Integrated Approach of Multiple Face Detection for Video Surveillance

Tae-Kyun Kim, Sung-Uk Lee, Jong-Ha Lee, Seok-Cheol Kee and Sang-Ryong Kim
Human Computer Interaction Lab., Samsung AIT, KOREA.
taekyun@sait.samsung.co.kr

Abstract

For applications such as video surveillance and human computer interface, we propose an efficiently integrated method to detect and track faces. Various visual cues are combined to the algorithm: motion, skin color, global appearance and facial pattern detection. The ICA (Independent Component Analysis)-SVM (Support Vector Machine) based pattern detection is performed on the candidate region extracted by motion, color and global appearance information. Simultaneous execution of detection and short-term tracking also increases the rate and accuracy of detection. Experimental results show that our detection rate is 91% with very few false alarms running at about 4 frames per second for 640 by 480 pixel images on a Pentium IV 1GHz.

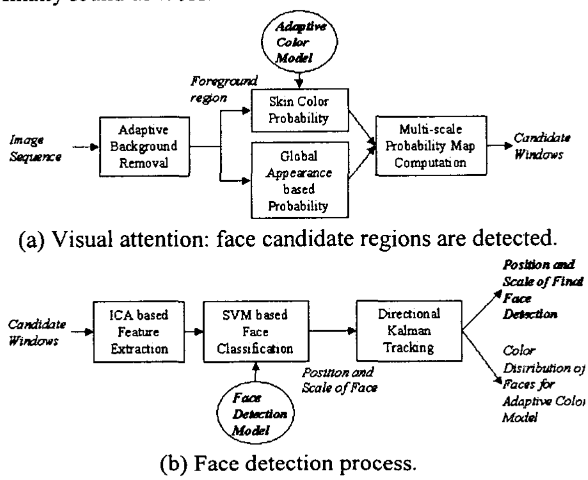
1. Introduction

Although there are various approaches to face detection, it is still difficult to detect multiple faces reliably in variations of condition and real-time. Current computing power should be combined with various visual cues such as motion, color and pattern to enhance the performance of face detection. Previous methods[1,4] utilised multi-modal information in this sense. Stereo information has been increasingly used, and skin color has been regarded as an essential clue. Neural network based methods were used for pattern matching in [1] and this pattern matching step is also essential for reliable detection. Recently, SVM has great attention for the face pattern classifier[2,6]. It is beneficial to have good generalization performance for new test data. By extracting features using ICA, better results were obtained compared to using raw images[2]. The methods[1,2,3,5,6] of pattern matching without facial feature localisation have the advantage of detecting very small faces, important in security systems needing to capture distant faces.

In this paper, we contribute two major aspects. For more reliable face pattern detection, the feature expansion based on ICA is proposed with SVM. Secondly, we propose a new integrated approach of real-time detection and tracking in crowded and variable environments using adaptive background removal, the probability map of skin color and global appearance, short-term tracking and pattern detection. Experiments were performed on both still grey images and color video sequences.

2. Overview of the Approach

As shown in Figure 1, the first step is background removal. By applying adaptive image difference techniques, the foreground region is extracted. The multi-scale probability map of the extracted region is made by using a skin color model and global appearance. HS(Hue and Saturation) Gaussian model is adapted to the color distribution of final detection. Pattern detection is executed on windows which have high probability on the map. The reliably matched windows are followed by Kalman tracking. Parameters of Kalman filter are updated with priority of concurrent detection result. Successively tracked objects are finally found as faces.



(a) Visual attention: face candidate regions are detected.

(b) Face detection process.

Figure 1. Architecture of the proposed algorithm.

3. Adaptive Background Removal

To segment foreground regions, we introduce a hybrid method using both intensity and color based difference images. Intensity difference provides relative robustness to minute and gradual changes of lighting condition. In spite of its effectiveness, regions of interest may be partially segmented. Because foreground regions which have same intensity of corresponding background are also removed.

On the other hand, color difference makes more noise, but provides an intact foreground. Using this complement, we determine seed regions by shape analysis on the intensity difference image. The seed regions are used as the bases of foreground segmentation in the color difference image. Binarization of difference images is executed. Figure 2 shows the process of foreground segmentation.

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

Figure 2. (a) Input image. (b) Intensity based differential image, determined seeds. (c) Color based differential image. (d) Output image

For reference frame updates, we use consecutive and gradual background adaptation. Let *B(x,y)* represent a background points on a segmentation result image as shown in Figure 2. (d), the adapted new background $R'(x,y)$ is obtained by

$$R'(x,y) = \beta R(x,y) + (1-\beta)B(x,y), \qquad (1)$$

where *R(x,y)* is current background. β is a weighting factor. Under typical office and home illumination condition, $\beta = 0.9$ yields the most effective adaptation.

4. Multi-scale Probability Map

The probability map having a scale parameter of face is defined by the product of two probability density functions(pdf).

$$P_{total}(x,y,n) = P_{skin}(x,y,n) \cdot P_{global}(x,y,n), \qquad (2)$$

where n is the size of face to be detected and $P_{skin}(x,y,n)$ is the pdf defined by skin color information and $P_{global}(x,y,n)$ is the pdf defined by global appearance. By decreasing n, the probability is calculated and pattern matching is performed on the position with high probability. If a face is detected, the rectangular region of the detected face is ignored in the next search iteration for speeding up.

The pdf for skin color is defined using a two-dimensional Gaussian function, $g(i,j;\mathbf{u},\Sigma)$ and face size n.

$$P_{skin}(x,y,n) = \frac{\sum_{i=x-n/2}^{i=x+n/2}\sum_{j=y-n/2}^{j=y+n/2} g(Hue(i,j), Sat(i,j); \mathbf{u}, \Sigma)}{n^2}, \qquad (3)$$

where $\mathbf{u}$ and Σ are the mean and the covariance matrix respectively of face skin cluster in Hue and Saturation color space. These are continuously updated by those of the color distribution of final detected faces for robustness to illumination changes.

The probability function of global appearance implies that normally faces are found in the upper position of a body. However, this can make a limitation of face detection in case that a segmentation result is not enough to distinguish each top of body and body is positioned upside down. Note that this can be easily removed from the proposed algorithm if this limitation can not be allowed in a given application.

Let B_{ij} represent the *j-th* point of equally divided points of the *i-th* boundary of the segmented foreground region. The local feature point, M_{ij} is the point such that

$$\frac{1}{\Delta j}\frac{|\Delta\angle \mathbf{n}(B_{ij})|}{\Delta j} = 0 \quad and \quad 0 \le \angle \mathbf{n}(B_{ij}) \le \pi, \qquad (4)$$

where $\angle \mathbf{n}(B_{ij})$ is the angle of normal vector of B_{ij}. The pdf is defined as the mixture of skew Gaussian distributions like the following:

$$P_{global}(x,y,n) = \sum_{i=1}^{N} g(x_i, y_i; \mathbf{u}_i, \Sigma)$$

$$\mathbf{u}_i = \begin{pmatrix} u_{ix} \\ u_{iy} \end{pmatrix} = \begin{pmatrix} m_{ix} + n\cos(\angle(-\mathbf{n}(m_i))) \\ m_{iy} + n\sin(\angle(-\mathbf{n}(m_i))) \end{pmatrix}, \quad \Sigma = \begin{pmatrix} n^2 & 1.5n^2 \\ 1.5n^2 & (1.5n)^2 \end{pmatrix},$$

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} = \begin{pmatrix} \cos(\angle \mathbf{n}(m_i)) & -\sin(\angle \mathbf{n}(m_i)) \\ \sin(\angle \mathbf{n}(m_i)) & \cos(\angle \mathbf{n}(m_i)) \end{pmatrix} \cdot \begin{pmatrix} x - u_{ix} \\ y - u_{iy} \end{pmatrix}, \qquad (5)$$

where m_i is the local feature point, N is the total number of feature points and n is a face size.

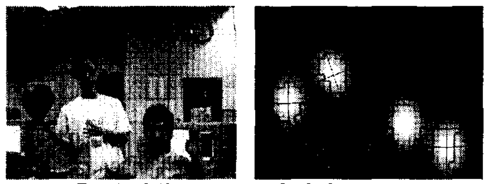


Figure 3. Probability map of global appearance

5. Pattern Matching

A hybrid learning scheme that combines ICA and SVM is used in the proposed algorithm. In low-level feature extraction, ICA produces statistically independent image bases. In high-level pattern classification, SVM classifies the ICA features as a face or non-faces. In [2], it was showed that by using ICA features better result was obtained than by training SVM directly on the raw image data. In this paper, we expanded ICA features from not only face but also face-like class and also used the residual error in each class to dramatically enhance detection performance. Let $\mathbf{X}$ be an input matrix whose rows are training images.

$$\mathbf{U} = \mathbf{W}\mathbf{X}, \qquad (6)$$

where $\mathbf{W}$ is a weight matrix and $\mathbf{U}$ is the output matrix of ICA. The rows of output, $\mathbf{U}$, are also images. To control the number of independent basis images, main eigenvectors of the training images are used to train ICA basis. Let $\mathbf{P_m}$ denote the matrix containing the first m eigenvectors in its columns. The PCA representation of the zero-mean images $\mathbf{X}$ is defined as $\mathbf{R_m} = \mathbf{XP_m}$ and reconstruction of $\mathbf{X}$ is obtained by $\mathbf{X_{rec}} = \mathbf{R_m P_m}^T$. ICA is performed on $\mathbf{P_m}$.

$$\mathbf{WP_m}^T = \mathbf{U} \;\Rightarrow\; \mathbf{P_m}^T = \mathbf{W}^{-1}\mathbf{U}. \tag{7}$$

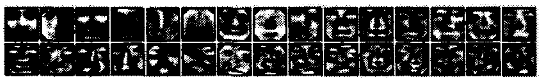
Therefore ICA reconstruction is obtained by

$$\mathbf{X_{rec}} = \mathbf{R_m P_m}^T \Rightarrow \mathbf{X_{rec}} = (\mathbf{XP_m})(\mathbf{W}^{-1}\mathbf{U}). \tag{8}$$

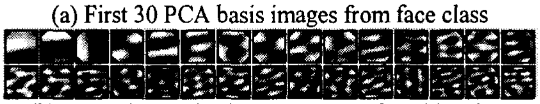
And the ICA representation of $\mathbf{X}$ is given by

$$\mathbf{B_m} = \mathbf{XP_m W}^{-1}. \tag{9}$$

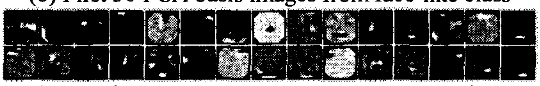
Various face, face-like and non-face images were collected and normalised to 20×20 to train ICA and SVM. Normalised images were histogram equalized and dimensions were reduced to be 300 by oval masking. The face images included artificially rotated, shifted, resized faces. The face-like images were initially selected as the non-face images which have small Euclidean distance to the average face image. Figure 4 shows the sample sets of learned PCA and ICA basis images.



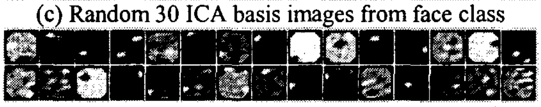
(a) First 30 PCA basis images from face class



(b) First 30 PCA basis images from face-like class



(c) Random 30 ICA basis images from face class



(d) Random 30 ICA basis images from face-like class

Figure 4. Learned basis images

As shown in Figure 4 (b), basis images of the face-like class consist of a set of horizontal edges. It is different from the basis images of random non-face class, which is a set of random edges. And the ICA basis images in Figure 4 (c), (d) appear more locally than the PCA basis images. Here, we used 50 coefficients of the ICA bases which span the face class and 50 coefficients of the ICA bases which span the face-like class. We added two more features, reconstruction errors in two classes. We used Euclidean distance measure for errors, assuming isotropic Gaussian data distribution like [3]. After extracting features of a given pattern, we classify it as a face or not using the trained polynomial SVM.

$$f(\mathbf{x}_{test}) = sign\left(\sum_{i=1}^{l} y_i \lambda_i K(\mathbf{x}_{test}, \mathbf{x}_i) + b\right), \tag{10}$$

where $\mathbf{x}_{test}$ is a test vector with 102 dimensions, y_i and $\mathbf{x}_i$ are a training label and vector respectively, λ_i and b are constant, K is a polynomial kernel with degree 2 and the number of support vectors, l, is 467.

6. Tracking with Directional Kalman Kernel

If a face is detected, Kalman tracking starts with the mean position and size of the detected face using the probability map of skin color described in Section 4. Gaussian kernel parameters, mean and covariance matrix, are continuously updated like [8]. However, if a newly detected face overlaps with the tracked face, the parameters would be initialized with those of the detected face. Successively tracked objects for more than 2 frames are finally found as faces.

The directional Kalman kernel, $\overline{G}(x,y)$, is proposed for efficient tracking even when multiple faces are moving fast. A kernel with a biased probability to the direction of motion not only tracks an object that moves fast but also gives the least disturbance to the other kernels. The kernel is defined by

$$\begin{aligned}
&\overline{\sigma}_x^2 = 2(\sigma_x^2 + \Delta\mu_x^2)\ ,\ \overline{\sigma}_y^2 = 2(\sigma_y^2 + \Delta\mu_y^2),\\
&f(x,y,\sigma_x,\sigma_y) = \frac{1}{S}\cdot\exp\left\{\frac{-(Z_x^2 - 2\sigma_{xy}Z_xZ_y + Z_y^2)}{2(1-\sigma_{xy}^2)}\right\},\\
&Z_x = \frac{x-\mu_x}{\sigma_x},\quad Z_y = \frac{x-\mu_y}{\sigma_y},\\
&\overline{G}(x,y) = \begin{cases} f(x,y,\sigma_x,\sigma_y) & \text{if case 1}\\ f(x,y,\overline{\sigma}_x,\overline{\sigma}_y) & \text{if case 2}\end{cases},
\end{aligned} \tag{11}$$

where, (μ_x, μ_y) is the mean and (σ_x, σ_y) is the standard deviation, case 1 implies the position on the negative side of motion and case 2 does the position on the positive side of the motion. S is the constant that makes the total sum of probabilities to be one.

7. Experimental Result

We tested the detection performance of the proposed algorithm in both still grey images and color video sequences. For the quantitative evaluation of the pattern detection step described in Section 5, two sets of still grey images were used. This test examined all image windows and classified them as a face or not. Set A contained 400 high-quality images with one face per image from Olivetti DB. Set B contained 36 images of mixed quality, with a total of 172 faces from CMU DB. Set A involved 1684800 pattern windows, while Set B involved 6178110. Table 1 shows that the proposed ICA feature expansion largely increases the detection performance.

Table 1. Test result of the pattern detection step

	Set A		Set B	
	Detect Rate	False Alarm	Detect Rate	False Alarm
Conventional*	97.2%	3	84.8%	127
Proposed**	98.5%	4	90.1 %	62

* : 50 ICA features from face class, ** : 100 ICA features from face and face-like class and 2 residual errors

The proposed integrated algorithm was also quantitatively evaluated using 7000 640×480 stills from color video sequences. The video sequences were collected under various and changing conditions. The average execution time of the integrated algorithm is 250ms on a Pentium IV 1GHz PC. Table 2 shows two results of sequential steps. The first is face candidate detection through the background removal, color and global appearance. The latter is face detection through the pattern matching and tracking. Final face detection rate is 91.2%. Some result images are shown in Figure 5.

Table 2. Test result of the integrated algorithm

	Detect Rate	False Alarms
Face candidate	97.66 %	
Face pattern	93.42 %	7

8. Conclusion

We have presented a novel method of using multi-modal cues for real-time detection and tracking of multiple faces in video sequence. Reliability of detection was guaranteed by using the ICA-SVM pattern matching and the tracking that was executed at the same time. The methods of background removal, adaptive skin color model, global appearance and the directional Kalman tracking was also proposed for efficient, real-time performance. Experiments on both still grey images and color videos show that the proposed method is sufficiently reliable and runs in real-time. Our future work will focus on multi-view face detection, which allows the surveillance system to detect more natural faces.

References

[1] T. Darrell, G. Gordon, M. Harville, J. Woodfill, "Integrated person tracking using stereo,color, and pattern detection", *CVPR*, pp. 601-609, Santa Barbara, June 1998.

[2] 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.

[3] Kah-Kay Sung and Tomaso Poggio, "Example-based learning for view-based human face detection", *IEEE Trans. on PAMI*, vol. 20, no. 1, pp. 39-51, Jan. 1998.

[4] Rein-Lien Hsu, Abdel-Mottaleb M. and Anil K. Jain, "Face detection in color images", *ICIP*, vol. 1, pp. 1046-1049, Greece, Oct. 2001.

[5] H.A. Rowley and T. Kanade, "Neural network-based face detection", *IEEE Trans. on PAMI*, vol. 20, no. 1, pp. 23-38, Jan. 1998.

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

[7] M.S. Bartlett and T.J. Sejnowski, "Independent component representations for face recognition", *SPIE Conf. On Human Vision and Electronic Imaging III*, vol. 3299, pp. 528-539, San Jose, Jan. 1998.

[8] K. Schwerdt and J.L. Crowley, "Robust face tracking using color", *FG2000*, pp.90-95, France, Mar. 2000.

Figure 5. Some results of our face detection algorithm