Automatic Eyeglasses Removal from Face Images

Chenyu Wu[1,2] Ce Liu[1,2] Heung-Yeung Shum[2] Ying-Qing Xu[2] Zhengyou Zhang[3]

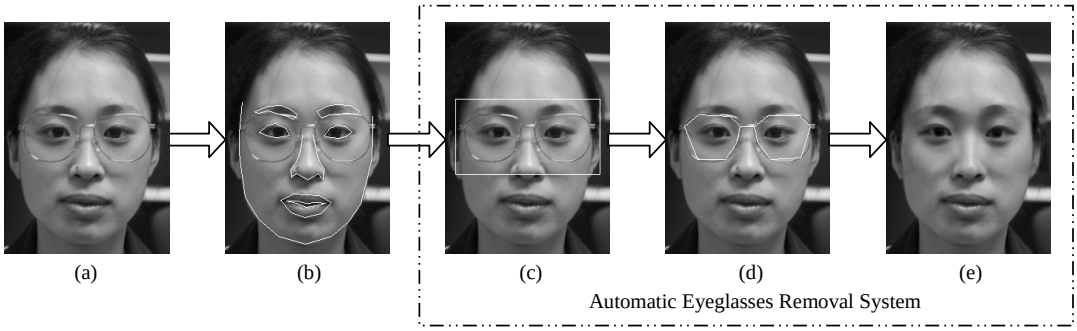

Figure 1. Automatic eyeglasses removal system overview. (a) An input face image. (b) Face localization by ASM. (c) Glasses recognition by a classifier. (d) Glasses localization by MCMC. (e) Glasses removal based on a set of training examples.

Abstract

In this paper, we present a system that can automatically remove eyeglasses from an input face image. Our system consists of three modules: eyeglasses recognition, localization and removal. Given a face image, we first use an eyeglasses classifier, to determine if a pair of eyeglasses is present. Then we apply a Morkov chain Monte Carlo method to accurately locate the glasses by searching for the global optimum of the posteriori. Finally, a novel example-based approach is developed to synthesize an image with eyeglasses removed from the detected and localized face image. Rather than applying conventional image processing techniques to the input image, we propose a statistical analysis and synthesis approach employing a database containing pairs of face images, one with eyeglasses while the other without. Experiments demonstrate that our approach produces good quality of face images with eyeglasses removed.

1. Introduction

In the last decade, face analysis and synthesis has become one of the most active research topics in computer vision and pattern recognition, where statistical learning-based methods have been successfully used. In face recognition and detection, eigenface [14], neural network [10, 12] and support vector machine (SVM) [9] are some of the typical approaches. Deformable models such as [7] and Active Shape Model (ASM) [2] have been demonstrated to be effective to localize faces. Recently, face hallucination [8] and facial sketch generation [1] can synthesize high-resolution face images and good-quality face sketches, respectively. Most of above methods need to extract facial features such as eye, eyebrows and nose with a non-glasses assumption on human faces, despite the fact that many people wear glasses. A person wearing glasses is likely to be missed by a face detector training on faces without glasses, or be mistaken to others by an identification system. Therefore, it is of great importance to analyze glasses for face detection, recognition and synthesis.

There are good reasons why people avoided dealing with eyeglasses. First, the appearance of glasses frames is so diverse due to various material properties such as metal and plastic. Second, the reflectance property of glasses differs significantly from that of human skin. Sometimes the reflection on the glasses is the brightest area on face. Third, faces are always well separated with the background whereas the glasses are stuck to the face and mixed by eyebrows. Thus tracking or stereo techniques may not be of help for eyeglasses detection.

There has been some recent work on glasses recognition, localization and removal. Jiang et al. [4] studied detecting glasses on facial images by a glasses classifier. Wu et al. [15] devised a sophisticated glasses classifier based on SVM. The recognition rates reported in their articles are close to 90%. Jing and Mariani [5] employed a deformable contour method to detect glasses under a Bayesian framework. In their work, 50 points are used to define the shape of glasses, and the position of glasses is found by maxi-

[1] State Key Lab of Intelligent Technique and System, Department of Automation, Tsinghua University, 100084, Beijing, China
Email: wuchenyu00@mails.tsinghua.edu.cn
[2] Visual Computing Group, Microsoft Research China, Sigma Building, Zhi-Chun Road, 100080, Beijing China.
Email: lce@misrchina.research.microsoft.com
[3] Vision Technology Group, Microsoft Research, One Microsoft Way, Redmond WA 98052-6399, USA.
Email: zhang@microsoft.com

mizing the posteriori. Saito et al. did the glasses removal work by using principal component analysis (PCA) [11]. The eigen-space of eyeglassless patterns is learnt by PCA to retain their principal variance. Projecting a glasses pattern into this space will get the corresponding non-glasses one.

In this paper we present a system that can automatically recognize, locate and remove eyeglasses from an input face image. Our system consists of three modules: glasses recognition, localization and removal. The first two modules are essential to make the glasses removal automatic. For recognition, we perform classification according to the reconstruction errors in the eigen-spaces which are learnt through principal component analysis (PCA). For localization, we use an Active Shape Model and learn the prior by PCA and the likelihood in a nonparametric way. We apply a Markov chain Monte Carlo (MCMC) technique to search for a global optimum of the glasses localization. For removal, in order to capture the global correspondence of glasses and non-glasses patterns, we again model their joint distribution in eigen-space by PCA. A solution is searched for in the eigen-space such that its glasses part is the closest to the input. The non-glasses part of the solution is naturally the result, and is intelligently pasted on the glasses region. This enables us to capture global properties of glasses such as symmetry, contours and illumination.

This article is organized as follows. Section 2 describes feature extraction and classifier designing for glasses recognition. Section 3 shows how to model the posteriori of the glasses frame given an face image, and how to find the global optimum of the posteriori by MCMC. The details of the glasses removal module, together with a number of removal results, are provided in Section 4. We conclude this paper with a discussion in Section 5.

2. Glasses Recognition

In our system we choose orientation pattern as the feature and devise a reconstruction error based classifier to discriminate glasses and non-glasses patterns.

2.1 Feature extraction

Although there exist many cues of eyeglasses from human knowledge, most of them are not robust discriminative features. The symmetry of glasses can not be selected as features since the face area around eyes is also symmetric. It is the same to the loop-like pattern which would be confused by the eyebrows and eyes, in particular for face images with heavy shadows. However, we observe that there exists more prominent orientation in glasses region than other face regions. It implies that in the glasses region the *anisotropy*, which indicates how concentrated local orientations are, is strong. Therefore we select the orientation and anisotropic measure, computed as [6, 1],

$$\theta(u,v) = \frac{1}{2}\arctan\frac{\iint_{\Omega(u,v)} 2I'_x I'_y dxdy}{\iint_{\Omega(u,v)}(I'^2_x - I'^2_y)dxdy} + \frac{\pi}{2} \qquad (1)$$

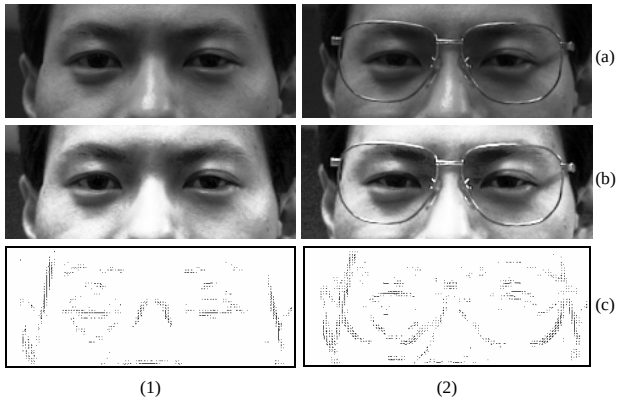

Figure 2. Feature extraction. (a) Original intensity image. (b) Histogram equalized image. (c) Feature image with orientations and anisotropic measure. (1) to (2) are non-glasses and glasses, respectively.

$$\chi(u,v) = \frac{(\iint_{\Omega(u,v)}(I'^2_x - I'^2_y)dxdy)^2 + (\iint_{\Omega(u,v)} 2I'_x I'_y dxdy)^2}{(\iint_{\Omega(u,v)}(I'^2_x + I'^2_y)dxdy)^2} \qquad (2)$$

where I'_x and I'_y are x and y components of the gradient vectors $G(x,y) = (I'_x, I'_y)$ respectively, and $\Omega(u,v)$ is the neighborhood of (u,v) with size 3×3 chosen in our implementation. $\theta(u,v) \in [0,\pi)$ and $\chi(u,v) \in [0,1]$ are orientation and anisotropic measures at (u,v) respectively. $\chi(u,v) \in [0,1]$ counts the concentration of local gradients in the neighborhood of (u,v). If local gradients uniformly distribute between 0 and π, then $\chi(u,v)$ is small. On the contrary, if all the local gradients point the same way, then $\chi(u,v)$ is strong and the pixel (u,v) is most likely to be on a meaningful edge.

Since $\theta(u,v) \in [0,\pi)$, the difference between orientations $\pi - \varepsilon$ and $0 + \varepsilon$ (ε is a very small value) will be the most distinct, but in fact they are all horizontal edges and very close in visual appearance. To solve this problem, we transfer the polar coordinates into a Cartesian one by doubling the orientation angle

$$\begin{cases} x(u,v) = \chi(u,v)\cos(2\theta(u,v)) \\ y(u,v) = \chi(u,v)\sin(2\theta(u,v)) \end{cases} \qquad (3)$$

We list some results in Figure 2 for histogram equalization and feature extraction, where the feature image is displayed as an orientation field in vector form. We notice that the glasses and non-glasses image appear distinct in the feature image as we expected .

2.2 Classifier Designing

The ideal feature space in pattern recognition should make the two classes far away on their centers in comparison with their own variations, such that they can be easily separated by a linear hyperplane. But the glasses and non-glasses patterns are so close on the centers no matter what feature spaces are chosen due to the position uncertainty of

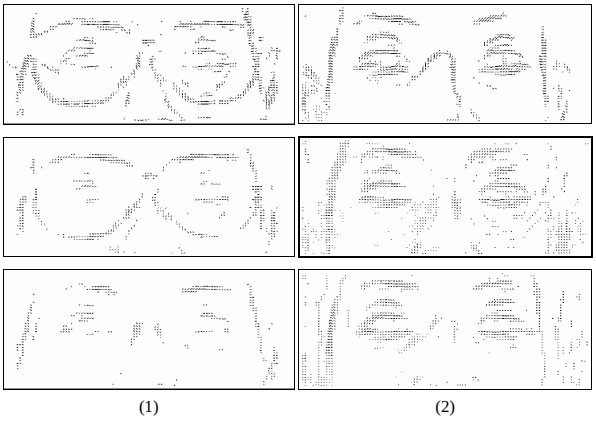

Figure 3. Results of reconstruction by glasses and non-glasses eigen-spaces. From (a) to (c) are input feature image, reconstructed by glasses and reconstructed by non-glasses respectively. (1) and (2) are results of glasses and non-glasses images respectively.

glasses. After applying PCA to these two classes, we get two groups of means μ_F and μ_G, and eigenvectors $\Phi_F = [\phi_F^{(1)}, \phi_F^{(2)}, \cdots, \phi_F^{(n_F)}]$ and $\Phi_G = [\phi_G^{(1)}, \phi_G^{(2)}, \cdots, \phi_G^{(n_G)}]$ for non-glasses and glasses pattern respectively. We find that μ_F and μ_G are very close in feature space while the inner products between the two groups of eigenvectors are always very close to zero, *i.e.* $<\phi_F^{(i)}, \phi_G^{(j)}> \approx 0$ in most cases. This indicates that glasses and non-glasses patterns vary in nearly orthogonal directions.

To discriminate these two patterns in feature space, we design a classifier based on the *reconstruction error* by eigen-space. We project the input feature vector I into the two eigen-spaces, reconstruct it by their eigenvectors and compute the error respectively. Mathematically,

$$\begin{cases} I_F = \Phi_F \Phi_F^T (I - \mu_F) + \mu_F \\ I_G = \Phi_G \Phi_G^T (I - \mu_G) + \mu_G \end{cases} \quad (4)$$

where I_G and I_F are reconstructed feature images by glasses and non-glasses eigen-space respectively. Thus the verification function $V(I)$ is defined as

$$V(I) = \text{sgn}(|I - I_F| - \lambda_G |I - I_G|), \quad (5)$$

where $\text{sgn}(\cdot)$ is a signal function and λ_G balances the priors of each classes. If $V(I) = 1$, we classify the input image to glasses. Otherwise, it will be classified to non-glasses.

A reconstruction example is illustrated in Figure 3. It is shown that the input feature image can be best reconstructed by its own eigen-space, which demonstrates the criterion of our classifier. Then we do a leave-one-out experiment to test our method. A recognition rate of 94.2% is achieved in our glasses recognition subsystem with 220 samples. Almost all missed glasses by our algorithm are the non-frame ones which have weak edge features, and the false alarms are always with strong shadow features.

3. Glasses Localization

To locate the precise position of glasses, we use a deformable contour model [7] or ASM [2] to describe the geometric information and position of the glasses. We define 15 key points on the glasses frame, denoted by $W = \{w_i = (x_i, y_i), i = 1, \cdots, n\}$ where $n = 15$. In the following computation, W is regarded as a long vector with dimension $2n$. Based on the Bayesian law, to locate the position is to find an optimal W^* in glasses region I_G by maximizing the posteriori, or the product of the prior and likelihood

$$W^* = \arg\max_W p(W|I_G) = \arg\max_W p(I_G|W)p(W). \quad (6)$$

Based on (6), we should first learn the prior $p(W)$ and estimate likelihood $p(I_G|W)$ respectively, and then design an optimization mechanism to search for the optimal solution.

3.1 Prior Learning

Physically the prior distribution of W comprises two independent parts: the *inner* parameters, or the position invariant shape W', and the *outer* parameters, or the position relevant variables such as the orientation, scale and central points. Mathematically, each key point can be described by

$$w_i = sR_\theta w_i' + C_{xy} \quad (7)$$

where s is the scale parameter, R_θ is the rotation matrix controlled by θ, and C_{xy} is the centroid of the key points. $W' = \{w_i' = (x_i', y_i'), i = 1, \cdots, n\}$ is the centralized and scale normalized key points. It is obvious that $\sum_{i=1}^n {w'}_i = \mathbf{0}$ and $\sum_{i=1}^n w_i = C_{xy}$. Since the position invariant shape w' in the *inner* parameter space and the position correlated s, θ and C_{xy} in the *outer* parameter space are statistically independent, we can decouple the prior distribution into

$$p(W) = p(W')p(s)p(\theta)p(C_{xy}). \quad (8)$$

We shall model the *inner* and *outer* priors in different ways.

Just as in the work of ASM and many other points distribution models, the distribution of the *inner* parameters W' is simply assumed Gaussian and learnt by PCA. The distribution has the form Suppose

$$p(W') = \frac{1}{Z} \exp\{-(W' - \mu_{W'})^T B \Lambda^{-1} B^T (W' - \mu_{W'})\} \quad (9)$$

where $\mu_{W'}$ is the mean shape, $B = [\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_m]$ and $\Lambda = diag(\sigma_1^2, \sigma_2^2, \cdots, \sigma_m^2)$ are eigenvectors and eigenvalues of W' analyzed by a set of training examples $\{W_k', k = 1, \cdots, M\}$ respectively, and Z is the normalization constant. The principal components $\{\mathbf{b}_i\}$ denote the main variations of the shape W'. Let $u = B^T(W' - \mu_{W'})$ be the variable in eigen-space, then varying u along the principal components is always much more meaningful than varying a single point.

The *outer* parameters, the scale, orientation and central point are all in low dimension(s). We simply use histograms to represent their distributions.

Figure 4. The results of glasses localization by MCMC.

3.2 Likelihood Learning

In analysis rather than synthesis, the likelihood $p(I_G|W)$ is used to measure if the local features $F_G^{(i)}$ on point w_i is similar to those of that key point in terms of the appearance. Usually it can be simplified to

$$p(I_G|W) = p(F_G|W) = \prod_{i=1}^{n} p(F_G^{(i)}|w_i), \tag{10}$$

where $F_G = \{F_j * I_G | j = 1, \cdots, l\}$ is the feature image of the intensity image, and $\{F_j\}$ are the local filters. What we should learn is the local feature distribution $p(F_G^{(i)}|w_i)$ for each key point.

Since the key points of glasses are defined on the frame, they are distinct in edge and orientation, as we already described in glasses recognition. We choose the responses of local edge detectors as the features, including a Laplacian operator, the first and second order Sobel operators with 4 directions. They are all band-pass to capture the local space-frequency statistics. For each key point w_i, we can get the training 9-d features $\{F_G^{(i)}(k)\}_{k=1}^{M}$. The likelihood of each key point is obtained by Parzen window method,

$$p(F_G^{(i)}|w_i) = \frac{1}{M} \sum_{k=1}^{M} G(F_G^{(i)}; F_G^{(i)}(k), \sigma_i), \tag{11}$$

where $G(\cdot; F_G^{(i)}(k), \sigma_i)$ is the Gaussian kernel function centered at the example $F_G^{(i)}(k)$ with variance σ_i.

3.3 MAP Solution by Markov Chain Monte Carlo

After the prior and likelihood models are built, we should find the optimal W^* by maximizing the posteriori under the MAP criterion. However the objective function, *i.e.* the posteriori, is very complex with many local maximums. Thus traditional deterministic gradient ascent algorithm will be stuck at local optima. Recently, Markov chain Monte Carlo (MCMC) has been used in solving Bayesian inference problems such as image segmentation [13]. We choose Gibbs sampling in optimization due to the low rejecting ratio. Since the key points W has been decoupled to the *inner* and *outer* parameters, the solution space is simplified to $X = \{u, s, \theta, C_{xy}\}$. Suppose $X = (x_i)_{i=1}^{k}$, the Markov chain dynamics in Gibbs sampling is given by

$$\begin{aligned} x_1^{(t+1)} &\sim p(x_1|x_2^{(t)}, x_3^{(t)}, \cdots, x_k^{(t)}) \\ x_2^{(t+1)} &\sim p(x_2|x_1^{(t+1)}, x_3^{(t)}, \cdots, x_k^{(t)}) \\ &\vdots \\ x_k^{(t+1)} &\sim p(x_k|x_1^{(t+1)}, x_2^{(t+1)}, \cdots, x_{k-1}^{(t)}) \end{aligned} \tag{12}$$

By sequentially flipping each dimension, the Gibbs sampler walks through the solution space with the target posterior probability density given by (6), (9) and (11). The optimal solution X^* is obtained in a sequence of independent samples after several sweeps, 20 for instance. Some localization results are shown in Figure 4.

The intuitive explanation of Gibbs sampling is that we can keep other dimensions unchanged and flip current dimension to a better (not the best) position with a certain probability. The components of X are independent, at least the variations in shape and position are uncorrelated. Therefore searching in this space is much more efficient than merely moving one key point once in previous localization methods. Our method is also insensitive to the initialization, guaranteed by MCMC global convergence. Unlike other MCMC algorithms, it takes no more than 0.5 second for the Gibbs sampler to find a globally optimal solution.

4. Glasses Removal

Once the glasses localization module accurately finds the position of the glasses, we shall do glasses removal to

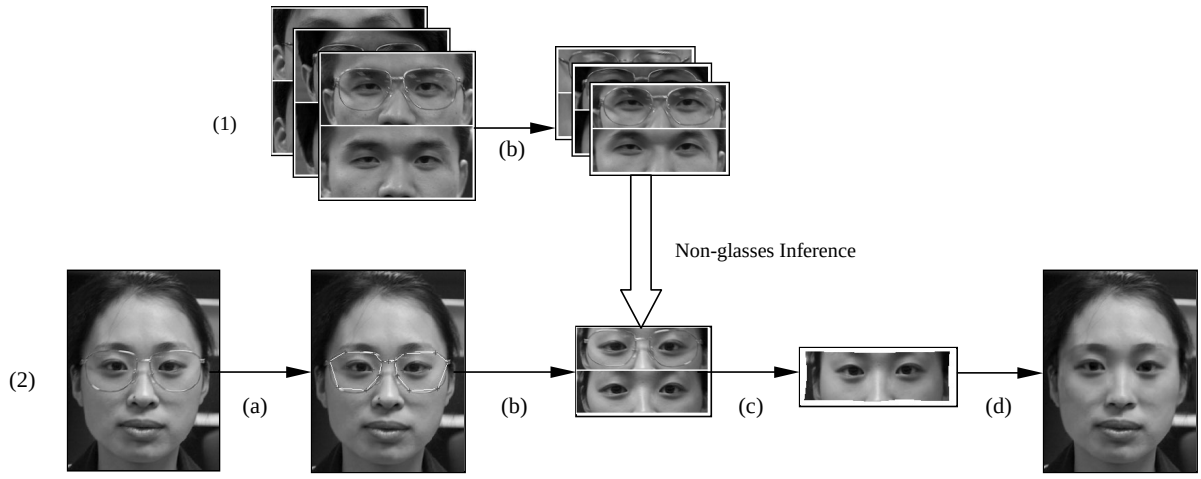

Figure 5. The flowchart of glasses removal module. (1) Learning procedure. (2) Automatic removal procedure. (a) Preceding work until the localization of the glasses. (b) Warping the glasses region to a normal template. (c) Inverse warping to the original image domain. (d) Paste the inferred non-glasses image onto the input with boundary blended.

replace the glasses pattern with the best non-glasses pattern fit. Since there is too much uncertainty to infer the lost information under the glasses, we adopt learning-based method for non-glasses pattern inference, *i.e.* to "guess" the underlying non-glasses pattern from a set of training glasses and non-glasses pairs. This kind of learning-based methods has been successfully applied to facial sketch synthesis [1] and super-resolution [8, 3]. Before learning, it is significant to calibrate each training sample with respect to a normal template, on which the learning machine learns the precise statistical relationship between glasses and non-glasses patterns, avoiding extra position and scale uncertainties involved. Given a new glasses sample, we should first warp it to the normal template, infer the non-glasses pattern, inversely warp the inferred one to the original face image, and finally paste it on the glasses area, as illustrated in Figure 5. We adopt the warping method introduced in [2].

Let us denote the warped glasses and non-glasses image by I'_G and I'_F, respectively. Given the glasses image I'_G, our task is to infer the non-glasses pattern I'_F based on a set of training pairs $\{I'_G(i), I'_F(i)_{i=1}^M\}$. After warping, the difference between these two patterns is only in the appearance of glasses (the training pair of each individual is taken under nearly the same lighting condition). We again choose PCA to learn the correspondence between the pairs. Let $Y^T = [I'^T_G \ I'^T_F]$, and the training examples become $\{Y(i)\}_{i=1}^M$. Through singular value decomposition (SVD), we can get principal components matrix $\Psi = [\psi_1 \ \psi_2 \ \cdots \ \psi_h]$ with ψ_j the jth eigenvector, eigenvalues $\{\sigma_i^2\}_{i=1}^h$ and mean μ_Y. Let V be the hidden variable lying in the eigen-space $V \in \mathbf{R}^h$. Then $V = \Psi^T(Y - \mu_Y)$ records the main variation of Y. On the contrary Y can be well approximated by V: $Y = \Psi V + \mu_Y + \varepsilon_Y$, where ε_Y is noise. Therefore the distribution of Y, *i.e.* the joint distribution of I'_G and I'_F can be replaced by V with a Gaussian form

$$p(V) = \frac{1}{Z'} \exp\{-V^T \Lambda^{-1} V\}, \tag{13}$$

where $\Lambda = \text{diag}[\sigma_1^2, \cdots, \sigma_h^2]$ and Z' is the normalization constant. Once given I'_G, we should infer the optimal V^* based on the Bayesian law

$$V^* = \arg\max_V p(V|I'_G) = \arg\max_V p(I'_G|V)p(V), \tag{14}$$

then the second part of Y^* corresponding to V^* is the desirable result. Since I'_G is the first half component of Y, we model the likelihood as a soft constraint

$$p(I'_G|V) = \frac{1}{Z''} \exp\{-\|\Psi_1 V + \mu_{Y_1} - I'_G\|^2/\lambda\}, \tag{15}$$

where $\Psi_1 = [\ \text{I}\ 0\ ]\Psi$ is the first half of Ψ, $\mu_{Y_1} = [\ \text{I}\ 0\ ]\mu_Y$ is the first half of μ_Y, λ scales the variance and Z'' is the normalization constant. This likelihood enforces the glasses part of the reconstructed Y^* close to the observed I'_G.

To maximize the posteriori in (14) is equivalent to

$$V^* = \arg\min_V \{\lambda V^T \Lambda^{-1} V + \|\Psi_1 V + \mu_{Y1} - I'_G\|^2\}, \tag{16}$$

with straightforward solution

$$V^* = (\Psi_1^T \Psi + \lambda \Lambda^{-1})^{-1} \Psi_1^T (I'_G - \mu_{Y1}). \tag{17}$$

Finally the optimal non-glasses pattern I'^*_F is calculated by $I'^*_F = \Psi_2 V^* + \mu_{Y2}$, where $\Psi_2 = [0\ \text{I}\ ]\Psi$ is the second half of Ψ and $\mu_{Y2} = [\ 0\ \text{I}\ ]\mu_Y$ is the second half of μ_Y. Then I'^*_F is inversely warped to the glasses region of I_G and pasted onto the face image with the abutting area blended around the boundary. Some glasses removal results are displayed in Figure 6. Note that different types of glasses have been used in our experiment. Even though the synthesized non-glasses images are a little bit blurred and different from the original non-glasses ones, the glasses have been successfully removed from the input image.

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

Figure 6. The results of glasses removal. (a) Input faces with glasses. (b) Removal results. (c) Original non-glasses faces of (a).

5. Discussion and Conclusion

In this paper, we have designed an automatic glasses removal system which includes glasses recognition, localization and removal modules. The orientation and anisotropic measures are extracted as features in recognition, and a reconstruction-error based classifier is designed due to the special distribution of glasses and non-glasses patterns. We apply MCMC or Gibbs sampling to accurately locate the glasses by searching for the global optimum of the posteriori. Finally a learning based approach is developed to synthesize a non-glasses image with the glasses removed from the detected and localized face image, by maximizing the joint distribution of the glasses and non-glasses patterns in eigen-space. A recognition rate of 94.2%, and the localization and removal results shown in Figures 4 and 6 demonstrate the effectiveness of our system. In addition, running through the three subsystems takes less than 5 seconds on a regular PC.

For future work, we shall incorporate more sophisticated methods in the glasses removal module to solve the blurring problem. We also plan to push our system to glasses removal in a video sequence. Finally, we can realize an automatic glasses wearing system by exchanging the training pairs in our current system.

References

[1] 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. of 8th ICCV*, 2001.

[2] T. Cootes and C. Taylor. Statistical Models of Appearance for Computer Vision. Technical report, University of Manchester, 2000.

[3] W. T. Freeman and E. C. Pasztor. Learning low-level vision. *Proc. of 7th ICCV*, pages 1182–1189, 1999.

[4] X. Jiang, M. Binkert, B. Achermann, and H. Bunke. Towards Detection of Glasses in Facial Images. *Proc. of ICPR*, pages 1071–1973, 1998.

[5] Z. Jing and R. Mariani. Glasses Detection and Extraction by Deformable Contour. *Proc. of ICPR*, 2000.

[6] M. Kass and A. Witkin. Analyzing orientated pattern. *Computer Vision, Graphics and Image Processing*, 37:362–397, 1987.

[7] T. K. Leung, M. Burl, and P. Perona. Finding Faces in Cluttered Scenes Using Random Labeled Graph Matching. *Proc. of 5th ICCV*, pages 637–644, 1995.

[8] C. Liu, H. Y. Shum, and C. Zhang. A Two-Step Approach to Hallucinating Faces: Global Parametric Model and Local Nonparametric Model. *Proc. of CVPR*, 2001.

[9] E. Osuna, R. Freund, and F. Girosi. Training Support Vector Machines: An Application To Face Detection. *Proc. of CVPR*, pages 130–136, 1997.

[10] H. Rowley, S.Baluja, and T.Kanade. Neural Network-Based Face Detection. *IEEE Transactions on PAMI*, 20(1), January 1998.

[11] Y. Saito, Y. Kenmochi, and K. Kotani. Esitmation of Eyeglassless Facial Images Using Principal Component Analysis. *Proc. of ICIP*, pages 197–201, 1999.

[12] K. K. Sung and T.Poggio. Example-based Learning for View-based Human Face Detection. *IEEE Transactions on PAMI*, 20(1):39–51, 1998.

[13] Z. W. Tu and S. C. Zhu. Image Segmentation by Data Driven Markov Chain Monte Carlo. *Proc. of 8th ICCV*, 2001.

[14] M. Turk and A. Pentaland. Eigenface for Recognition. *Journal of Cognitive Neurosciences*, pages 71–86, 1991.

[15] C. Y. Wu, C. Liu, and J. Zhou. Eyeglass Existence Verification by Support Vector Machine. *Proc. of 2nd Pacific-Rim Conference on Multimedia*, 2001.