Pattern Recognition 46 (2013) 519–528

Contents lists available at SciVerse ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Robust gender classification using a precise patch histogram ☆

Huang-Chia Shih *

Human–Computer Interaction Multimedia Laboratory, Department of Electrical Engineering, Yuan Ze University, Chungli, Taoyuan, Taiwan, ROC

ARTICLE INFO

Article history:
Received 17 April 2012
Received in revised form
12 July 2012
Accepted 3 August 2012
Available online 14 August 2012

Keywords:
Biometric analysis
Gender classification
Active appearance model
Local binary patch
Bayesian classifier
Face recognition
Human–computer interaction

ABSTRACT

This study proposed a precise facial feature extraction method to improve the accuracy of gender classification under pose and illumination variations. We used the active appearance model (AAM) to align the face image. Images were modeled by the patches around the coordinates of certain landmarks. Using the proposed precise patch histogram (PPH) enabled us to improve the accuracy of the global facial features. The system is composed of three phases. In the training phase, non-parametric statistics were used to describe the characteristics of the training images and to construct the patch library. In the inference phase, the choice of feature patch from the library needed to approximate the patch of the testing image was based on the maximum a posteriori estimation. In the estimation phase, a Bayesian framework with portion-oriented posteriori fine-tuning was employed to determine the classification decision. In addition, we developed the dynamic weight adaptation to obtain a more convincing performance. The experimental results demonstrated the robustness of the proposed method.

1. Introduction

Recently, the biometric analysis of the human face has been shown to reveal a large amount of physical and psychological information. The use of biometrics in gender classification has resulted in the field of biometrics expanding at a rapid rate. Biometrics can reveal a substantial amount of high-level semantic information, including gender, age, ethnicity, and emotion. Generally speaking, gender classification is divided into two main categories: geometry-based and appearance-based. The geometry-based category is focused on extracting the geometric feature points from the facial image and describes the shape structure of the face. Saatci et al. [1] presented an algorithm to determine the gender and expression of facial images by using active appearance models (AAMs) [2,3] for feature extraction and support vector machines (SVMs) for classification. Mäkinen et al. [4] illustrated a systematic evaluation on gender classification, and showed how face alignment influences the accuracy of gender classification using AAM. Based on the AAM, a pose and shape-independent texture feature extraction for face recognition is proposed in [5].

The appearance-based gender classification methods can be divided into two categories: texture-oriented and statistics-oriented. The texture-oriented approach utilizes different texture descriptors to characterize the gender of a facial image, and utilizes a machine learning strategy to recognize the gender. Many texture characteristics have been applied in gender classification, such as the local binary pattern (LBP) [6,7], local Gabor binary mapping pattern (LGBMP) [8], edge histogram [9], and wavelet transform [10,11]. Baluja et al. [12] demonstrated a feature describing the relationship between the gray-scale values of two pixels using five different types of pixel comparison operators. The Adaboost algorithm [13] was applied to identify the sex of a person from a low resolution facial image.

Among the all-texture features, LBP can be treated as a general approach to the conventionally divergent structural and statistical models of texture analysis. The methodology of the LBP-based face description approach is well-established in both face analysis and its applications. It is robust to monotonic gray-scale changes such as in illumination variations. The basic methodology for the LBP-based face description is as proposed by Ahonen et al. [14]. A notable example is the illumination-invariant face recognition system proposed by Li et al. [15], which combines the LBP features of near-infrared images and Adaboost learning. Hadid and Pietikäinen [16] proposed the spatiotemporal LBPs description for gender recognition from video sequences. Zhang et al. [17] used the LBP-based feature in terms of 40 Gabor filters with different scales and orientations for face recognition. Zhao et al. [18] adopted the spatiotemporal LBP descriptors to represent and recognize the mouth regions and visual lipreading, respectively.

The statistics-based approach usually acquires satisfactory results for the classification scheme. It focuses on using different features that are quantified into a probability to characterize a

☆This work was supported in part by the National Science Council (NSC) of Taiwan, under grant NSC 100-2628-E-155-007-MY2.
* Tel.: +886 3 4638800x7122; fax: +886 3 4639355.
E-mail address: hcshih@saturn.yzu.edu.tw

0031-3203/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2012.08.003

facial image as to gender using its visual characteristics. Aghajanian et al. [19] proposed a patch-based framework to determine the ambiguous inside of an object and roughly replace each patch from the pre-defined library and frequency parameters in order to provide the Bayesian posteriori probability. Toews et al. [20] presented the combination of the local scale-invariant feature transform (SIFT) and the object class invariant model for detecting, localizing and classifying the visual gender specific traits. Li et al. [21] used another patch-based feature representation called Spatial Gaussian Mixture Models to describe the image spatial information, while taking the local and global scales into consideration for image misalignment.

In this paper we proposed a robust facial feature description method with a statistical classifier to determine the gender, which represents facial characteristics locally and globally to provide a posteriori probability with more confidence. The active appearance model (AAM) is used to align face images. Facial images are modeled by the patches around the coordinates of landmarks. A so-called precise patch histogram (PPH) will be extracted after the AAM landmark points are determined. We applied non-parametric statistics to describe the characteristics of the training images and construct the patch library in the training phase. We also exploited the relationship between the PPH features of the testing images and those of the library-images to predict the gender in the testing phase. In the present study we proposed a Bayesian framework in which we marginalized the feature patches to determine the classification. It was evident that the accuracy of the global facial features was improved using the proposed PPH feature.

The major contributions of this paper include (1) The proposed patch-based feature acquisition, which provides more precise local and global facial features (i.e., PPH) for gender classification. (2) The flexible library selection approach based on the eigenface with *k*-means clustering provides a huge range of patch choices for model estimation in both the training and the testing process. (3) The Bayesian-oriented gender determination framework with portion-oriented posteriori fine-tuning.

The remainder of the paper is organized as follows: Section 2 provides a brief overview of the system. In Section 3, we illustrate how to encode the PPH based on the AAM facial feature points. In Section 4, the algorithms of the library selection and the patch-based gender classification are explained. Section 5 presents the experimental results of the system, and finally, we draw our conclusions and provide a discussion in Section 6.

2. System overview

The flow-diagram of the gender classification system is shown in Fig. 1 as follows: (1) apply the characteristics of the eigenface with *k*-means clustering for the library selection, (2) employ AAM to extract the landmarks from which the feature patches can be extracted, (3) choose the representative samples from the training image database to be used as a patch library, (4) classify the gender by using the inference procedure and the Bayesian estimation to fine tune the classification.

In the training phase, the posteriori probability of each feature patch of the library-images is obtained by the training inference. The patch is then ranked using the values of the posteriori probabilities. The library selection procedure is the critical procedure of the proposed gender classification system, and is based on the number of assigned libraries with patch ranks. In the testing phase, the posteriori probability of each feature patch of the input testing image is determined using the on-line inference process. Finally, a well-defined Bayesian estimation algorithm is used to marginalize the overall posteriori probabilities of the feature patches to make the gender decision. In addition, the portion-oriented posteriori fine-tuning method is used to enhance the results of the classification.

The eigenface [22] determination is based on the principle component analysis (PCA) for finding an appropriate feature representation in a low-dimensional space. There are two primacy procedures: (a) *Eigenspace generation*: given I face normalization image patches $\mathbf{x}_i$ which converted into the column vector type. (b) *Projection onto the eigenspace*: each of the training image patches $\mathbf{x}_i$ is projected onto the eigenspace as a weighting vector which calculated using the eigenspace and the average patch vector with inner product. After the *k*-means clustering, we select the centroid of each cluster and the feature point with the minimum Euclidean distance as the qualified library image.

3. Precise facial feature extraction

The patch-based gender classification is mainly composed of three components: testing images, patch library and training images. In the training phase, a statistical model is built to describe the characteristics of the training images and is used to construct the patch library. In the testing phase, all patches of the testing images are approximated by the patches in the library. We then compute the posteriori probability for each patch of the approximate image and apply it for ranking in the training phase.

3.1. Patch feature extraction using AAM

To extract more precise facial features, we need to align the facial images in advance. The AAM algorithm provides a good fit for locating the pre-defined facial model. Thus, we used the AAM to fit the facial images. In this paper there are 28 AAM landmark points that need to be determined and utilized. Each patch will be

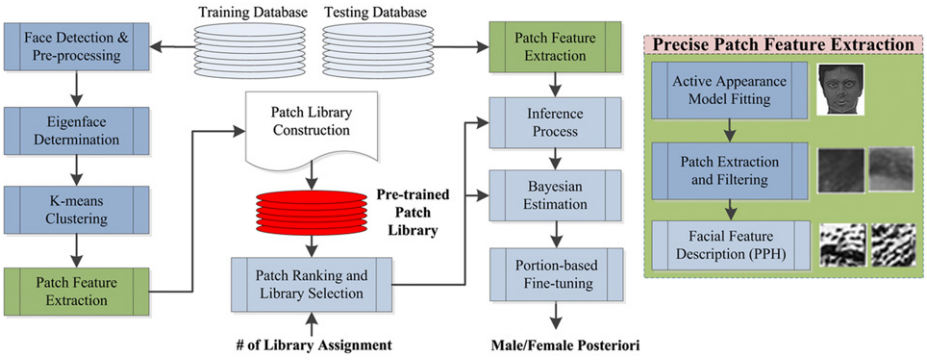

Fig. 1. The flow-diagram of the gender classification system.

unified at a size of 30×30 pixels. The selected patch regions may overlap each other. However, the patch regions will be treated independently of the facial features. Based on our observations, the location of the last landmark point from the AAM algorithm is usually mismatched position-wise, and this results in the performance of the system decaying seriously. To deal with this problem, only the first 27 patch features are used in both the training and the testing phase.

The active appearance model (AAM) has been successfully used to represent the appearance and shape variations of the human face. In typical gender classification approaches, uniform sampled grids lose the intrinsic properties of the facial image. To overcome this problem, we aligned the precise face location using the AAM before sampling the patch features. This means that the AAM parameters are found by the best-matched location between the model instance and the input image. The representation of the facial appearance is obtained by using the different linear models for texture and shape feature [3]. The parameters of the model are statistically learned from the training images.

3.1.1. Statistical model of facial appearance

The AAM is a statistical model of the shape and texture of the target object, where the shape model describes the shape of the object and the texture model describes the gray-level information of the object. We used the labeled face images as the training data of the AAM, and collected a labeled dataset of facial images with landmark points for training the parameters of the AAM model. These landmark points were selected as the salient points in the face region. The model was built such that landmarks and edges were defined for a set of example faces.

Generally speaking, the accuracy of the model fitting increases the larger the number of landmark points used. However, the computation cost of the model fitting process tends to increase as well. Therefore, the number of landmark points will be decided by the different manifold requirements/specifications. In the present study, we used 28 landmark points, as shown in Fig. 2. The positions of the landmark points are arranged around the regions of the eyebrows, eyes, nose, mouth, and chin. However, we did not assign any landmarks in the other regions such as ears due to the occlusion issue.

3.1.2. Shape model

The shape information is defined by a mesh with a particular vertex. A set of ordered N_L landmarks $\mathbf{x}_i$, $i=1, 2,\ldots,N_L$ is used to represent the shape model. We define a shape $\mathbf{s}$ as a vector containing the coordinates of N_L landmark points of a face image. Let $\mathbf{s}=(\mathbf{x}_1\ \mathbf{x}_2\ \ldots\ \mathbf{x}_i\ \ldots\ \mathbf{x}_{N_L})^{\mathrm{T}}$ denote the landmark coordinates arranged in a shape model, where $\mathbf{x}_i \in \mathbb{Z}^2$.

A shape model is constructed by using the coordinates of the labeled points of the training images. Normally, the training meshes are first applied to the generalized Procrustes analysis (GPA) for normalization. Then, the principal component analysis (PCA) is used to reduce the dimension of the aligned shape feature. We aligned the locations of the corresponding points on different training faces using the AAM to represent the instance $\mathbf{s}_p$ of a particular shape as the mean shape $\bar{\mathbf{s}}$ and the linear combination of t eigenshape s_i as:

$$\mathbf{s}_p = \bar{\mathbf{s}} + \sum_{i=1}^{t} \mathbf{p}_i \mathbf{s}_i, \tag{1}$$

where coefficient $\mathbf{p}_i$ constitutes the shape parameter vector $\mathbf{p}=(p_1,p_2,\ldots,p_t)^T$ representing the matrix of the first t eigenvectors. The mean shape $\bar{\mathbf{s}}$ and the shape variations $\mathbf{s}_i$ are statistically learned using the training dataset.

Fig. 2. Extract the 28 lined patches from the AAM landmark points.

3.1.3. Texture model

To construct a texture model, all the aligned training images are first warped into the shape-normalized space. The shape normalized space is determined by the mean shape $\bar{\mathbf{s}}$ of the shape model. The texture of AAM is defined by the gray level information at pixels $\mathbf{x}=(x,y)^{\mathrm{T}}$ which lie inside the mean shape $\bar{\mathbf{s}}$. Then, we sample the gray level information $\boldsymbol{g}$ of the warping images at the mean shape region (i.e., $\mathbf{g}(\mathbf{x}) \in \bar{\mathbf{s}}$). Before applying PCA on the texture data, we minimize the effect of lighting variation by normalizing g. Let $\bar{\mathbf{g}}(\mathbf{x})$ define the mean of the normalized texture feature, scaled and offset so that the sum is zero and the variance is unity. Similar to shape, $\mathbf{q}=(q_1,q_2,\ldots,q_k)^{\mathrm{T}}$ is a vector of the texture parameters representing a texture instance

$$\mathbf{g}_q(\mathbf{x}) = \bar{\mathbf{g}}(\mathrm{x}) + \sum_{i=1}^{k} \mathbf{q}_i \mathbf{g}_i(\mathbf{x}). \tag{2}$$

We iteratively estimate $\bar{\mathbf{g}}(\mathbf{x})$ until the model converges, then apply PCA to the normalized texture data.

3.1.4. Model search

The parameters $\lambda=(\mathbf{p}^T\mathbf{q}^T)^T$ of the generative model need to be estimated in order to fit the target model (i.e., testing image). Appearance is defined as the intensities of a facial image at a set $\bar{A}$ of positions $\mathbf{x}$ in a shape-normalized space $A \in \mathbb{R}^2$. The process of model fitting is generally done by minimizing the error between the modeled appearance and the target image. The error at position $\mathbf{x} \in \bar{A}$ between the generated texture and the target image is

$$\xi(\mathbf{x},\lambda) = \mathbf{g}_q(\mathbf{x}) - I_t[\boldsymbol{\psi}(\mathbf{x},\mathbf{p})], \tag{3}$$

where $\boldsymbol{\psi}(\mathbf{x},\mathbf{p})$ is a warping function that maps positions $\mathbf{x} \in A$ of the model to positions $\mathbf{x}' \in I_t$ of the target image. The desired parameter $\hat{\lambda}$ is obtained by minimizing the sum of the squared error of all positions $\mathbf{x}$,

$$\hat{\lambda} = \underset{\lambda}{\operatorname{argmin}} \left\{ \sum_{\mathbf{x} \in \bar{A}} [\xi(\mathbf{x},\lambda)]^2 \right\}. \tag{4}$$

Many optimization algorithms have been proposed for parameters searching. In this paper, we adopt the so-called AAM-API method proposed by Cootes et al. [2,23]. Its performance is sufficient for gender classification.

3.2. Face description using a precise patch histogram

Here, we propose a so-called precise patch histogram (PPH) to describe the local and global features of the face in a more precise manner. The PPH is an LBP-based approach which is a multi-resolution texture operator that takes the local and global texture structures into consideration. As shown in Fig. 3, this PPH effectively provides a description of the face, whereas the LBP descriptions for the histogram contain four types of textures from different levels, including the pixel-level within the LBP block, the block-level within the patch, the region-level within the facial part, and the global description in the facial image. It is worth noting that if the histogram-based method is used, it is not

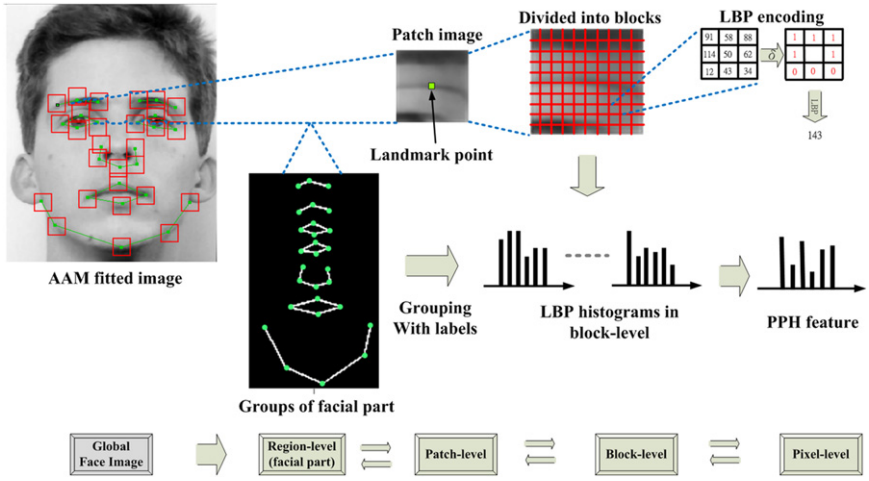

Fig. 3. Extracting a precise patch histogram (PPH).

necessary for the patch regions to be rectangular. In addition, the histogram-based method is tolerant for regions with a partial overlap.

The proposed face description method is illustrated in Fig. 3, which is an LBP-based approach for texture classification. First, the 27 facial landmarks are determined by AAM and then facial patches are extracted. Second, the precise patch grid is extracted based on the locations of the landmark points. Then, the LBP is used to describe the characteristics of the patch. Assume that the notation (P, R) is used to represent the LBP code for the pixel and its neighborhoods which means P sampling points on a circle of radius of R. The value of the LBP code of a pixel g_c is given by

$$LBP(P,R) = \sum_{p=0}^{P-1} \delta(g_p - g_c)2^p, \tag{5}$$

where

$$\delta(x) = \begin{cases} 1 & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Each feature patch is viewed as an independent component that provides independent information about the gender (male or female). Here, we group the patches that belong to the same facial part, and label the ordering with a semantic weight. Based on the face model used in this paper, the facial part of a face image includes left eyebrow, right eyebrow, left eye, right eye, nose, mouth, and chin. The number of LBP code occurrences in an image is collected into a histogram (i.e., PPH). Classification is then performed by computing the histogram similarities between the testing image and the patch library. We observed that using a similar approach for facial image representation results in a loss of spatial information, while modifying the texture information while retaining their locations is a more reliable approach. It is also the reason for keeping the group information of the facial part for each patch. The feature description method groups the local characteristics of the face into a global view. Thus, the proposed PPH description method will not only preserve the local facial feature but also take the global characteristics into consideration. This way it preserves the local feature in a more robust way against variations in pose or illumination compared to the holistic methods.

Gender is denoted as a random variable C with class label that belongs to either *Male* or *Female*. The testing image $\mathbf{Z}$ is represented by the patches centered around the landmark points of the AAM fitted model $\mathbf{Z}=[z_1, z_2,\ldots, z_\theta]$, where θ is the number of patches. The resolution of the patch is $M \times N$ (which we set at 30×30). We choose I_t training images with corresponding gender labels (male or female). Meanwhile, the training images are divided into a regular grid of overlapping patches of the same size as the testing images for use in fine-tuning process. In addition, we assign the size of the patch library, and the system will determine the images equally from these two classes to construct the library. The pre-trained patch library is considered as a set of feature patches. There are θ feature patches for each image.

4. Patch-based gender classification (PGC)

Consideringv the aspects of efficiency and simplicity, we utilize the Bayesian estimation approach for the patch-based gender classification. The gender classification consists of two major phases: the training phase and the inference phase. In the training phase, we build a model with non-parametric statistics to describe the characteristics of the training images and embed them into the patch library. In the inference phase, the test images are approximated by the patches of the library. We can obtain the posteriori probability for each feature patch for the approximated image based on the ranking results in the training process. The portion-oriented fine-tuning method is used to verify the gender class.

We used the Bayesian estimation method to predict the gender from the facial images. The main goal of the Bayesian classification is to compute the posteriori probability $P(C|\mathbf{Z})$, which can be derived from prior $P(C)$ and the class-conditional densities $P(\mathbf{Z}|C)$. From the training image we extract θ overlapping feature patches. Since each training image can be represented by these feature patches, the training set can be decomposed into θ patch groups as $\mathbf{X}^p$, where $p=1,2,...,\theta$. For gender classification problem, each patch group has two classes, expressed as $\mathbf{X}^p=\mathbf{X}^p_{male}\cup\mathbf{X}^p_{female}$. Here, we suppose that the feature patch $\mathbf{x}^p_{c,i}\in\mathbf{X}^p_c$, where $\mathbf{x}^p_{c,i}$ denotes the pth feature patch in the ith training image with class label $c\in\{male, female\}$. The Bayesian posteriori probability over class label C, given the training samples $\mathbf{X}$, can be obtained by using the Bayes' rule as

$$P(C=c|\mathbf{Z},\mathbf{X}) = \frac{P(\mathbf{Z}|C=c,\mathbf{X})\cdot P(C=c)}{P(\mathbf{Z})} = \prod_{p=1}^{\theta} \frac{P(\mathbf{z}_p|C=c,\mathbf{X})}{P(\mathbf{z}_p)} \cdot P(C=c), \tag{7}$$

where $\mathbf{z}_p$ is an individual testing patch of input test image $\mathbf{Z}$ which provides independent information about gender based on the training set $\mathbf{X}$. Based on Eq. (7), we can use the correlation between $\mathbf{z}_p$ and $\mathbf{X}$ to determine the posteriori $P(C=c|\mathbf{Z}, \mathbf{X})$. The gender determination function $\Lambda(\cdot)$ can be formulated by maximizing a posteriori as

$$\Lambda(\mathbf{Z}) = \underset{c}{argmax}\, P(C=c|\mathbf{Z},\mathbf{X}). \tag{8}$$

Here, we constructed the library with θ library subsets for each class as W_c^1, …, W_c^θ, each with L feature patches as $W_c^p = \{\boldsymbol{\omega}_{c,l}^p | l=1,2,\ldots,L\}$. $P(\mathbf{z}_p|\boldsymbol{\omega}_{c,l}^p)$ together with the relational information of the parameter space and the training set can be used to predict the gender probability of $\mathbf{z}_p$ for $p=1,2,\ldots,\theta$. We then obtain the class-conditional densities $P(\mathbf{z}_p\,|C=c, \mathbf{X})$, based on the integration of the joint density $P(\mathbf{z_p}, \boldsymbol{\omega}_{c,l}^p|C=c, X)$ over a set of parameters $\{\boldsymbol{\omega}_{c,l}^p\}$ as

$$\begin{aligned} P(\mathbf{z}_p|C=c,\mathbf{X}) &= \int P(\mathbf{z}_p,\boldsymbol{\omega}_{c,l}^p\big|C=c,\mathbf{X})\; d\boldsymbol{\omega}_{c,l}^p \\ &= \int P(\mathbf{z}_p|\boldsymbol{\omega}_{c,l}^p)\cdot P(\boldsymbol{\omega}_{c,l}^p\big|C=c,\mathbf{X})\; d\boldsymbol{\omega}_{c,l}^p \end{aligned} \tag{9}$$

4.1. Patch library construction

In the training process, we compute the accumulated frequency of all possible patches from the patch library. In the inference process, we exploit the accumulated information of the library patches containing the highest similarity with the corresponding test patch. Here, we need to take the variations of all possible library compositions into consideration in order to guarantee the accuracy of the gender classification.

The use of the integration extends over the parameter space, where $\boldsymbol{\omega}_{c,l}^p$ denotes all the parameters related to the pth feature patch in the library. Eq. (9) links the required class-conditional densities $P(\mathbf{z}_p|\;C=c, \mathbf{X})$ to the chain probability of $P(\mathbf{z}_p|\boldsymbol{\omega}_{c,l}^p)$ and $P(\boldsymbol{\omega}_{c,l}^p|\;C=c, \mathbf{X})$ for parameter set $\{\boldsymbol{\omega}_{c,l}^p\}$. $P(\boldsymbol{\omega}_{c,l}^p|\;C=c, \mathbf{X})$ denotes the characteristic of the assembled training feature patches in X converted into the parameter space $\{\boldsymbol{\omega}_{c,l}^p\}$, where for $c=\{male, female\}$, $p=1,2,\ldots,\theta$ and $l=1,2,\ldots,L$. The role of the parameter space is similar to the patch library, which represents a specific distribution for class c.

Testing patch z_p in terms of $P(\mathbf{z}_p|\boldsymbol{\omega}_{c,l}^p)$ and $P(\boldsymbol{\omega}_{c,l}^p|\;C=c, X)$ is interpreted in the *inference process* and *training process*, respectively. The first term is a testing-library inference which is an on-line process. The second term is a library-training inference which is an off-line process. Based on the *maximum a posteriori* (MAP) estimation, we obtain the maximum posteriori over $P(C=c|\mathbf{Z}, \mathbf{X})$ to determine the gender of the input image.

4.2. Training process

Training set $\mathbf{X}$ contains two labeled training sets, $\{\mathbf{X}_{male}, \mathbf{X}_{female}\}$ which are then further decomposed into θ labeled patch groups as $\mathbf{X}_{male}=\mathbf{X}_{male}^p|p=1,2,...,\theta$ and $\mathbf{X}_{female}=\mathbf{X}_{female}^p|p=1,2,...,\theta$. Two labeled training sets can be described as $\mathbf{X}_{male}=\{\mathbf{x}_{c,i}^p|c=male,\ p=1,2,\ldots,\theta,\ i=1,2,\ldots,\ I_t\}$ and $\mathbf{X}_{female}=\{\mathbf{x}_{c,i}^p|c=female,\ p=1,2,\ldots,\theta,\ i=1,2,\ldots,I_t\}$. We use the Bayesian rule to calculate the posteriori density $P(\boldsymbol{\omega}_{c,l}^p|C=c,\ \mathbf{x}_{c,i}^p)$ over the parameters space $\boldsymbol{\omega}_{c,l}^p$, and let $P(\boldsymbol{\omega}_{c,l}^p|C=c,\mathbf{x}_{c,i}^p)=P(\boldsymbol{\omega}_{c,l}^p|\mathbf{x}_{c,i}^p)$ for a specific class c as

$$P(\boldsymbol{\omega}_{c,l}^p|\mathbf{x}_{c,i}^p) = \frac{P(\mathbf{x}_{c,i}^p|\boldsymbol{\omega}_{c,l}^p)P(\boldsymbol{\omega}_{c,l}^p)}{P(\mathbf{x}_{c,i}^p)}. \tag{10}$$

The prior $P(\boldsymbol{\omega}_{c,l}^p)$ denotes the weight of the lth patch in the library subset W_c^p which is referred by the frequency of the patches being selected in the training step, where $P(\boldsymbol{\omega}_{c,l}^p)$ is normalized using $P(\boldsymbol{\omega}_{c,l}^p)/\Sigma_l P(\boldsymbol{\omega}_{c,l}^p)$.

Frist, we acquire the likelihood function $P(\mathbf{x}_{c,i}^p|\boldsymbol{\omega}_{c,l}^p)$ for each library patch $\boldsymbol{\omega}_{c,l}^p$, which finds the posteriori density $P(\boldsymbol{\omega}_{c,l}^p|\;C=c, \mathbf{X})$ for class c. To calculate the likelihood $P(\mathbf{x}_{c,i}^p|\boldsymbol{\omega}_{c,l}^p)$ of the labeled training sets $\mathbf{X}_{male}$ and $\mathbf{X}_{female}$ for $\boldsymbol{\omega}_{c,l}^p$. Then, we use the relationship of the similarity that exists between the corresponding feature patches $\mathbf{x}_{c,i}^p$ and $\boldsymbol{\omega}_{c,l}^p$.

The MAP algorithm is used to search for the best match from the training feature patch $\mathbf{x}_{c,i}^p$ for all the candidate library feature patches $\boldsymbol{\omega}_{c,l}^p$ in W_c^p. For every training feature patch $\mathbf{x}_{c,i}^p$, we determine the most similar library feature patch in W_c^p as

$$\hat{\boldsymbol{\omega}}_{c,l}^p = \underset{\omega_{c,l}^p}{\text{argmin}}\; \Phi(\boldsymbol{\omega}_{c,l}^p,\mathbf{x}_{c,i}^p) \quad \forall \mathbf{x}_{c,i}^p, \tag{11}$$

where $\boldsymbol{\omega}_{c,l}^p \in W_c^p$, and W_c^p denotes the pth library subset, and $\Phi(\cdot)$ is the dissimilarity between two patches. The Bhattacharyya distance measure is used in this paper.

4.3. Inference process

In this step, we determine the highest similarity between the testing feature patch $\mathbf{z}_p$ and the candidate library patches in the patch library based on the MAP estimation. The MAP algorithm is also employed to determine the best matching feature patch $\mathbf{z}_p$ for all the candidate library feature patches $\boldsymbol{\omega}_{c,l}^p$ in W_c^p. We then search all possible test patches of $\mathbf{Z}=\{\mathbf{z}_p\}$ for the one with the most matches. For every $\boldsymbol{\omega}_{c,l}^p$, we find the most similar test feature patch as

$$\hat{\mathbf{z}}_p = \underset{\mathbf{z}_p}{\text{argmin}}\; \Phi(\mathbf{z}_p,\boldsymbol{\omega}_{c,l}^p) \quad \forall \boldsymbol{\omega}_{c,l}^p. \tag{12}$$

We may have the matching feature patch in the library for each feature patch $\mathbf{z}_p$. The conditional probability $P(\boldsymbol{\omega}_{c,l}^p|\mathbf{z}_p)$ is formulated by a Gaussian form in terms of the dissimilarity measure between the two feature patches. The Bayes' rule is used to compute the posteriori density $P(\mathbf{z}_p|\boldsymbol{\omega}_{c,l}^p)$ over the parameters $\{\boldsymbol{\omega}_{c,l}^p\}$

$$P(\mathbf{z}_p|\boldsymbol{\omega}_{c,l}^p) = \frac{P(\boldsymbol{\omega}_{c,l}^p|\mathbf{z}_p)P(\mathbf{z}_p)}{P(\boldsymbol{\omega}_{c,l}^p)}, \tag{13}$$

where the prior $P(\mathbf{z}_p)$ denotes the weight of the pth testing patch acquired in the inference process.

Based on the male and female training images, we have the accumulated distribution of both males and females in the same library space. Finally, we integrate all the facial image patches in order to provide a posteriori probability and then determine the gender.

4.4. Portion-oriented posteriori fine-tuning

We use the accumulated information of the library patch which has the highest similarity with the corresponding test patch in the inference process. Because the facial alignment is used in this system, we can determine the position-related patches for gender classification. This results in a higher classification accuracy because more precise face characteristics are employed. However, it should be noted that the AAM fitting algorithm is not always very accurate due to the quality of face image. Here, we illustrate a portion-oriented posteriori fine-tuning method to improve the performance of the system. It is the same as the aforementioned training process, but images are divided into a series of regular grid non-overlapping patches with the same size. Fig. 4(a) and (b) shows an example of the original image and the image constructed from the patches with the highest similarity from the patch library. Obviously, the location

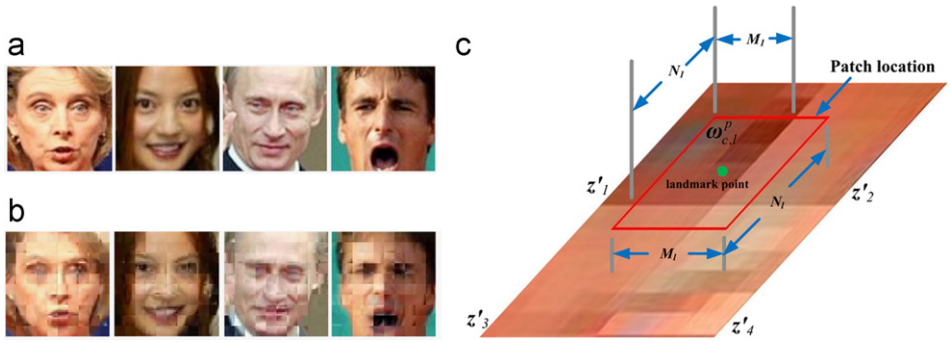

Fig. 4. Portion-oriented class-conditional density estimation: (a) original images, (b) patch fitted images and (c) portion-oriented fine tuning.

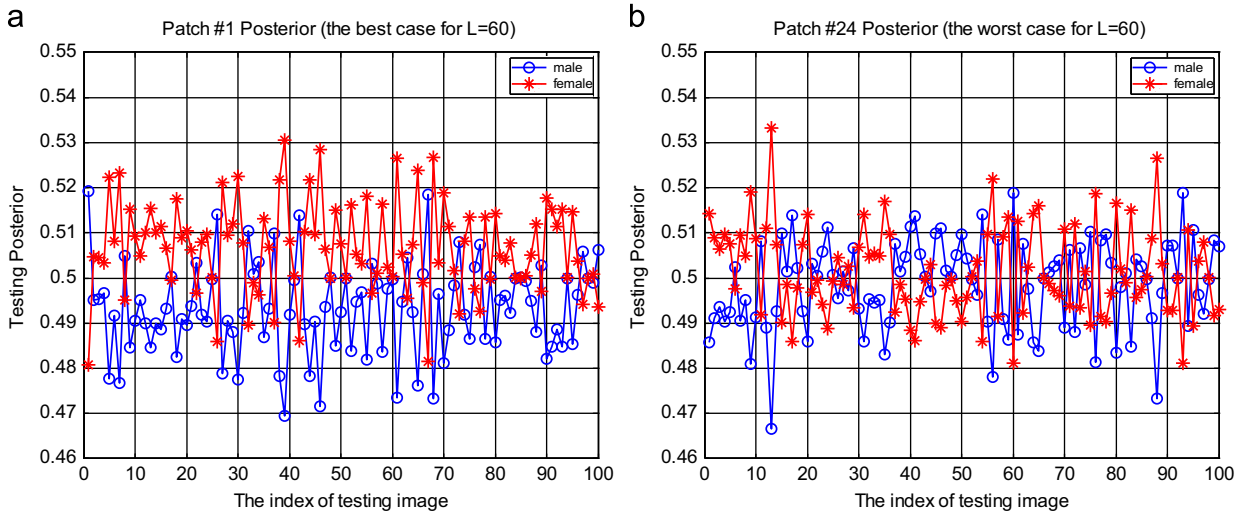

Fig. 5. The first 100 testing images' posteriors with the library selecting $L=60$, where the discriminant power: (a) the best case is the ♯1 patch and (b) the worst case is the ♯24 patch.

of the test patch can be anywhere in the fitted patch image. Therefore, we present a method that computes the posteriori probability of gender estimation based on the portion of the AAM fitted patch. As shown in Fig. 4(c), the landmark point will decide the location of the patch measuring $M_l \times N_l$. The class-conditional densities $P(\mathbf{z}_p \mid C=c, \mathbf{X})$ should be adjusted according to the portion of the overlapping four patches $\{\mathbf{z}'_1,\mathbf{z}'_2,\mathbf{z}'_3,\mathbf{z}'_4\}$. The landmark point located in patch $\mathbf{z}'_1$, and the overlapping area measures $M_1 \times N_1$. Hence, the highest density percentage will be determined by the class probability of the first patch, and by integrating the class-likelihoods of each block involved in the fitted patches weighted by a Gaussian template centered at the location of the landmark point. Thus, the class of gender can be determined by the portion-oriented posteriori probability.

5. Experimental results and discussions

This section assesses the performance of the proposed patch-based gender classification (PGC). First of all, we adopted the two times AAM model fitting (it performed sufficient fitting results in the experiment). Then, we extracted 27 patches using the AAM landmark points as 30×30 blocks and grouping each patch to a different facial part. The 27 facial feature patches are encoded as PPH. The accuracy of model fitting depends on the iteration times to optimization and the visual quality of the face image. The experiments were performed in three subjects: (1) the effectiveness of the different number of libraries to select from, (2) the comparison between the two methods of weight adaptation, and (3) the improvement using the portion-oriented posteriori fine-tuning.

5.1. Preliminary

In the experiments, we used the facial images of the *Labeled Faces in the Wild* (LFW) dataset [24] and the Color FERET dataset [25]. The LFW dataset contains more than 13 000 images of faces collected from the web, including 5749 people. The faces have a large range of variation include lighting, expression, pose, race, gender, background, etc. The images of the FERET dataset also include uniform illumination conditions and clean backgrounds. Here, we only use the regular frontal facial images (i.e., the *fa* partition) in the FERET database, containing 1364 images.

First, we collected all facial images from these two databases, and then manually classified these images into two gender classes which worked out to about 1862 male facial images and 1503 female facial images. Second, the Adaboost face detector was used to extract the facial regions. Finally, the facial images were processed using normalization and quality enhancement.

5.2. The effectiveness of the number of libraries

We chose about 800 female and 800 male images to train the classifier. Here, we assumed that the number of patch libraries was assigned ($L=\{60, 120, 240, 300\}$) with independent clustering from 27 patches. The classification results for the female

testing images from a different number of patch libraries are shown in Figs. 5–8. The case of two distributions with more distinguishable tendencies would indicate that the stronger discriminant power was reached. The sub-figures (a) of Figs. 5–8 show that the best discriminant power for classification is where the x-axis indicates the first 100 female images and the y-axis represents the posteriori for every image under $L=60$, 120, 240, and 300, respectively. On the other hand, the sub-figures (a) of Figs. 5–8 show that the worst discriminant power for classification is in the cases of $L=60$, 120, 240, and 300, respectively. Fig. 9 shows the average classification rate of over 300 testing images while 800 training images were used in the library training step. The experimental results show that the higher the number of libraries selected, the stronger the discriminant ability the test posteriori received.

For the Bayesian estimation, each single patch contained two specific posteriors, one for females and one for males. Based on our observation, the average PGC rate was subject to dynamic changes. This means that some faces may have a good discriminant property in some patches but a worst one in others. Therefore, the real gender posteriori for the testing step should consider summarizing or apply dynamic weights for the 27 posteriori patches. Here, we adopted a flexible weight adaptation method to overcome this problem. The results are shown in the next section.

5.3. Comparisons on the weight adaptation

The classification results for 27 patches with uniform weights have been shown in the previous experiments. In the training phase, we can obtain the discriminant power for the extracted patches. Obviously, each patch has a specific PGC rate. Here, we tried to test the gender discrimination using adaptive weights. For different library selections, the prior weight $P(z_p)$ indicates the weight of the pth testing patch. It can be assigned by the average weights instead of the frequency of the designated patches selected in the inference process. Because we have two labels (i.e., female and male), we can determine two patch weights and then take the average of the weights of the different cases in the library selected. The average weights for selecting the different libraries are shown in Fig. 10.

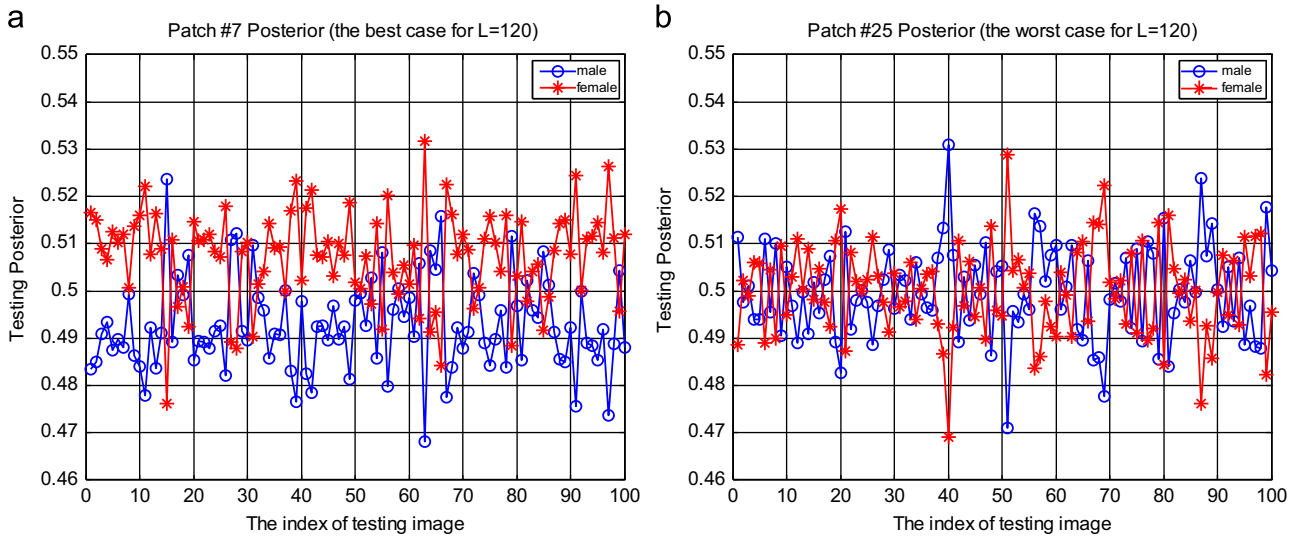

Fig. 6. The first 100 testing images' posteriors with the library selecting $L=120$, where the discriminant power: (a) the best case is the ♯7 patch and (b) the worst case is the ♯25 patch.

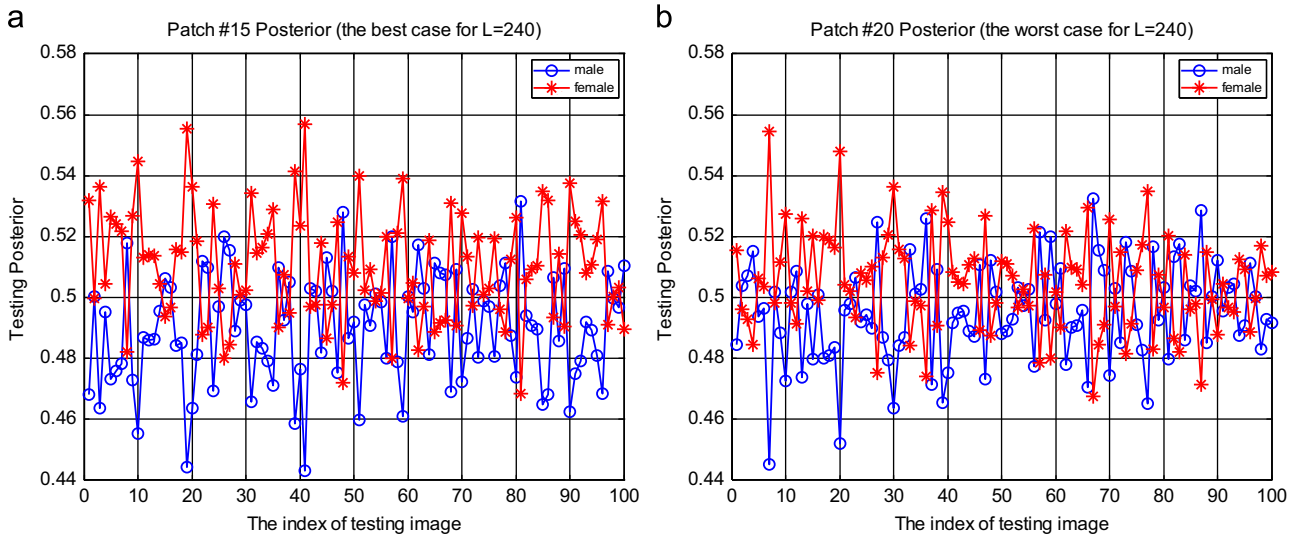

Fig. 7. The first 100 testing images' posteriors with the library selecting $L=240$, where the discriminant power: (a) the best case is the ♯15 patch and (b) the worst case is the ♯20 patch.

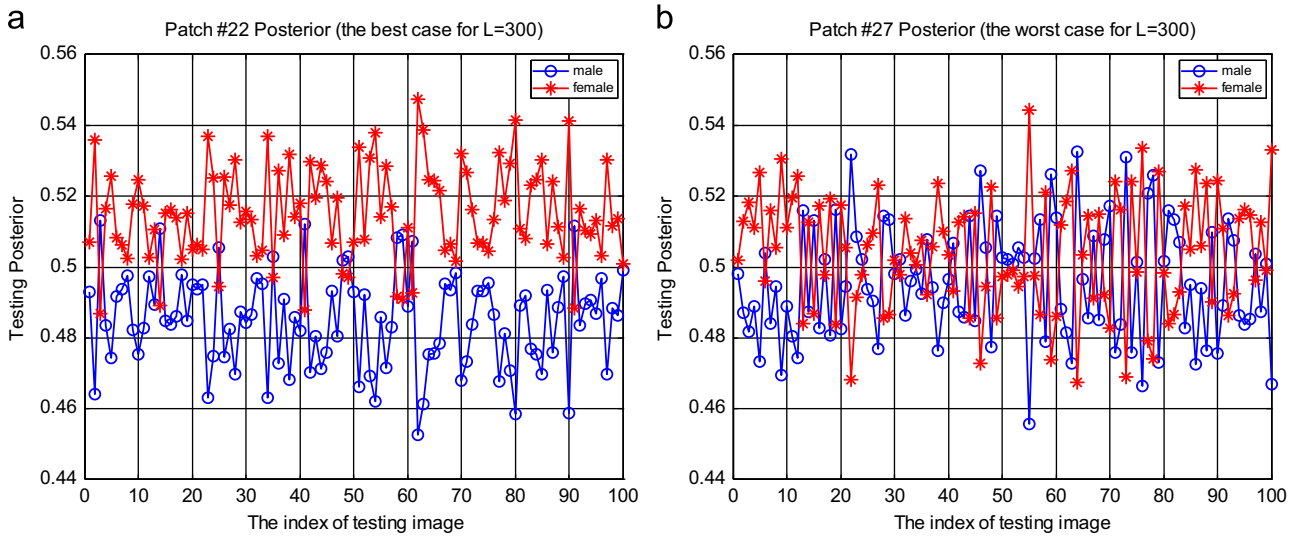

Fig. 8. The first 100 testing images' posteriors with the library selecting $L=300$, where the discriminant power: (a) the best case is the ♯22 patch and (b) the worst case is the ♯27 patch.

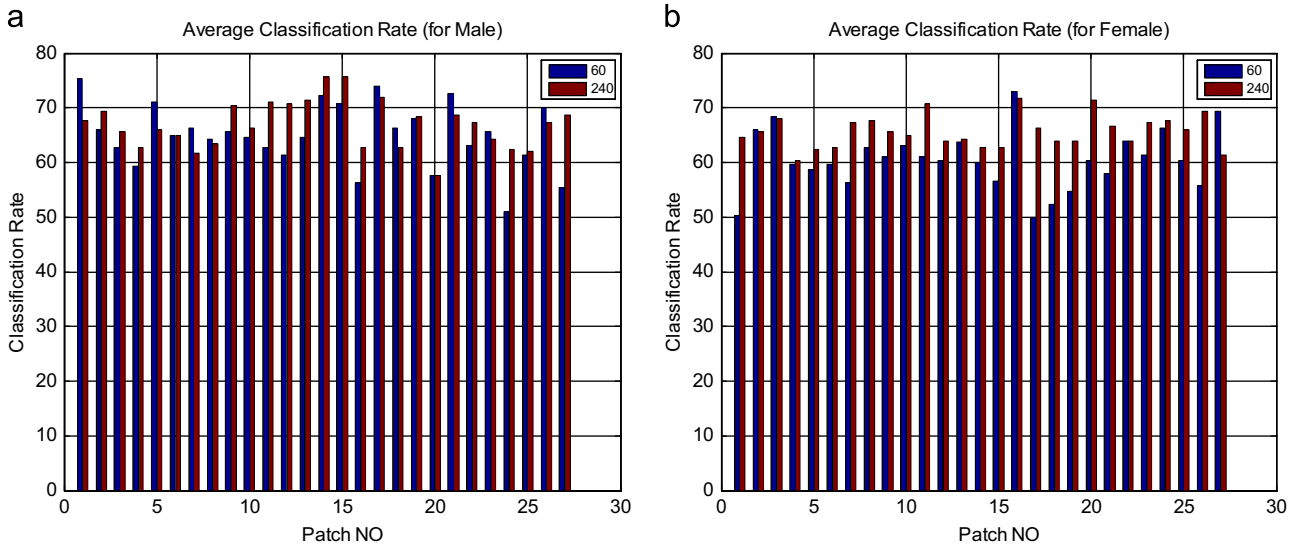

Fig. 9. The average classification rate of 27 patches of (a) male and (b) female PGC for testing 800 training images, and applying 300 testing images.

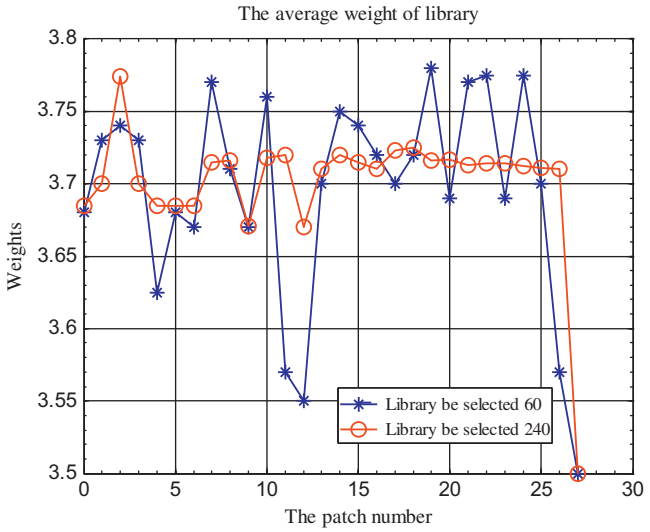

Fig. 10. The average weights for 27 patches.

Table 1
The PGC results of the training/testing dataset for overall 27 patches.

♯Library	Uniform weights (%)	Adaptive weights (%)
$L=60$	96.5/75.2	97.0/80.7
$L=240$	99.5/77.3	99.3/84.7

Based on the experimental results, the patches surrounding the eyes and mouth usually have the better discriminant ability. We also found that the overall performance of the PGC applying adaptive weights is better than that of the uniform weight by roughly 5–7%. The quantitative comparisons of the PGC of the training and testing dataset are shown in Table 1.

5.4. *The improvement using portion-oriented posteriori fine-tuning*

In the present paper, the library selection methods employed the eigenface approach followed by *k*-means clustering and random selection. We tested every single image on the 27 patches

Table 2
Accuracy comparisons with/without fine-tuning.

♯Library	Accuracy without fine-tuning		Accuracy with fine-tuning	
	Random selection (%)	With clustering (%)	Random Selection	With clustering
$L=60$	79.2	80.1	82.4	83.9
$L=120$	81.7	82.5	82.2	84.2
$L=240$	82.3	84.2	84.4	86.5

using the Bayesian estimation, and each patch contained a specific posterior for a female and a male, respectively. Based on our observation, the results of the random selection do not change with the library selected. On the other hand, for the property of the PPH facial feature description, the real facial features are usually sensitive to the results of the AAM fitting. Therefore, the proposed portion-oriented posteriori fine-tuning method was applied to understand this sensitivity. The results shown in Table 2 indicate that the performance of classification with portion-oriented fine-tuning is apparently higher than that without fine-tuning higher by about 2% compared to that without fine-tuning. Portion-oriented fine-tuning is especially required when the accuracy of the AAM fitting is very unstable. However, we need to construct another patch library in which the training images are divided into a regular grid of overlapping patches. In fact, it gains the off-line computational loadings in order to enhance the system's performance. As part of our future work, we will develop a more comprehensive algorithm.

6. Conclusions

In this study a precise facial feature extraction method was employed to classify gender. The accuracy of the global facial features was improved by using the proposed precise patch histogram (PPH). The proposed PPH finds the correlation between the patch features and the location information regarding the face. We also proposed a library images selection scheme based on the *k*-means clustering. A Bayesian framework with portion-oriented fine-tuning was used to marginalize the feature patches to make the classification. In addition, a dynamic weight adaptation was developed to obtain a more convincing performance. The experimental results showed that the portion-oriented posteriori fine-tuning method was applied to understand this sensitivity.

Acknowledgment

The authors would like to thank the anonymous reviewers for their valuable suggestions and comments, which were crucial in improving this paper. The authors also would like to thank Prof. Chung-Lin Huang at the Asia University for the constructive comments and useful suggestions that led to improvement in the quality of this paper.

References

[1] Y. Saatci, C. Town, Cascaded classification of gender and facial expression using active appearance models, in: Proceedings of the Seventh International Conference on Automatic Face and Gesture Recognition (FGR'06), April 2006, pp. 393–400.
[2] T.F. Cootes, G.J. Edwards, C.J. Taylor, Active appearance models, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (6) (2001) 681–685.
[3] I. Matthews, S. Baker, Active appearance models revisited, International Journal of Computer Vision 60 (2) (2004) 135–164.
[4] E. Mäkinen, R. Raisamo, Evaluation of gender classification methods with automatically detected and aligned faces, IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (3) (2008) 541–547.
[5] T. Gernoth, A. GooBen, R.-R. Grigat, Face recognition under pose variations using shape-adapted texture features, in: Proceedings of the IEEE International Conference on Image Processing, 2010, pp. 4525–4528.
[6] H.C. Lian, B.L. Lu, Multi-view gender classification using multi-resolution local binary patterns and support vector machines, International Journal of Neural Systems 17 (16) (2007) 479–487.
[7] Y. Fang, Z. Wang, Improving LBP features for gender classification, in: Proceedings of the International Conference on Wavelet Analysis and Pattern Recognition, August 2008, pp. 373–377.
[8] B. Xia, H. Sun, B.L. 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.
[9] A.R. Ardakany, A.M. Joula, Gender recognition based on edge histogram, International Journal of Computer Theory and Engineering 4 (2) (2012) 127–130.
[10] J. Li, B.-L. Lu, A Framework for Multi-view Gender Classification, Neural Information Processing, Springer-Verlag, Berlin, Heidelberg, 2008.
[11] X.-M. Leng, Y.-D. Wang, Gender classification based on fuzzy SVM, in: Proceedings of the Seventh International Conference on Machine Learning and Cybernetics, July 2008, 1260–1264.
[12] S. Baluja, H. Rowley, Boosting sex identification performance, International Journal of Computer Vision 71 (1) (2007) 111–119.
[13] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in: Proceedings of the Conference on Computer Vision and Pattern Recognition, December 2001, pp. 511–518.
[14] T. Ahonen, A. Hadid, M. Pietikäinen, Face description with local binary patterns: application to face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (12) (2006) 2037–2041.
[15] S.Z. Li, R.F. Chu, S.C. Liao, L. Zhang, Illumination invariant face recognition using near-infrared images, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (4) (2007) 627–639.
[16] A. Hadid, M. Pietikäinen, Combining appearance and motion for face and gender recognition from videos, Pattern Recognition 42 (11) (2009) 2818–2827.
[17] W. Zhang, S. Shan, W. Gao, X. Chen, H. Zhang, Local gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition, in: Proceedings of the IEEE International Conference on Computer Vision, 2005, pp. 786–791.
[18] G. Zhao, M. Barnard, M. Pietikäinen, Lipreading with local spatiotemporal descriptors, IEEE Transactions on Multimedia 11 (7) (2009) 1254–1265.
[19] J. Aghajanian, J. Warrell, S.J.D. Prince, P. Li, J.L. Rohn, B.Baum, Patch-based within-object classification, in: Proceedings of the IEEE 12th International Conference on Computer Vision, 2009, pp. 1125–1132.
[20] M. Toews, T. Arbel, Detection, localization, and sex classification of faces from arbitrary viewpoints and under occlusion, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (9) (2009) 1567–1581.
[21] Z. Li, X. Zhou, Thomas S. Huang, Spatial Gaussian mixture model for gender recognition, in: Proceedings of the IEEE International Conference on Image Processing, November 2009, pp. 45–48.
[22] M. Turk, A. Pentland, Eigenfaces for recognition, Journal of Cognitive Neuroscience 3 (1) (1991) 71–86.
[23] [Online] AAM Library: 〈http://www2.imm.dtu.dk/~aam/〉.
[24] G.B. Huang, M. Ramesh, T. Berg, E. Learned-Miller, Labeled faces in the wild: A database for studying face recognition in unconstrained environments, University of Massachusetts, Amherst, Technical Report 070-49, October, 2007.
[25] P.J. Phillips, H. Moon, S.A. Rizvi, P.J. Rauss, The FERET evaluation methodology for face-recognition algorithms, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (10) (2000) 1090–1104.

Huang-Chia Shih received the B.Eng degree with the highest honors in electronic engineering from the National Taipei University of Technology, Taipei, Taiwan in 2000 and the M.S. and PhD. degrees in electrical engineering (EE) from the National Tsing Hua University in 2002 and 2008, respectively.

He has been an Assistant Professor in the department of EE, Yuan Ze University, Taoyuan, Taiwan, since August 2010. His research interests are content-based multimedia processing, pattern recognition, and human-computer interaction (HCI). His recent research has been concerned with fusion of biometrics for human recognition and attention-level human activity analysis in HCI.

Dr. Shih served as visiting scholar in Department of EE, University of Washington from September 2006 to April 2007. He served as visiting professor in John von Neumann Faculty of Informatics, Obuda University in Hungary from July 2011 to September 2011. Dr. Shih has received the best doctoral dissertation from Image Processing and Pattern Recognition Society, Taiwan in 2009.