MVA2000 IAPR Workshop on Machine Vision Applications, Nov. 28–30, 2000, The University of Tokyo, Japan

3—30

Real-time Facial Feature Detection for Person Identification System

Sung Uk Lee, Yu Shin Cho, Seok Cheol Kee, Sang Ryong Kim *
HCI Lab
Samsung Advanced Institute of Technology

Abstract

In this paper, we present an approach to real-time facial feature detection. Facial regions are segmented using Gabor filter responses with M-style grid matching. M-style grid matching method has been shown more effective than Gabor bunch graph matching method in many aspects such as frontal face detection against expression, in-plane/depth rotation, and various illumination environments. In addition, this approach can be implemented with low computational complexity. The center positions of both eyes are detected, from the segmented face region, by iterative binary thresholding with perfect contour tracing. Comparing with other pattern matching methods, it is shown that our scheme is faster and more effective eye detection method. Off-line simulation results using the test image set taken under office illumination (fluorescence) are over 99% successful segmentation rate of facial region (Face Detection Rate : FDR), and 99% effective eye center position detection rate of facial region (Eye position Detection Rate : EDR)

We have implemented the real-time system on Pentium-III 550MHz PC, and the system is capable of finding a pair of facial feature points on 240 by 320 images at 220ms per image. 97% of FDR, and 85% of EDR are real-time performance of the online system. The measured computational complexity is as low as about 32WMOPS.

1 Introduction

We have developed a person identification system, what we call **MagicGate** to authorize an entrance into secured area. Several HCI technologies, such as speech recognition, speaker verification, face detection, and face verification modules are integrated into the system. Anyone who wants to gain an entrance into secured area should speak 4-digit PID code, and the face image is segmented at the same time. When the PID recognized, speaker verification and face verification [1] modules begin the verification process. In this paper, we introduce the face (facial feature) detection method used in the application system.

Accuracy and speed are contrary to each other in finding faces on real office, not coordinated, background. Actually we've had some experiments with several methods. Using PCA [2] is the one of the mainly considered method, but it needs well-aligned faces to train model DB and has some critical defects in expression, rotation, size, and illumination variations. So, we've decided to use Gabor filter responses, well known for effectiveness in description of local feature, from fiducial points in face images[3]. It is also known that it has some invariant attributes against the defects of PCA and it is flexible to background change. Typical face detection method using Gabor kernels is *Elastic Bunch Graph Matching*[4]. In this method, facial features as fiducial points must be detected correctly to make appropriate graph. To detect facial features, it needs too much computational complexity. Because Gabor filter responses from all points on input image are compared to all Jets in the Bunch respectively. Therefore it is useful to verify or recognize face, but not for detection. Figure 1 shows examples of Bunch Graph Matching. There are some groups of dot that have same intensities, it means that many points of background can be considered as a facial features. So to speak, in many locations on background, each filter responses from a point considered a facial feature such as eyes, nose, and lip. All possible combinations of each detected facial point are selected and tested to consist of facial graph. Figure 1 shows the successfully detected graph and the wrong case. But the step of fine search should be followed despite the successful detection of the graph, in order to process face authentication. Fine search means that Gabor filter responses should be computed from most of all points on input image, and also it needs too much complexity.

Using topographical constraints[5] is another efficient approach. The Gabor filter responses are also used, but coarse lattice on the whole input image is organized to find the adequate grid set for rep-

*Address: San 14, Nongseo-Ri, Kiheung-Eup, Yongin City, Kyungki-Do, 449-712, Korea. E-mail: {sulee, sckee}@sait.samsung.co.kr

resenting the facial region. After best fit of a grid set that is representative of facial region is selected, the step of fine search also regulates all points of the grid for corresponding to each fiducial point. This approach is more effective one, but it needs too much computational complexity in fiducial point detection. To represent the facial region well, the process of spreading test points on input image, and selecting the suitable grid set is significant. Therefore, to complete the defects, we propose new grid form, M-style grid, and reference face model DB for facial region segmentation. A hybrid shape analysis method is also proposed to detect the facial features as fiducial points where Gabor filter responses are computed.

The process of facial region segmentation is described in next section, and the process of detecting the center points of both eyes as fiducial points is described in section 2. We show the performance of proposed method in section 4.

2 Facial Region Detection using M-Style Grid Matching

Face detection, facial region segmentation module is divided into two parts. One is model DB that is consist of trained Gabor filter response vectors computed from reference face models. The face models are classified into several groups, and the Gabor filter responses from all grid points of the correctly masked M-style grid on each face model are averaged. The average response vectors of each group are the reference feature vectors of model DB. Another is the matching procedure. In this procedure, the lattice of test points on whole input image is organized, and Gabor filter response vectors on each point are computed. All available M-style grid sets are compared with model DB to detect the face. Figure 2 illustrates the functional flow of the proposed system. In figure 3, an example of the lattice for grid structure and M-style grid selection is shown.

2.1 M-style Grid

Gabor kernel is well known for its biological relevance and technical properties[3]. The Gabor kernels are similar shape as the receptive fields of simple cells in the primary visual cortex. So Gabor filter has multi-scale and multi-orientation kernels. The response describes a small patch of gray values in an image $I(\vec{x})$ around a given pixel $\vec{x} = (x, y)$. It is defined as a convolution

$$J_i(\vec{x}) = \int I(\vec{x}')\Psi_j(\vec{x} - \vec{x}')d^2\vec{x}' \quad (1)$$

$$\Psi_j(\vec{x}) = \frac{k_j^2}{\sigma^2}\exp(-\frac{k_j^2 x^2}{2\sigma^2}) \exp[\exp(i\vec{k_j}\vec{x}) - \exp(-\frac{\sigma^2}{2})] \quad (2)$$

with a family of Gabor filters in the shape of plane waves with wave vectors k_j, restricted by a Gaussian envelope function. We employ a discrete set of 5 different frequencies, index $\nu = 1, \ldots, 4$ and and 8 orientations, index $\mu = 0, \ldots, 7$ with index $j = \mu + 8\nu$.

$$\vec{k}_j = (\frac{k_{jx}}{k_{jy}}) = (\frac{k_\nu \cos\varphi_\mu}{k_\nu \cos\varphi_\mu}), \quad (3)$$

where

$$k_\nu = 2^{-\frac{\nu}{2}}\pi,$$
$$\varphi_\mu = \mu\frac{\pi}{8}.$$

Gabor kernels provide robustness against varying brightness and contrast in the image, and have the characteristics of the local face area as a value, so it's more effective than using the original face image[4].

To compare the similarity of the values, we used similarity function.

$$S_a(J, J') = \frac{\sum_j a_j a'_j}{\sqrt{\sum_j a_j^2 \sum_j a'^2_j}} \quad (4)$$

The size of input image is 240×320, and the target face size is 40~80 pixel wide (the criterion is ocular distance). But we reduce the input image by half for preprocessing. The resolution as low as 120×160 is not good for face recognition or verification, but it has some good attributes to detect a face. It is also effective on noise reduction, and lower computational complexity is needed. So, we modify the size of kernels by $k_\nu = 2^{-\frac{\nu}{2}}\pi$, instead of $k_\nu = 2^{-\frac{\nu+2}{2}}\pi$, to improve the efficiency of facial region detection. As the size of kernels modified, the required convolution window size is changed 64×64 to 24×24.

As shown in Figure 3, M-style grid consists of 20 (5×4) points. Each 5 points that form M-style consist a·row. The width of the grid is a quarter of inter ocular distance, the height is 1/3 of perpendicular distance between the center of mouth and inter ocular line. In comparison with normal grid structure (square-style), the proposed grid structure has advantage of accuracy at same size, same resolution of Gabor kernels.

2.2 Model DB

To generate the model DB, face model images are classified into several groups according to the characteristics such as lights, backgrounds, wearing glasses. Figure 4 shows the example groups of face models. Let the model DB has N groups of face models, model DB can be defined by $R(i) = \{G_i\}$, for

$i = 0, \ldots, N$. The Averaged vectors of a face model group can be defined as (5) where $\vec{p}$ represents a grid point, $L(\vec{p})$ is the real coordinate on face model, and J_j is a Gabor filter response vector defined in (1).

$$G_i(\vec{p}) = \frac{1}{M}\sum_{n=1}^{M} g_n(\vec{p}), \tag{5}$$

where

$$g_n(\vec{p}) = J_j(L(\vec{p}))$$

M is the number of face models in a group. Figure 5 describes the process of generating averaged filter response vectors from a face model group.

3 Eye Detection

We employ the iterative binary thresholding with perfect contour tracing for eye center point detection. Figure 6 shows the steps of eye point detection process. After histogram normalization, 256 gray levels of input image changes to binary one repeatedly. The threshold is stepwise increased by 8 from 48. In each step, the binary image is processed by the continuous operations of erosion, dilation, and dilation. By erosion, noisy pixels are removed, and successive dilation enriches the edge information of eye region. After a stage of operations, all segmented areas are traced by perfect contour tracing. Eye candidates are selected by the criterion of good-shape condition. The good-shape condition tests each contour whether circle or ellipse, large or small, and whether it's center has adequate location as an eye or not. Finally, a pair of contour that has the relation of appropriate position selected to both eyes. The centers of the pair considered as eye positions. Figure 7 shows well-segmented facial regions and successfully detected eye points.

4 Experimental Results

As shown in Figure 4, we have classified the face model images into 5 groups according to the characteristics such as simple background, backlight, unalloyed light with complicated background and wearing glasses, read as follows except for glasses, and polluted eye region by noisy spots. Each group has 10 models of face. To verify the performance, we consist 6 groups of averaged Gabor filter response vectors for model DB. 5 vectors are computed from the face model groups, the last is the average of whole 5 vectors. The size of grid structure from model DB training is also applied to test image. The test image sets have various light condition of office (fluorescence) light and some direct rays of the sun from window. The target face to detect is the frontal face that has the variation of ± 15 degrees, and the size varies from 30 to 100 pixels. The ocular distance between the left center to the right center is the basis of face size. The off-line simulation results are over 99% successful segmentation rate of facial region (Face Detection Rate : FDR), and 99% effective eye center position detection rate of facial region (Eye position Detection Rate : EDR). The application system, **MagicGate**, on Pentium-III 550MHz PC, has the efficiency of 97% FDR, and 85% EDR on 240 by 320 images. The processing time is 220ms per image. The measured computational complexity is as low as about 32WMOPS. We summarize the performance evaluation in table 1.

5 Conclusion

In this paper, we propose an efficient method for face detection. To make a real-time application, we supposed that the target face to detect is frontal face, and the size and rotation variations are restricted. But M-style grid matching of Gabor filter responses provides robustness against illumination and background variations. The results show that the proposed approach is quite practical and useful. As further research, we will propose an efficient method that has strong invariant attributes under size and pose variations using additional information of skin color and deformable templates.

References

[1] K.C.Chung, S.C.Kee and S.R.Kim, *Face Recognition using Principal Component Analysis of Gabor Filter Responses*, International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, pp.53-57, Sep. 1999.

[2] M. Turk and A. Pentland. *Eigenfaces for recognition*, Journal of Cognitive Neuroscience, 3:72-86, 1991.

[3] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Wurtz, and W. Konen. *Distortion invariant ob-ject recognition in the dynamic link architecture*, IEEE Trans. on Computers, 42(3):300-311, 1993.

[4] L. Wiskott, J. M. Fellous, N. Kruger, and C. V. der Malsburg. *Face recognition by elastic graph matching*, IEEE Trans. On Pattern Recognition and Machine Intelligence, 19:775-779, 1997.

[5] L. Wiskott, *The Role of Topographical Constraints in Face Recognition*, Pattern Recognition Letters, 20(1):89-96.

Figure 1: Examples of Bunch Graph Matching

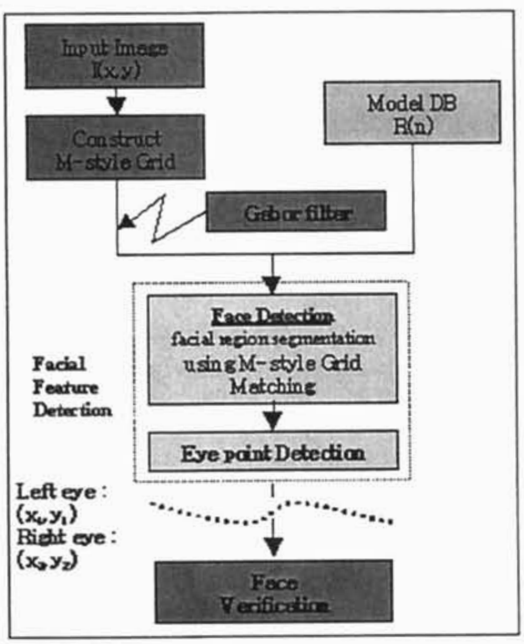

Figure 2: Functional flow of facial feature detection

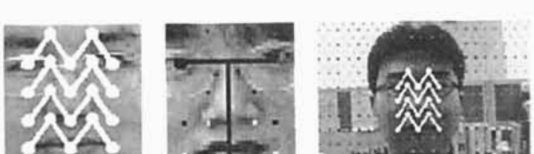

Figure 3: M-style grid

Figure 4: Example of Face model groups

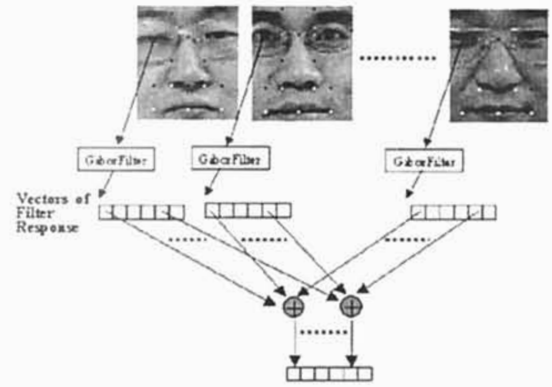

Figure 5: Generating averaged filter response vectors from a face model group

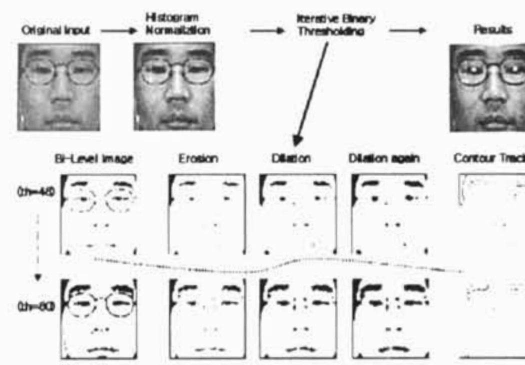

Figure 6: Functional flow of Eye Center Point Detection

Figure 7: Segmented facial regions and detected eye points

Table 1: Performance evaluation : accuracy, speed, and computational complexity

Module	Performance
Facial Region Segmentation	Detection size : 30 ~ 100 pixels Light : office illumination (fluorescence) Rotation : -15~+15 degree (frontal face based) Accuracy : Offline test – 99% (6 faults in 557 images) Speed : Online(Real-time) test – 480ms/image (in Pentium II 266MHz PC) Computational Complexity : 24.9WMOPS MAC : 22040064*1weight = 22 WMOPS Divide : 67664 *18weight = 1.2 WMOPS SQRT : 5147*32weight = 1.7WMOPS
Eye Point Detection	Object : segmented facial region Accuracy : Offline test – 99% (3 faults in 441 test regions) Speed: Online(Real-time) test – 120ms/image (in Pentium II 266MHz PC) Computational Complexity : 8WMOPS MAC and IF operation = 2004480*4 = 8WMOPS
Magic Gate	Detection size : 40 ~ 80 pixels Light : office illumination (fluorescence) Rotation : -10~+10 degree (frontal face based) Accuracy : 97% of FDR, 85% of EDR Speed : Online(Real-time) test – 220ms/image (in Pentium III 550MHz PC)