2006 IEEE International Symposium on Signal Processing and Information Technology

Pain Recognition Using Artificial Neural Network

Md. Maruf Monwar and Siamak Rezaei
Computer Science
University of Northern British Columbia,
3333 University Way, Prince George,
BC V2N 4Z9, Canada
Tel: 1-250-960-6263
Fax: 1-250-960-5544
E-mail: {monwar,siamak}@unbc.ca

Abstract - *Facial expressions are a key index of emotion. To make use of the information afforded by facial expression for emotion science and clinical practice, reliable, valid, and efficient methods of measurement are critical. Enabling computer systems to recognize facial expressions and infer emotions from them is a challenging research topic. In this paper, we present an efficient video analysis technique for recognition of a specific expression, pain, from human faces. We employ an automatic face detector and facial feature tracker for face detection and feature extraction respectively. The face detector uses skin color modeling approach. For pain recognition, location and shape features of the detected faces are computed. These features are then used as inputs to the artificial neural network which uses standard error back-propagation algorithm for classification of painful and painless faces.*

Keywords - P*ain recognition, skin color model, location features, shape features, error back-propagation.*

I. INTRODUCTION

In recent years, tremendous amount of research has been carried out in the field of automatic expressions (such as, pain, anger, sadness etc.) recognition from video sequence and still has significant potential for further research and development. This coupled with its vast array of commercial applications (like in Medical system, in Psychological research etc.) make it an attractive area of research. To make use of the information afforded by facial expression for emotion science and clinical practice, reliable, valid, and efficient methods of measurement are critical. Until recently, selecting a measurement method meant choosing among one or another human-observer-based coding system or facial electromyography (EMG). While each of these approaches has advantages, they are not without costs. Human-observer-based methods are time consuming to learn and use, and they are difficult to standardize, especially across laboratories and over time. Facial EMG requires placement of sensors on the face, which may inhibit facial action and which rules out its use for naturalistic observation. An emerging alternative to these methods is automated facial image analysis using computer vision. Computer vision is the science of extracting and representing meaningful information from digitized video and image and recognizing perceptually meaningful patterns. In this paper, we propose a method for automatically inferring pain in video sequences.

We have focused our research toward developing a sort of supervised pain recognition scheme that does not depend on excessive geometry and computations like deformable templates. Instead, standard error back-propagation algorithm of artificial neural network is used which seemed to be an adequate method to be used in a pain recognition due to its simplicity, speed and learning capability [1]. Although pain recognition is a fundamental step in a fully automated facial expression analysis system, the first important step in pain recognition is detecting faces in video sequences and for that to know where the skin regions are. This is called skin region detection. Most of the existing expression recognition methods use gray intensity values to detect faces [2][3] in video stream. However, it is a well-known fact that the majority of images acquired today are colored and the skin color features should be important sources of information for discriminating faces from the background. In this system, color is modeled as a Gaussian function in chromatic color space [4]. Where intensity plays no role and whole information is provided by hue, saturation in other word in pure color (r, g). Each detected skin regions are then tested for presence of face in the next steps. Approximate face locations are detected using a proper height-width proportion of general face. Once rough face locations are detected, they are verified by an eye template-matching scheme. There is biological evidence that eyes play the most important role in human face detection [5][6]. Once a face is detected, the facial features are extracted. We observe that most facial feature changes that are caused by a pain are in the areas of eyes, brows and mouth. In this paper, two types of facial features in these areas are extracted: location features and shape features. The standard back-propagation in the form of a three-layer neural network with one hidden layer is used to recognize painful faces. The inputs to the network were normalized extracted location and shape features.

The block diagram of the proposed system is shown in Fig.1.

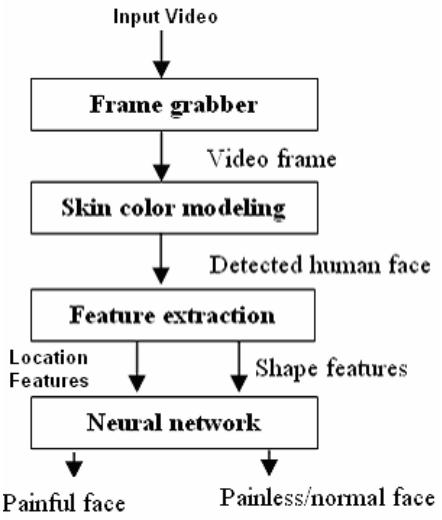

Fig. 1. Block diagram of the pain recognition system

II. SKIN COLOR MODELING FOR FACE DETECTION

A. Image Acquisition

Painful and normal video files, collected from the Psychophysiology Laboratory Database of University of Northern British Columbia, Canada, are the sources of inputs of our system. There are two video files for every subject and a total of 38 subjects with different colors, ethnicities, ages and genders are considered. In one file, the subject is in normal mood and in the other file, the subject is in painful mood due to moving hands or pressing something or shaking heads etc. The resolutions of the videos are 96 X 96. These video files are first read and the numbers of frames of each video are determined. Middle frame of the videos are then stored as image in the database for further processing. The reason for taking middle frame is that, in almost all the pain videos, the expression for pain begins after some time from the starting of the videos and ends some time before the ending of the videos. So, by taking the middle frame, it is ensured that the expression for pain in a pain video will be captured. The videos are roughly 1 to 1.5 second long.

B. Skin Color Model

In order to segment human skin regions from non-skin regions based on color, we need a reliable skin color model that is adaptable to people of different skin colors and to different lighting conditions [8]. The common RGB representation of color images is not suitable for characterizing skin-color. In the RGB space, the triple component (r, g, b) represents not only color but also luminance. Luminance may vary across a person's face due to the ambient lighting and is not a reliable measure in separating skin from non-skin region [9]. Luminance can be removed from the color representation in the chromatic color space. Chromatic colors [10], also known as "pure" colors in the absence of luminance, are defined by a normalization process shown below:

$$r = R/(R+G+B) \quad \text{and} \quad b = B/(R+G+B)$$

Color green is redundant after the normalization because $r + g + b = 1$. If two points $P_1[r_1,g_1,b_1]$ and $P_2[r_2,g_2,b_2]$, are proportional, i.e.,

$$\frac{r_1}{r_2} = \frac{g_1}{g_2} = \frac{b_1}{b_2}$$

P_1 and P_2 have the same color but different brightness.

Chromatic colors have been effectively used to segment color images in many applications [11]. It is also well suited in this case to segment skin regions from non-skin regions. The color distribution of skin colors of different people was found to be clustered in a small area of the chromatic color space. Although skin colors of different people appear to vary over a wide range, they differ much less in color than in brightness. In other words, skin colors of different people are very close, but they differ mainly in intensities [8]. With this finding, we could proceed to develop a skin-color model in the chromatic color space.

A total of 68 skin samples from 68 color images taken from same number of videos (normal and painful) were used to determine the color distribution of human skin in chromatic color space and generate the statistical skin-color model. Our samples were taken from persons of different ethnicities: Asian, Caucasian and African and from different ages and genders with varying illumination condition. As the skin samples were extracted from color images, the skin samples were filtered using a low-pass filter to reduce the effect of noise in the samples.

Fig. 2 and Fig. 3 illustrate the training process, in which a skin-color region is selected and its RGB representation is stored. It was verified, using training data, that skin colors are clustered in color space, as illustrated in Fig. 3(a).

Fig. 2(a). Selected skin region in RGB image

Fig. 2(b). Selected skin region in Chromatic color

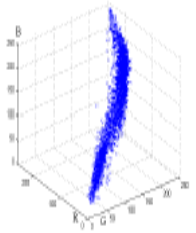

Fig. 3(a). Cluster in color space (RGB)

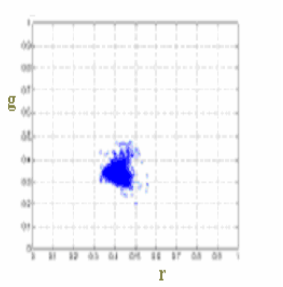

Fig. 3(b). Cluster in chromatic space (r,g)

The color histogram reveals that the distribution of skin-color of different people are clustered in the chromatic color space and a skin color distribution can be represented by a Gaussian model N(m, C), where:

Mean, $m = E\{ x$, where $x = (r\ b)^T$

Covariance, $\sum = \begin{bmatrix} \sigma_{rr} & \sigma_{rg} \\ \sigma_{gr} & \sigma_{gg} \end{bmatrix}$

Fig. 4 shows the Gaussian Distribution N(m, C) fitted by our data.

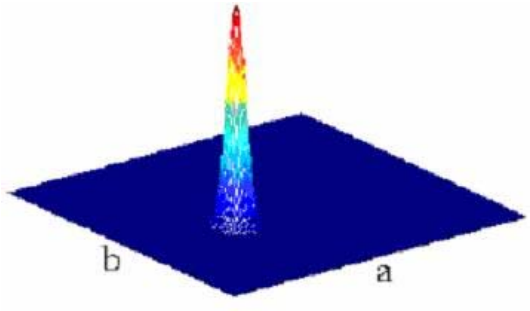

Fig. 4. Fitting skin color into a Gaussian Distribution

With this Gaussian fitted skin color model, we can now obtain the likelihood of skin for any pixel of an image. Therefore, if a pixel, having transform from RGB color space to chromatic color space, has a chromatic pair value of (r,b), the likelihood of skin for this pixel can then be computed as follows:

$$\text{Likelihood} = P(r,b) = \exp[-0.5(x-m)^T C^{-1}(x-m)],\ [\text{ where, } x = (r,b)^T]$$

Hence, this skin color model can transform a color image into a gray scale image such that the gray value at each pixel shows the likelihood of the pixel belonging to the skin. With appropriate thresholding, the gray scale images can then be further transformed to binary images showing skin regions and non-skin regions.

C. Skin Region Segmentation:

Our main goal in this segmentation process is to remove the background of the image from skin regions using previously discussed skin color model. First, input image is converted to chromatic color space. Using Gaussian model, a grayscale image of skin likelihood pixels is constructed and skin pixels have some set of constant values for each r, g and b component. Every pixel in normalized image has three values and they are normalized-red, normalized-green and normalized-blue.

Segmentation process extracts these normalized components and constructs two images (Fig. 6(c) and Fig. 6(d)). Each of these images is converted into black and white image by applying different threshold for normalized input image (Fig. 6(e) and Fig. 6(f)) such that r = 0.41-0.50 and g = 0.21- 0.30. Finally, we perform an 'AND' operation between these two black and white images where white pixels are skin and blacks are non skin pixel. In this approach, due to noise and distortion in input image, color information of some skin pixels acts like non skin region and generate non contiguous skin color region. To solve this problem, first morphological closing operator is used to obtain skin-color blobs (Fig. 6(g)). A median filter was also used to eliminate spurious pixels (Fig. 6(h)). Boundaries of skin-color regions are determined using a region growing algorithm in the binary image. Regions with size less than 1% of image size are eliminated [5]. At the end of the segmentation process, black and white skin regions of images are multiplied by the original RGB image and we then get the skin region (Fig. 6(i)) of face. Fig. 5 illustrates a simple block diagram for the segmentation process and Fig. 6 shows an example of segmentation and face location detection process performed on a painful image.

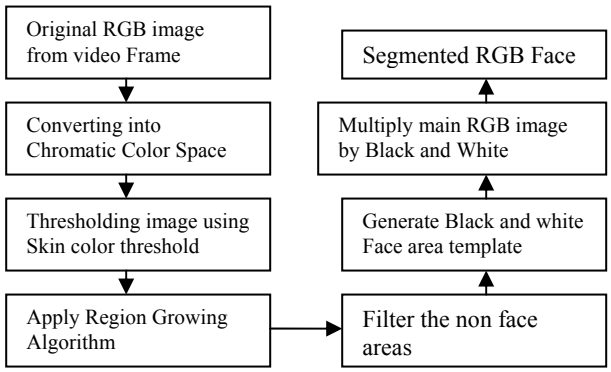

Fig. 5. Block Diagram for Face Segmentation

An example of segmentation and face location detection process is performed on a painful image is given bellow:

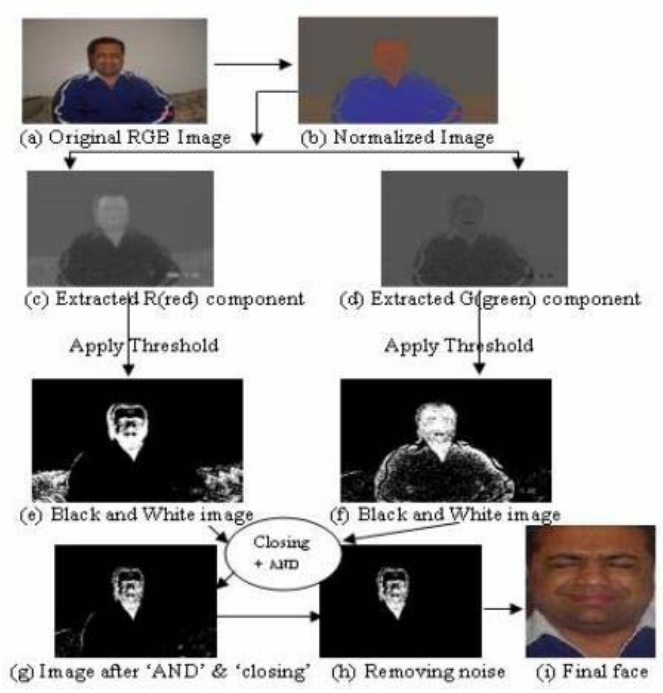

Fig. 6. Segmentation and approximate face location detection process.

D. Face Detection

To reduce some search space for eyes template matching, bounding rectangles of all connected areas from black-white template are taken into consideration and center of the areas have been calculated. These are the mass point of the template areas. Now calculation of the height and width of the bounding rectangle have been performed. If height-width proportions satisfy for face like shape, keep those areas otherwise they are removed. In our system we use the largest area for face and remove the smallest area of skin regions. Thus the template with approximate face area is multiplied by the original image and we get face. Then to consider only the meaningful portions of the face we use a mask image. A bitwise 'AND' operation is used to apply the mask image with the original face image. Features in the image which coincide with the white areas on the mask image will be displayed.

The original video frame, obtained gray level image, the mask image and the resultant image is shown in Fig. 7.

7(a) 7(b) 7(c) 7(d)

Fig. 7. (a) Original video frame (b) Gray level image (c) Mask image and (d) Resultant image

III. FACIAL FEATURES EXTRACTION

After detecting human faces from video frames, we need to detect reliable facial features. We observe that most facial feature changes that are caused by a pain are in the areas of eyes, brows and mouth. In this paper, two types of facial features in these areas are extracted: location features and shape features. The idea for extracting features presented here is similar to that taken by Yang et al. [6] and Ying-li Tian et al. [7] and is an attempt to make the feature extraction robust to the available video sequences for this research.

A. Location Feature Extraction

In this system, six location features are extracted for pain recognition. They are the two eye centers, two eyebrow inner endpoints and two corners of the mouths.

Eye Centers and Eyebrow Inner Endpoints: To find the eye centers and eyebrow inner endpoints inside the detected frontal or near frontal face, we have developed an algorithm that searches for two pairs of dark regions which correspond to the eyes and the brows by using certain geometric constraints such as position inside the face, size and symmetry to the facial symmetry axis. Similar to paper [6], the algorithm employs an iterative thresholding method to find these dark regions. Fig. 8 shows the iterative thresholding method to find eyes and brows. Generally, after four iterations, all the eyes and brows are found. If satisfactory results are not found after 15 iterations, we think the eyes or the brows are occluded. Unlike the work of Yang *et al.* to find one pair of dark regions for the eyes only, we find two pairs of parallel dark regions for both the eyes and eyebrows. By doing this, not only are more features obtained, but also the accuracy of the extracted features is improved. As shown in Fig. 8(b), the right brow and the left eye is wrongly extracted as the two eyes in Yang's approach. Fig. 8(d) shows the correct positions are extracted for all the eyes and eyebrows in our method. Then the eye centers and eyebrow inner endpoints can be easily determined.

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

Fig. 8. Iterative thresholding of the face to find eyes and brows. (a) grey-scale face image, (b) threshold = 45, (c) threshold = 55, (d) threshold = 65

Mouth Corners: After finding the positions of the eyes, the location of the mouth is first predicted. Then the vertical position of the line between the lips is found using an integral projection of the mouth region proposed by Yang *et al.* [6].

Finally, the horizontal borders of the line between the lips are found using an integral projection over an edge-image of the mouth. After Yang *et al.* , the following steps are use to track the corners of the mouth: 1) Find two points on the line between the lips near the previous positions of the corners in the image 2) Search along the darkest path to the left and right, until the corners are found. Finding the points on the line between the lips can be done by searching for the darkest pixels in search windows near the previous mouth corner positions. Because there is a strong change from dark to bright at the location of the corners, the corners can be found by looking for the maximum contrast along the search path [7].

B. Location Feature Representation

After extracting the location features, all the faces are normalized to a 90 X 90 pixels. We transform the extracted features into a set of parameters. We represent the face location features by 5 parameters, which are shown in Fig. 9. These parameters are the distances between the eye-line and the corners of the mouth, the distances between the eye-line and the inner eyebrows, and the width of the mouth (the distance between two corners of the mouth).

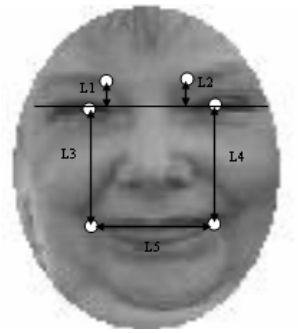

Fig. 9. Face location feature representation for expression recognition

C. Shape Feature Extraction

In order to extract the mouth shape features, first an edge detector is applied to the normalized face to get an edge map. Unlike Ying-li Tian et al. [7], which divided the edge map into 3 X 3 zones, here the edge map is divided into 2 X 2 zones as shown in Fig. 10(b). The eyes and mouth shape features are computed from zonal shape histograms of the edges in the mouth and eyes region. To place the 2 X 2 zones onto the face image, the upper two zones are placed at the locations of the eyes ad the lower two portions are placed at the location of mouth. The coarsely quantized edge directions are represented as local shape features and more global shape features are presented as histograms of local shape (edge directions) along the shape contour. The edge directions are quantized into 4 angular segments (Fig. 10(c)). Representing the whole mouth as one histogram does not capture the local shape properties that are needed to distinguish pain expressions. Therefore we use the zones to compute four histograms of the edge directions. Hence, the eyes and mouth is represented as a feature vector of 16 components (4 histograms of 4 components). An example of the histogram of edge directions corresponding to the lower right zone is shown in Fig. 10(d).

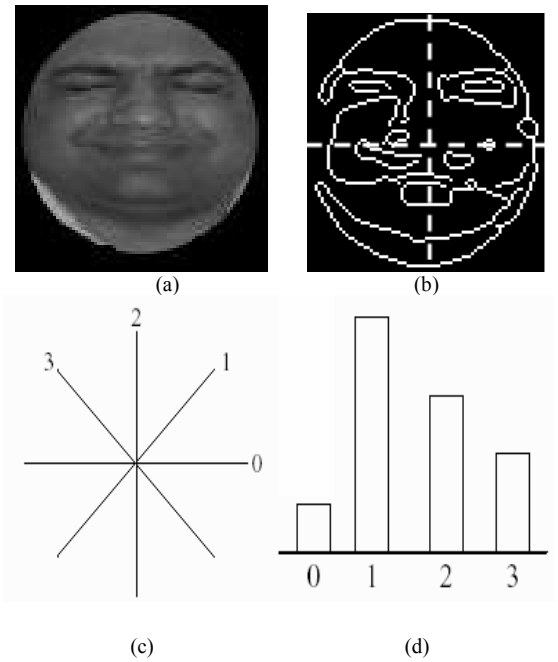

Fig. 10. Zonal-histogram features. (a) normalized face, (b) zones of the edge map of the normalized face, (c) four quantization levels for calculating histogram features, (d) histogram corresponding to the middle zone of the mouth.

IV. PAIN RECOGNITION

The proposed system is applied on a wide variety of painful and normal video sequences collected from Psychophysiology Laboratory Database of University of Northern British Columbia (UNBC), Canada. The videos include different lightening conditions and with different backgrounds. It is found that the system successfully detects skin region of the images collected from video analysis. However, it is important to note that not all detected regions contain faces. Some corresponds to parts of human body, while other corresponds to objects with colors similar to those of skin. We implemented the entire algorithm in MATLAB 7.0 on a PENTIUM-IV windows XP workstation.
We used a neural network-based recognizer having the structure shown in Fig. 11. The standard back-propagation in the form of a three-layer neural network with one hidden layer was used to recognize facial expressions. The inputs to the network were the 5 location features (Fig. 9) and the 16 zone components of shape features of the eyes and mouth

regions (Fig. 10). Hence, a total of 21 features were used to represent the amount of pain in a face image. The outputs were a set of two values – painful face or painless face. We tested various numbers of hidden units and found that 10 hidden units gave the best performance.

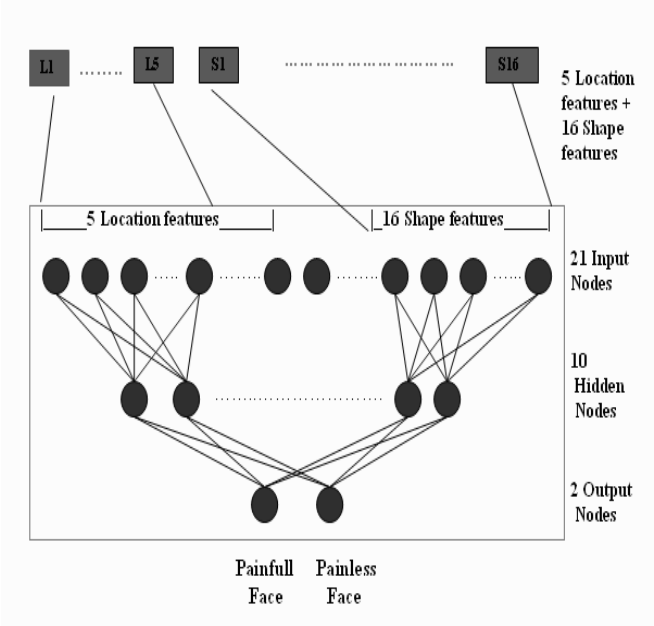

Fig. 11. Neural network-based recognizer for facial expressions

V. EXPERIMENT RESULTS

Sixty eight different videos are used for this experiment. Some videos contain the same person but in different orientation. To examine the accuracy of our proposed pain recognition system, we have tested the system with several hidden layer units. It works best for 10 hidden layer units. The program had some problems with videos which contain partially or fully bald headed people. In those cases, it showed the wrong output because of the wrongly detection of face areas using skin color modeling technique. Also it is very hard to find correct output for older person when the color of the hair and the color of the skin are almost the same. The following table shows the effect of the number of the hidden units used in the neural network.

Table 1. Effect of various numbers of hidden units on system accuracy

Number of hidden Units	Painless face recognition rate	Painful face recognition rate	Average accuracy
5	71.23%	45.76%	58.5%
10	95.32%	87.99%	91.67%
15	87.90%	84.21%	86.01%
20	80%	72.65%	76..33%

VI. CONCLUSIONS

We have presented a pain recognition approach in this paper. For this, we have addressed the problems of how to find a human face in a video sequence and how to represent and recognize pain expressions presented in those faces. Skin color modeling approach is used for face detection and neural network-based recognizer is used for pain recognition in those faces. Though some false recognition occurs, the overall pain recognition performance of the proposed system is still quite satisfactory. Detecting al types of faces correctly from the video sequences and recognizing all the expressions from real time video sequences are our future plan.

ACKNOWLEDGMENT

This research has been conducted in the Computer Science program of UNBC, Canada. We would like to thank Professor Ken Prkachin at Department of Psychology, UNBC and his co-investigator Dr. Patty Solomon at the McMaster University, Canada (who collected the video database) for providing the video sequences for this research.

REFERENCES

[1] P. Sinha (1994), “Object Recognition via Image Invariants: Case Study Investigative Ophthalmology and Visual Science”, vol. 35, pp. 1735-1740.

[2] H. A. Rowley. S. Bluja & T. Kanade (1998), “Neural Network-based Face Detection”, IEEE Transaction on Pattern. Analysis & Machine Intelligence, vol. 20, no. 1, pp. 39-51.

[3] M.C. Burl, T.K. Leung & P. Perona (1995), “Face Localization via Shape Statistics”, in proceedings of the 1st International Workshop on Face and Gesture Recognition, Zurich, Switzerland.

[4] Demir Gökalp, “Skin Color Based Face Detection”, Department of Computer Engineering, Bilkent University, Turkey.

[5] R.S. Feris, T. E. de Campos & R. M. C. Junior (2000), “Detection and Tracking of Facial Features in Video Sequences”, in proceedings of the Mexican Internaional Conference on Artificial Intelligence: Advances in Artificial Intelligence, Mexico, pp. 127–135.

[6] J. Yang, R. Stiefelhagen, U. Meier & A. Waibe(1998), “Real-time Face and Facial Feature Tracking and Applications”, in Proceedings of the Auditory-visual Speech Processing (AVSP 98), NSW, Australia.

[7] Yingli Tian & Lisa Brown (2003), “Real World Real-time Automatic Recognition of Facial Expressions”, IEEE Workshop on Performance Evaluation of Tracking and Surveillance, Graz, Austria.

[8] Jie Yang & Alex Waibel (1996), "A Real-Time Face Tracker", in proceedings of the 3rd IEEE Workshop on Application of Computer Vision, Sarasota, Florida, USA.

[9] J. Cai, A. Goshtasby & C. Yu (1998), “Detecting Human Faces in Color Images”, in proceedings of the International Workshop on Multimedia Database Management Systems.

[10] G. Wyszecki & W.S. Styles (1982), “Color Science: Concepts and Methods, Quantitative Data and Formulae”, Second edition, John Wiley & Sons, New York, USA.

[11] Y. Gong & M. Sakauchi (1995), “Detection of Regions Matching Specified Chromatic Features”, Computer Vision and Image Understanding vol. 61, no. 2, pp. 263–269.