Automatic Eyeglasses Removal from Face Images

Chenyu Wu, Ce Liu, Heung-Yeung Shum, *Member*, *IEEE*, Ying-Qing Xu, and
Zhengyou Zhang, *Senior Member*, *IEEE*

Abstract—In this paper, we present an intelligent image editing and face synthesis system that automatically removes eyeglasses from an input frontal face image. Although conventional image editing tools can be used to remove eyeglasses by pixel-level editing, filling in the deleted eyeglasses region with the right content is a difficult problem. Our approach works at the object level where the eyeglasses are automatically located, removed as one piece, and the void region filled. Our system consists of three parts: eyeglasses detection, eyeglasses localization, and eyeglasses removal. First, an eye region detector, trained offline, is used to approximately locate the region of eyes, thus the region of eyeglasses. A Markov-chain Monte Carlo method is then used to accurately locate key points on the eyeglasses frame by searching for the global optimum of the posterior. Subsequently, a novel sample-based approach is used to synthesize the face image without the eyeglasses. Specifically, we adopt a statistical analysis and synthesis approach to learn the mapping between pairs of face images with and without eyeglasses from a database. Extensive experiments demonstrate that our system effectively removes eyeglasses.

Index Terms—Intelligent image editing, find-and-replace, eye region detection, eyeglasses localization, eyeglasses removal.

--- ✦ ---

1 INTRODUCTION

A N important application for computer vision is intelligent image editing which allows users to easily modify or transfer an image with minimum amount of manual work. Tools such as Intelligent Scissor [6] and Jet Stream [20] were devised to segment images with user-specified semantics. Most image editing operations are pixel-based and time-consuming. For example, to remove scratches from a face image, a user needs to painstakingly mark all the pixels that are considered as damaged. To repair the damaged image areas, image inpainting approaches (e.g., [2]) can be used to fill in the blank region by using information from nearby pixels.

Often people need to edit human face images. A good example is to remove red eyes commonly seen in images taken with a flash. Interesting global operations on human faces include changing various lighting effects [28] and modifying facial expressions [17]. In this paper, our goal is to automatically remove eyeglasses from a human face image (Fig. 1). Because of significant variations in the geometry and appearance of eyeglasses, it is very useful to construct a face image without eyeglasses on which many

existing face analysis and synthesis algorithms can be applied. For instance, we may start with a face image with eyeglasses, create its corresponding face image without eyeglasses, then generate a cartoon sketch for this face [4], and, finally, put an eyeglasses template back to complete the cartoon sketch.

Conventional image editing tools can be used to painstakingly mark all pixels of eyeglasses frame from a face image. These pixels are then deleted. It is, however, difficult to fill in this rather large deleted region with the right content. We propose in this paper a novel find-and-replace approach at the object level instead of at the pixel level. In our approach, we find the eyeglasses region and replace it with a synthesized region with eyeglasses removed. The synthesized new region is obtained by using the information from the detected eyeglasses region from the given face image and, more importantly, based on a learned statistical mapping between pairs of face images with and without eyeglasses. This find-and-replace approach can work with other facial components (e.g., beard and hair) as well. It is also worthwhile to emphasize that our approach is used for the purpose of image synthesis, not for recognition.

1.1 Previous Work

Our approach is related to several statistical learning methods developed recently in computer vision. For instance, a statistical technique [8] was proposed to learn the relationship between high resolution and low resolution image pairs. Similarly, image analogy [10] learns the mapping relationship between a pair (or pairs) of training images, which is then applied to the new input and synthesized image pair. These statistical approaches generally use a number of training data to learn a prior model of the target and then apply Bayesian inference to

- C. Wu is with Robotics Institute, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213. E-mail: chenyuwu@cmu.edu.
- C. Liu is with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Apt. 12A2, 550 Memorial Drive, Cambridge, MA 02139. E-mail: celiu@mit.edu.
- H.-Y. Shum and Y.-Q. Xu are with Microsoft Research Asia, Sigma Center, No. 49 Zhichun Road, Haidian District, Beijing 100080. E-mail: {hshum, yqxu}@microsoft.com.
- Z. Zhang is with Microsoft Corp., One Microsoft Way, Redmond, WA 98052-6399. E-mail: zhang@microsoft.com.
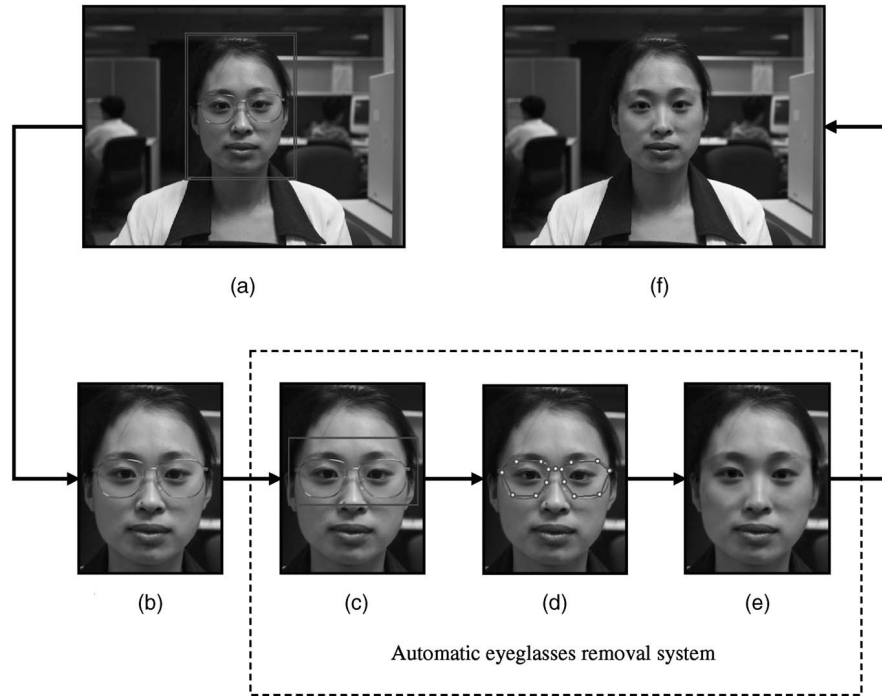
Fig. 1. The automatic eyeglasses removal system consists of three parts: detection, localization, and removal. The removal system can be used for object-level image editing. (a) An input image is first scanned by a face detector and the face wearing eyeglasses is selected by the user. (b) The cropped face image. (c) Eye area detection to roughly estimate the region of eyeglasses. (d) Eyeglasses localization by MCMC, which yields an accurate position of the glasses. (e) Eyeglasses removal based on a set of training examples, which removes the glasses by inferring and pasting the glasses-free pattern. (f) The final result on the original image.

synthesize new images by maximizing a posteriori. Such approaches have also been applied to human faces in face hallucination [1], [15] and face cartoon sketch generation [4]. We briefly review related work in face analysis and statistical learning before introducing our approach.

Many statistical learning-based methods have been successfully deployed. For instance, neural networks [22], [25], support vector machines (SVM) [19], wavelets [24], and decision trees [27] have been applied to face detection. Deformable shape models [30], [13] and *active shape models* (ASM) [5] have been demonstrated to be effective in localizing faces. Meanwhile, appearance models such as *active appearance models* (AAM) [5] are developed to include texture information for better localization. These object detection and alignment algorithms are useful in automatically locating user-intended objects in a cluttered image.

In facial image analysis, people usually focus on facial parts such as eyes, nose, mouth and profile. For example, in the neural network-based face detection work [22], specific feature detectors are devised corresponding to each facial part in order to improve detection. In ASM [5], facial key points (landmarks) are defined as the edge points and corners of various parts and profile. In face sketch generation work [4], a specific eye model is designed to sketch the eye. Unfortunately, the presence of occluders such as eyeglasses tends to compromise the performance of facial feature extraction and face alignment. However, face detection is not very much affected by eyeglasses except for extreme cases of sunglasses and eyeglasses with significant reflection where the eyes are completely occluded.

Several researchers have worked on eyeglasses recognition, localization, and removal recently. Jiang et al. [11] used a glasses classifier to detect glasses on facial images. Wu et al. [29] devised a sophisticated eyeglasses classifier based on SVM, with a reported recognition rate close to 90 percent. Jing and Mariani [12] employed a deformable contour method to detect glasses under a Bayesian framework. In their work, 50 key points are used to define the shape of glasses, and the position of glasses is found by maximizing the posteriori. Saito's eyeglasses removal work [23] is based on principal component analysis (PCA). The eigen-space of eyeglasses-free patterns is learned by PCA to retain their principal variance. Projecting a glasses pattern into this space results in the corresponding glasses-free pattern. However, the joint distribution between glasses and glasses-free patterns is not discussed.

1.2 Our Approach

Fig. 1 shows how our eyeglasses removal system is used for intelligent image editing. Suppose that we start with an image where a face with eyeglasses is present. Using conventional face detector (not discussed in this paper), we crop the face image out as the input to our system. Our system consists of three modules: eyeglasses detection, eyeglasses localization, and eyeglasses removal. Once the eyeglasses region is detected and key points on eyeglasses frame are located, we apply a novel approach to synthesizing the face image with eyeglasses removed. The eyeglasses-free face image then replaces the original face image to output an edited new image.
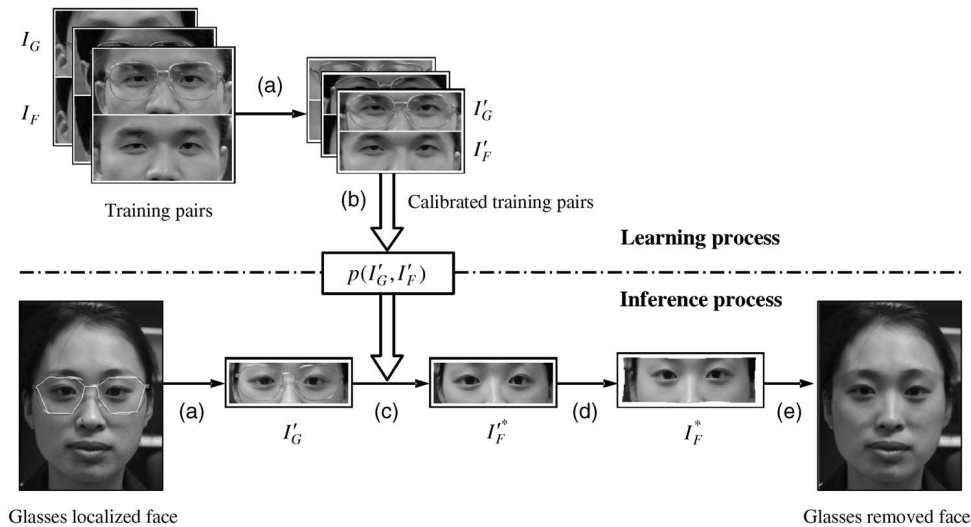
Fig. 2. The eyeglasses removal system consists of two parts: The learning process is shown in the upper half and the inference process shown in the bottom half. (a) By warping the training pairs of images to a face template and then an eyeglasses template, we obtain a calibrated training pairs of regions with and without eyeglasses. (b) Learning the correspondence or joint distribution of glasses and glasses-free regions. (c) Inferring the best-fit glasses-free region from the calibrated region with glasses based on the learned distribution. (d) Inverse warping to the original image. (e) Paste the inferred glasses-free region onto the input image with boundary blended.

Central to our system is a sample-based approach that learns the statistical mapping between face images with eyeglasses and their counterparts without eyeglasses. From a number of training pairs of face images with and without eyeglasses, we model their joint distribution effectively in an eigenspace spanned by all training pairs. With the learned joint distribution, given a face image with eyeglasses, we can obtain in closed form its corresponding face image without eyeglasses.

The success of our approach depends on how well we construct the eigenspace on which the joint distribution is constructed. We propose an eyeglasses frame template and a face template to calibrate all training data. By warping all training pairs to the templates, we reduce the geometrical misalignment between various eyeglasses frames. The warped and calibrated pairs of face images with and without eyeglasses would constitute a more compact eigenspace. Similar to ASM in face alignment, we construct an active eyeglasses shape model to accurately locate these key points. We also propose an eyeglasses detection scheme (similar to boosting-based face detection [27], [14]) to initialize the process of eyeglasses localization.

The remainder of this paper is organized as follows: Section 2 describes the sample-based approach to eyeglasses removal. Section 3 discusses how to locate the key points on the eyeglasses frame. Eyeglasses region detection is also briefly discussed. Experimental results on eyeglasses detection, localization and removal are shown in Section 4. We summarize this paper in Section 5.

2 EYEGLASSES REMOVAL

An overview of eyeglasses removal is illustrated in Fig. 2. Given an input face image with eyeglasses, it is compared with a training set of face image pairs with and without eyeglasses to infer the corresponding face image without eyeglasses. Note that each face image is calibrated to a face template and an eyeglasses template to reduce the geometrical misalignment among different face images.

2.1 A Sample-Based Approach

We denote the calibrated pair of glasses and glasses-free images by I'_G and I'_F, respectively. Based on the maximum a posteriori (MAP) criterion, we may infer the optimal I'_F from I'_G by

$$I'^{*}_F = \arg\max_{I'_F} p(I'_F | I'_G)$$

$$= \arg\max_{I'_F} p(I'_G | I'_F) p(I'_F) \quad (1)$$

$$= \arg\max_{I'_F} p(I'_G, I'_F). \quad (2)$$

Because of the high dimensionality of I'_F and I'_G, modeling the conditional density or likelihood $p(I'_G | I'_F)$ is difficult. Instead, we choose (2) as the objective function.

We estimate the joint distribution $p(I'_F, I'_G)$ by introducing a hidden variable V which dominates the main variance of I'_F and I'_G

$$p(I'_F, I'_G) = \int p(I'_F, I'_G | V) p(V) dV$$

$$= \int p(I'_F | V) p(I'_G | V) p(V) dV. \quad (3)$$

The second line of (3) assumes that I'_F and I'_G are conditionally independent given the hidden variable. But, how does one choose the hidden variable V? A popular method is to set it as the principal components of I'_F and I'_G. Let $Y^T = [I'^T_G \; I'^T_F]$ be a long vector with two components I'_G and I'_F, and the training examples become $\{Y(i), i = 1, \cdots, M\}$. Through singular value decomposition (SVD), we can get the principal components matrix $\Psi = [\psi_1 \, \psi_2 \, \cdots \, \psi_h]$ with ψ_j the jth eigenvector, the eigenvalues $\{\sigma_i^2, i = 1, \cdots, h\}$, and the mean μ_Y. A number of

principal components are chosen such that the sum of eigenvalues corresponding to the principal components accounts for no less than 97 percent of the sum of the total eigenvalues. PCA yields a linear dimensionality reduction to Y by

$$Y = \boldsymbol{\Psi}V + \mu_Y + \varepsilon_Y, V = \boldsymbol{\Psi}^T(Y - \mu_Y), \tag{4}$$

where ε_Y is a Gaussian noise. By PCA, $V \in \mathbb{R}^h$ and $Y \in \mathbb{R}^{2m}$ with the condition $h \ll 2m$. Let

$$\boldsymbol{\Psi} = \begin{bmatrix} \boldsymbol{\Psi}_G \\ \boldsymbol{\Psi}_F \end{bmatrix}, \ \mu_Y = \begin{bmatrix} \mu_G \\ \mu_F \end{bmatrix}, \ \varepsilon_Y = \begin{bmatrix} \varepsilon_G \\ \varepsilon_F \end{bmatrix}. \tag{5}$$

We have

$$\begin{aligned} I'_G &= \boldsymbol{\Psi}_G V + \mu_G + \varepsilon_G \\ I'_F &= \boldsymbol{\Psi}_F V + \mu_F + \varepsilon_F \end{aligned}, \tag{6}$$

which indicates

$$\begin{aligned} p(I'_G|V) &= \frac{1}{Z_G}\exp\left\{-\frac{\|I'_G - (\boldsymbol{\Psi}_G V + \mu_G)\|^2}{\sigma_G^2}\right\} \\ p(I'_F|V) &= \frac{1}{Z_F}\exp\left\{-\frac{\|I'_F - (\boldsymbol{\Psi}_F V + \mu_F)\|^2}{\sigma_F^2}\right\} \end{aligned}, \tag{7}$$

where σ_G and σ_F are the variances of ε_G and ε_F, and Z_G and Z_F are normalization constants, respectively. The distribution of the hidden variable is also Gaussian

$$p(V) = \frac{1}{Z_V}\exp\{-V^T \boldsymbol{\Lambda}_V^{-1} V\}, \tag{8}$$

where $\boldsymbol{\Lambda}_V = \mathrm{diag}[\sigma_1^2, \sigma_2^2, \cdots, \sigma_h^2]$ is the covariance matrix and Z_V is the normalization constant.

As mentioned above, in PCA at least 97 percent of the total variance of Y is retained in V, which implies

$$\frac{\sigma_G^2 + \sigma_F^2}{\mathrm{trace}(\boldsymbol{\Lambda}_V)} \le \frac{3}{97} \approx 3.09\%, \tag{9}$$

where $\mathrm{trace}(\boldsymbol{\Lambda}_V) = \sum_{i=1}^h \sigma_i^2$ is the total variance of V. In other words, $p(V)$ captures the major and always global uncertainty of Y, while $p(I'_G|V)$ and $p(I'_F|V)$ just compensate the approximation error from V to Y (4). Mathematically,

$$\mathrm{entropy}\{p(I'_G|V)\} + \mathrm{entropy}\{p(I'_F|V)\} \ll \mathrm{entropy}\{p(V)\}. \tag{10}$$

Let us come back to the optimization problem. From (2) and (3), we have

$$I'^*_F = \arg\max_{I'_F} \int p(I'_F|V)p(I'_G|V)p(V)dV. \tag{11}$$

Obviously, the function to integrate $p(I'_F|V)p(I'_G|V)p(V)$ has a sharp peak around $\boldsymbol{\Psi}^T(Y - \mu_Y)$, formed by the constraint from I'_F and I'_G. Therefore, maximizing the integration can be approximated by maximizing the function to be integrated,

$$\{I'^*_F, V^*\} = \arg\max_{I'_F, V} p(I'_F|V)p(I'_G|V)p(V). \tag{12}$$

This naturally leads to a two-step inference

$$\begin{aligned} a. \quad & V^* = \arg\max_V p(I'_G|V)p(V); \\ b. \quad & I'^*_F = \arg\max_{I'_F} p(I'_F|V). \end{aligned} \tag{13}$$

This approximate inference is straightforward. From (10), we know that $p(V)$ can approximate $p(I'_F, I'_G)$ with an error less than 3 percent. So, given I'_G, we may directly find the best hidden variable V^* with its glasses component closest to I'_G. The optimal glasses-free part I'_F best fit for I'_G is naturally the glasses-free component of V^*.

To maximize the posteriori in (13) a is equivalent to minimizing the energy

$$\begin{aligned} V^* = \arg\min_V \{&\sigma_G^2 V^T \boldsymbol{\Lambda}_V^{-1} V + \\ &(\boldsymbol{\Psi}_G V + \mu_G - I'_G)^T(\boldsymbol{\Psi}_G V + \mu_G - I'_G)\}. \end{aligned} \tag{14}$$

Since the objective function is a quadratic form, we can get a closed-form solution:

$$V^* = (\boldsymbol{\Psi}_G^T \boldsymbol{\Psi}_G + \sigma_G^2 \boldsymbol{\Lambda}_V^{-1})^{-1} \boldsymbol{\Psi}_G^T(I'_G - \mu_G). \tag{15}$$

To ensure numerical stability, the inverse $(\boldsymbol{\Psi}_G^T \boldsymbol{\Psi}_G + \sigma_G^2 \boldsymbol{\Lambda}^{-1})^{-1}$ is computed by the standard SVD algorithm. In the global solution (15), all the matrices can be computed and stored offline so that the solution is very fast. Finally, the optimal glasses-free region I'^*_F is calculated by maximizing (13)

$$\begin{aligned} I'^*_F &= \boldsymbol{\Psi}_F V^* + \mu_F \\ &= \boldsymbol{\Psi}_F(\boldsymbol{\Psi}_G^T \boldsymbol{\Psi}_G + \sigma_G^2 \boldsymbol{\Lambda}_V^{-1})^{-1} \boldsymbol{\Psi}_G^T(I'_G - \mu_G) + \mu_F. \end{aligned} \tag{16}$$

Then, I'^*_F is inversely warped to the glasses region of I_G, denoted by I^*_F. Finally, I^*_F is pasted on the face with the abutting area blended around the boundary.

2.2 Experimental Setup

Collecting a good training database is crucial for sample-based learning approaches. Unfortunately, there are few face image databases available that contain pairs of corresponding face images with and without eyeglasses. We have found a very small sample set of 20 pairs of such images from the FERET face database [21]. Moreover, we have captured 247 image pairs of people in our lab. All images contain upright frontal faces wearing conventional eyeglasses. Some typical samples are shown in Fig. 3. The extreme appearance of eyeglasses such as sunglasses and pure specular reflections are not taken into account in our system.

In the training procedure, we manually labeled the pair of glasses and glasses-free samples as shown in Fig. 3. Each sample, wearing glasses or not, is labeled by seven landmarks as key points to represent the face. As shown in Figs. 4e and 4f, the image pair is normalized with these seven landmarks by affine transform Figs. 4c and 4d to be frontal parallel and then cropped as a normalized pair of glasses and glasses-free regions. Each eyeglasses frame uses 15 additional key points to represent its shape. These 15 points, along with four key points of eye corners, are used to warp the normalized pair of glasses and glasses-free regions to the calibrated training pair that are used to learn the correspondence. These two steps of warping with
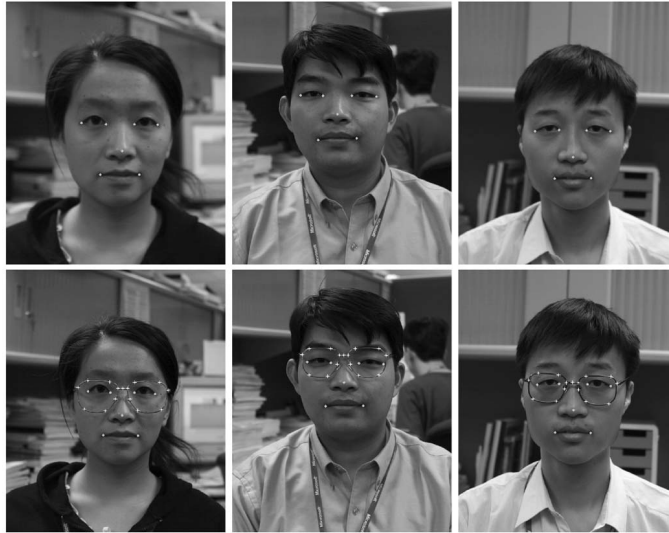
Fig. 3. Some face images with and without glasses sample in the data set. Top: seven key points to align face are marked in red. Bottom: 15 key points are marked in yellow on the eyeglasses frame, in addition to several key points on face.
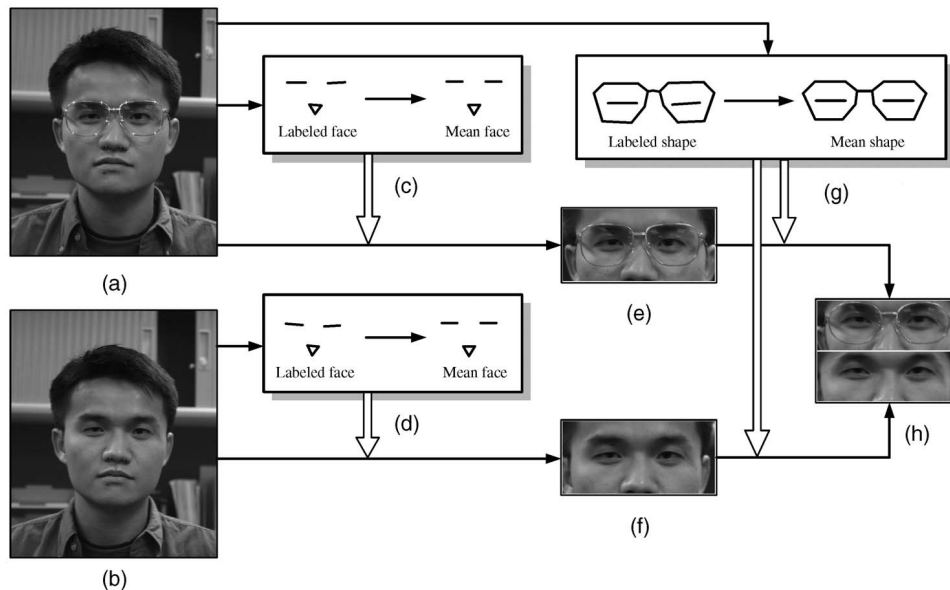

Fig. 4. A training pair of glasses and glasses-free regions is first normalized to a face template and then warped to an eyeglasses template to construct a calibrated training pair. Different transforms are applied to the training pair ((a) and (b)) to generate normalized pair of images ((e) and (f)). However, the same warping field is applied to normalized images pair.

respect to the face template (seven key points) and eyeglasses template (15 key points on the frame and four on the face) are necessary to calibrate each training sample pair to reduce the misalignment due to geometrical variations of faces and eyeglasses. In this paper, we use the warping method of *thin plate splines* [3], [5], which has been frequently used in statistical shape analysis and computer graphics because of its smooth deformations.

To begin the inference procedure shown in Fig. 2 with a new face image, all the key points must be automatically localized. This is the topic for the next section.

3 EYEGLASSES LOCALIZATION

To accurately locate eyeglasses, we use a *deformable contour model* [13], [30] or *active shape model* (ASM) [5] to describe the geometric information such as shape, size, and position of

the glasses. We denote all key points on the eyeglasses frame by $W = \{(x_i, y_i), i = 1, \cdots, n\}$, where $n = 15$. In the following, W is a long vector with dimensionality of $2n$. Based on the Bayesian rule, to locate the position is to find an optimal W^* in the eyeglasses region I_G by maximizing the posterior, or the product of the prior and likelihood

$$W^* = \arg \max_W p(W|I_G) = \arg \max_W p(I_G|W)p(W). \quad (17)$$

We shall learn the prior $p(W)$ and likelihood $p(I_G|W)$, respectively, and then design an optimization mechanism to search for the optimal solution.

3.1 Prior Learning

The prior distribution of W is composed of two independent parts: *internal* parameters or the position invariant shape w, and *external* parameters or the position relevant
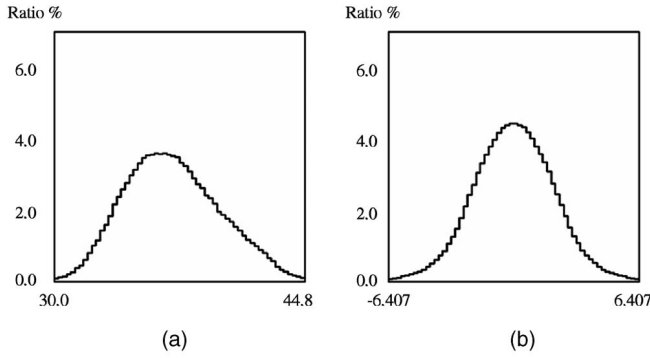
Fig. 5. The discretized histograms of *external* parameters: (a) scale and (b) orientation. They have been smoothed by a Gaussian filter.

variables such as the orientation θ, scale s, and centroid C_{xy}. With a linear transition matrix $T_{(s,\theta)}$ to scale and rotate the shape w, we get

$$W = T_{(s,\theta)}w + C_{xy}, T_{(s,\theta)} = s\begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}. \quad (18)$$

The position invariant shape $w = \{(x_i', y_i'), i = 1, \cdots, n\}$ is the centralized, scale, and orientation normalized key points. Since w in the *internal* parameter space and the position correlated s, θ, and C_{xy} in the *external* parameter space are statistically independent, we can decouple the prior distribution into

$$p(W) = p(w)p(s)p(\theta)p(C_{xy}). \quad (19)$$

We shall model the *internal* and *external* priors in different ways.

Similar to ASM and many other points distribution models, the distribution of the *internal* parameters w is also assumed Gaussian. We use PCA to reduce the dimensionality and learn the distribution. Suppose

$$u = \mathbf{B}^T(w - \mu_w), \quad (20)$$

where $\mathbf{B} = [b_1, \cdots, b_m]$ is the matrix in which the column vectors are principal components, the bases of the reduced feature space, $u \in \mathbb{R}^m$ is the controlling variable and μ_w is the mean shape. We also obtain the eigenvalue $\mathbf{\Lambda} = \text{diag}(\sigma_1^2, \cdots, \sigma_m^2)$ scales the variance in each dimension of the feature space. The principal components matrix B and the variances $\mathbf{\Lambda}$ are chosen to explain the majority,

97 percent, for instance, of the eigenvalues of the covariance matrix of a set of training examples $\{w_k, k = 1, \cdots M\}$. We may directly use u to approximate w

$$w = \mathbf{B}u + \mu_w, p(u) = \frac{1}{Z}\exp\{-u^T\mathbf{\Lambda}^{-1}u\}. \quad (21)$$

The principal components b_i denote the main variations of the shape w. This implies that it is more meaningful to vary along the principal components of the shape, the whole set of individual points than a single point in estimating the shape.

The *external* parameters, the scale, orientation, and central point are all in low dimension(s) (1D or 2D). We simply use histograms to represent their distributions as shown in Fig. 5.

3.2 Likelihood Learning

The likelihood $p(I_G|W)$ is used to measure if the local features $F_G^{(i)}$ on point (x_i, y_i) are similar to those of the key point in terms of the appearance. Under the assumption of independency, it can be simplified to

$$p(I_G|W) = p(F_G|W) = \prod_{i=1}^{n} p(F_G^{(i)}|(x_i, y_i)), \quad (22)$$

where $F_G = \{F_j * I_G, j = 1, \cdots, l\}$ is the feature images filtered by linear filters $\{F_j\}$. What we should learn is the local feature distribution $p(F_G^{(i)}|(x_i, y_i))$ for each key point.

Since the key points of eyeglasses are defined on the eyeglasses frame, they are distinct in edge and orientation. We choose the responses of local edge detectors as the features, including a Laplacian operator, the first and second order Sobel operators with four orientations. These operators are all band-pass filters to capture the local space-frequency statistics. We find that the results obtained with these filters are comparable with those using wavelets filters such as Gabors, and these filters have the additional benefit of being more efficient. Before filtering, we use a Gaussian filter to smooth out the noise. All these filters are displayed in Fig. 6, and an example of the filtering results is shown in Fig. 7.

For each key point, we can get the training vector-valued features $\{F_G^{(i)}(j)|j = 1, \cdots, M\}$. Then, the likelihood of each key point can be estimated by a Gaussian mixture model, i.e.,

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} \quad \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \\ -2 & -1 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \quad \begin{bmatrix} -2 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

Laplacian   Sobel(1) 0° Sobel(1) 45° Sobel(1) 90° Sobel(1) 135°

$$\begin{bmatrix} 1 & 3 & 1 \\ 3 & 4 & 3 \\ 1 & 3 & 1 \end{bmatrix} \quad \begin{bmatrix} -4 & 8 & -4 \\ -5 & 10 & -5 \\ -4 & 8 & -4 \end{bmatrix} \quad \begin{bmatrix} 6 & -4 & -3 \\ -4 & 10 & -4 \\ -3 & -4 & 6 \end{bmatrix} \quad \begin{bmatrix} -4 & -5 & -4 \\ 8 & 10 & 8 \\ -4 & -5 & -4 \end{bmatrix} \quad \begin{bmatrix} -3 & -4 & 6 \\ -4 & 10 & -4 \\ 6 & -4 & -3 \end{bmatrix}$$

Gaussian   Sobel(2) 0° Sobel(2) 45° Sobel(2) 90° Sobel(2) 135°

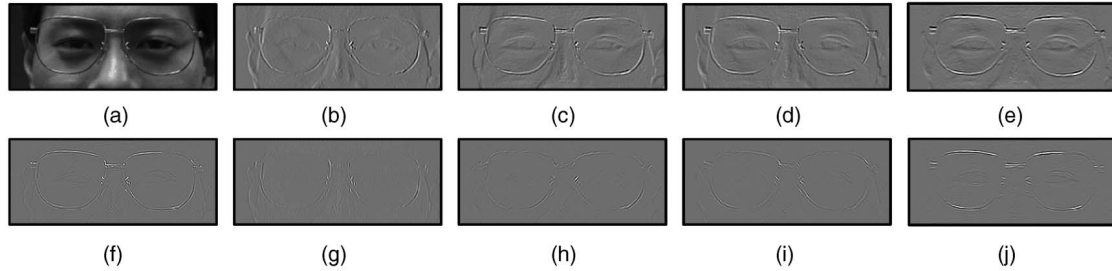Fig. 6. The filters whose responses are used as features in likelihood modeling.

Fig. 7. The local features captured by different small band-pass filters from (b) to (j). (a) Original, (b) Sobel(1) $0°$, (c) Sobel(1) $45°$, (d) Sobel(1) $90°$, (e) Sobel(1) $135°$, (f) Laplacian, (g) Sobel(2) $0°$, (h) Sobel(2) $45°$, (i) Sobel(2) $90°$, (j) Sobel(2) $135°$.
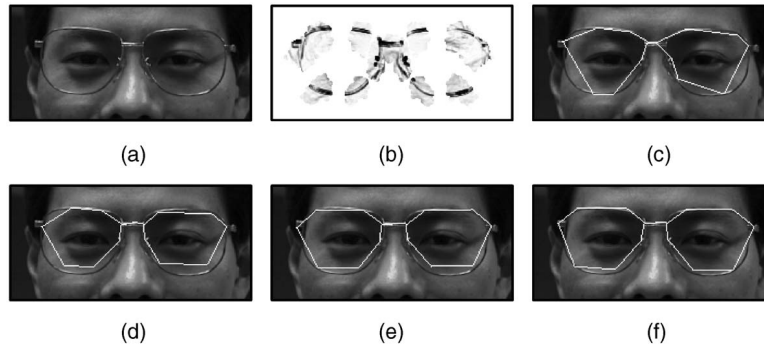

Fig. 8. Localization procedure. (a) Input inage. (b) Salient map. (c) Initialization. (d) After one iteration. (e) After five iterations. (f) After 10 iterations.

$$p(F_G^{(i)}|x_i, y_i) = \sum_{k=1}^{K} \alpha_k^{(i)} G(F_G^{(i)}; \mu_k^{(i)}, \Sigma_k^{(i)}), \qquad (23)$$

where weights $\{\alpha_k^{(i)}\} : \sum_k \alpha_k^{(i)} = 1$, means $\{\mu_k^{(i)}\}$ and covariance matrices $\{\Sigma_k^{(i)}\}$ are learned using EM algorithm.

3.3 MAP Solution by Markov-Chain Monte Carlo

After the prior and likelihood models are learned, we should find the optimal W^* by maximizing the posterior under the MAP criterion. However, the posterior is too complex to be globally optimized. Conventional deterministic gradient ascent algorithms tend to get stuck at local optima. Markov-chain Monte Carlo (MCMC) is a technique with guaranteed global convergence and, thus, is chosen to locate the eyeglasses in our system. MCMC has been successfully used recently in Markov random field learning [31], face prior learning [16], structure from motion [7], and

image segmentation [26]. Specifically, we choose Gibbs sampling [9] over Metropolis-Hastings sampling [18] in optimization because it has a low rejection ratio and it does not require the design of a sophisticated proposal probability. Since the key points W have been decoupled to *internal* and *external* parameters by (18) and the *internal* parameter w has been reduced to the controlling variable u by (21), the solution space is simplified to $X = \{u, s, \theta, C_{xy}\}$. Suppose $X = (x_1, x_2, \cdots, x_k)$, the Markov chain dynamics in Gibbs sampling is given by

$$
\begin{aligned}
x_1^{(t+1)} &\sim & p(x_1|x_2^{(t)}, x_3^{(t)}, \cdots, x_k^{(t)}) \\
x_2^{(t+1)} &\sim & p(x_2|x_1^{(t+1)}, x_3^{(t)}, \cdots, x_k^{(t)}) \\
&\vdots & \\
x_k^{(t+1)} &\sim & p(x_k|x_1^{(t+1)}, x_2^{(t+1)}, \cdots, x_{k-1}^{(t+1)}),
\end{aligned}
\qquad (24)
$$

where the conditional density is directly computed from the joint density $p(x_1, x_2, \cdots, x_k)$. By sequentially flipping each dimension, the Gibbs sampler walks through the solution space with the target posterior probability density
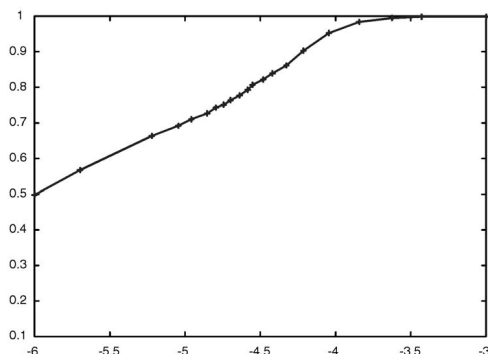

Fig. 9. The ROC curve of eye area detector testing on 1,386 face images containing eyeglasses or not. The unit of the horizontal axis is the power of 10.
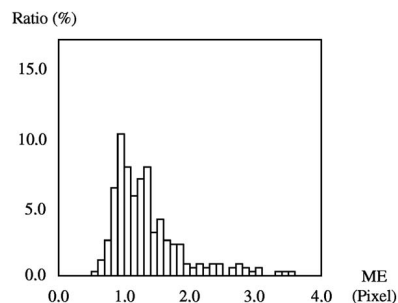

Fig. 10. The distribution of mean error in eyeglasses localization for the test data.
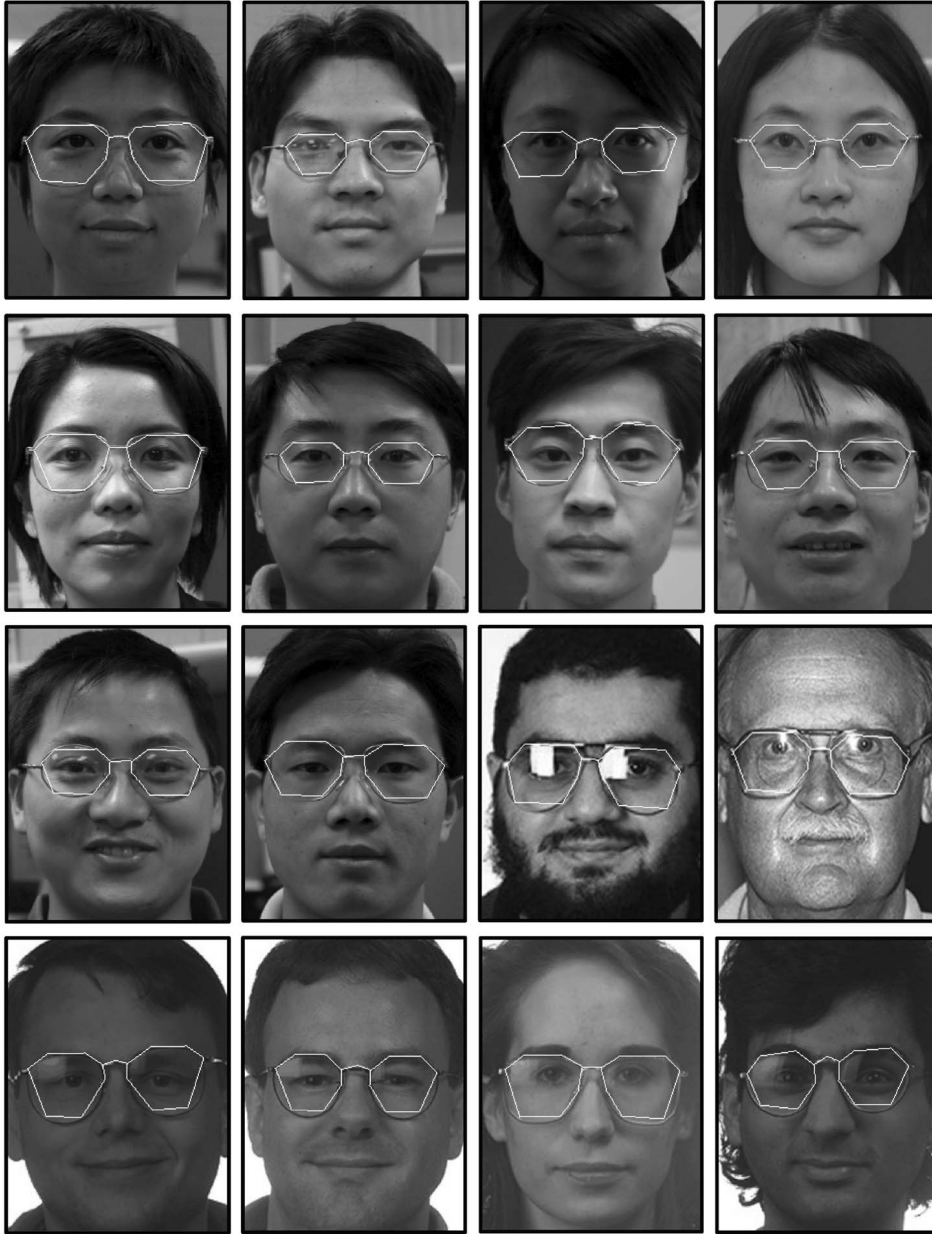
Fig. 11. The results of eyeglasses localization by MCMC.

given by (17), (21), and (23). After R jumps in the Markov chain, the optimal solution X^* is obtained in the samples $\{X^{(1)}, X^{(2)}, \cdots, X^{(R)}\}$ drawn by the Gibbs sampler. Finally, W^* is computed by (18).

To save computational cost, we generate a *salient map* $\{p(x_i, y_i | F_G), i = 1, \cdots, n\}$ for each key point (x_i, y_i)

$$p(x_i, y_i | F_G) \propto p(F_G^{(i)} | x_i, y_i) p(x_i, y_i) \propto p(F_G^i | x_i, y_i). \quad (25)$$

The second "$\propto$" exists because each key point has nearly the same opportunity to appear in the image. The salient maps can be regarded as the likelihood maps as a preprocessing step before Gibbs sampling. They can also be used for initialization by sampling $p(x_i, y_i | F_G)$ or choosing the maximum probability point.

The mechanism of the proposed method is illustrated in Fig. 8. First of all, the salient maps (Fig. 8b) of 15 key points

are computed from the input (Fig. 8a) which takes only 0.2 seconds on a Pentium IV 2G Hz computer with 512M memory. Initialization is done by simply looking for the optimal likelihood points in the salient map. Then, the Gibbs sampler flips the parameters sequentially and completes an iteration by flipping all parameters. The localization results after 1, 5, and 10 iterations are shown in Figs. 8d, 8e, and 8f, respectively. Clearly, Gibbs sampling converges quickly to the global optimum.

3.4 Eyeglasses Region Detection

It is helpful to know where the eyeglasses region is approximately located in the face image before the face localization step is invoked. Our key observation is that the eyeglasses region can be located by finding the eye area including eyes and eyebrows without considering eyeglasses. Although eyeglasses regions vary significantly from
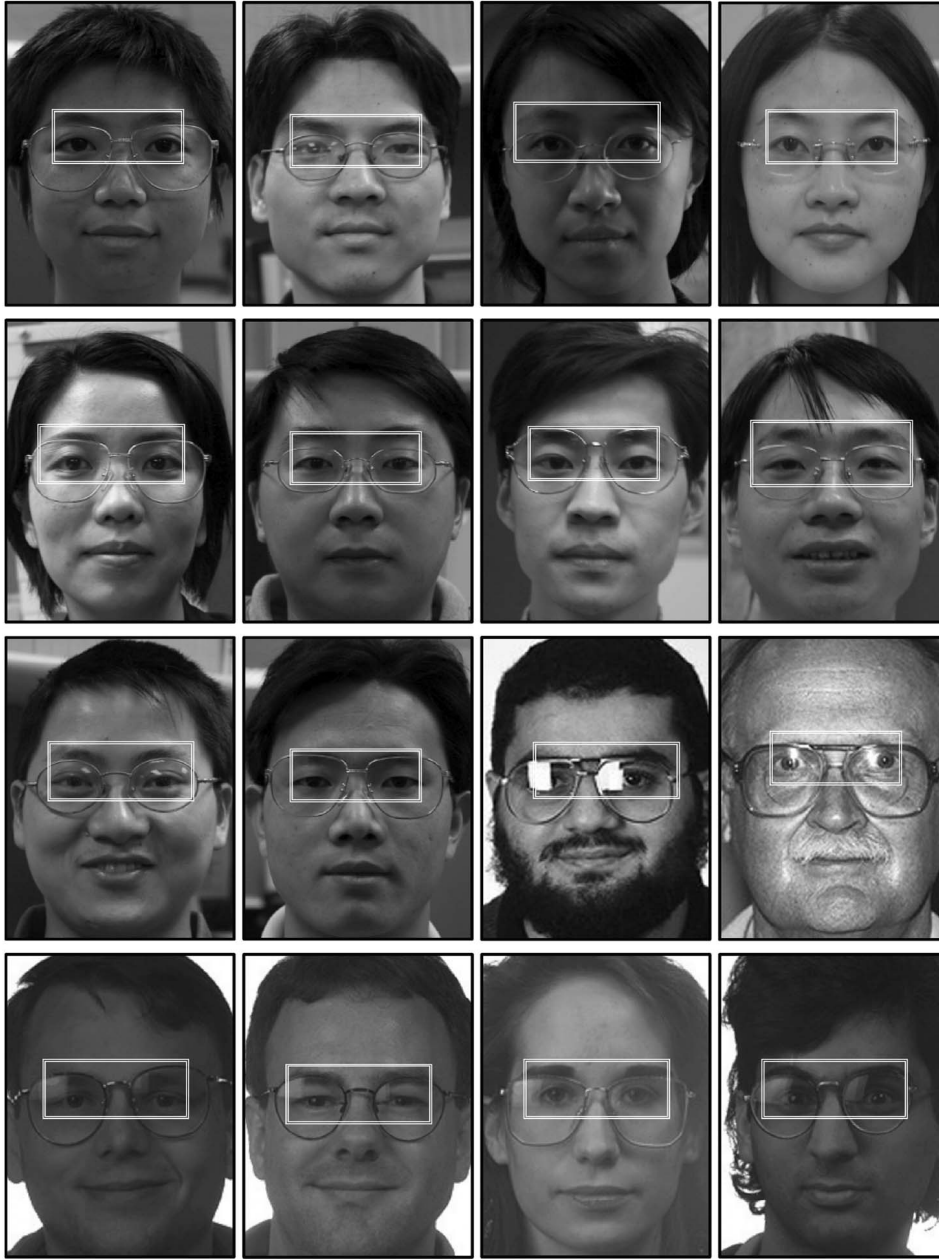
Fig. 12. The results of eye area detection.

one to another, and the relative positions between eyeglasses and eyes change from case to case, eyes and very often (parts) of eyebrows are visible in all face images (except sunglasses where eyes may be completely blocked). Therefore, we do not train a detector for all possible eyeglasses patterns. Rather, we train a detector for the eye area.

Our eye area detector is in principle very similar to the face detector [27]. The eye area is defined as a 30×12 patch cropped out from a face image. A dense Gaussian pyramid is built upon this patch as well. Then, a boosting classifier (i.e., a cascade of eye area classifiers) scans the pyramid shifting by two pixels and verify if each 30×12 patch is an eye area. A brief introduction to boosting classifier can be found in the Appendix. We have adopted the KLBoosting classifier introduced in [14] and use a cascade of KLBoosting classifier to quickly rule out the most dissimilar

negative samples. The training results in a 7-level cascade of classifiers, by retaining the false negative ratio no more than 0.5 percent while forcing the false alarm ratio under 30 percent in each level.

For training data, we have collected 1,271 face images covering all possible cases of shape, gender, age, race, pose, and lighting. For each face image, we manually cropped out an eye area. By randomly perturbing each eye area in scale and orientation as in [22], the number of the positive samples is significantly increased. we obtain 12,710 positive samples. All other possible 30×12 patches in a dense Gaussian pyramid for each face image are chosen as negative samples, with the criterion that the negative patch must be within the face and should not overlap the positive patch too much. We have generated 8.2×10^7 negative samples from face images in our experiment.
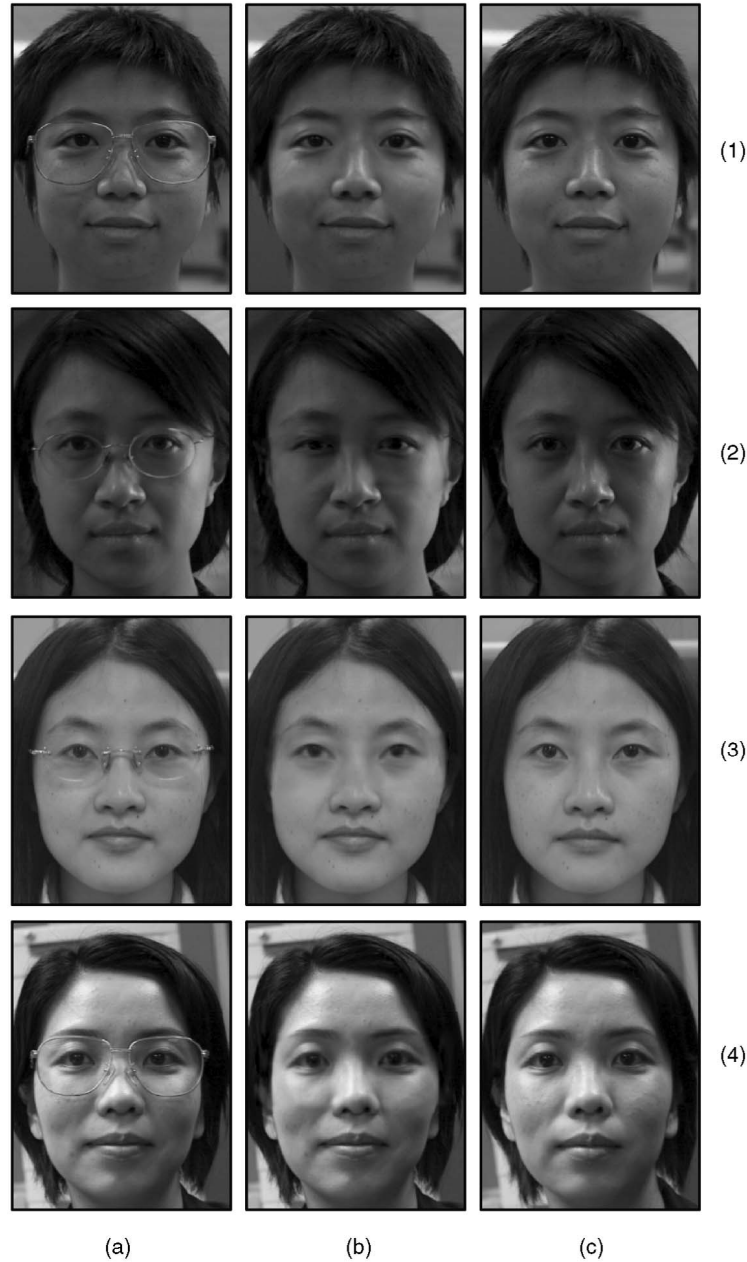
Fig. 13. The results of eyeglasses removal from samples (1) to (4). (a) Input images with glasses. (b) Synthesized results. (c) Images of (a) taken without glasses.

4 EXPERIMENTAL RESULTS

4.1 Eyeglasses Detection

We have also collected another test data set consisting of 1,386 face images that are not in the training data. The ROC curve is shown in Fig. 9 to demonstrate the eye area detector performance. Although this is slightly worse than the latest face detection result [14], it is good enough for our purpose with the detection rate of 96 percent and the false alarm at 10^{-4}. Some typical detection results for the test data are presented in Fig. 12. The average time to detect an eye area from a 96×128 image takes 0.1 second. These results demonstrate that our algorithm is effective and efficient. Note that the location of the eyeglasses pattern is defined as twice as big as the detected eye area.

4.2 Eyeglasses Localization

Eyeglasses localization results are shown in Fig. 11. The number of Gibbs sampling iterations is set to be 100 to ensure for most cases that satisfactory results can be reached. These results demonstrate that our method works for a variety of samples with difference face profiles, illuminations, eyeglasses shapes, and highlight distributions. For quantitative analysis, we plot the distribution of the *mean error*(ME), namely, the average difference between the result from automatic localization and that from manual labeling, computed using the test data in Fig. 10. In this figure, 90.3 percent of ME lie within two pixels, which also demonstrates the effectiveness of our algorithm.

In our localization experiment, we have collected 264 samples for training and 40 for testing. More than 95 percent
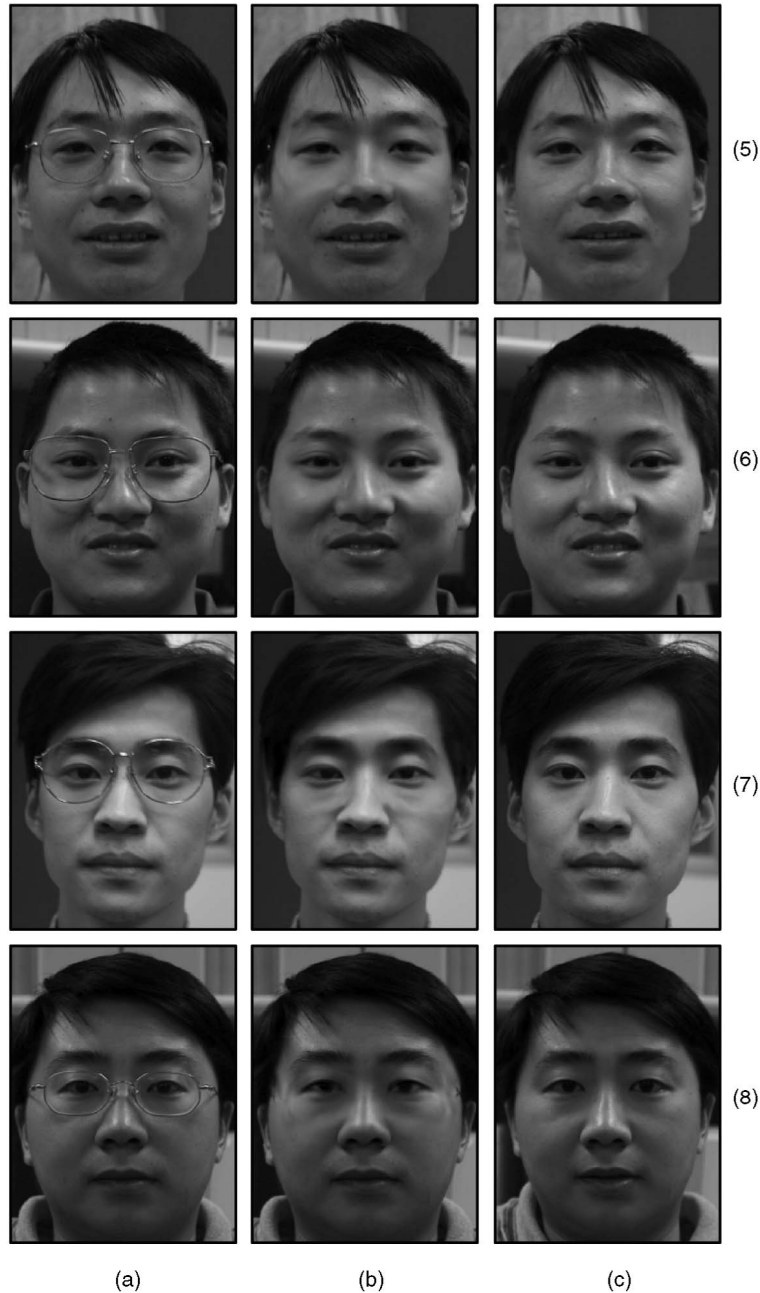
Fig. 14. The results of eyeglasses removal from samples (5) to (8). (a) Input images with glasses. (b) Synthesized results. (c) Images of (a) taken without glasses.

samples in our experiment can be detected and localized well. Much of the detection and localization error is due to the fact that we train our system based on a limited number of training data.

4.3 Eyeglasses Removal

Some results with eyeglasses removal are shown in Figs. 13, 14, and 15, labeled from (1) to (12). In these three figures, column a represents input images with eyeglasses, column b synthesized images with glasses removed using our system, and column c images of the same people taken without eyeglasses. We recommend the reader to view the pdf file for better visualization of results. Note that different types of glasses from our own database and the FERET database have

been used in our experiment. Even though the synthesized glasses-free images appear slightly blurred and somewhat different from those that would have been taken without eyeglasses, the visual quality of images in column b with eyeglasses removed is acceptable. Residues of eyeglasses frame can be seen in Fig. 14 sample 8b because this part of the frame was not modeled in the eyeglasses template with 15 key points. Some irregularities can also be observed in Fig. 14 (samples 5, 8), and Fig. 15 (sample 11) around the glasses boundaries, due to the modeling error of PCA.

Comparison between our results and those from image inpainting [2] are shown in Fig. 16. The user specifies the regions of frame, highlights, and shadows brought by the
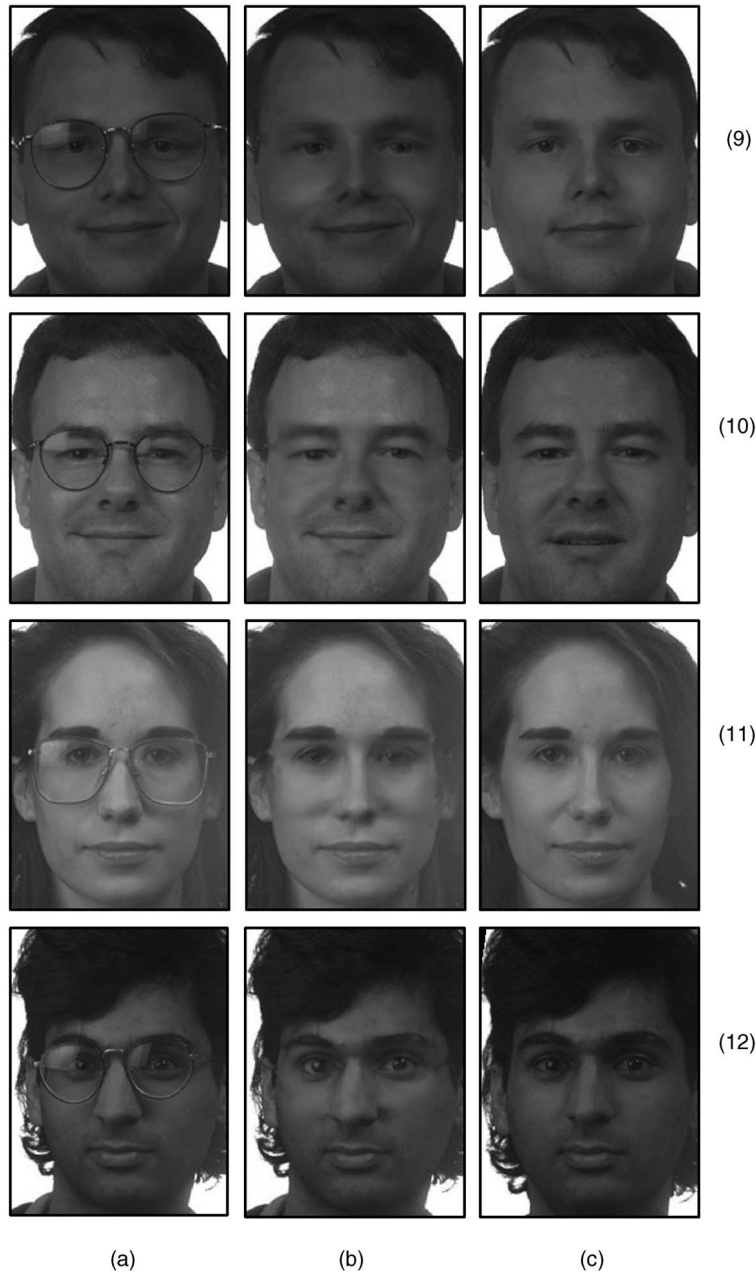
Fig. 15. The results of eyeglasses removal from samples (9) to (12). (a) Input images with glasses. (b) Synthesized results. (c) Images of (a) taken without glasses.

eyeglasses marked in red as illustrated in Fig. 16b. This procedure is tedious and sometimes it is hard to decide which region to be marked. The inpainting results (Fig. 16c) are reasonable for the upper example but unacceptable for the lower example where the lenses are semitransparent. The lenses can be either removed with the eyes disappeared or remained to occlude the eyes. The removal results generated in our system (Fig. 16d) are better. Our system also removes the eyeglasses quickly without any user intervention.

Two failure cases are shown in Fig. 17. The significant highlights from these two images were not present in the training data nor modeled in our system. Our system was unable to correctly infer the underlying information in the

eye region occluded by the highlights. It is difficult for human being to imagine such missing information as well.

Finally, we have done a leave-one-out experiment to validate our system by comparing the synthesized image with glasses removed and the image taken without glasses. For lack of a good measure of visual similarity, we choose two simple distances, namely, RMS errors on gray-level images and on Laplacian images. The distributions of these two errors tested on 247 samples are shown in Fig. 18. We may observe that the RMS errors on Laplacian images are fairly small, which implies that the synthesized results match the ground truth (the images taken without glasses) well in terms of local image features.
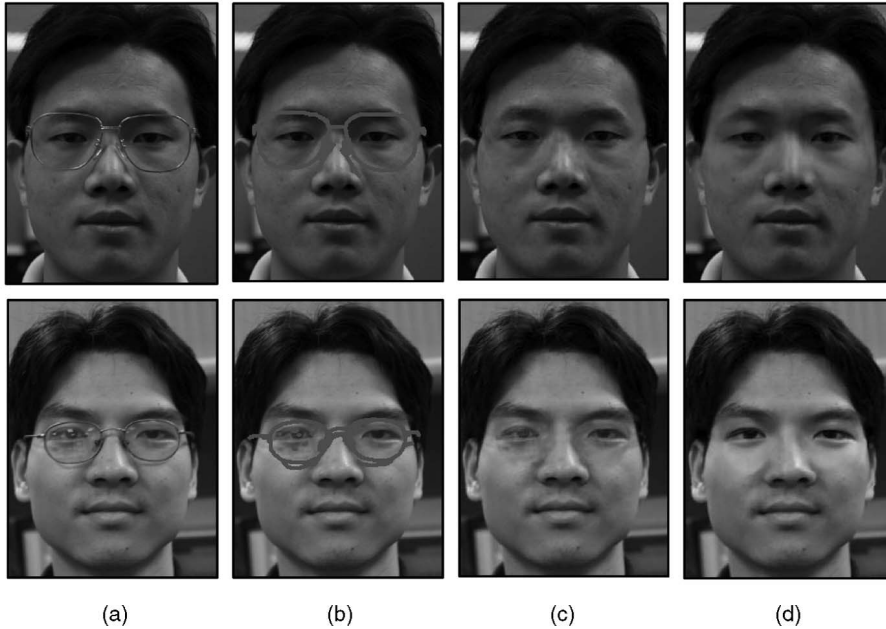
(a)   (b)   (c)   (d)

Fig. 16. The comparison between our method and image inpainting [2]. Note that inpainting cannot remove the semitransparent effect inside the eyeglasses frame. (a) Eyeglass inage, (b) region marked by user, (c) inpainting result, (d) result of our system.
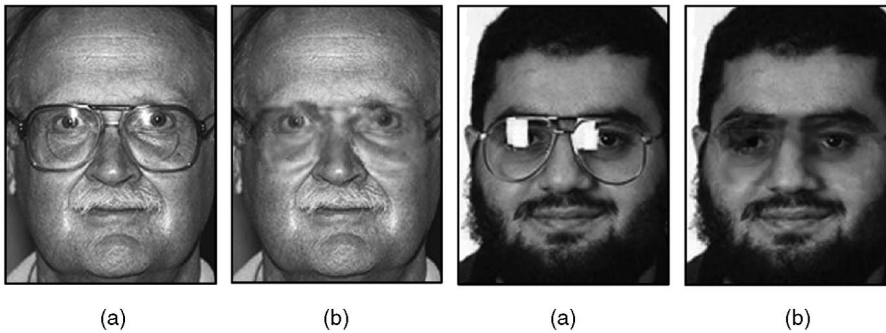

(a)   (b)   (a)   (b)

Fig. 17. Two failure cases. Although the eyeglasses frame are successfully removed, the pixels inside the frame are not synthesized well. Note that the significant highlights in these two images were not present in the training data. (a) Eyeglass inage, (b) removal result, (c) eyeglass image, (d) removal result.

5   SUMMARY

In this paper, we have presented a system to automatically remove eyeglasses from face images. Our system consists of three modules, eye area detection, eyeglasses localization and removal, by integrating statistical learning techniques such as boosting, MCMC, and subspace analysis. Extensive experiments have been carried out to demonstrate the effectiveness of the proposed system.

In our system, we have adopted a *find-and-replace* approach for intelligent image editing at the object level instead of the pixel level. Such an approach makes sense for two reasons. First of all, both object detection and localization can be done very well based on existing statistical learning and computational methods, as demonstrated in our paper by the eye area detection and eyeglasses localization systems. Second, the prior knowledge plays an essential role in object-level image editing. When important information is missing, it is difficult for the user fill in the object pixel by pixel, even with the help of tools like image inpainting. Based on statistical learning techniques, however, we can model the prior knowledge

from training samples and use Bayesian inference to find the missing information, as we have demonstrated in reconstructing eyeglasses-free region from eyeglasses region. Finally, this *find-and-replace* approach is not limited in automatic eyeglasses removal, but can be widely applied to editing a variety of visual objects.

APPENDIX

BOOSTING CLASSIFIER

The mechanism of boosting is used to train a binary classifier based on the labeled data. Suppose that we are given a set of labeled samples $\{x_i, y_i\}_{i=1}^{N}$, where $x_i \in \mathbb{R}^d$ and $y_i \in \{-1, 1\}$ and are asked to give a decision y to any $x \in \mathbb{R}^d$. It is convenient for us to get some 1D statistics from the data, using a mapping function $\phi^T x \colon \mathbb{R}^d \to \mathbb{R}^1$, where $\phi \in \mathbb{R}^d$ is a linear feature. Once we have a set of linear features $\{\phi_i\}_{i=1}^{k}$, we may obtain the histograms of positive and negative samples $h_i^+(\phi_i^T x)$ and $h_i^-(\phi_i^T x)$ with certain weights along each feature ϕ_i. At a specific point $z = \phi_i^T x$, if $h_i^+(z) > h_i^-(z)$, it will be more likely from this evidence that
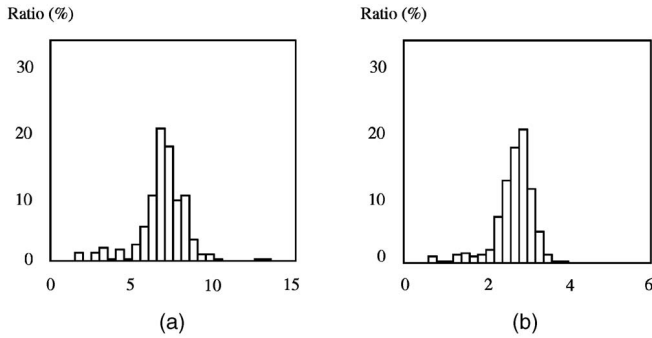
Fig. 18. Quantitative study of the eyeglasses removal system on 247 samples: histogram of error distribution. (a) RMS error on images. (b) RMS error on Laplacian images.

x is a positive sample. Therefore the classification function is defined as

$$F(x) = \text{sign}\left[\sum_{i=1}^{k} \alpha_i \log \frac{h_i^+(\phi_i^T x)}{h_i^-(\phi_i^T x)}\right] \qquad (26)$$

with parameters $\{\alpha_i\}$ to balance the evidence from each feature. $\text{sign}(\cdot) \in \{-1, 1\}$ is an indicator function.

There are two terms to learn in the classification function (26), the feature set $\{\phi_i\}$, and combining coefficients $\{\alpha_i\}$. Simple Harr wavelet features have been used for computational efficiency in face detection [27]. Recently, a data-driven approach [14] is proposed to compose a minimum number of features. The coefficients $\{\alpha_i^*\}$ can be either empirically set for incrementally optimal like AdaBoost [27] or optimized as a whole by a greedy algorithm [14].

The boosting scheme is used to gradually find the set of most discriminating features. From the previous learned classifier, the weight of misclassified samples is increased while that of recognized samples reduced and then a best feature is learned to discriminate them. Assume that at step $(k-1)$ the weights of the positive and negative samples are $W_{k-1}(x_i^+)$ and $W_{k-1}(x_i^-)$, respectively. Then, at step k, we may reweight the samples by

$$\begin{aligned} W_k(x_i^+) &= \frac{1}{Z^+} W_{k-1}(x_i^+) \exp\{-\beta_k y_i^+ F_{k-1}(x_i^+)\} \\ W_k(x_i^-) &= \frac{1}{Z^-} W_{k-1}(x_i^-) \exp\{-\beta_k y_i^- F_{k-1}(x_i^-)\}, \end{aligned} \qquad (27)$$

where Z^+ and Z^- are normalization factors for $W_k(x_i^+)$ and $W_k(x_i^-)$, respectively, and sequence β_k controls how fast to adapt the weight computed from the training error [14].

ACKNOWLEDGMENTS

REFERENCES

[1] S. Baker and T. Kanade, "Hallucinating Faces," *Proc. Fourth Int'l Conf. Autoamtic Face and Gesture Recognition,* Mar. 2000.

[2] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballsester, "Image Inpainting," *Proc. SIGGRAPH 2000 Conf.,* pp. 417-424, 2000.

[3] F.L. Bookstein, "Principle Warps: Thin-Plate Splines and the Decomposition of Deformations," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 11, no. 6, pp. 567-585, June 1989.

[4] H. Chen, Y.Q. Xu, H.Y. Shum, S.C. Zhu, and N.N. Zheng, "Example-Based Facial Sketch Generation with Non-Parametric Sampling," *Proc. Eighth IEEE Int'l Conf. Computer Vision,* July 2001.

[5] T. Cootes and C. Taylor, "Statistical Models of Appearance for Computer Vision," technical report, Univ. of Manchester, 2000.

[6] E. Mortensen and W. Barrett, "Intelligent Scissor for Image Composition," *Proc. SIGGRAPH 1995 Conf.,* pp. 191-198, 1995.

[7] D.A. Forsyth, J. Haddon, and S. Ioffe, "The Joy of Sampling," *Int'l J. Computer Vision,* vol. 41, nos. 1/2, pp. 109-134, 2001.

[8] W.T. Freeman and E.C. Pasztor, "Learning Low-Level Vision," *Proc. Seventh IEEE Int'l Conf. Computer Vision,* pp. 1182-1189, 1999.

[9] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distribution and the Bayesian Restoration of Images," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 6, no. 6, pp. 721-741, June 1984.

[10] A. Hertzmann, C.E. Jacobs, N. Oliver, B. Curless, and D.H. Salesin, "Image Analogies," *Proc. SIGGRAPH '01 Conf.,* Aug. 2001.

[11] X. Jiang, M. Binkert, B. Achermann, and H. Bunke, "Towards Detection of Glasses in Facial Images," *Proc. Int'l Conf. Pattern Recognition,* pp. 1071-1973, 1998.

[12] Z. Jing and R. Mariani, "Glasses Detection and Extraction by Deformable Contour," *Proc. Int'l Conf. Pattern Recognition,* 2000.

[13] T.K. Leung, M.C. Burl, and P. Perona, "Finding Faces in Cluttered Scenes Using Random Labeled Graph Matching," *Proc. Fifth IEEE Int'l Conf. Computer Vision,* pp. 637-644, 1995.

[14] C. Liu and H.Y. Shum, "Kullback-Leibler Boosting and Its Applications in Face Detection," *Proc. Computer Vision and Pattern Recognition Conf.,* 2003.

[15] C. Liu, H.Y. Shum, and C.S. Zhang, "A Two-Step Approach to Hallucinating Faces: Global Parametric Model and Local Non-parametric Model," *Proc. IEEE Conf. Computer Vision and Pattern Recogntion,* Dec. 2001.

[16] C. Liu, S.C. Zhu, and H.Y. Shum, "Learning Inhomogeneous Gibbs Model for Faces by Minimax Entropy," *Proc. Eighth IEEE Int'l Conf. Computer Vision,* July 2001.

[17] Z.C. Liu, Y. Shan, and Z. Zhang, "Expressive Expression Mapping with Ratio Images," *Proc. SIGGRAPH '01 Conf.,* Aug. 2001.

[18] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, "Equations of State Calculations by Fast Computing Machines," *J. Chemical Physics,* vol. 21, pp. 1087-1091, 1953.

[19] E. Osuna, R. Freund, and F. Girosi, "Training Support Vector Machines: An Application To Face Detection," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 130-136, 1997.

[20] P. Pérez, A. Blake, and M. Gangnet, "JetStream: Probabilistic Contour Extraction with Particles," *Proc. Int'l Conf. Computer Vision (ICCV),* vol. 2, pp. 524-531, 2001.

[21] P. Philips, H. Moon, P. Rauss, and S. Rizvi, "The FERET Evaluation Methodology for Face-Recognition Algorithms," *Proc. Computer Vision and Pattern Recognition Conf.,* pp. 137-143, 1997.

[22] H.A. Rowley, S. Baluja, and T. Kanade, "Neural Network-Based Face Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 20, no. 1, Jan. 1998.

[23] Y. Saito, Y. Kenmochi, and K. Kotani, "Esitmation of Eyeglassless Facial Images Using Principal Component Analysis," *Proc. IEEE Int'l Conf. Image Processing,* pp. 197-201, 1999.

[24] H. Schneiderman and T. Kanade, "A Statistical Method for 3D Object Detection Applied to Faces and Cars," *Proc. Seventh IEEE Int'l Conf. Computer Vision,* May 2000.

[25] K.K. Sung and T. Poggio, "Example-Based Learning for View-Based Human Face Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 20, no. 1, pp. 39-51, Jan. 1998.

[26] Z.W. Tu and S.C. Zhu, "Image Segmentation by Data Driven Markov Chain Monte Carlo," *Proc. Eighth Int'l Conf. Computer Vision,* July 2001.

[27] P. Viola and M. Jones, "Robust Real-Time Face Detection," *Proc. Eighth IEEE Int'l Conf. Computer Vision,* July 2001.

[28] Z. Wen, T. Huang, and Z. Liu, "Face Relighting with Radiance Environment Maps," *Proc. Computer Vision and Pattern Recognition Conf.,* 2003.

[29] C.Y. Wu, C. Liu, and J. Zhou, "Eyeglass Existence Verification by Support Vector Machine," *Proc. Second IEEE Pacific-Rim Conf. Multimedia,* Oct. 2001.

[30] A.L. Yuille, D.S. Cohen, and P. Hallinan, "Feature Extraction from Faces Using Deformable Templates," *Int'l J. Computer Vision,* vol. 8, no. 2, pp. 99-112, 1992.

[31] S.C. Zhu, Y.N. Wu, and D. Mumford, "Filter, Random Field and Maximum Entropy," *Neural Computation,* vol. 9, no. 7, Nov. 1997.

Chenyu Wu received the BS degree in automation and the ME degree in pattern recognition from the Department of Automation, Tsinghua University in 2000 and 2003, respectively. She is currently a doctoral student in the Robotics Institute, Carnegie Mellon University. Her research interests include medical application especially image-guided surgical tools, robotics, computer vision, computer graphics, and pattern recognition.

Ce Liu received the BS degree in automation and the ME degree in pattern recognition from the Department of Automation, Tsinghua University in 1999 and 2002, respectively. From 2000 to 2002, he was a visiting student at Microsoft Research Asia. After that, he worked at Microsoft Research Asia as an assistant researcher. He is currently a doctoral student at Massachusetts Institute of Technology in the Computer Science and Artificial Intelligence Lab (CSAIL). His research interests include computer vision, machine learning, and pattern recognition.

Heung-Yeung Shum received the PhD degree in robotics from the School of Computer Science, Carnegie Mellon University in 1996. He worked as a researcher for three years in the vision technology group at Microsoft Research Redmond. In 1999, he moved to Microsoft Research Asia where he is currently a senior researcher and the assistant managing director. His research interests include computer vision, computer graphics, human computer interaction, pattern recognition, statistical learning, and robotics. He is on the editorial boards of *IEEE Transactions on Circuit System Video Technology (CSVT)*, *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, and *Graphical Models*. He is the general cochair of 10th International Conference on Computer Vision (ICCV 2005) in Beijing. He is a member of the IEEE.

Ying-Qing Xu received the PhD degree from the Chinese Academy of Sciences in 1997. He is a researcher at Microsoft Research Asia. His research interests include statistical learning, computer vision, and computer graphics. He has authored and coauthored 40 papers in computer vision and computer graphics.

Zhengyou Zhang received the BS degree in electronic engineering from the University of Zhejiang, China, in 1985, the MS degree in computer science from the University of Nancy, France, in 1987, the PhD degree in computer science from the University of Paris XI, France, in 1990, and the Doctor of Science diploma from the University of Paris XI, in 1994. He is a senior researcher at Microsoft Research, Redmond, Washington. He was with INRIA (French National Institute for Research in Computer Science and Control) for 11 years until he joined Microsoft Research in March 1998. In 1996-1997, he was on sabbatical as an invited researcher at the Advanced Telecommunications Research Institute International (ATR), Kyoto, Japan. His current research interests include 3D computer vision, dynamic scene analysis, vision and graphics, facial image analysis, multisensory technology, and human-computer interaction. He is a senior member of the IEEE and is an associate editor of the *IEEE Transactions on Pattern Analysis and Machine Intelligence* and the *International Journal of Pattern Recognition and Artificial Intelligence*. He has published more than 100 papers in refereed international journals and conferences, and has coauthored the following books: *3D Dynamic Scene Analysis: A Stereo Based Approach* (Springer, Berlin, Heidelberg, 1992), *Epipolar Geometry in Stereo, Motion and Object Recognition* (Kluwer Academic Publishers, 1996), and *Computer Vision* (textbook in Chinese, Chinese Academy of Sciences, 1998). He is an area chair and a demo chair of the International Conference on Computer Vision (ICCV2003), October 2003, Nice, France, and a program chair of the Asian Conference on Computer Vision (ACCV2004), January 2004, Jeju Island, Korea.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.