International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:6, No:3, 2012

Facial Expressions Recognition from Complex Background using Face Context and Adaptively Weighted sub-Pattern PCA

Md. Zahangir Alom, Mei-Lan Piao, Md. Ashraful Alam, Nam Kim, Jae-Hyeung Park

Abstract**—**A new approach for facial expressions recognition based on face context and adaptively weighted sub-pattern PCA (Aw-SpPCA) has been presented in this paper. The facial region and others part of the body have been segmented from the complex environment based on skin color model. An algorithm has been proposed to accurate detection of face region from the segmented image based on constant ratio of height and width of face (δ= 1.618). The paper also discusses on new concept to detect the eye and mouth position. The desired part of the face has been cropped to analysis the expression of a person. Unlike PCA based on a whole image pattern, Aw-SpPCA operates directly on its sub patterns partitioned from an original whole pattern and separately extracts features from them. Aw-SpPCA can adaptively compute the contributions of each part and a classification task in order to enhance the robustness to both expression and illumination variations. Experiments on single standard face with five types of facial expression database shows that the proposed method is competitive.

Keywords**—**Aw-SpPC, Expressoin Recognition, Face context, Face Detection, PCA

I. Introduction

HUMAN being uses much more input information than the spoken words during a conversation with another human: the ears to hear the words and the tone of the voice, the eyes to recognize movements of the body and facial muscles, the nose to smell where somebody has been, and the skin to recognize physical contact. In the following we will concentrate on facial expressions. Facial expressions are not only emotional states of a user but also internal states affecting his/her interaction with a dialogue system, e.g. helplessness, disgust, anger, unhappy and neutral. If a system wants to know about the user's internal state by observing the face, it first has to localize the face and then recognize the facial expression. Face localization aims to determine the image position of a single face. The literature shows various methods. In [1] a combination of skin color and luminance is used to find the face in a head-shoulder image. A combination of facial components like eyes and nostrils found by SVMs and their geometric relation is used in [2]. They used this method to detect faces in frontal and near-frontal views of still grey level images. A probabilistic face detection method for faces of different pose, with different expression and under different lighting conditions is the mixture of factor analyzers used by [3].

Authors are with the School of Electrical & Computer Engineering Department, Chungbuk National University, Cheongju, South Korea. phone: +82-43-276-9957; fax: 043-274-6206; e-mail: zahan.cse@gmail.com,{mrpark,alam}@osp.chungbuk.ac.kr,namkim@chungbuk.ac.kr, jh.park@cbnu.ac.kr.

Only color information is used by [4] to form a statistical model for person detection in web images. The task of facial expression recognition is to determine the emotional state of a person. A common method is to identify facial action units (AU). These AU were defined by Paul Ekman in [5]. In [6] a neural-network is used to recognize AU from the coordinates of facial features like lip corners or the curve of eye brows. In this paper we are proposing a universal and golden ratio based face and facial action unit's detection (FAUD). As a classical self-organized learning method, principle component analysis (PCA) is widely used in the field of data compression and feature extraction [6]. There are two basic approaches to the computation of principal components: batch and adaptive methods. The batch methods include the method of eigen decomposition and the method of singular value decomposition (SVD), while the adaptive methods are exemplified by Hebbian-based neural networks, such as generalized Hebbian algorithm (GHA) and adaptive principal components extraction (APEX) etc. [1,5]. Despite these different implementations of PCA, their essences are the same, namely, to explain the variance–covariance structure of the data through a few linear combinations of the original variables. Currently, PCA has also become one of the most popular appearance-based algorithms applied to face recognition [7-10]. However, due to utilizing only the global information of face images, this method is not very effective under different facial expression, illumination condition and pose, etc. In this paper, Aw-SpPCA algorithm has been used that is base on the concepts of confine illumination conditions, facial expressions variations to local areas Divide a face image into several sub-images, and carry out PCA computation on each local area independently. And also to emphasize different parts of human face has different discrimination capabilities adaptively compute the contribution factor of each local area, and incorporate contribution factor into final classification decision.The rest of this paper is organized as follows. Section II presents the skin color detection system and facial features extraction like mouth and eye. Section III describes on Aw-SpPCA algorithm. Experimental results are given in Section IV. Conclusion and future direction of this work is given in Section V.

II. Skin color detection and features extraction

It has been proposed in the method, only pixels that are significant for facial expressions recognition are used to create an eigenspace. These significant pixels are selected automatically by a training set of face images showing facial expressions. There is no assumption about the spatial relation of these pixels in contrast to [15] where only an oval region of the face is used to omit background and hair. The procedures to detect the facial features are given in the following sections.

International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:6, No:3, 2012

A. Implementation of Proposed Method

The detection system based on color and feature is faster and accurately find human, many researchers combine these two methods to obtain real human face in a picture. However, the traditional color-based method is hard to detect the skin-color for the case of different lighting condition, and the typical feature-based method has high computation complexity. In this section, we propose a new lighting compensation scheme to overcome the problem of color-based method and simplify the feature-based detection. A system overview of our face detection algorithm is illustrated in Fig. 2, and the details are explained as follows

1) Color We adopt skin-color detection in order to apply the real-time system as the first step of face detection. Due to YCbCr color space transform is faster than other approaches [5][6], this transform has been select to detect human skin. However, the luminance of every image is different and it results that every image has different color distribution. Therefore, the proposed system lighting compensation is based on luminance to modulate the range of skin-color distribution. First, we compute the average luminance Y_{aveg} of input image.

$$Y_{aveg} = \sum Y_{i,j} \tag{1}$$

Where $Y_{i,j} = 0.3R + 0.6G + 0.1B$, is normalized to the range (0,255) and i, j are the index of pixel. According to Y_{aveg}, we can determine the compensated image $C_{i,j}$ by following equations:

$$R'_{ij} = (R_{ij})^{T} \tag{2}$$

$$G'_{ij} = (G_{ij})^{T} \tag{3}$$

$$C_{ij} = \{R'_{ij}, G'_{ij}, B_{ij}\} \tag{4}$$

where

$$\tau = \begin{cases} 1.4, & Y_{aveg} < 64 \\ 0.6, & Y_{aveg} > 192 \\ 1, & otherwise \end{cases} \tag{5}$$

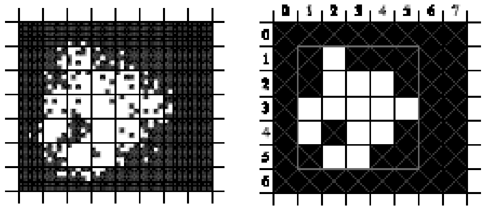
Note that we only compensate the color of R and G to reduce computation. Due to chrominance C_r can represent human skin well, we only consider C_r factor for color space transform to reduce the computation. C_r is defined as follow:

$$C_r = 0.5R' - 0.419G' - 0.081B \tag{6}$$

In Eq. (5) we can see that R and G are important factors due to their high weight. Thus, we only compensate R and G to reduce computation. According to C_r and experimental experience, we define the human skin by a binary matrix:

$$S_{ij} = \begin{cases} 0, & 10 < C_r < 45 \\ 1, & otherwise \end{cases} \tag{7}$$

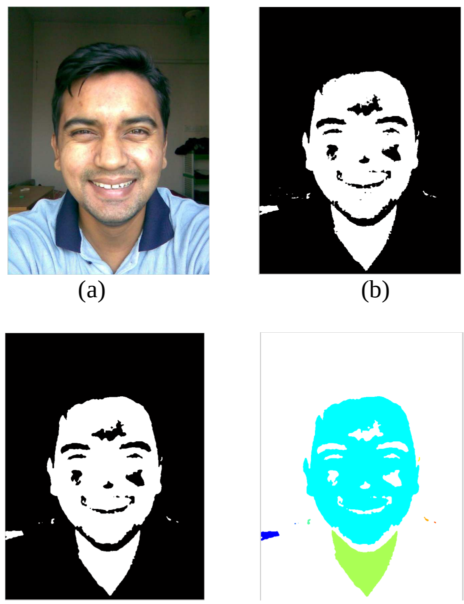
where "0" is the white point, and "1" is black point. Fig. 2 shows the compensation effect on bright and dark image respectively. We can see that result shown in Fig. 2(d) is good enough for skin-color detection.



(a) (b)

Fig. 1 Mask for removing noise

2) High frequency noisy removing: To remove high frequency noise fast, we have applied a low pass filter by a 5 × 5 mask. First, we segment S_{ij} into 5×5 blocks, and calculate how many white points in a block. Then, every point of a 5 × 5 block is set to white point when the number of white points is greater than half number of total points. On the other hand, if the number of black points is more than a half, this 5 × 5 block is modified to a complete black. Fig. 1(b) shows an example that we remove high frequency noise from Fig. 1(a). Although this fast filter will bring block effect, it can be disregarded due to that our target is to find where human skin is?



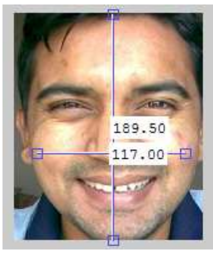
(a) (b)

(c) (d)

Fig. 2 a) Lighting compensation b) face skin detection c) skin with noise removal d) face candidate

3) Find out the skin-color blocks There are several skin color regions may be human face will be in S_{ij} after performing the low pass filter. In order to mark those regions, we stored for vertices of rectangle for every region. First, we find the leftmost, rightmost, upmost, and down most points. By these four points, we create a rectangle around this region. Fig. 2(c) shows an example after applying that store (1, 1), (1, 5), (5, 1), and (5, 5) to describe the candidate region. Thus, we can get several skin-color blocks called candidate blocks to detect facial feature.

4) Height and Width ration checking After the step of face localization, we can get several regions which may be human face. Then, the feature of height to width ratio, mouth, and eyes are detected sequentially for every candidate block. Because any of these three features detection can reject the candidate blocks, low computation module has high priority to process. Height to width ratio is a very fast and simple detection. Let the size of candidate block is *hxw*. We have calculated the distance between the ending position of the height and width of faces. We define that if the height to width ratio $(h:w)$ is exactly 1.618 or nearest to the mentioned value then it is face else it should be not a face. The candidate block from the segmented image is discarded. Note that the range is determined by experiments. If the ratio is between 1.55 and 1.7 is considered as face region.



(a)

Fig. 3 Ratio of face features (δ = h/w =1.618)

5) Mouth Detection After determining the height to width ratio for the candidate blocks, a more complex detection will be applied to find mouth feature. We use θ proposed by [7] to find the mouth pixels. The θ value is calculated for all of the pixels of the candidate blocks. The θ is defined as:

$$\theta = \cos^{-1}\left(\frac{0.5(R' - G' - B)}{\sqrt{(R' - G') + (R' - B)(G' - B)}}\right) \tag{8}$$

The pixel will be determined to be part of mouth by a binary matrix M:

$$M_{pq} = \begin{cases} 0, \ \theta < 90 \\ 1, \ otherwise, \end{cases} \tag{9}$$

where “0” means that pixel is mouth. Fig. 4(a) and (b) is an example for mouth pixel detection. In Fig. 4(b), the mouth pixel is presented by white point. Then, we use vertical based histogram to determine whether or not it is a mouth in this block. We calculate how many mouth pixels are in the same y-coordinates, and use w_h to store the value of different y-coordinates. Fig. 4(c) illustrates an example of the histogram of Fig. 4(b). Note that the maximum value of w_h is denoted by w_{max} and the y-coordinate of w_{max} is represented by h_m. Thus, we define if w_{max} is less than 1/6 block width w, this block will be rejected. For example, in Fig. 4(c) w_{max} is more than (1/3) w, we can know that the mouth feature is embedded in this block.

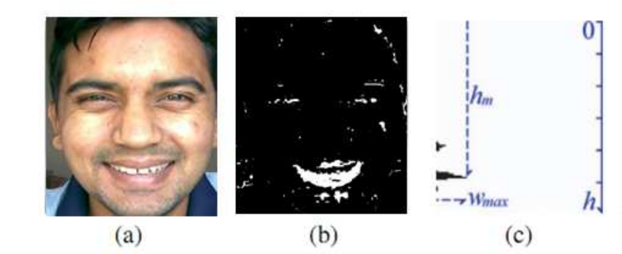


Fig. 4 Mouth Segmentation

6) Eyes Detection After mouth detection stage, we know that the y-coordinate of mouth is h_m and the y-coordinate of eyes must smaller than h_m according to our definitions. This information is used to detect human eyes in the smaller region. The region is defined by the y-coordinate 0 to $h_m - w_{\max}$. Because the y-coordinate of mouth is must larger than eyes, the considered height of region must be less than h_m. An example of detecting region is shown in Fig. 5(a). Due to the deeper lineaments around human eyes, we can detect the existence of human eyes by the luminance which is slightly darker than average skin-color. The pixels which around human eyes is defined by $E_{\hat{h}w}$:

$$E_{\hat{h}w} = \begin{cases} 0, \ 65 < Y < 80 \\ 1, \ otherwise, \end{cases} \tag{10}$$

where $E_{\hat{h}w} = h_m - w_{\max}$. Fig. 5(b) shows an example that we find out the pixels around eyes.

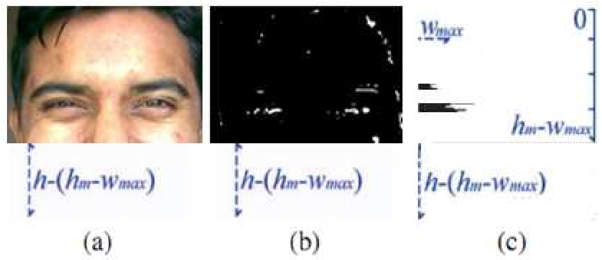


Fig. 5 Eye Segmentation

Then, the vertical based histogram, illustrated in Fig. 5(c), shows the distribution of $E_{\hat{h}w}$. In this histogram, we assume the candidate block has human eyes if there exist a α value greater than a threshold β. Here we let $\alpha = 0.5 w_{\max}$ and $\beta = w_{\max}$. When we finish the eyes detection, we regard the blocks which pass three feature detections are human face

6) Cropping the desired part of face: After selecting the specific face part of human from the image. The desired part has cropped from the selected face. We have calculated the desired part according to the following Figure. The rectangular part of face like of the Fig. 6(d) from the input face image with background has been cropped. According to the above mentioned processes, the database for train and test has been created. Finally, Aw-SpPCA has been applied for classification and recognition of facial expression.

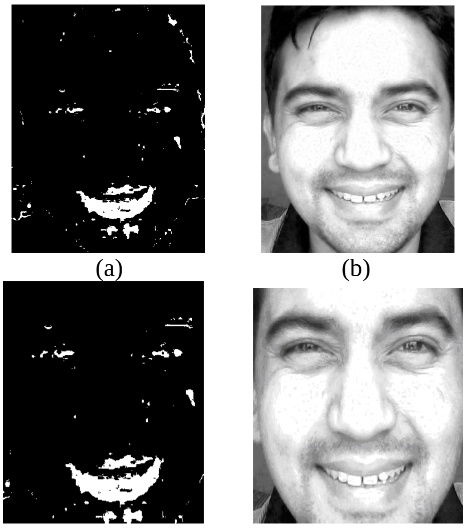


Fig. 6 (a) possible face region segmentation (b) Crop the segmentated region (c) Desired part segmentation and (d) Crop the desigred part of face

Algorithm: An algorithm for classifying facial expression.

Firstly, a low dimensional face space is created by using the train images that are utilized. This is done by performing adaptively weighted sub-pattern PCA (As-SpPCA) in the training image set and taking the principal components (i.e. eigen vectors with greater eigen values). The projected versions of all the train images are also created in this process.

Secondly, as a result-the test images also are projected on the face space; the selected principal components are used to represent the test images.

Thirdly, the Euclidian distance of a projected test image from all the projected train images are calculated and the minimum value is chosen in order to find out the train image which is most similar to the test image. The test image is assumed to fall in the same class that the closest train image belongs to.

Fourthly, in order to determine the intensity of a particular expression, its Euclidian distance from the mean of the projected neutral images is calculated. The more the distance - according to the assumption - the far it is from the neutral expression. As a result, it can be recognized as a stronger the expression.

III. AW-SPPCA ALGORITHM

There are three main steps in Aw-SpPCA algorithm: (1) partition face images into sub-patterns, (2) compute contributions of each sub-pattern, and (3) classify an unknown image.

A. Partition of Images

A face image can be partitioned to a set of equally or unequally sized sub-images in the Aw-SpPCA algorithm, depending on user options, due to the mPCA's inherent limitation all sub-images partitioned in the modular PCA (mPCA) are strictly confined to equal size. In this paper without loss of generality, we still adopt equally sized partition for a face image. Suppose there are N $W_1 x W_2$ images belonging to m expression of one person in the training set, these expressions possess $N_1, N_2, N_3 .... N_m$ face images, respectively. Each image is first divided into L equally sized sub-images in a non overlapping way which are further concatenated into corresponding column vectors with dimensionality of $(W_1 x W_2)/L$, then we collect these vectors at the same position of all facial expression images to form a specific sub-pattern's training set, in this way, L separate sub-pattern sets are formed. This process is illustrated in Fig. 7.

B. Computing Using Algorithm

A gallery set and a probe set for each sub-pattern are generated and thus possess corresponding L sub-patterns' gallery and probe sets, respectively. The gallery set is identical to the sub-pattern's training set, but the probe set is generated by both the ''sub-pattern median face'' and the ''sub-pattern mean face'' of each person in this sub-pattern's training set rather than by one validation set independent from the gallery set as usual.

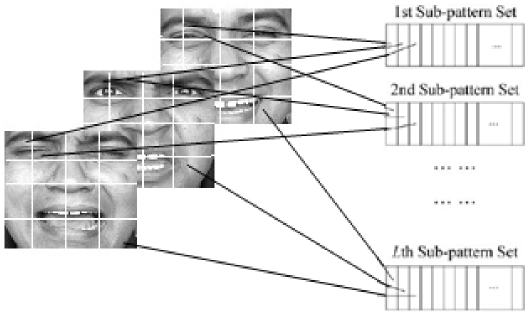


Fig. 7 Face images sub-pattern sets construction

To determine contributions made by different parts to facial expression classification the selected sub-pattern median and mean of facial expression from the training set. A process of computing the contribution consists of the two following steps. In the first step, we compute sub-pattern median and mean of facial expression image and define a similarity between two samples; the second step is to compute so-needed contributions.

Step 1: For the jth sub-pattern, so-called sub-pattern median of face of the ith person is first computed by

$$I_{ij\_median} = Median\ (I_{ij1}, I_{ij2}, ...\ .....\ I_{ijNi}) \tag{11}$$

and similarly the sub-pattern mean of face by

$$I_{ij\_mean} = \frac{1}{N_i}\sum_{k=1}^{N_i} I_{ijk} \quad (12)$$

where I_{ijk} denotes the column vector corresponding to the vectorized ith pattern's jth sub-image in the kth facial image of the person. And then the conventional PCA is applied to the jth sub-pattern's gallery set, and the respective projection matrix U_j is constructed by selecting first M' eigenvectors associated with the first largest M' eigenvalues. The similarity between sub-pattern samples x and y is defined as

$$Similarity\ (x, y) = -(x-y)^T U_j U_j^T (x-y) \quad (13)$$

Step 2: The contribution of sub-pattern to classification as follows: For a sub-pattern sample from the probe set, the similarities between it and every sample in this sub-pattern's gallery set are first computed, then the gallery samples are ranked in the descending order of the obtained similarities, and the identity of the top 1 sample in the rank list is considered as the recognition result. The result is true if the resulted identity and the probe's identity are matched, else false. After the computation is completed for all probe set samples of the jth sub-pattern, we denote by C_j the number of how many probe set samples of the jth sub-pattern are correctly classified. Finally, the contributions made by the jth sub- pattern to classification is defined as

$$W_j = C_j / 2M \quad (14)$$

C. Expression classification

According to this process, in order to classify an unknown expression image p, the image is also first partitioned into L sub-patterns in the same way previously applied to the training images. Then in this image's each sub-pattern, the unknown sub-pattern sample's identity is determined in a similar way described in Section B, Step 2. Since one classification result for the unknown sample is generated independently in each sub pattern, there will be total L results from L sub-patterns.

To combine L classification results from all sub-patterns of this face image p, a distance matrix is constructed and denoted by $D(p) = (d_{ij})_{NxL}$ with the size of NxL, where d_{ij} denotes the distance between the corresponding jth sub-patterns of the p and the ith person, and d_{ij} is set to W_j if the computed identity of the unknown sample and the ith expression's identity are identical, 0 otherwise. Consequently, a total confidence value p finally belongs to the ith expression is defined as

$$TC_i(p) = \sum_{j=1}^{L} d_{ij} \quad (15)$$

And the final identity of this p is determined by

$$Identity(p) = \arg\max(TC_i(p)), \quad 1 \le i \le N \quad (16)$$

IV. Experiments

A. Experimental database

Since the main purpose of this work is facial expression recognition (not face detection), therefore, the sample pictures are taken under special consideration to ease up the face detection process. Each picture is taken under the condition that, only face is the largest skin colored continuous object in the images. There are two sets of pictures. One is used for training purpose and another is used for testing. The training pictures are located into the ".\Images" Folder. Cropped versions of these images are placed into the ".\Images\Cropped" folder. Every picture in this set is sorted in an ascending order of expressional intensity. The pictures are classified in the following expressional classes:

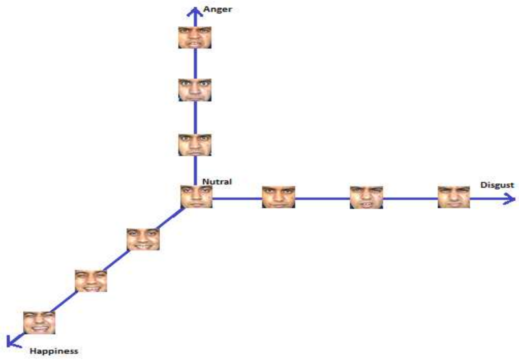


Fig. 8 Example figure for three expressions.

1. Image001 to Image013 = Happy
2. Image014 to Image024 = Disgust
3. Image025 to Image034 = Anger
4. Image035 to Image050 = Neutral

Another image set is used for testing purpose. These images are taken in quite an arbitrary fashion. It also includes some expressions that are not contained in the training set like "Surprise" and "Confused".

B. Experimental Results

TABLE I
PERFORMANCE ANALYSIS

No.	Type of Gesture	No. of Input Images	Recognized	Result (%)
1	Happy	13	12	92.3
2	Disgust	11	10	90.9
3	Anger	10	9	90
4	Neutral	7	7	100

V. Conclusion

In this paper, an accurate and high speed facial expression detection system has been proposed from color images with complex background. The color and feature-based detections have been adopted to find skin-color fast and selected candidate blocks carefully. The lighting compensation is used to improve the performance of color-based scheme, and reduce the computation of feature-based scheme. One of the major contribution of this paper is proposed method that cropped exact face region from background image based on height and width ratio (δ = 1.618) successfully. This method is also effective for dark or bright vision, close eyes, open mouth, wearing glasses, and half-profile face. Aw-SpPCA statistical algorithm has been used for classification of the train data for facial expression recognition. The experimental results indicate that this proposed approach not only is effective but also outperforms under different facial expression and illumination condition. This proposed approach can firstly detect the facial expression with the high average recognition rate 93.3%. In future work, the proposed approach can be applied to detect face, facial expression or other objects recognition for 3D objects using integral imaging technique.

ACKNOWLEDGMENT

This work is supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2011-0030815).This work is partially supported by the IT R&D program of MKE/KEIT. [KI001810039169, Development of Core Technologies of Holographic 3D Video System for Acquisition and Reconstruction of 3D Information]

REFERENCES

[1] Chai, D.; Ngan, K. N.: Locating Facial Regions of a Head–and–Shoulders Color Image, in Proceedings of the International Conference on Automatic Face and Gesture Recognition, 1998, S. 124–129.

[2] Heisele, B.; Poggio, T.; Pontil, M.: Face Detection in Still Gray Images, in MIT AI Memo, AIM–1687, 2000.

[3] Yang, M.; Ahuja, M.; Kriegman, D.: Face Detection using a Mixture of Factor Analyzers, in Proceedings of the International Conference on Image Processing, Bd. 3, 1999, S. 612–616.

[4] Jones, M.; Rehg, J.: Statistical Color Models with Application to Skin Detection, in Proceedings of Computer Vision and Pattern Recognition, 1999, S. I:274–280.

[5] Ekman, P.; Friesen, W.: The Facial Action Coding System: A Technique for the Measurement of Facial Movement, in Consulting Psychologists Press, Palo Alto, CA, 1978.

[6] Tian, Y.; Kanade, T.; Cohn, J.: Recognizing Action Units for Facial Expression Analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Bd. 23, Nr. 2, 2001, S. 97–115.

[7] S. Gundimada, Li Tao, and v. Asari, “Face detection technique based on intensity and skin color distribution,” in 2004 International Conference on Image Processing, Otc. 2004, vol. 2, pp. 1413–1416.

[8] K. P. Seng, A. Suwandy, and L.-M. Ang, “Improved automatic face detection technique in color images,” in IEEE Region 10 Conference TENCON 2004, Nov. 2004, vol. 1, pp. 459–462.

[9] Y. Araki, N. Shimada, and Y. Y. Shirai, “Detection of faces of various directions in complex backgrounds,” in 16th International Conference on Pattern Recognition, 2002. Proceedings, Object 2002, vol. 1, pp. 409–412.

[10] S. Becker, M. Plumbley, Unsupervised neural network learning procedures for feature extraction and classification, J. Appl. Intell. 6 (3) (1996) 185–205.

[11] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Machine Intell. 19 (7) (1997) 711–720.

[12] S.C. Chen, Y.L. Zhu, Subpattern-based principle component analysis, Pattern Recogn. 37 (1) (2004) 1081–1083.

[13] R. Gottumukkal, V.K. Asari, An improved face recognition technique based on modular PCAapproach, Pattern Recogn. Lett. 25 (4) (2004) 429–436.

[14] I.T. Jolliffe, Principal Component Analysis, second ed., Springer, Berlin, New York, 2002.

[15] M. Kirby, L. Sirovich, Application of the KL procedure for the characterization of human faces,IEEE Trans. Pattern Anal. Machine Intell. 12 (1) (1990) 103–108.

[16] A.M. Martinez, R. Benavente, The AR face database, CVC Technical Report #24, June 1998.

[17] http://www.uk.research.att.com/facedatabase.html M. Turk, A. Pentland, Eigenfaces for recognition, J. Cognitive Neurosci. 3 (1) (1991) 71–86.