MVA2002 IAPR Workshop on Machine Vision Applications, Dec. 11- 13, 2002, Nara- ken New Public Hall, Nara, Japan

3—20

Hierarchical Combination of Face/Non-face Classifiers Based on Gabor Wavelet and Support Vector Machines

Young Hoon Sung, Tae-Kyun Kim, and Seok Cheol Kee
HCI_LAB, Samsung Advanced Institute of Technology
{yhsung, tkkim, sckee}@sait.samsung.co.kr

Abstract

In this paper, we propose a real-time face and eye detection algorithm for video surveillance and human computer interface. Different types of face/non-face classifiers are hierarchically combined for reliability and real-time performance. Each classifier is connected by consisdering the accuracy and translation/scale sensitivity of the previous classifier. First, face candidates are extracted using similarity matching of Gabor filter responses in M-style grid. Then, a hierarchy of SVM classifiers trained on PCA subspaces is applied to the candidates. Two steps of SVMs are trained on different number of PCA features and resolution images. The combination of classifiers based on different types of features, frequency and intensity, significantly reduces the false positives due to complementary characteristics of their domains in classification. Coarse-to-fine search by three steps of detection also reduces run-time complexity. Our system can speed up conventional SVM classifier by a factor of 40 resulting in comparable face detection rate (FDR). In addition, the center positions of both eyes are efficiently detected by iterative binary thresholding method with contour tracing in the accurately localized face region. Experimental results on the test set taken under office illumination show the accuracy with FDR of 98%, 0.5% false positives, and eye detection rate (EDR) of 99%. Our current system detects a face in runtime of 250ms.

1 Introduction

Face detection emerged as an important area of computer vision and pattern acknowledgement with a vast number of applications such as video surveillance, human-computer interface, and identity authentication. Numerous algorithms related to face detection have been proposed and developed within past 20 years. The number of the products of face detection algorithm is increasing nowadays for the security at the airports and gateways. However, the performance of previous face detection system is not enough for fully automatic surveillance. To increase automatism and efficiency of the security system and Human-Computer interface, we propose a novel system that has a hierarchical combination of the face and non-face classifiers.

As approaches to solve the problem of face detection and classification, linear subspace analysis has been mainly adopted like Principal Component Analysis (PCA) or Fisher Linear Discriminant Analysis (FLDA) because of its simplicity and efficiency [3]. Support Vector Machines (SVM) is a very powerful su pervised learning algorithm, that is rooted in statistical learning theory [1,4,5,6,8]. By minimizing the sum of the empirical risk and the complexity of the hypothesis space,
SVM gives good generalization performance on many pattern recognition problems.

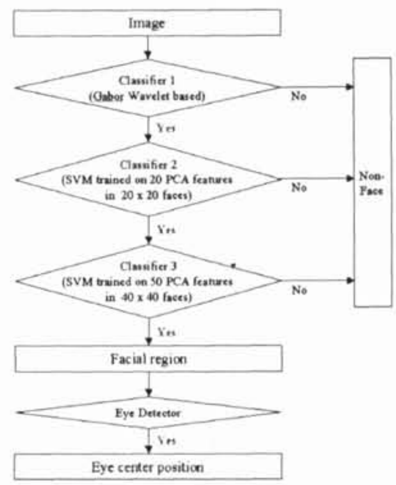


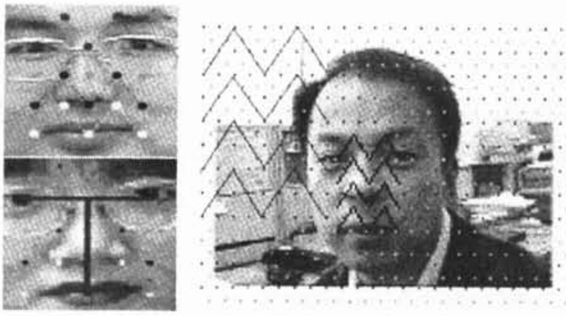
Figure 1. Overview of the proposed algorithm

In this paper, we propose a hierarchical learning scheme that integrates the gabor wavelet with SVM, and apply it to face detection problem. Through the gabor wavelet matching in M-grid, we extract the candidate region of faces [2,7]. And we use PCA features of normalized images to get a larger margin of separation and fewer support vectors of SVM. The combination of the grid matching of Gabor wavelet and the SVM classifiers based on PCA features of images significantly reduces the false positives due to different types of features. Figure 1 shows the overview of the proposed system.

The rest of the paper is organized as follows. In section 2, we present gabor wavelet filtering using M-style grid matching for finding a face candidate in image. In section 3, we give a brief overview of PCA and SVM, and present the application to our system. Combination of Gabor and SVM based classifiers is described in section 4. In section 5, we explain eye detection algorithm with the boundary of search. In section 6, we show the experimental results and finally gives the conclusion in section 7.

2 M-style grid Gabor matching for face candidate

To extract the candidate region of continuous input im-



(a) composition (b) M-grid similarity matching

Figure 2. M-style grid

ages, many researches related with color or shape of the facial images have been developed. Pixel-level skin color can greatly reduce the search space rapidly prior to higher level classification. But skin color of the human is affected by environmental parameters in which camera is built. Also skin color is changed according to the direction and strength of lighting source. In this paper, we use shape information by similarity matching of Gabor filter responses in M-style grid [2].

Gabor wavelets are used for image analysis because of their biological relevance and computational properties [7]. The Gabor kernels have similar shape as the receptive fields of simple cells in the primary visual cortex. Gabor filter has multi-scale and multi-orientation kernels. The response describes a small patch of gray values in an image $I(\mathbf{x})$ around a given pixel $\mathbf{x} = (x, y)$. It is defined as a convolution with a family of Gabor filters in the shape of plane waves with wave vectors $\mathbf{k}_j$, restricted by a Gaussian envelop function.

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

$$\Psi_j(\mathbf{x}) = \frac{k_j^2}{\sigma^2}\exp(-\frac{k_j^2x^2}{2\sigma^2})\left[\exp(i\mathbf{k}_j\mathbf{x}) - \exp(-\frac{\sigma^2}{2})\right] \quad (2)$$

We employ a discrete set of 5 different frequencies and 8 orientations. Gabor kernels provide robustness against varying brightness and contrast of images. To compare the similarity of the values in Gabor Wavelet based classifier, we used similarity function below.

$$S(J, J') = \frac{\sum_j a_j a_j'}{\sqrt{\sum_j a_j^2 \sum_j a_j'^2}}. \quad (3)$$

The system to extract the face candidate is composed with two parts: the one is training of the model face images containing the various changes of the face, the other is the similarity comparison of the input image and trained model facial image using M-style grid matching. Figure 2 depicts the composition of M-style grid and it's selection for extracting the face candidate. Scalable M-style grid structure is adapted to an input image and training of model face DB. Gabor filter response of each point at the grid structure is compared with the mean of Gabor responses of the model face DB. As shown in Figure 2, M-style grids consist of 20 points which form M shape. The horizontal interval of the grid points is 1/4 of the inter ocular distance and the vertical interval is 1/3 of the perpendicular distance between mouth and inter ocular line.

(a) First 20 PCA bases from 20x20 faces

(b) First 50 PCA bases from 40x40 faces

Figure 3. Learned basis images

3 Feature reduction and hierarchical combination of SVM classifier

In the section 2, we extracted the face candidate region by using M-style grid Gabor matching. To detect the face from the candidate region, a hierarchy of SVM classifiers trained on PCA subspaces is applied to the candidates.

3.1 Feature extraction using PCA

Each image pattern of dimension $I \times J$ can be considered as a column vector $\mathbf{x}$ in a $N = I \times J$ dimensional space. Facial images are not randomly distributed in higher dimensional image space and thus can be described by a relatively low dimensional subspace. As a suitable mean to reduce the dimensionality of the face space, PCA is adopted [3]. The central idea of PCA is to find a low dimensional subspace which captures most of variation of images and therefore allows the best least square approximation.

Consider a data set $\chi = \{x_1, x_2, ..., x_M\}$ of N dimensional vectors.

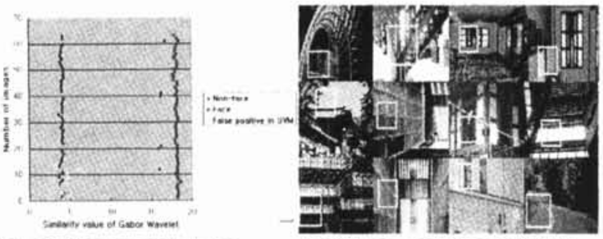
$$\boldsymbol{\mu} = \frac{1}{M}\sum_{m=1}^{M}\mathbf{x}_m, \quad (3)$$

$$\Sigma = \frac{1}{M}\sum[\mathbf{x}_m - \boldsymbol{\mu}][\mathbf{x}_m - \boldsymbol{\mu}]^T. \quad (4)$$

where Σ is a $N \times N$ symmetric matrix. This matrix charaterizes the scatter of the data set. A non-zero vector u_k for which

$$\Sigma\mathbf{u}_k = \lambda_k\mathbf{u}_k. \quad (5)$$

is an eigenvector of the covariance matrix. It has the corresponding eigenvalue λ_k. If $\lambda_1, \lambda_2, ..., \lambda_K$ are the K largest eigenvalues, distinct eigenvalues then matrix, $U = [\mathbf{u}_1\mathbf{u}_2...\mathbf{u}_K]$ represents the K dominant eigenvectors. These eigenvectors are mutually orthogonal and span a K dimensional subspace called the principal subspace. When the data are face images these eigenvec



(a) similarity matching (b) false positive in SVM

Figure 4. Similarity matching value of Gabor Wavelet about face/non-face samples and false positive in SVM

tors are often referred to as eigenfaces. Figure 3 shows the learned basis images of PCA..

3.2 SVM classification with PCA feature

After obtaining PCA features, we build the SVM training set $\{\mathbf{x}_i, y_i\}_{i=1}^{n}$, where $y_i = \{1,-1\}$ is the class type of $\mathbf{x}_i$, which is PCA features of each training image. n is the total number of training images. SVM learn a hyperplane which separates the data according to the class with a large margin. A test pattern, $\mathbf{x}_{test}$, is classified as a face or not by using the trained SVM [1].

$$f(\mathbf{x}_{test}) = sign(\sum_{i=1}^{l} y_i \lambda_i \mathbf{K}(\mathbf{x}_{test}, \mathbf{x}_i) + b), \quad (6)$$

where y_i and $\mathbf{x}_i$ are a class label and a training feature vector respectively, λ_i and b are constants which are decided by learning, $\mathbf{K}$ is a polynomial kernel and l is the number of support vectors.

3.3 Hierarchy of SVM classifier for face detection

SVM classifier in our system consists of two cascaded SVMs that are trained on different number of PCA features of low/or high resolution images. The first level of our hierarchical SVM classifier consists of SVM that is trained on 20 PCA features of 20 x 20 pixel resolutin face images. On the second level, we utilize the SVM that is trained on 50 PCA features of 40 x 40 face images. In SVM classifier, run-time complexity is proportional to the number of the support vectors(SVs). One main goal of our system is to speed up a classifier without loss of classification performance. By propagating only pattern that was classified as face in the first level classifier having a small number of SVs, we quickly reduce the amount of data in the second level classifier.

4 Hierarchy of M-style grid Gabor matching and SVM classifier

Although SVM is known as a superior classifier, it is difficult to adapt it for real-time implementation because run time complexity is very high. Currently, many researches have been proposed and developed to solve this problem [6]. We combined different types of face and non-face classifier hierarchically for reliability and real time performance. The combination of classifiers based on different types of features significantly reduces the false positives due to complementary characteristics of frequency(here, it is obtained from Gabor wavelet) and

(a) (b) (c) (d) (e) (f)

Figure 5. Detection results of the hierarchical steps

intensity (PCA features) domains in classification.

The bottom level of our hierarchy consists of the Gabor wavelet based classifier which is robust to the variation of illumination, expression, and in-plane/depth rotation. It also can extract the face candidate rapidly due to M-grid search as shown in Figure 2, (b). For example of the implemented system, the size of input image is 320x240, and the target face size has 40~80 pixel wide of ocular distance. The processing time is 170ms per image on Pentium-4 1.6GHz. In addition, it can filter non-face patterns which is false positives in SVM classifier based on PCA features. Figure 4 (a) shows that false positive patterns of SVM classifier is nearer to the non-face patterns in Gabor Wavelet matching being far from the face patterns. We built the system hierarchically for the mutual complementary characteristics of classifiers as mentioned above. We extract the face candidate with Gabor wavelet based classifier. In the next step, we apply the non-linear SVM that is trained on the 20 PCA feature in 20x20 face images to the face candidate extracted in the previous step. Finally, we use the SVM with a second degree polynomial kernel that is trained on 50 PCA feature in 40x40 face images. This classifier is highly sensitive to translation and scale. Figure 5 shows the results of the each step of our system about two different face. Figure 5 (a, d) is the result of first step of our hierarchical system, (b, e) is that of second step and (c, f) is final result of face detection. We detect face very accurately in the final level of our system (c,f).

5 Eye Detection

The iterative binary thresholding with contour tracing in the restricted region is employed for eye center point detection. Because facial region extracted in our system is accurate, the region to search the eye center position can be strictly constrained. In the restricted region, 256 gray levels of an input image is converted to binary one repeatedly with a threshold. 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. After a stage of morphological filtering, all segmented areas are traced by contour tracing. Then, good-shape condition is given to decide whether each contour is from eyes The condition tells that whether the shape of contour is circle or ellipse, large or small, and whether it's center is adequately located as an eye or not. Figure 6 shows well-segmented facial regions and successfully detected eye points.

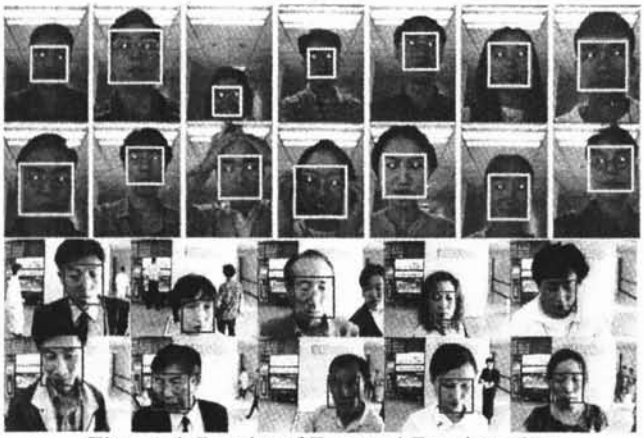


Figure 6. Results of Face and Eye detection

6 Experimental Results

For the bottom level of classifier, Gabor grid mathing, the face model was divided into 5 groups according to the various characteristics such as simple background, backlight, wearing glass, etc. When a test pattern is given, it's Gabor response is compared with all 5 means of Gabor responses of the face groups and the minimum matching value is selected to decide whether it is a face or not. In the second level of our hierarchical system, SVM was trained on 20 PCA features of 3,175 20x20 faces and 10,000 randomly selected non-face patterns. And for the final level, SVM was trained on 50 PCA features of the 40x40 faces and non-faces with the same number of patterns. The number of SVs of the second and third SVM is 187 and 270 respectively.

The test image sets include various light conditions of office light and some direct rays of sun from window. The target face to detect is the frontal face that has the variation of ± 15 degrees in depth and plane, and the size varies from 40 to 80 pixels of the ocular distance. 400 images with 200 faces and 200 non faces was tested in Pentium-4 1.6GHz. Table 1 shows the comparison of the previous method and the proposed algorithm for face detection.

Table 1.Test results of the face detection methods

	FDR	FPR	Speed
Gabor Wavelet	98%	48.25%	170ms
SVM classifier	98.5%	3%	11497ms
Proposed	**98%**	**0.5%**	**250ms**

* : FDR : face detection rate, FPR : false positive rate

We reduced the FPR largely and achieved a speed up by a factor of 40 compared to the conventional SVM classifier. Also, we detected the eye center position with 99% detection rate within ± 10 pixel accuracy. Figure 6 shows the accurate detection result of face and eyes and Figure 7 shows the decrease of false positives about non-face images. Gabor wavelet based classifier and SVM based one show many false positives in Figure 7 (a) and (b). But our system reduces the false detection largely due to complementary characteristics of frequency and intensity domains in classification as shown in Figure 7 (c).

(a) false positives in M-grid Gabor wavelet classifier

(b) false positives in SVM classifier

(c) results of the proposed hierarchical classifier

Figure 7. False positives

7 Conclusion

In this paper, we proposed an efficient method for face detection which consists of hierarchy of Gabor Wavelet and SVM classifier base on different types of features. This largely increased the detection accuracy and can extract facial region 40 times faster than conventional SVM classifier. Also our system provides robustness against illumination and background variation. The results show that the proposed approach is quite practical and useful.

References

[1] V. Vapnik, *Statistical learning theory*. John Wiley and Sons, New York, 1998.

[2] S. U. Lee, Y. S. Cho, S. C. Kee and S. R. Kim, "Real-time Facial Feature Detection for Person Identification System," *IAPR Workshop on Machine Vision Applications*, Tokyo, Japan, Nov 2000.

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

[4] Yuan Qi, David Doermann, and Daniel DeMenthon, "Hybrid independent component analysis and support vector machine learning scheme for face detection," *ICASSP*, vol.3, pp. 1481-1484, Utah, May 2001.

[5] Edgar Osuna, Rovert Freund and Federico Girosi, "Training Support vector machines: an application to face detection," *CVPR*, pp. 130-136, San Juan, June 1997.

[6] B. Heisele, T. Serre, S. Mukherjee, T. Poggio, "Feature reduction and Hierarchy of Classifiers for Fast Object Detection in Video Images," *CVPR*, vol2, pp. 18-24, Hawaii, December 2001.

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

[8] Qi. Yuan, D. David, and D. Daniel, "Hybrid independent component analysis and support vector machine learning scheme for face detection," *ICASSP*, pp. 1481-1492, May 2001.