Journal of Visualization, Vol. 9, No. 3 (2006)331-338

A Real Time Implementation of a Pose Invariant Face Recognition Engine with a Novel Face Segmentation Algorithm

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

*1 Sri Jayachamarajendra College of Engineering, Mysore, India. E-mail: rsswamy@rediffmail.com
*2 Sri Jayachamarajendra College of Engineering, Mysore, India.

Received 23 December 2004
Revised 23 January 2006

Abstract : In this paper we propose a fast and efficient algorithm for segmenting a face suitable for recognition from a video sequence. We first obtain a coarse face region using skin color, then using dynamic template matching the face is efficiently segmented at varying scale and pose in real time. We have also developed and tested some heuristics which localizes only the face region, even when subjects are wearing skin color dress. The segmented face is then handed over to a recognition algorithm based on principal component analysis and linear discriminant analysis. The on-line face detection, segmentation and recognition algorithm takes an average of 0.06 sec on a 3.2 GHz P4 machine.

Keywords : Face detection, Face recognition, Principal Component Analysis, Linear Discriminant Analysis.

1. Introduction

In literature it is found that, most of the face recognition work is carried our 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 (Ming Hsuan et al., 2002). Though most of these techniques are efficient they are computationally expensive for real time applications.

Skin color has proved to be a fast and robust cue for human face detection, localization and tracking (Vezhnevets V et al., 2003). Skin color based face detection and localization however have the following drawbacks: a) it gives only a coarse face segmentation, b) it gives spurious results when the background is cluttered with skin color regions, c) it is unsuitable when the subjects are wearing skin color dress or has patches of skin color on the dress. Appearance based Holistic approaches based on statistical pattern recognition tools such as principal component analysis and linear discriminant analysis provides a compact non-local 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 i.e., aligned properly to achieve a good recognition rate. The shape of the face differs from person to person. Faces may be long, round, square or even triangular and the size of the

face may vary depending on the distance of the person from the camera. Segmenting a face uniformly, invariant to shape and pose, suitable for recognition, in real time is therefore very challenging. Thus face segmentation 'on-line' 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 handed over to a recognition algorithm based on principal component analysis and linear discriminant analysis to recognize the person on-line.

2. Background Scene Modeling and Foreground Region Detection

For an indoor/office environment it was found that a single Gaussian model (Grimson et al., 1999) of the background scene works reasonably well. Hence a single Gaussian model of the background is used. 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. Background PDF is updated using a simple adaptive filter (Wern et al., 1997). The means for the succeeding frames is computed using Eq. (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 the changes in lighting.

3. Skin Color Modeling

In the foreground regions skin color regions are detected. Segmentation of skin color region becomes robust only if the chrominance component is used in analysis and research has shown that skin color is clustered in a small region of the chrominance plane (Vezhnevets V et al., 2003). Hence, the C_bC_r plane (chrominance plane) of the YC_bC_r color space is used to build the model where Y corresponds to luminance and C_bC_r corresponds to the chrominance plane. Skin color distribution in the chrominance plane is modeled as a unimodal Gaussian (Vezhnevets V et al., 2003). A large database of labeled 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.

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

Here, c is a color vector, and 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:

$$\mu_s = \frac{1}{n}\sum_{j=1}^{n} c_j \tag{3}$$

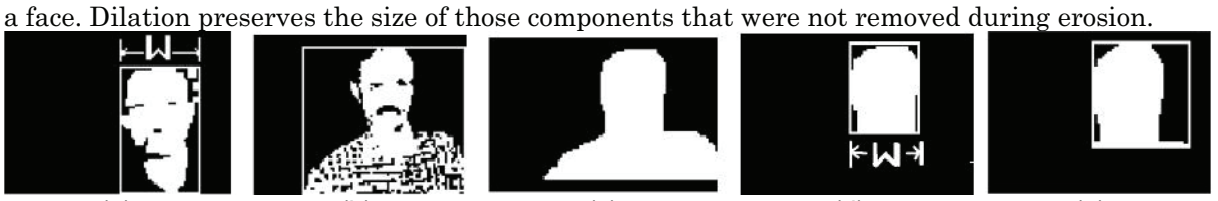
$$\Sigma_s = \frac{1}{n-1}\sum_{j=1}^{n} (c_j - \mu_s)(c_j - \mu_s)^T \tag{4}$$

Where n is the total umber of skin color samples with color vector c_j. The probability $p(c \mid skin)$ can be used directly as a measure of how "skin-like" the pixel color is. Alternately, the Mahalanobis distance λs computed using Eq. (5), from the color vector c to mean μ_s, given the covariance matrix Σ_s can be used to classify a pixel as a skin pixel (Vezhnevets V et al., 2003).

$$\lambda_s(c) = (c - \mu_s)\sum_s^{-1}(c - \mu_s) \quad (5)$$

4. Connected Component Analysis and Course Face Region Extraction

Skin pixel classification may give rise to false detection of non-skin tone pixels, which should be eliminated. An 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.



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

Fig. 1. Extraction of Face region.

Now a shape-based connected operator such as *solidity* is used. *Solidity* of the connected component is defined as the ratio of its area to the area of the min-max box (rectangular box) enclosing the skin cluster.

$$Solidity = \frac{A}{D_x D_y} \quad (6)$$

Where, A is the area of the connected component inside the rectangular min-max box, while D_x and D_y are the width and height of the rectangular box respectively. If the solidity is more than some set threshold, the cluster may be said to contain only the face regions. Figure 1(a) shows the binary image of skin pixel classification, in this case the solidity is more than a set threshold. However problems arise when a person is wearing skin color dress or if the dress contains some patches of skin color. This situation is shown in Fig. 1(b) where the person is wearing a shirt with skin tone pixels, these skin tone pixels on the shirt also gets classified as face region. When the subjects are wearing skin color dress the solidity of the skin cluster will be less then the set threshold (see Fig. 1(b)). Yang et al. (1998) have proposed to merge the skin regions, fit an ellipse and test its fitness each time a region is merged. Fitting an ellipse and testing its fitness when every skin pixel is merged is not suitable for real time as it is computationally expensive. Here we propose a method to extract only the head region when the subject is wearing a skin color dress. If the solidity of the skin tone cluster is less than the set threshold, then the following steps are performed to extract only the face region:

1) The skin tone regions with areas less than a set threshold (i,e., less than the size of smallest face to be detected in a given frame) are first eliminated.
2) The remaining region of the skin tone pixels is processed with vertical morphological filter. This increases the solidity of the face region pixels, filling holes and regions around the neck. The result of this step is as shown in Fig. 1(c).
3) Region merging is then performed from the top (tip of the head) of the skin color cluster downwards until the *solidity* of the merged pixel is more than the set threshold. The result of this step is as shown in Fig. 1(d). If region merging is continued further down solidity reduces (see Fig. 1(e)).

5. 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 also (see Fig. 2(a)). Thus, skin color based segmentation provides only coarse face segmentation, and cannot be used directly for recognition. The face presented for recognition can be a full-face as shown in Fig. 2(b) or closely cropped face

which includes internal structures such as eye-brows, eyes, nose, lips, and chin region as shown in Fig. 2(c). It can be seen from Fig. 2(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.

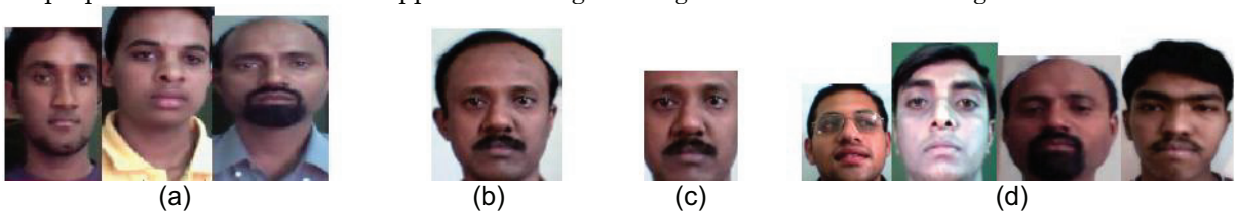


Fig. 2. (a) Face segmented using skin color regions, (b) Full face, (c) Closely cropped face, (d) Faces of various shape.

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 Fig. 3. The face region enclosed within this rectangle is then segmented.

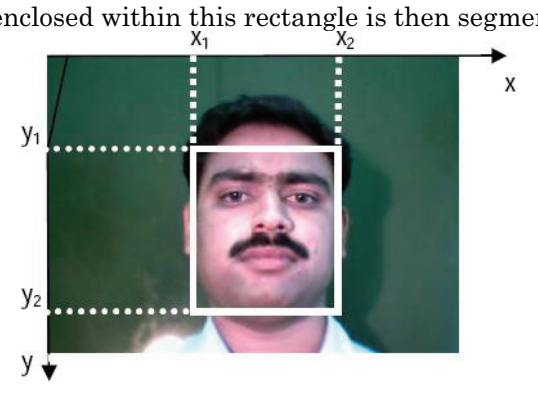


Fig. 3. Rectangular boundary defining the face region.

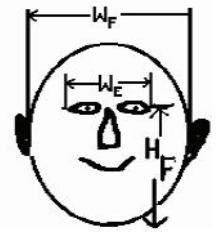


Fig. 4. Sketch of a face to define feature ratios.

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 to eyes W_E (extreme corner eye points, (see Fig. 4) 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 center 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.

5.1 Pruning of Ears

For some subjects the ears may be big and extending outwards prominently, while for others 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 region is shown in Fig. 5.

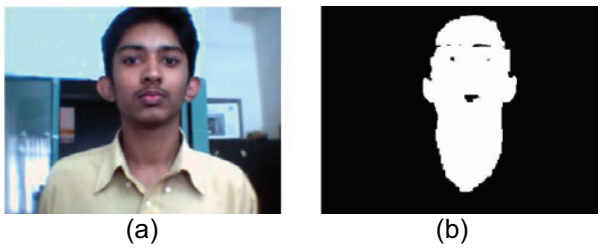


Fig. 5. Subject with big ears and corresponding skin cluster.

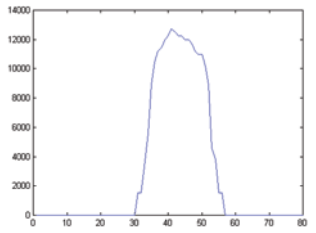


Fig. 6. Vertical projection of Fig. 5(b).

The vertical projection of the skin tone regions of Fig. 5(b) is obtained. The plot of this projection is shown in Fig. 6. 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 Fig. 7.

5.2 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 Fig. 7. The projection of left vertical (LV) and right (RV) on to the x-axis gives x_1 and x_2 respectively as shown in Fig. 7. The distance between these two vertical lines gives the width of the face W_F.

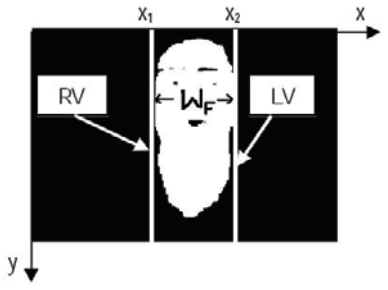


Fig. 7. Skin tone cluster without ears.

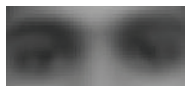


Fig. 8. Template.

5.3 Rectangular Boundary Definitions 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 Fig. 8 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 Fig. 7), 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:

$$W_k = \gamma \times W_F \qquad k = 1, 2, 3.\ldots\ldots 6 \tag{7}$$

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 Fig. 9. 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

$$R_k = \frac{\langle I_T D_k \rangle - \langle I_T \rangle \langle D_k \rangle}{\sigma(I_T)\sigma(D_k)} \tag{8}$$

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 Fig. 9). 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 x 80 is used. However segmentation of the face is made in the original higher resolution image. Let $R_k(\tilde{x}_d, \tilde{y}_d)$ 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 i.e., 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

$$R_{opt}(x^*, y^*) = \max(R_{k\max}(\tilde{x}_d, \tilde{y}_d)) \qquad k = 1,2,\ldots\ldots 6 \tag{9}$$

Where (x^*, y^*) is the image coordinates which give R_{opt}. If R_{op} is less then a set threshold, the current

frame is discarded and the next frame is processed. Thus the required point on the image y_1 is given by

$$y_1 = y^* \tag{10}$$

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

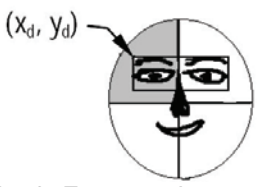


Fig. 9. Four quadrants of skin regions.

Fig. 10. Average face template.

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 from person to person and the ratio H_F/H_E varies 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 Fig. 10. The center point (x_{cen}, y_{cen}) between the two eyes is found by the center of the optimal dynamic template. From this centre point, height of the face H_E is computed by

$$H_{F_k} = (1.1 + \beta) \times W_E^* \qquad k = 1,2,....,10 \tag{11}$$

Where, β is a constant which vary 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_E (k = 1,2,....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 Fig. 11. Correlation coefficient ∂_k, k = 1,2,.....10 with these segmented face and average face template is given by Eq. (12).

$$\partial_k = \frac{\langle I_{seg} AF \rangle - \langle I_{seg} \rangle \langle AF \rangle}{\sigma(I_{seg})\sigma(AF)} \tag{12}$$

Where, I_{seg} is the segmented and normalized face images, AF the average face template as shown in Fig. 11, $\langle\rangle$ the average operator, $I_{seg}AF$ 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 Fig. 12. For real time requirements the mean variance of the average face template are computed ahead of time and used as constants for the computation of the correlation coefficient ∂_k.

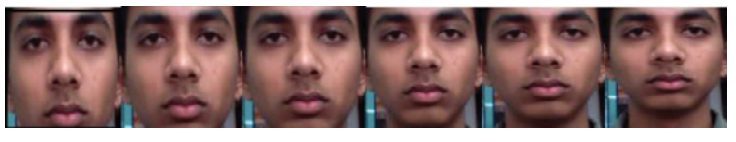


Fig. 11. Some samples of segmented faces with different values of coefficient H_{Fk} normalised to same size.

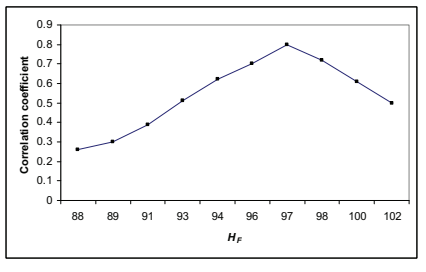


Fig. 12. Plot of correlation ∂_k versus H_F.

The height (number of pixels) of the face H_E corresponding to the maximum correlation coefficient $\partial_{\max} = \max(\partial_k)$ k = 1,2,.....10 is added to the y-coordinates of the centre point between the two eyes in odder 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 Fig. 13. The segmented face is displayed at the top right corner window labelled 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 the two eyes including the eye-brows. The green rectangle indicates the face regions to be segmented using the proposed method.

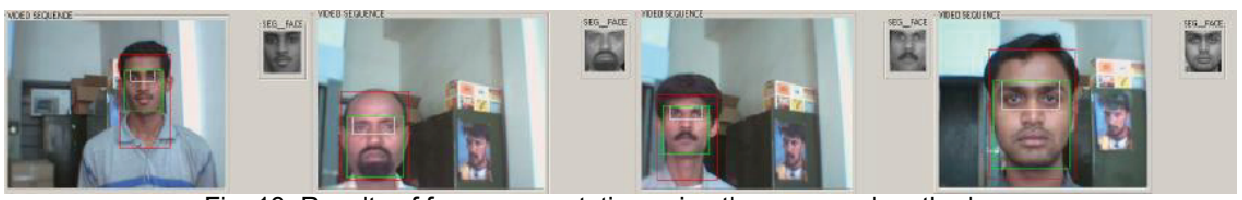
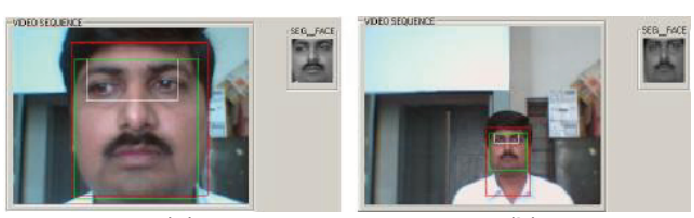


Fig. 13. Results of face segmentation using the proposed method.

5.4 Face Segmentation with Scale and Pose Variations

The result of the face segmentation with scale variations is as shown in Fig. 14. It can be observed that the proposed face segmentation is invariant to large scale variations. The smallest face that can be segmented by the proposed method is 3.5 % of the frame size as shown in Fig. 14(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 the face segmentation with pose variation are as shown in Fig. 15.



(a) (b)

Fig. 14. Largest and smallest face images segmented by the proposed method.

Fig. 15. Result of face segmentation with pose variations.

6. 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 (Turk et al., 1991). The features extracted by PCA may not be necessarily good for discriminating among classed 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 (Belhumer et al., 1996).

7. Experimental Results

A data base of 450 images of 50 individuals consisting of 9 images of each individual with pose, lighting and expression variations captured in our lab was used for training the face recognition algorithm. The result of the on-line 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.

Table 1. Recognition rate of the on-line face recognition system.

Recognition rate of the on-line face recognition system	
PCA features	LDA features
90%	98%

8. Conclusion

We have been able to develop an on-line face recognition system which captures image sequence from a camera, detects, tracks, segments efficiently and recognize 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 form the segmented face a recognition rate of 98 % was achieved. Further LDA features provide better recognition accuracy compared to PCA features.

References

Belhumer., B., Hespanha., J. and Kriegman., D., Eigen Vs Fisherfaces: Recognition using class specific linear projection, Proceeding of Fourth European conference on Computer Vision, ECCV'96, (1996-4), 45-56.

Grimson. W. E. L. and Stauffer., C., Adaptive background mixture models for real-time tracking, Computer Vision and Pattern Recognition, 2 (1999-6), Fort Collins, Colorado.

Ming, Hsuan, David J., Kriegman, and Narendra, Ahuja, Detecting faces in images: A Survey, IEEE transactions on Pattern analysis and machine intelligence 24-1 (2002-1).

Srikantaswamy, R. and Sudhaker Samuel, R. D., A Real Time Face recognition Engine with a Novel Face Segmentation Approach, Proceedings of International conference on Robotics, Vision, Image, Signal Processing, (2005-7), University Sains Malaysia.

Turk. M and Pentland. A, Eigenfaces for recognition, Journal of Cognitive Neuroscience, (1991-3).

Vezhnevets, V., Sazonov, V. and Andreeva, A., survey on Pixel-based Skin colour detection Techniques, Proc. Graphicon-2003, (Moscow, Russia), (2003-9), 85-92.

Wren.C, Azabayejani. A, Darrell. T and Pentland. A, Pinder: Real-time tracking of the human body, IEEE Transactions on Pattern Analysis and Machine Intelligence,(1997), 780-785.

Yang. Y. and Ahuja, Detecting Human faces in colour images, International Conference on Image Processing, (1998), 127-130.

Author Profile

R. Srikantaswamy: He received his M.Tech in Industrial Electronics in 1995 form University of Mysore, India. He is working as an Assistant Professor in the Department of Electronics and Communication. His research interests include Computer vision and Pattern Recognition, Neural networks and Image Processing.

R. D. Sudhaker Samuel: He received his M.Tech degree in Industrial Electronics in 1986 from the University of Mysore, India and his Ph.D. in Computer Science and Automation-Robotics in 1995 from Indian Institute of Science, Bangalore, India. He is working as a Professor and the Head in the Department of Electronics and Communication, Sri Jayachamarajendra of Engineering, Mysore, India. His research interests include Industrial Automation, VLSI design, Robotics, Embedded systems and Biometrics