S. Huang, "Age synthesis and estimation via faces: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 1955–1976, 2010.

[94] Jin-Li Suo, Song Chun Zhu, Shiguang Shan, and Xilin Chen, "A compositional and dynamic model for face aging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 3, pp. 385–401, 2010.

[95] Yun Fu, Guodong Guo, and Thomas S. Huang, "Age synthesis and estimation via faces: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 1955–1976, 2010.

[96] Y. Fu G. Guo, C.R. Dyer and T.S. Huang, "Is gender recognition affected by age?," *IEEE Internatinal Conference on Computer Vision Workshops*, pp. 2032–2039, 2009.

[97] Unsang Park, Yiying Tong, and Anil K. Jain, "Age-invariant face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 947–954, May 2010.

[98] Atsunobu Suzuki, Takahiro Hoshino, Kazuo Shigemasua, and Mitsuru Kawamurab, "Decline or improvement?: Age-related differences in facial expression recognition," *Biological Psychology*, vol. 74, no. 1, pp. 75 –84, 2007.

[99] Feng Gao and Haizhou Ai, "Face age classification on consumer images with gabor feature and fuzzy lda method," in *International Conference on Advances in Biometrics*, 2009, pp. 132–141.

[100] Hang Qi and Liqing Zhang, "Age classification system with ica based local facial features," in *International Symposium on Neural Networks: Advances in Neural Networks*, 2009, pp. 763–772.

[101] Guodong Guo, Yun Fu, Charles R. Dyer, and Thomas S. Huang, "Image-based human age estimation by manifold learning and locally adjusted robust regression," *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1178–1188, 2008.

[102] Vladimir N. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag New York, Inc., 1995.

[103] Xin Geng, Zhi-Hua Zhou, Yu Zhang, Gang Li, and Honghua Dai, "Learning from facial aging patterns for automatic age estimation," in *ACM International Conference on Multimedia*, 2006, pp. 307–316.

[104] A. Lanitis, C. Draganova, and C. Christodoulou, "Comparing different classifiers for automatic age estimation," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 34, no. 1, pp. 621–628, February 2004.

[105] Xiaoguang Lu and Anil K. Jain, "Ethnicity identification from face images," in *SPIE International Symposium on Defense and Security : Biometric Technology for Human Identification*, 2004, pp. 114–123.

[106] L. Farkas, "Anthropometry of the head and face," in *Second ed. New York: Raven Press*, 1994.

[107] Xiaogang Wang and Xiaoou Tang, "Face photo-sketch synthesis and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 11, pp. 1955–1967, 2009.