Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2007, Article ID 51648, 6 pages
doi:10.1155/2007/51648

Research Article

A Novel Face Segmentation Algorithm from a Video Sequence for Real-Time Face Recognition

R. Srikantaswamy[1] and R. D. Sudhaker Samuel[2]

[1] *Department of Electronics and Communication, Siddaganga Institute of Technology, Tumkur 572103, Karnataka, India*
[2] *Department of Electronics and Communication, Sri Jayachamarajendra College of Engineering, Mysore, India*

Received 1 September 2006; Accepted 14 April 2007

Recommended by Ebroul Izquierdo

The first step in an automatic face recognition system is to localize the face region in a cluttered background and carefully segment the face from each frame of a video sequence. In this paper, we propose a fast and efficient algorithm for segmenting a face suitable for recognition from a video sequence. The cluttered background is first subtracted from each frame, in the foreground regions, a coarse face region is found using skin colour. Then using a dynamic template matching approach the face is efficiently segmented. The proposed algorithm is fast and suitable for real-time video sequence. The algorithm is invariant to large scale and pose variation. The segmented face is then handed over to a recognition algorithm based on principal component analysis and linear discriminant analysis. The online face detection, segmentation, and recognition algorithms take an average of 0.06 second on a 3.2 GHz P4 machine.

1. INTRODUCTION

In literature, it is found that most of the face recognition work is carried out on still face images, which are carefully cropped and captured under well-controlled conditions. The first step in an automatic face recognition system is to localize the face region in a cluttered background and carefully segment the face from each frame of a video sequence. Various methods have been proposed in literature for face detection. Important techniques include template-matching, neural network based, feature-based, motion-based and face-space methods [1]. Though most of these techniques are efficient, they are computationally expensive for real time applications. Skin colour has proved to be a fast and robust cue for human face detection, localization, and tracking [2]. Skin colour based face detection and localization however has the following drawbacks: (a) it gives only a coarse face segmentation, (b) it gives spurious results when the background is cluttered with skin colour regions. Further, appearance based holistic approaches based on statistical pattern recognition tools such as principal component analysis and linear discriminant analysis provides a compact nonlocal representation of face images, based on the appearance of an image at a specific view. Hence, these algorithms can be regarded as picture recognition algorithm. Therefore, face presented for recognition to these approaches should be efficiently segmented, that is, aligned properly to achieve a good recognition rate. The shape of the face differs from person to person. Segmenting a face uniformly, invariant to shape and pose, suitable for recognition, in real-time is therefore very challenging. Thus, face segmentation "online" in "real-time" sense from a video sequence still emerges as a challenging problem in the successful implementation of a face recognition system. In this work, we have proposed a method which accommodates these practical situations to segment a face efficiently from a video sequence. The segmented face is then handed over to a recognition algorithm based on principal component analysis and linear discriminant analysis to recognize the person online.

2. BACKGROUND SCENE MODELING AND FOREGROUND REGION DETECTION

As the subject enters the scene, the cluttered background is first subtracted from each frame to identify the foreground regions. The system captures several frames in the absence of any foreground objects. Each point on the scene is associated with a mean and distribution about that mean.

This distribution is modeled as a Gaussian. This gives the background probability density function (PDF). A pixel $P(x, y)$ in the scene is classified as foreground if the Mahanalobis distance of the pixel $P(x, y)$ from the mean μ is greater than a set threshold. This threshold is found experimentally. Background PDF is updated using a simple adaptive filter [3]. The means for the succeeding frame is computed using (1), if the corresponding pixel is classified as a background pixel,

$$\mu_{t+1} = \alpha P_t + (1 - \alpha)\mu_t. \tag{1}$$

This allows compensating for changes in lighting conditions over a period of time. Where α is the rate at which the model is compensated for changes in lighting. For an indoor/office environment it was found that a single Gaussian model [4] of the background scene works reasonably well. Hence, a single Gaussian model of the background is used.

3. SKIN COLOUR MODELING

In the foreground regions, skin colour regions are detected. Segmentation of skin colour region becomes robust only if the chrominance component used in analysis and research has shown that skin colour is clustered in a small region of the chrominance plane [2]. Hence, the C_bC_r plane (chrominance plane) of the YC_bC_r colour space is used to build the model where Y corresponds to luminance and C_b-Cr corresponds to the chrominance plane. Skin colour distribution in the chrominance plane is modeled as a unimodal Gaussian [2]. A large data base of labelled skin pixels of several people both male and female has been used to build the Gaussian model. The mean and the covariance of the database characterize the model. Let $c = [C_b \;\; C_r]^T$ denote the chrominance vector of an input pixel. Then the probability that the given pixel lies in the skin distribution is given by

$$p(c \mid \text{skin}) = \frac{1}{2\pi\sqrt{\Sigma_s}} e^{-(1/2)(c-\mu_s)^T\Sigma_s^{-1}(c-\mu_s)}. \tag{2}$$

Here, c is a color vector, μ_s and Σ_s are the mean and covariance, respectively, of the distribution parameters. The model parameters are estimated from the training data by

$$\begin{aligned}\mu_s &= \frac{1}{n}\sum_{j=1}^{n} c_j, \\ \Sigma_s &= \frac{1}{n-1}\sum_{j=1}^{n}(c_j - \mu_s)(c_j - \mu_s)^T,\end{aligned} \tag{3}$$

where n is the total number of skin colour samples with colour vector c_j. The probability $p(c \mid \text{skin})$ can be used directly as a measure of how “skin-like” the pixel colour is. Alternately, the Mahalanobis distance λ_s, computed using (4), from the colour vector c to mean μ_s, given the covariance matrix Σ_s, can be used to classify a pixel as skin pixel [2],

$$\lambda_s(c) = (c - \mu_s)^T\Sigma_s^{-1}(c - \mu_s). \tag{4}$$

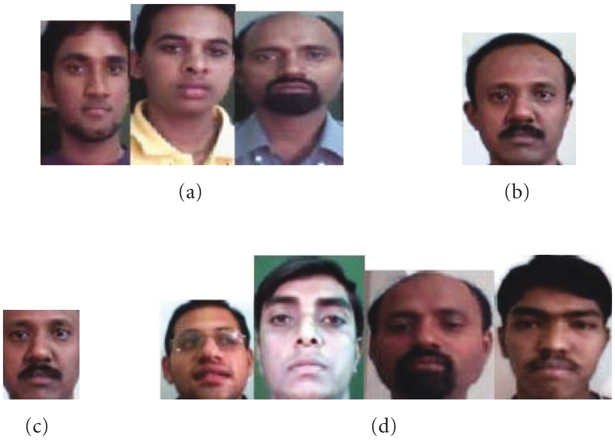


FIGURE 1: (a) Face segmented using skin colour regions (b) full face (c) closely cropped face (d) faces of various shapes.

Skin pixel classification may give rise to some false detection of nonskin tone pixels, which should be eliminated. A, iteration of erosion followed by dilation is applied on the binary image. Erosion removes small and thin isolated noise like components that have very low probability of representing a face. Dilation preserves the size of those components that were not removed during erosion.

4. DYNAMIC TEMPLATE MATCHING AND SEGMENTATION OF FACE REGION SUITABLE FOR RECOGNITION

Segmenting a face, using a rectangular window enclosing the skin tone cluster will result in segmentation of the face along with the neck region (see Figure 1(a)). Thus, skin colour based face segmentation provides only coarse face segmentation, and cannot be used directly for face recognition. The face presented for recognition can be a full face as shown in Figure 1(b) or closely cropped face which includes internal structures such as eye-brows, eyes, nose, lips, and chin region as shown in Figure 1(c). It can be seen from Figure 1(d) that the shape of the face differs from person to person. Here, we propose a fast and efficient approach for segmenting a face suitable for recognition.

Segmenting a closely cropped face requires finding a rectangle on the face image with the top left corner coordinates (x_1, y_1) and bottom right corner coordinates (x_2, y_2) as shown in Figure 2. The face region enclosed within this rectangle is then segmented.

From a database of about 1000 frontal face images created in our lab, a study on the relationship between the following facial features were made. (i) The ratio of distance between the two eyes W_E (extreme corner eye points, see Figure 3) to the width of the face W_F excluding the ear regions. (ii) The ratio of the distance between the two eyes W_E to the height of the face from the centre of the line joining two eyes to the chin H_F. It was found that the ratio W_E/W_F vary in the range 0.62–0.72 while the ratio H_F/W_E vary in the range 1.1–1.3.

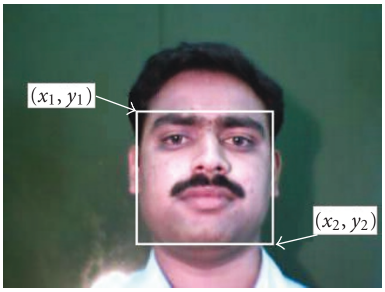


Figure 2: Rectangular boundary defining the face region.

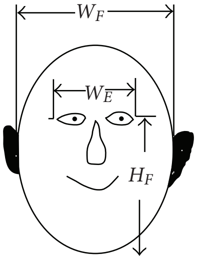


Figure 3: A sketch of face to define feature ratios.

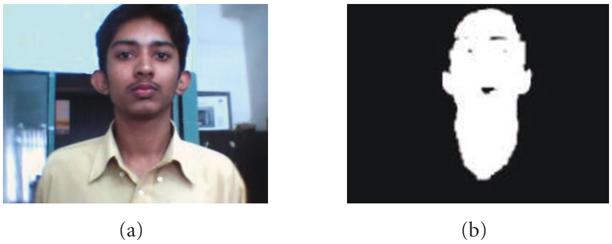


Figure 4: Subject with big ears and the corresponding skin cluster.

4.1. Pruning of ears

For some subjects, the ears may be big and extending outward prominently, while for other it may be less prominent. To obtain uniform face segmentation, the ear regions are first pruned. An example of the face with ears extending outward and its corresponding skin tone regions is shown in Figure 4.

The vertical projection of the skin tone regions of Figure 4(b) is obtained. The plot of this projection is shown in Figure 5. The columns which have skin pixels less than 20% of the height of the skin cluster are deleted. The result of this process is shown in Figure 6.

4.2. Rectangular boundary definitions x_1 and x_2

After the ears are pruned, the remaining skin tone regions are enclosed between two vertical lines as shown in Figure 6. The projection of left vertical (LV) and right vertical line (RV) on the x-axis gives x_1 and x_2, respectively, as shown in Figure 6. The distance between these two vertical lines gives the width of the face W_F.

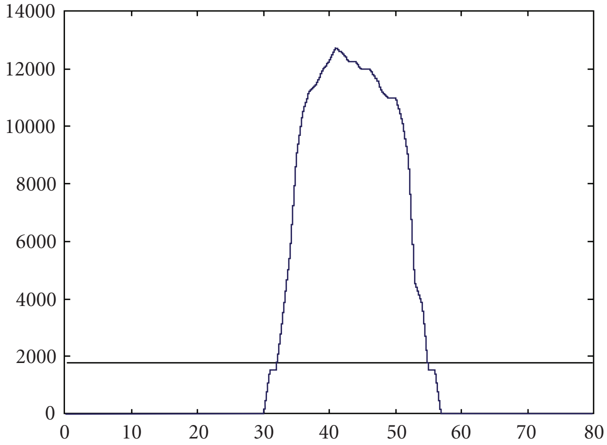


Figure 5: Vertical projection of Figure 4(b).

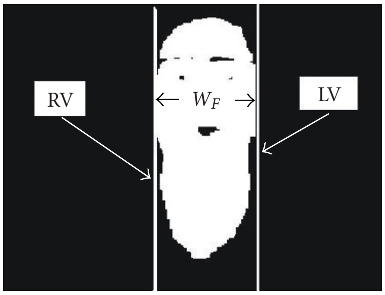


Figure 6: Skin tone cluster without ears.

4.3. Rectangular boundary definition y_1 and y_2

To find y_1, the eye brows and eye regions must be localized. Template matching is used to localize the eyes and eye brow regions. A good choice of the template containing eyes along with eyebrows should accommodate (i) variations in facial expressions, (ii) variations in structural components such as presence or absence of beard and moustache, and (iii) segmentation of faces under varying pose and scale by using a pair of eyes as one rigid object instead of individual eyes. Accordingly, a normalized average template containing eyes including eyebrows as shown in Figure 7 has been developed after considering several face images. The size of the face depends on its distance from the camera, and hence a template of fixed size cannot be used to localize the eyes. Here, we introduce a concept called dynamic template. After finding the width of the face W_F (see Figure 6), the width of the template containing eyes and eyebrows is resized proportional to the width of the face W_F keeping the same aspect ratio. The resized template whose width is proportional to the width of the face is what we call a dynamic template. As mentioned earlier, the ratio W_E/W_F vary in the range 0.62–0.72. Therefore, dynamic templates D_k with widths W_k are constructed, where W_k is given by

$$W_k = \gamma_k \times W_F \quad k = 1, 2, 3, \ldots, 6, \tag{5}$$

Figure 7: Template.

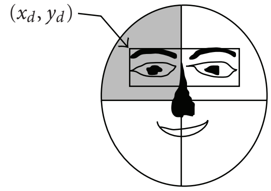


Figure 8: Four quadrants of skin tone regions.

where γ varies from 0.62 to 0.72 in steps of 0.02 keeping the same aspect ratio. Thus, six dynamic templates $D_1, D_2, \ldots, D_6$ with widths $W_1, W_2, \ldots, W_6$ are constructed. Let (x_d, y_d) be the top left corner coordinates of the dynamic template on the image as shown in Figure 8. Let $R_k(x_d, y_d)$ denote the correlation coefficient obtained by template matching when the top left corner of dynamic template D_k is at the image co-ordinates (x_d, y_d). The correlation coefficient R_k is computed by

$$R_k = \frac{\langle I_T D_k \rangle - \langle I_T \rangle \langle D_k \rangle}{\sigma(I_T)\sigma(D_k)}, \tag{6}$$

where I_T is the patch of the image I which must be matched to D_k, $\langle\rangle$ is the average operator, $I_T D_k$ represents the pixel by pixel product, and σ is the standard deviation over the area being matched. For real time requirements, (i) template matching is performed only within the upper left half region of the skin cluster (shaded region in Figure 8). (ii) The mean and the standard deviation of the template D_k is computed only once for a given frame. (iii) A lower resolution image of size 60 × 80 is used. However, segmentation of the face is made in the original higher resolution image. Let $R_{k_{\max}}(\tilde{x}_d, \tilde{y}_d)$ denote the maximum correlation obtained by template matching with the dynamic template D_k at the image coordinates $(\tilde{x}_d, \tilde{y}_d)$. Let R_{opt} denote the optimum correlation, that is, maximum of $R_{k_{\max}}$, $k = 1, 2, 3, \ldots, 6$ obtained with dynamic templates D_k, $k = 1, 2, 3, \ldots, 6$. Let W_k^* denote the width of the dynamic template D_k which give R_{opt}. The optimal correlation is given by

$$R_{\text{opt}}(x^*, y^*) = \max R_{k_{\max}}(\tilde{x}_d, \tilde{y}_d) \quad k = 1, 2, \ldots, 6, \tag{7}$$

where (x^*, y^*) is the image coordinates which give R_{opt}. If R_{opt} is less than a set threshold, the current frame is discarded and the next frame is processed. Thus, the required point on the image y_1 is then given by

$$y_1 = y^*. \tag{8}$$

The distance between the two eyes W_E^* is given by the width of the optimal dynamic template which give R_{opt}, therefore $W_E^* = W_k^*$.

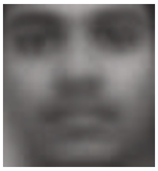


Figure 9: Average face template.

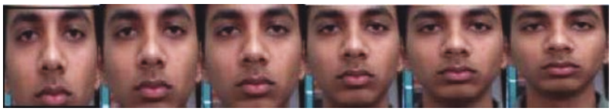


Figure 10: Some samples of segmented faces with different values.

After finding x_1, y_1, and x_2, we now need to estimate y_2. As mentioned earlier, the height of the face varies form person to person and the ratio H_F/W_E vary in the range 1.1–1.3. Several face images, about 450, were manually cropped from images captured in our lab and an average of all these face images forms an average face template as shown in Figure 9. The centre point $(x_{\text{cen}}, y_{\text{cen}})$ between the two eyes is found by the centre of the optimal dynamic template. From this centre point, height of the face H_{F_k} is computed by

$$H_{F_k} = (1.1 + \beta) \times W_E^*, \quad k = 1, 2, \ldots, 10, \tag{9}$$

where β is a constant which varies from 0 to 0.2 in steps of 0.02. The face regions enclosed within the boundary of the rectangle formed using the coordinates x_1, y_1, x_2 and the heights $H_{F_k}(k = 1, 2, \ldots, 10)$ are segmented and normalized to the size of the average face template. Some of the faces segmented and normalized by this process are shown in Figure 10. Correlation coefficient ∂_k, $k = 1, 2, \ldots, 10$ with these segmented faces and the average face template is given by (10),

$$\partial_k = \frac{\langle I_{\text{seg}}\text{AF}\rangle \langle I_{\text{seg}}\rangle \langle A_F\rangle}{\sigma(I_{\text{seg}})\sigma(\text{AF})}, \tag{10}$$

where I_{seg} is segmented and normalized face images, AF is the average face template as shown in Figure 9, $\langle\rangle$ is the average operator, $I_{\text{seg}}\text{AF}$ represents the pixel by pixel product, and σ is the standard deviation over the area being matched. A plot of correlation coefficient ∂_k versus H_F is shown in Figure 11. For real-time requirement, the mean and the variance of the average face template are computed ahead of time and used as constants for the computation of the correlation coefficient ∂_k.

The Height (number of pixels) of the face H_{F_k} corresponding to the maximum correlation coefficient $\partial_{\max} = \max(\partial_k)$, $k = 1, 2, \ldots, 10$ is added to the y-coordinates of the centre point between the two eyes to obtain y_2. Finally, the face region enclosed within the boundary of the rectangle formed using the coordinates (x_1, y_1) and (x_2, y_2) is segmented. The results of the proposed face detection and segmentation approach are shown in Figure 12.

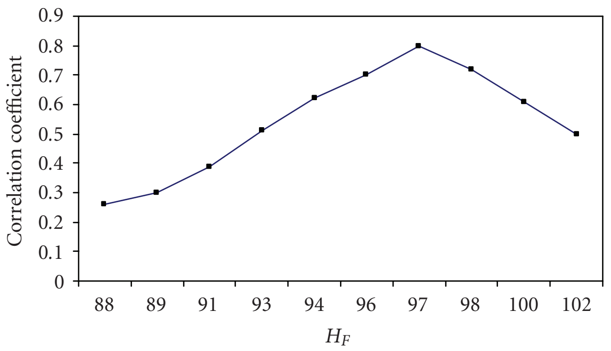
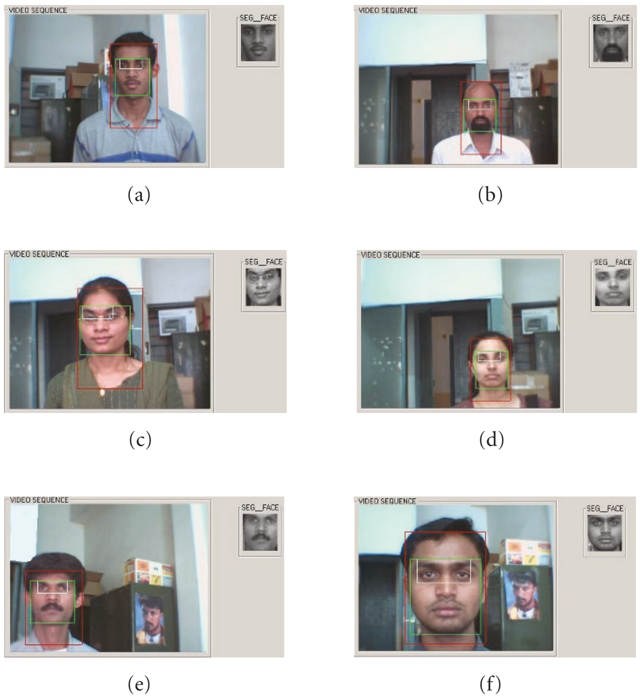


FIGURE 11: Plot of correlation coefficient of H_{F_k} normalized to the same size ∂_kversus H_F.



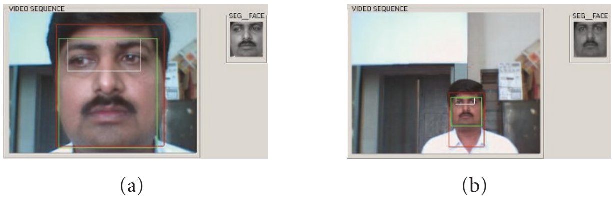
(a) (b) (c) (d) (e) (f)

FIGURE 12: Results of face segmentation using the proposed method.

The segmented face is displayed at the top right corner window labeled SEG_FACE of each frame. Observe that the background is cluttered with a photo of a face in it. The red rectangle indicates the coarse face localization based on skin colour. The white rectangle indicates the localization of two eyes including the eye brows. The green rectangle indicates the face regions to be segmented using the proposed method.

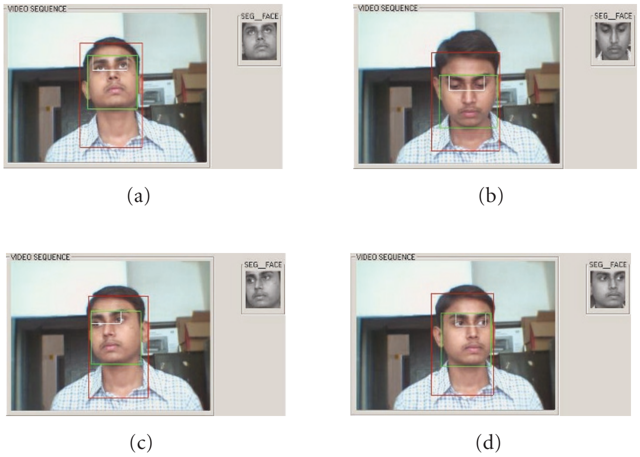
4.4. *Face segmentation with scale and pose variations*

The result of the face segmentation with scale variations is as shown in Figure 13. It can be observed that the proposed face segmentation is invariant to large scale variations.



(a) (b)

FIGURE 13: Largest and smallest face images segmented by the proposed method.



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

FIGURE 14: Result of face segmentation with pose variations.

The smallest face that can be segmented by the proposed method is 3.5% of the frame size as shown in Figure 13(b). However, the largest face that can be segmented depends on the size of the full face that can be captured when the subject is very close to the camera. The results of face segmentation with pose variations are shown in Figure 14.

5. FEATURE EXTRACTION

After the face is segmented, features are extracted. Principal component analysis (PCA) is a standard technique used to approximate the original data with lower dimensional feature vector. The basic approach is to compute the eigenvectors of the covariance matrix and approximate the original data by a linear combination of the leading eigenvectors [5]. The features extracted by PCA may not be necessarily good for discriminating among classes defined by a set of samples. On the other hand, LDA produces an optimal linear discriminant function which maps the input into the classification space which is well suitable for classification purpose [6].

6. EXPERIMENTAL RESULTS

A data base of 450 images of 50 individuals consisting of 9 images of each individual with pose, lighting, and expression

Table 1: Recognition rate of the online face recognition system.

Recognition rate of the online face recognition system	
PCA features	LDA features
90%	98%

variations captured in our lab was used for training the face recognition algorithm. The result of the online face recognition system using the proposed face segmentation algorithm is shown in Table 1. The entire algorithm for face detection, segmentation, and recognition is implemented in C++ on a 3.2 GHz P4 machine which takes an average of 0.06 seconds per frame to localize, segment, and recognize a face. The face localization and segmentation stage takes an average of 0.04 seconds. The face recognition stage takes 0.02 seconds to recognize a segmented face. The face segmentation algorithm is tolerant to pose variations of ± 30 degrees of pan and tilt on an average. The recognition algorithm is tolerant to pose variations of ± 20 degrees of pan and tilt.

7. CONCLUSION

We have been able to develop an online face recognition system which captures image sequence from a camera, detects, tracks, segments efficiently, and recognizes a face. A method for efficient face segmentation suitable for real-time application, invariant to scale and pose variations is proposed. With the proposed face segmentation approach followed by linear discriminant analysis for feature extraction from the segmented face, a recognition rate of 98% was achieved. Further LDA features provide better recognition accuracy compared to PCA features.

REFERENCES

[1] M.-H. Yang, D. J. Kriegman, and N. Ahuja, "Detecting faces in images: a survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34–58, 2002.

[2] V. Vezhnevets, V. Sazonov, and A. Andreeva, "A survey on pixel-based skin color detection techniques," in *Proceedings of the International Conference on Computer Graphics (GRAPH-ICON '03)*, pp. 85–92, Moscow, Russia, September 2003.

[3] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "Pfinder: real-time tracking of the human body," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 780–785, 1997.

[4] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '99)*, vol. 2, pp. 246–252, Fort Collins, Colo, USA, June 1999.

[5] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[6] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," in *Proceedings of the 4th European Conference on Computer Vision (ECCV '96)*, vol. 1, pp. 45–58, Cambridge, UK, April 1996.

R. Srikantaswamy received his M.Tech degree in industrial electronics in 1995 and Ph.D. degree in electronics in 2006 from University of Mysore, India. He is working as a Professor in the Department of Electronics and Communication, Siddaganga Institute of Technology, Tumkur, India. His research interests include computer vision and pattern recognition, neural networks, and image processing.

R. D. Sudhaker Samuel received his M.Tech degree in industrial electronics in 1986 from the University of Mysore, and his Ph.D. degree in computer science and automation (robotics) in 1995 from Indian Institute of Science, Bangalore, India. He is working as a Professor and Head of the Department of Electronics and Communication, Sri Jayachamarajendra College of Engineering, Mysore, India. His research interests include industrial automation, VLSI design, robotics, embedded systems, and biometrics.