M.Tech. (Computer Science) Dissertation Series

FACE DETECTION IN COLOR IMAGES

a dissertation submitted in partial fulfilment of the requirements for the M.Tech. (Computer Science) degree of the Indian Statistical Institute

By

Swarup Roy Chowdhury

under the supervision of

Prof. C.A. Murthy
Machine Intelligence Unit (MIU)

INDIAN STATISTICAL INSTITUTE
203, Barrackpore Trunk Road
Calcutta - 700035

ABSTRACT

FACE DETECTION IN COLOR IMAGES

By

Swarup Roy Chowdhury

Face recognition has received substantial attention from researchers in biometrics, computer vision, pattern recognition, and cognitive psychology communities because of the increased attention being devoted to security, man-machine communication, content-based image retrieval, and image/video coding. Three major tasks involved in face recognition systems are: (i) face detection, (ii) face modeling, and (iii) face matching. So Face Detection is an essential step in face recognition.

We have developed a face detection algorithm for color images that can detect faces with different sizes and various poses from both indoor and outdoor scenes. The goal of this dissertation is to detect all regions that may contain faces while maintaining a low false positive output rate. We first develop a skin color detector based on color analysis and the fuzzy set theory, whose performance is much better than the existing skin region detectors. We also develop a hair color detector, which makes possible the use of the hair part as well as the skin part in face detection. We design multiple head-shape models to cope with the variation of the head pose. We propose a fuzzy set theory based pattern-matching technique, and use it to detect face candidates by finding out patterns similar to the prebuilt head-shape models from the extracted skin and hair regions.

The utility of the proposed method is determination of faces at a faster rate, and accurately. The multiple head pose models allow it to detect faces of various poses. This method is not affected by local changes in a face and hence the method is not sensitive to change in facial expression and the results are good.

ACKNOWLEDGMENTS

This work has been possible due to a number of people. Firstly, I would like to thank my supervisor Prof. C.A. Murthy for his guidance and assistance throughout the course of this work. He has always been a great source of inspiration. His thoughts on the subject greatly influenced my own understanding of the subject and are reflected to some extent in the present work. He lent much of his time to review this work and helped me on a member of points. He broadened my view in research areas, and taught me how to focus on research problems.

Many thanks to Prof. Malay Kundu with whom we had a number of useful discussions. I would also like to thank my teachers Prof. C.A. Murthy, Prof. Malay Kundu, Prof. B.Chanda, Prof. D.P. Mondal for their excellent courses related to this work.

Also, I must thank all my batch mates and friends who stood by me during all good and hard times. I also thank all the professors of Machine Intelligence Unit (MIU), Indian Statistical Institute, Kolkata for their fruitful interactions. I am grateful to all the staff of MIU for providing all kinds of facilities whenever needed.

I would like to thank my parents, for all the happiness they have shared with me.

Last but not the least I am grateful to all the teachers and staff of ISI.

TO WHOM IT MAY CONCERN

This is to certify that **Mr. Swarup Roy Chowdhury** of M.Tech (Computer Science), second year student of Indian Statistical Institute, Kolkata has done his dissertation titled **"Face Detection in Color Images"** under my guidance. This dissertation partially fulfills the requirement of M.Tech (Computer Science) curriculum.

Date: 9.7.04.

Place: Kolkata

(Prof. C.A. Murthy)
Machine Intelligence Unit
Indian Statistical Institute, Kolkata

Table of Contents

1 Introduction

In recent years face recognition has received substantial attention from researchers in biometrics, pattern recognition, and computer vision communities (see surveys in [19], [20], [21]). This common interest among researchers working in diverse fields is motivated by our remarkable ability to recognize people (although in case of certain rare brain disability, e.g., prosopagnosia or face blindness [22], this recognition ability is lost) and the fact that human activity is a primary concern both in everyday life and in cyberspace. Besides, there are a large number of commercial, security, and forensic applications requiring the use of face recognition technology. These applications (see Fig. 1.1) include automated video surveillance (e.g., super bowl face scans and airport security checkpoints), access control (e.g., to personal computers and private buildings), mugshot identification (e.g., for issuing driver licenses), design of human computer interface (HCI) (e.g., classifying the activity of a vehicle driver), multimedia communication (e.g., generation of synthetic faces), and content-based image database management [23]. These applications involve locating, tracking, and recognizing a single (or multiple) human subject(s) or face(s).

Face recognition applications in fact involve several important steps, such as face detection for locating human faces, face tracking for following moving subjects, face modeling for representing human faces, face coding/compression for efficiently archiving and transmitting faces, and face matching for comparing represented faces and identifying a query subject. Face detection is usually an important first step. Detecting faces can be viewed as a two-class (face vs. non-face) classification problem, while recognizing faces can be regarded as a multiple-class (multiple subjects) classification problem within the face class.

Face detection involves certain aspects of face recognition mechanism, while face recognition employs the results of face detection.

We can consider face detection and recognition as the first and the second stages in a sequential classification system. The crucial issue here is to determine an appropriate feature space to represent a human face in such a classification system. We believe that a seamless combination of face detection, face modeling, and recognition algorithms has the potential of achieving high performance for face identification applications.

1.1 Challenges in Face Recognition

The human face has been considered as the most informative organ for communication in our social lives [24]. Automatically recognizing faces by machines can facilitate a wide variety of forensic and security applications. The representation of human faces for recognition can vary from a 2D image to a 3D surface. Different representations result in different recognition approaches. Extensive reviews of approaches to face recognition were published in 1995 [20], 1999 [25], and in 2000 [21]. A workshop on face processing in 1985 [26] presented studies of face recognition mainly from the viewpoint of cognitive psychology. Studies of feature-based face recognition, computer caricatures, and the use of face surfaces in simulation and animation were summarized in 1992 [24]. In 1997, Uwechue et al. [27] gave details of face recognition based on high order neural networks using 2D face patterns. In 1998, lectures on face recognition using 2D face patterns were presented from theory to applications [19]. In 1999, Hallinan et al. [28] described face recognition using both the statistical models for 2D face patterns and the 3D face surfaces. In 2000, Gong et al. [29] emphasized the statistical learning methods in holistic recognition approaches and discussed face recognition from the viewpoint of dynamic vision.

1.2 Challenges in Face Detection

As face detection involves certain aspects of face recognition mechanism, while face recognition employs the results of face detection so face detection from images is a key problem in human computer interaction studies and in pattern recognition researches. Many studies on automatic face detection have been reported recently. Most of them concentrate on quasifrontal view faces [3]–[7]. This is because the prior knowledge of the geometric relation with regard to the facial topology of frontal view faces can help the detection of facial features and it also makes the face modeling with a generic pattern possible. However, the quasi-frontal view assumption limits the kind of faces that can be processed.

A representative paradigm detects faces with two steps :-

1) Locating the face region [4], [9], [6], or assuming that the location of the face part is known [3], [5], [6], [7].
2) Detecting the facial features in the face region based on edge detection, image segmentation, and template matching or active contour techniques.

One disadvantage of step 1 is that the face location algorithm may not be powerful enough to find out all possible face regions while maintaining the false positive rates to be low. Another disadvantage is that the facial-feature-based approaches rely on the performance of feature detectors. For small faces or low quality images, the proposed feature detectors are not likely to perform well. Another paradigm is the visual learning or neural network approach [8], [10], [16], [11], [14]. Although the performance reported is quite good, and some of them can detect non-frontal faces, approaches in this paradigm are extremely computationally expensive. A relatively traditional approach of face detection is template matching and its derivations [15], [12], [13]. Some of them can detect non-frontal faces. This approach uses a small image or a simple pattern that represents the *average face* as the face model. It does not perform well for cluttered scenes. Face detection based on deformable shape models was also re

-ported [17]. Although this method is designed to cope with the variation of face poses, it is not suitable for generic face detection due to the high expense of computation.

(a)

(b)

Figure 1.1 : Applications using face recognition technology: (a) and (b) automated video surveillance; (c) and (d) access control;

(c)

(d)

Figure 1.1 (Cont'd)

2 Existing method on Face Detection using Fuzzy Methods

Authors Haiyuan Wu, Qian Chen and Masahiko Yachida described a method [18] to detect faces in color images based on the fuzzy set theory. They made two fuzzy models to describe the *skin color* and *hair color* respectively. In these models, they used a perceptually uniform color space to describe the color information to increase the accuracy and stableness. They used the two models to extract the skin color regions and the hair color regions. Then comparing them with the pre-built *head-shape models* by using a fuzzy set theory based pattern-matching method detected face candidates.

2.1 Detecting skin regions and hair regions

2.1.1 Perceptually uniform color space

The terms "*skin color*" and "*hair color*" is subjective human concepts. Because of this, the color representation should be similar to the color sensitivity of human eyes to obtain a stable output similar to the one given by the human visual system. Such a color representation is called the *perceptually uniform color system* or *UCS*. Many researchers have proposed conversion methods from *CIE's* (Commission Internationale de l'Eclarirage) *XYZ* color system to *UCS*. Among them, the L*u*v*, and L*a*b* color representations were proposed by G. Wyszecki. Although they are simple and easy to use, both of them are just rough approximations of *UCS*. The psychologist Farnsworth proposed a better *UCS* through psychophysical experiments in 1957 [2].

First *RGB* color information in images is converted to *CIE's XYZ* color system :-

$$\begin{cases} X = 0.619R + 0.177G + 0.204B \\ Y = 0.299R + 0.586G + 0.115B \\ Z = 0.000R + 0.056G + 0.944B \end{cases}, \quad \begin{cases} x = \dfrac{X}{X+Y+Z} \\ y = \dfrac{Y}{X+Y+Z} \end{cases} \qquad (2.1)$$

where Y carries the luminance information, and (x, y) describes the chromaticity. Then we convert the chromaticity (x, y) to the Farnsworth's *UCS* with a non-linear transformation. The result of this conversion is represented by a tuple value (u_f, v_f). The values of (u_f, v_f) of all visible colors are in the range of :-

$$u_f \rightarrow [0, 91]$$
$$v_f \rightarrow [0, 139] \qquad (2.2)$$

2.1.2 Skin color distribution model

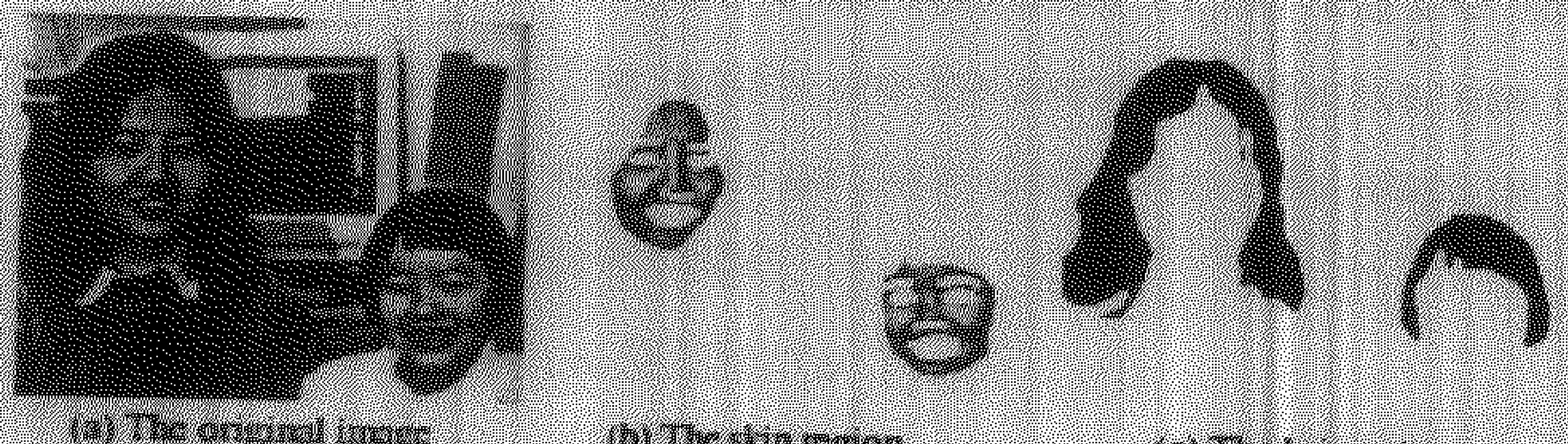
In conventional methods, all visible colors are divided into two groups: One is the "skin color" and the other is not. They assigned a value within [0.0, 1.0] to each point in the color space to indicate how much a visible color looks like the skin color. They called this value as *skin color likeness*, and used a table to describe the skin color likeness of all visible colors. They call it as "*Skin Color Distribution Model*", or simply SCDM. The SCDM is a fuzzy set of *skin color*. They used a large image set containing faces to investigate the distribution of color of the human skin region in order to build the SCDM. Fig.2.1(a) shows a sample image. The procedure to build the SCDM is as follows :-

1) Manually select skin regions in each image (See Fig.2.1(b)).

2) Prepare a table of 92 x 140 entries to record the 2-dimensional chromatic histogram of skin regions, and initialize all the entries with zero.

3) Convert the chromaticity value of each pixel in the skin regions to Farnsworth's *UCS*, and then increase the entry of the chromatic histogram corresponding to it by one.

4) Normalize the table by dividing all entries with the greatest entry in the table.



(a) The original image (b) The skin region selected by viewer. (c) The hair region selected by viewer.

Figure 2.1 : An image used to build the SCDM and HCDM

2.1.3 Hair color distribution model

They used a model similar to SCDM to describe the hair color. They called it as "*Hair Color Distribution Model*", or simply HCDM. The HCDM describes the hair color likeness of all visible colors. Because hair regions often show low brightness and the chromaticity estimation of low brightness color is not stable, they used the luminance information as well as chromaticity to describe the hair color. The HCDM is a function of three variables :- the luminance Y and the chromaticities (u_f, v_f).

$$HCDM(Y, u_f, v_f) = HCDM_y(Y) \times HCDM_c(u_f, v_f) \tag{2.3}$$

where $HCDM_y$ and $HCDM_c$ are histograms of the luminance and chromaticity of hair regions respectively. They first investigated the distribution of color on the hair region in sample images (See Fig.2.1(c)) in the luminance and chromaticity space, then they built $HCDM_y$ and $HCDM_c$ similarly as they built the SCDM.

2.1.4 Skin color detector and hair color detector

They used SCDM and HCDM to extract the skin color region and the hair color region respectively as follows. The results are the skin/hair color likeness of each pixel in the input image. They call them as *Skin Color Similarity Map* (or SCSM) and *Hair Color Similarity Map* (or HCSM).

$$\begin{cases} SCSM = SCS(p) = SCDM(u_f(p), v_f(p)) \\ HCSM = HCS(p) = HCDM(Y(p), u_f(p), v_f(p)) \end{cases} \tag{2.4}$$

where Y (p), and (u_f (p), v_f (p)) are the luminance and chromaticity of pixel p in the input image, SCS(p) and HCS(p) are the skin color likeness and the hair color likeness of pixel p, respectively.

2.2 Head-shape model

In this paper the authors ignored the detail of facial features and considered the face as a pattern composed of a skin part and a hair part. They abstract the appearance of faces in images into 5 kinds of pattern:- frontal view, left side view, right side view, left diagonal view, and right diagonal view. Accordingly they made five head-shape models. Each head-shape model is a two dimensional pattern consisting of m x n square cells.

They assign two properties to each cell :- The skin proportion M_F and the hair proportion M_H, which indicate the ratios of the skin area or of the hair area within the cell to the area of the cell.

They built the head-shape models with the following procedure :-

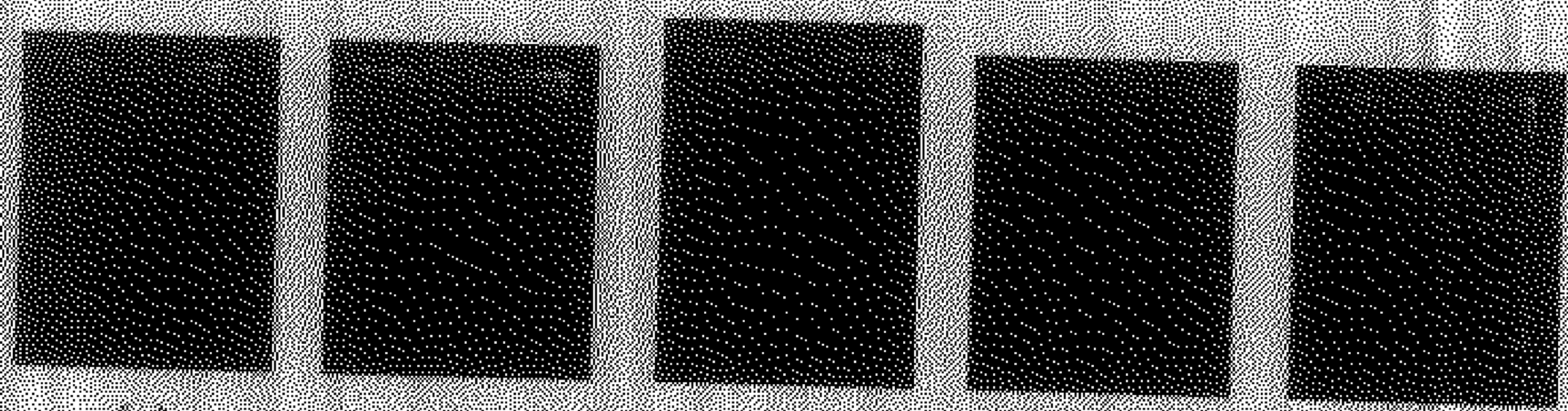
1) Collect images containing frontal faces, and the faces rotated to the left (and to the right) by 15, 30, 45, 60, 75, and 90 degree.

2) Manually select the rectangular face region, the skin part and the hair part in it, and then divide it into m x n square cells.

3) Use the frontal-view face images and the images containing faces rotated to the left and to the right by 15 degree to calculate the average M_F and M_H for each cell. Then use them to build the frontal-view head-shape model.

4) Similarly, use the images containing faces rotated to the left (or to the right) by 30 and 45 degree to build the left (or right) diagonal view head-shape model. Moreover, use the images containing faces rotated to the left (or to the right) by 60, 75, and 90 degree to build the left (or right) side-view head-shape model.

Fig.2.2 shows the head-shape models. The frontal-view model contains 10 x 13 cells, while the others contain 10 x 12 cells.



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

Figure 2.2 : The primitive head-shape models

2.3 Fuzzy Pattern Matching

In this research, they detected faces by comparing the head-shape models with the SCSM and HCSM. Because the head-shape models, SCSM and HCSM, describe different kinds of information, the image correlation or template matching approach cannot be applied directly. Here, they described a new pattern matching method for such cases based on the fuzzy set theory. They call it as *fuzzy pattern matching*. They first developed a method based on the fuzzy set theory to estimate the skin proportion and the hair proportion from the average SCS and the average HCS of a square region. They also used the fuzzy set theory to estimate the degree of similarity between the square regions in images and the cells in head-shape models.

2.3.1 Proportions of skin and hair color area

To compute the proportion of the area of the skin (or hair) color part in a square image region, they first calculated the average skin hair color similarity a_S (or the average hair color similarity a_h) in the square region :-

$$\begin{cases} a_s = \dfrac{\sum_{p \in region} SCS(p)}{n^2} \\ a_h = \dfrac{\sum_{p \in region} HCS(p)}{n^2} \end{cases} \tag{2.5}$$

where n is the size of the square region in pixels.

They defined two fuzzy sets R_S and R_H : R_S (or R_H) is the fuzzy set of $A_S \ni a_S$ (or $A_H \ni a_h$), which is defined by a fuzzy membership function μA_S (or μA_H) :-

$$\mu_{A_S} : R_S -> [0,1]; \quad \text{or} \quad \mu_{A_H} : R_H -> [0,1]; \tag{2.6}$$

R_S (or R_H) is used to describe the relationship between the average skin (or hair) color similarity a_S (or a_h) and the proportion of the skin (or hair) color part in a square region of the input image. Two S type standard functions were used to represent μA_S and μA_H. An S type standard function is defined by the following equation :-

$$S(x;a,b)=\begin{cases} 0 & x\le a \\ \dfrac{2(x-a)^2}{(b-a)^2} & a\langle x\le \dfrac{(a+b)}{2} \\ 1-\dfrac{2(x-b)^2}{(b-a)^2} & \dfrac{(a+b)}{2}\langle x\le b \\ 1 & b\langle x \end{cases} \qquad (2.7)$$

where $0 \le a \le 1$, $0 \le b \le 1$, and $a \le b$. The parameters 'a' and 'b' control the shape of the function (See Fig.2.3). When 'a' is close to 'b, the function will behave like a step function. If 'a' is set to a big value, the output of the function will decrease, and if 'b' is set to a small value, the output will increase.

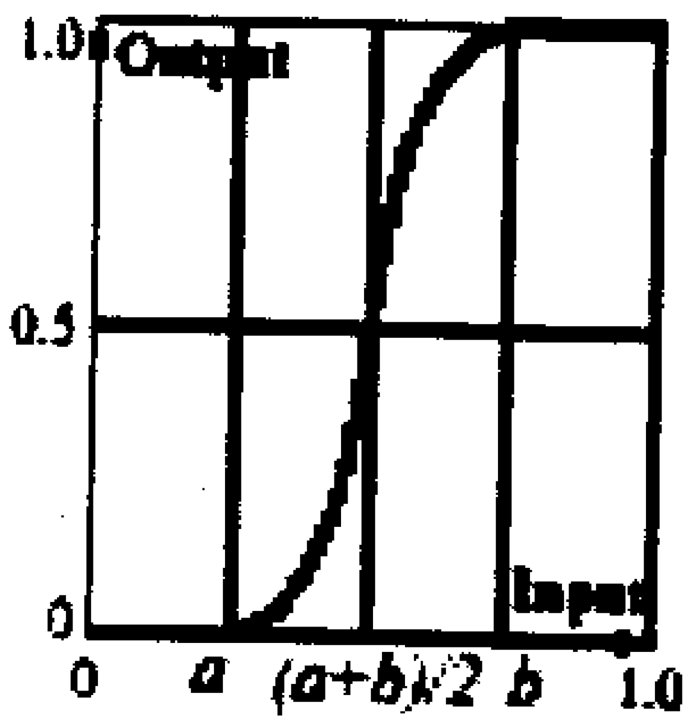


Figure 2.3 : The S type standard function

The values of (a, b) were chosen to be (0.0, 0.6) for μA_S , and (0.0, 0.75) for μA_H. These were determined through experiments so that the proportions of the skin color area and of the hair color area given by the functions become similar to the one given by human viewers. Thus, the skin color proportion (R_S) and hair color proportion (R_H) can be estimated by the following equations :-

$$\begin{cases} R_S = \mu_{A_S}(a_s) = S(a_s;0.0,0.6) \\ R_H = \mu_{A_H}(a_h) = S(a_h;0.0,0.75) \end{cases} \qquad (2.8)$$

2.3.2 Fuzzy pattern matching based on two-term fuzzy relation

To estimate the similarity between a square region in the input image and a cell in a headshape model, a method was used to compare the properties of the square region in the image (R_S and R_H) and the properties of the cell in a head-shape model (M_F and M_H). In the fuzzy set theory, the degree of similarity between two sets of real number x_1 and x_2 is described by *two-term fuzzy relation*. It can be expressed by :-

$$AE(x_1, x_2) = e^{-a|x_1 - x_2|^b} \tag{2.9}$$

where a > 0, and b > 0. The parameters a and b control the shape of the function (See Fig. 2.4).

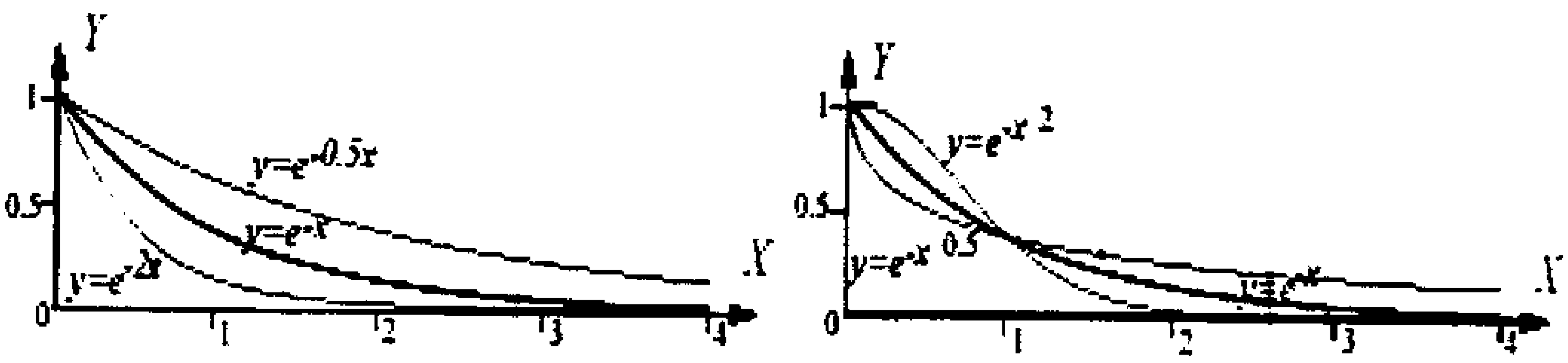


Figure 2.4 : The shape of the function e^{-ax^b}

To estimate the similarity between (R_S, R_H) and (M_F, M_H), they defined the distance between them as the following :-

$$|(R_S, R_H) - (M_F, M_H)| = \sqrt{(R_S - M_F)^2 + (R_H - M_H)^2} \tag{2.10}$$

Thus, the degree of similarity between square regions in the image and the cells in the headshape model can be calculated with the following equation, according to (2.9).

$$
match(square, cell) = AE(R_{S,H}, M_{F,H}) = e^{-a\left|(R_S, R_H)-(M_F, M_H)\right|^b} = e^{-((R_S-M_F)^2-(R_H-M_H)^2)^{0.5b}} \quad (2.11)
$$

They defined the matching degree between a rectangular part in the image and a head-shape model as the sum of the degree of similarity between all cells in the model and the corresponding square regions in the image rectangle :-

$$
Match(rect, \mathrm{mod}el) = \frac{\sum_{square \in rect} match(square, cell)}{m * n} \quad (2.12)
$$

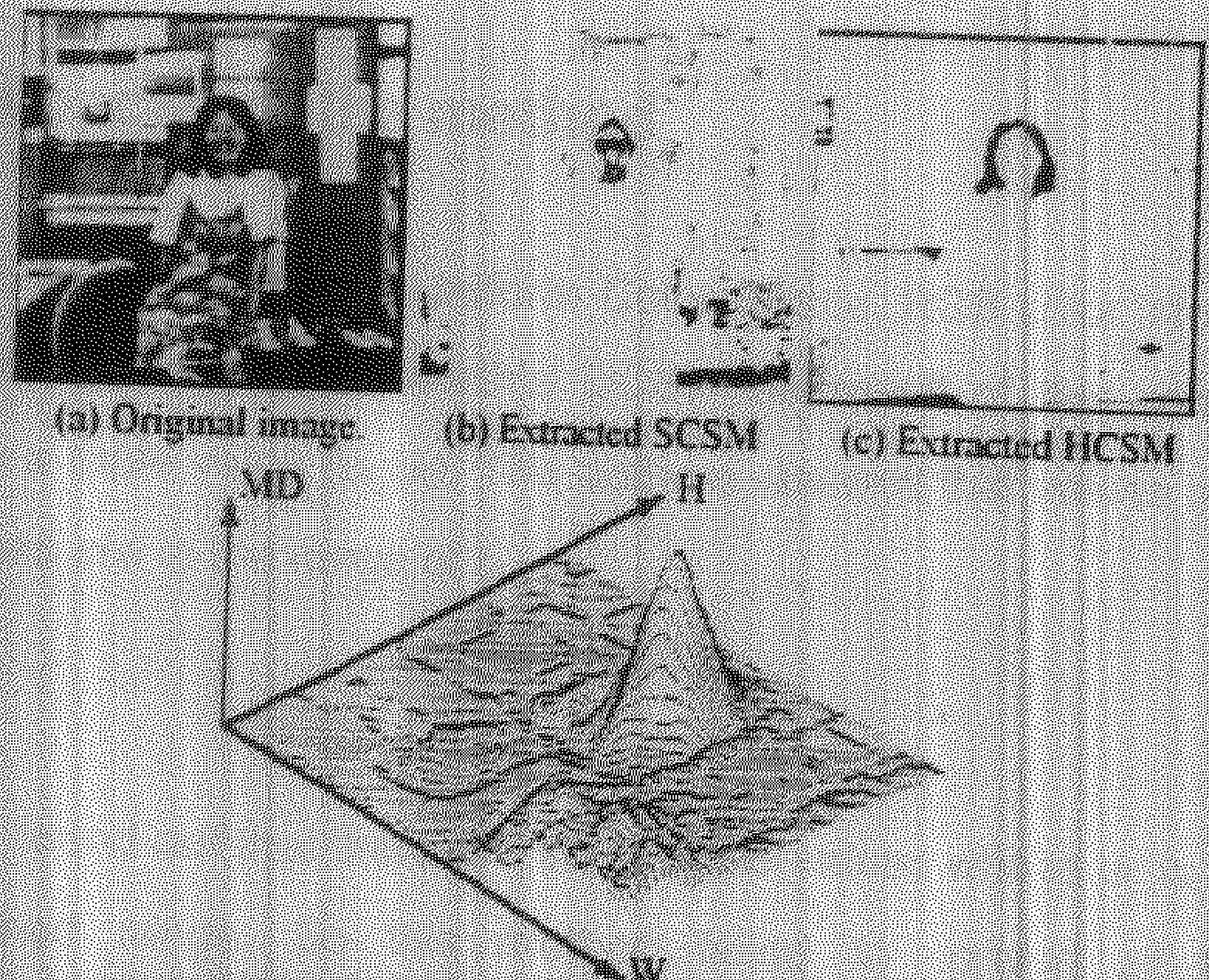
where m and n are the number of rows and columns of cells in a head-shape model.

To detect faces with different sizes and various poses, they compared all rectangular regions of the given size with the head-shape models. Each rectangular region is divided into m × n square sub-regions, each of them corresponding to a cell in the head-shape model. In this research, they let the size of the square sub-region vary from one pixel to N pixels, where

$$
N = \frac{\text{The height of image}}{\text{The number of cells in a column of the head shape model}} \quad (2.13)
$$

Therefore, there will be multiple matching degrees at the same position in an image. Each of them describes the matching degree of a rectangular region of a particular size. They selected the one with the highest matching degree as the matching degree at the position and build a matrix to record the matching degree at all positions in an image. They called this matrix as *map of matching degree*, or simply MMD. It also carries the information about the size of the rectangle region and the kind of the head-shape model giving that matching degree.

They let a and b in (11) vary from 0.1 to 10, and use them to estimate the MMDs for various images. Then they selected a = 0.5 and b = 1 that give the best MMDs. The "best" here means that the MMD shows high values in the face regions and low values in the none-face regions.



(a) Original image. (b) Extracted SCSM (c) Extracted HCSM

(d) The MMD obtained by comparing the extracted SCSM & HCSM.

Figure 2.5 : An MMD obtained by comparing the skin-color similarity map and the hair-color similarity map with the head-shape models.

Fig.2.5(a) shows an input image. Fig.2.5(b) and (c) are gray-scale images that indicate the SCSM and HCSM estimated from 2.5(a), respectively. Fig.2.5(d) shows the MMD. In the figure, the W-axis and the H-axis are the horizontal axis and the vertical axis of the image plane, and the MD-axis indicates the matching degree. One can see a "mountain" appearing at the position where the face exists. The matching degree (Match(rect, model)) in equation (2.12) can be considered as the likeness between the rectangular part (rect) and a face. They treated the rectangular image regions having the likeness greater than a given threshold as face candidates. They used 0.7 as the threshold in their research. All the local maximum of MMD greater than this threshold value are considered as face candidates. A higher threshold value will increase the reliability of the detected faces, but will also fail to detect some real faces. On the other hand, if a lower threshold is used, all faces in an image may be detected successfully, but many non-face regions may also be detected as faces.

2.4 Drawbacks of the method

1) This method used Farnsworth Uniform Color Space. RGB color information of a pixel was converted to Farnsworth's UCS. The result of this conversion is represented by a tuple value (u_f, v_f). The authors claimed that values of (u_f, v_f) of **all visible colors** are in the range of :-

$$u_f \text{ -> } [0, 91]$$

$$v_f \text{ -> } [0, 139]$$

We in our work found out that there are some RGB values whose u_f is not within the above-defined range. So this method failed to convert all the color information to Farnsworth's UCS.

2) **Occlusion** :- If a face is largely occluded, the cells in head-shape models corresponding to the occluded part of the face will give low output, thus the total matching degree may not be high enough to let the face be detected.

3) **Adjacent faces** :- If two or more faces are too close, the skin parts or hair parts of them may be merged together. The shape of the resulting skin-hair pattern may be very different from the one for a single face.

4) **Hairstyle** :- Faces with special hairstyles, such as skinhead, or wearing a hat, may fail to be detected. This is because the shape of the skin-hair pattern of such a face in the image may become quite different from our head-shape model.

3 Proposed Method for Face Detection

In our work, we implemented the existing method (Method-2) for face detection as explained in section 2, and tried to find out the problems with that method and improve that method. We tested the method on a database of images.

3.1 Problems with Farnsworth UCS

The term skin color and hair color are subject to human concepts. So color representation should be similar to the color sensitivity of human eyes to obtain a stable output similar to the one given by human visual system. Such a color representation is called the perceptually uniform color system or UCS. The authors in Method-2 claimed that values of (u_f, v_f) of **all visible colors** are in the range of :-

$$u_f \rightarrow [0, 91]$$
$$v_f \rightarrow [0, 139] \qquad (3.1)$$

We in our work found that there are some RGB values whose u_f is not within the above-mentioned range. For example the pixel having RGB value as (28,0,0) gives the value of u_f to be 104. There are several other RGB values, the u_f value of which exceeds the above-defined range.

The authors in Method-2 prepared skin color distribution model and hair color distribution model, which are tables of size 92 x 140 and 92 x 140 x 256 respectively. As the value of u_f exceed 92, so we increased the size of the models to 110 x 140 and 110 x140 x 256 respectively. This is no doubt a drawback of Method-2, which we removed in our work.

3.2 Method for finding Head Shape Model

In Method-2 the authors did not mention the method for finding the skin proportion M_F and the hair proportion M_H, in each cell of an head shape model. We in our work devised a method for finding M_F and M_H.

First of all we selected rectangular face region, the skin part, and the hair part in it, then divide it into m x n square cells. Then we identified the cells in the image, which are completely skin (or hair) cells. We named the pixels of all those identified cells as :-

$X_1, X_2, \ldots\ldots\ldots\ldots X_m,$ for hair cells

$Y_1, Y_2, \ldots\ldots\ldots\ldots Y_n,$ for skin cells

Then we determined M_F and M_H by the following equation for those cells, which have both skin and hair in it. First of all we found membership values of all the pixels in those cells for both skin($\mu_S(p)$) and hair($\mu_H(p)$) and as follows :-

$$\mu_H(p)=\frac{1}{m+n}\left(\sum_{i=1}^{m}\left(e^{-d^2(p,x_i)}\right)+\sum_{j=1}^{n}\left(1-e^{-d^2(p,y_j)}\right)\right)$$

$$\mu_S(p)=\frac{1}{m+n}\left(\sum_{i=1}^{n}\left(1-e^{-d^2(p,x_i)}\right)+\sum_{j=1}^{m}\left(e^{-d^2(p,y_j)}\right)\right) \qquad (3.2)$$

where, p is the pixel value of the concerned pixel.

3.3 Modification of S-function

The authors in Method-2 defined two fuzzy sets R_S and R_H : R_S (or R_H) is the fuzzy set of $A_S \ni a_S$ (or $A_H \ni a_h$), which is defined by a fuzzy membership function μA_S (or μA_H) :-

$$\mu_{A_S} : R_S -> [0,1]; \quad \text{or} \quad \mu_{A_H} : R_H -> [0,1]; \tag{3.3}$$

R_S (or R_H) was used to describe the relationship between the average skin (or hair) color similarity a_S (or a_h) and the proportion of the skin (or hair) color part in a square region of the input image. Two S type standard functions were used to represent μA_S and μA_H. An S type standard function is defined by the following equation :-

$$S(x;a,b) = \begin{cases} 0 & x \leq a \\ \dfrac{2(x-a)^2}{(b-a)^2} & a \langle x \leq \dfrac{(a+b)}{2} \\ 1 - \dfrac{2(x-b)^2}{(b-a)^2} & \dfrac{(a+b)}{2} \langle x \leq b \\ 1 & b \langle x \end{cases} \tag{3.4}$$

where $0 \leq a \leq 1$, $0 \leq b \leq 1$, and $a \leq b$. The parameters 'a' and 'b' control the shape of the function.

We, in our work introduced a generalized S-function [30], which gave better result than Method-2. The generalized S type function is defined by the following equation :-

$$S_k(x) = \frac{2^k}{2}\left(\frac{x}{c}\right)^k \qquad 0 \leq x \leq \frac{c}{2}$$

$$= 1 - \frac{2^k}{2}\left(\frac{c-x}{c}\right)^k, \qquad c/2 \leq x \leq c \tag{3.5}$$

For k = 2, Equation (3.5) reduces to that of Equation (3.4). Equation (3.5) can therefore be viewed as a generalization of Zadeh's S function.

In this work we chose 'c' to be 0.6 for μA_S, and 0.75 for μA_H. These were determined through experiments so that the proportions of the skin color area and of the hair color area given by the functions become similar to the one given by human viewers. We applied $S_k(x)$ on several input images for the above-mentioned value of c and for different values of k. We found that for k = 2.5 it gives better result.

3.4 Modification of distance function

In Method-2, to estimate the similarity between a square region in the input image and a cell in a headshape model, a method was used to compare the properties of the square region in the image (R_S and R_H) and the properties of the cell in a head-shape model (M_F and M_H). In the fuzzy set theory, the degree of similarity between square regions in the image and the cells in the headshape model can be calculated with the following equation,

$$\text{match(square, cell)} = e^{-a\left(\left(R_S - M_F\right)^2 + \left(R_H - M_H\right)^2\right)^{0.5b}} \tag{3.6}$$

We introduced a new distance function defined as :-

$$\text{match(square, cell)} = \frac{e^{-\left(\frac{d}{1+d}\right)^a} - e^{-a}}{1 - e^{-a}} \tag{3.7}$$

where,

$$d = \left(\left(R_S - M_F\right)^2 + \left(R_H - M_H\right)^2\right)^{0.5} \tag{3.8}$$

The distance function (equation 3.6) used in Method-2 consist of two parameters- 'a' and 'b' and there values have to be selected. The distance function we have used (equation 3.7), consist of only one parameter. So this is more advantageous than Method-2.

4 Experiments and Results

We built SCDM and HCDM, for our database of images, which consist of faces of 100 different people of a single community. We considered faces of both men and women while building SCDM and HCDM.

We used seven image sequences (five men and two women) to build the head-shape models.

We tested our method for still color images, which are not in the image database used for building SCDM and HCDM. The images consist of indoor scenes under fluorescent light, and under the mixed illumination of fluorescent light and sunlight, as well as outdoor scenes. The face size varied from 20 x 24 pixels to 200 x 240 pixels. The experimental results are summarized in Table 4.1.

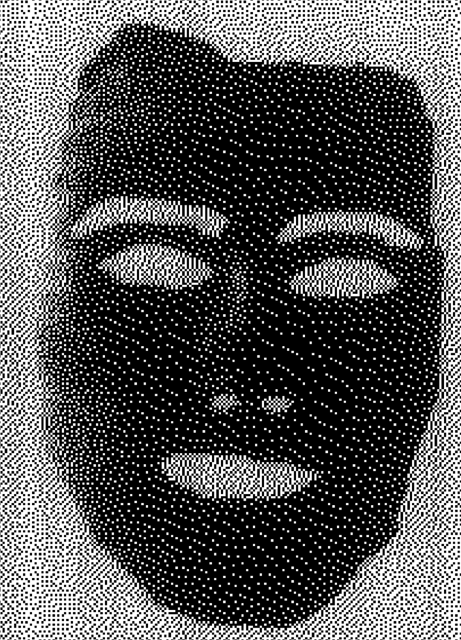
Face size (pixels)	Number of faces	Correctly detected faces	False positive	Detection rate (%)
Bigger than 100 x 120	10	9	1	90
Bigger than 50 x 60	10	8	2	80
Bigger than 20 x 24	10	6	4	60

Table 4.1 : Experimental Results

As Method-2 considers the range of u_f to be [0, 91], so this method fails for those images, which have the value for u_f exceeding 91. So this method fails completely. But as we removed this drawback, so our method detects faces for any image having any u_f value.

We compared our method with that of neural network method [31] on the same set of images and found our method to be faster and accurate in detecting faces. When neural network method takes 34 secs and detect 8 out of 10 faces, our method runs in less than a second and detect faces more accurately. So our method is faster and accurate in detecting faces.


(a) The original image


(b)Selected skin region

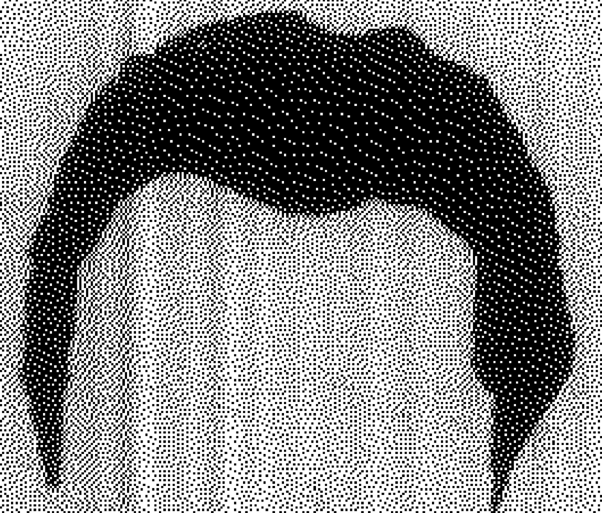
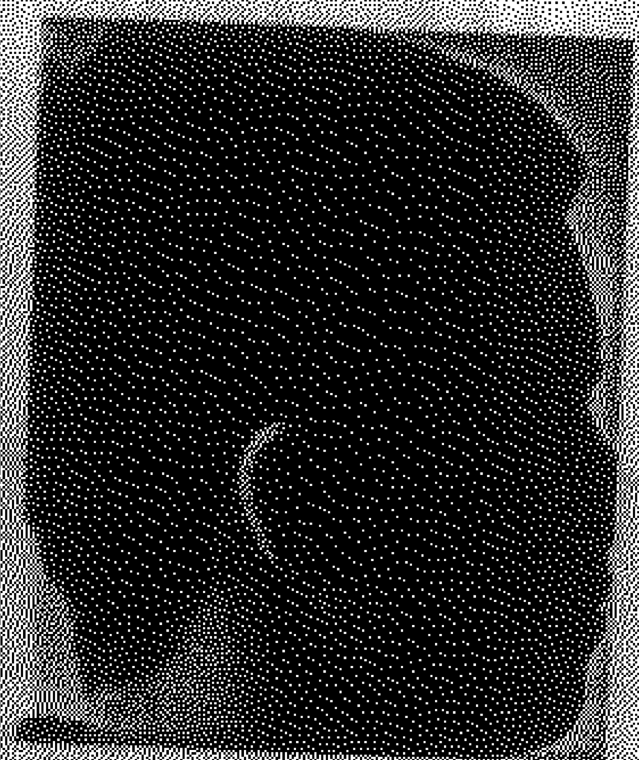

(c)Selected hair region

Figure 4.1 : An image used to build the SCDM and HCDM

Figure 4.1(a) shows an original image from which we selected the skin region (Figure 4.1(b)) and the hair region (Figure 4.1(c)) for building SCDM and HCDM.


(a)


(b)

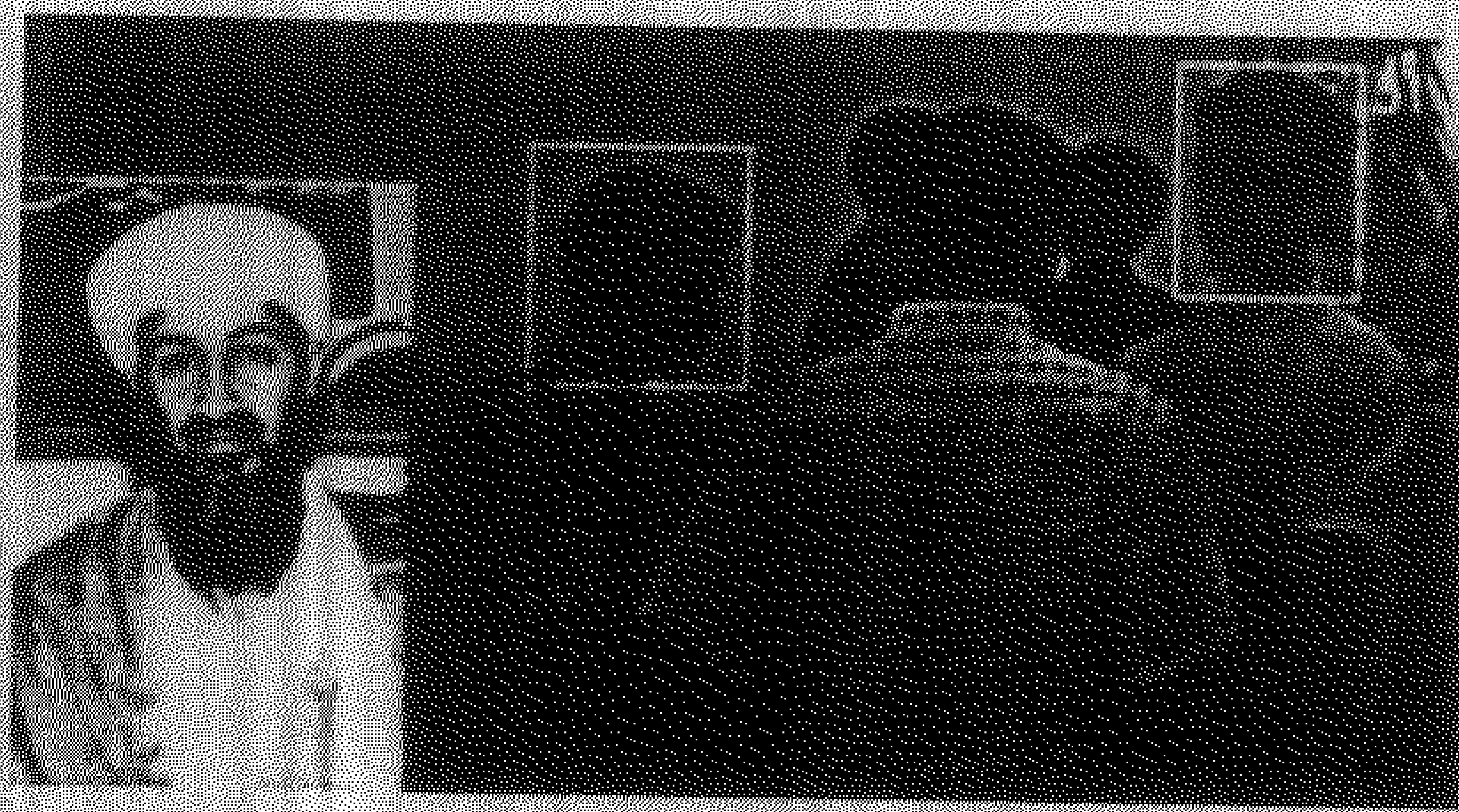

(c)


(d)

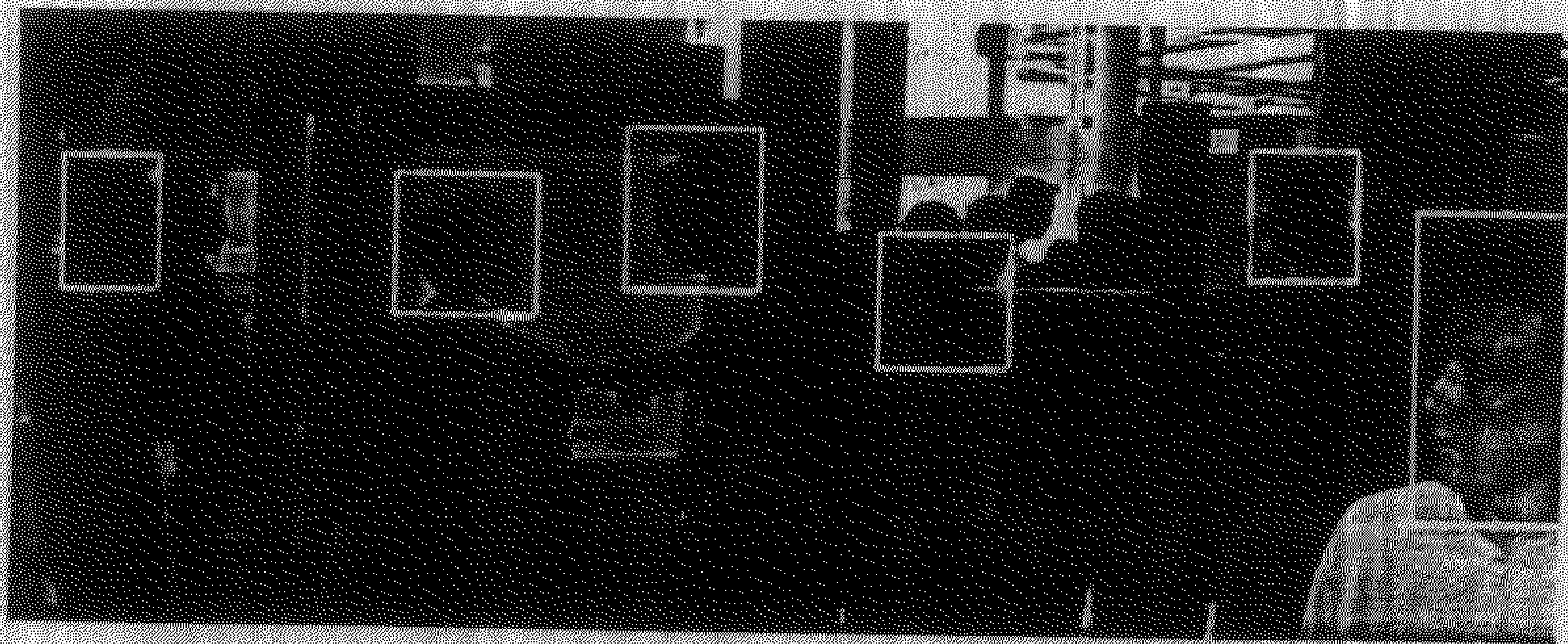

(e)

Figure 4.2 : The primitive head-shape models

Figure 4.2 shows the appearance of faces in images into 5 kinds of pattern:- frontal view (Figure 4.2(c)), left side view (Figure 4.2(e)), right side view (Figure 4.2(a)), left diagonal view(Figure 4.2(d)), and right diagonal view(Figure 4.2(b)). We used seven image sequences (five men and two women) to build the head-shape models.



(a)



(b)

Figure 4.3 : Experimental results of face-candidate detection.

Figure 4.3 show the results of our face detection method. Figure 4.3(a) contains the face of Osama Bin Laden with long beard. Our method failed to detect his face as his face is with long beard and does not match with any of our head shape model. It also failed to detect the face of the person who is not facing the camera. Figure 4.3(b) contain some faces which are very small in size and our method failed to detect those faces.

5 Conclusion, Discussion and Scope for further work

Face detection is a challenging problem and there is still a lot of work that need to be done in this area. Detection of human faces in color images is our scope of work.

In this report we have described an approach to detect the face in images. As our method is an upgraded version of Method-2, so this is more efficient in detecting faces than that of Method-2. Here we have increased the size of "Skin (and Hair) Color Distribution Model", so this method works for any image having any RGB value. Infact this method works for the images for which Method-2 fails. We also used an S function that we can control by changing the value of 'k'. So this is more generalized method than that of Method-2. Rest of the advantages of Method-2 is also applicable to our method.

Because we use a perceptually uniform chromatic system and the fuzzy set theory based models to describe the skin color and the hair color, our method can detect skin regions and hair regions much more accurately and stably than conventional approaches. It helps both increases the face detection rate and reduces the false positive rate. In Method-2 a new pattern recognition method was developed, called fuzzy pattern matching, which makes possible the pattern detection using a pattern description model carrying different kinds of information from the input image. This gives the flexibility in designing the head-shape model. Thus we could create head-shape models that describe the essential features of the head shape. All these make the face like patterns distinctive from others.

Compared with the existing face detection approaches, the skin color detection method is much more accurate and efficient. The multiple head pose models allow us to detect faces of various poses. By not looking at the details of facial features, our method will not be affected by small, local changes in a face. Therefore, the proposed approach is not sensitive to image noise or the change of facial expression and head-pose and is very robust. The experimental results showed that our approach could detect faces successfully in an uncontrolled environment with complex background. Compared with neural network based approaches, this method is much faster and the performance is good.

As this method fails for images of small size, so a new method can be devised to meet the above constraint. If two or more faces are too close, the skin parts or hair parts of them may be merged together. The shape of the resulting skin-hair pattern may be very different from the one for a single face. Faces with special hairstyles, such as skinhead, or wearing a hat, may fail to be detected. This is because the shape of the skin-hair pattern of such a face in the image may become quite different from our head-shape model. This method also fails for those faces with long beard. We tested our method for the image of Osama Bin Laden. The method failed to detect the face. This can be removed if we devise a method for detecting beard and then remove the beard so that we get the shape of the head as in our head shape model. Then our method can work in detecting these kinds of faces. So these can be the scope for further work.

BIBLIOGRAPHY

[1] Q. Chen, H. Wu and M. Yachida, "Face Detection by Fuzzy Pattern Matching", Proc. 5^{th} ICCV, pp. 591–596, 1995.

[2] G¨unter Wyszecki and W.S. Stiles, "COLOR SCIENCE", John Wiley & Sons, Inc, 1967.

[3] S.Y.Lee, Y.K.Ham and R.-H. Park, "Recognition of human front faces using knowledge-based feature extraction and neuro-fuzzy algorithm", Pattern Recognition, Vol.29, No.11, pp. 1,863-1,876, 1996.

[4] C.H.Lee, J.S.Kim and K.H.Park, "Automatic human face location in a complex background using motion and color information", PR, 29(11):pp.1877-1889, November 1996.

[5] X.Jia and M.S.Nixon, "Extending the feature vector for automatic face recognition", IEEE Trans. on PAMI, Vol.17, No.12, pp.1167-1176, 1995.

[6] E.Saber and A.Tekalp, "Face detection and facial feature extraction using color, shape and symmetry based cost functions", In ICPR'96, pp.654-658, 1996.

[7] P.Sinha, "Object recognition via image invariants: A case study", In Investigative Ophthalmology and Visual Science, Volume 35, pp.1735-1740, Sarasota, Florida, May 1994.

[8] P.Juell and R.Marsh, "A hierarchical neural network for human face detection", Pattern Recognition, Vol.29, No.5, pp.781-787, 1996.

[9] Y.H.Kwon and N.da Vitoria Lobo, "Face detection using templates", In ICPR'94, pp.764-767, 1994.

[10] A.J.Colmenarez and T.S.Huang, "Face detection with information based maximum discrimination", Proc. CVPR'97, pp.782-787, 1997.

[11] H.A.Rowley, S.Baluja and T.Kanade, "Neural network-based face detection", In CVPR'96, pp.203-208, 1996.

[12] M.Lew and N.Huijsmanns, "Information theory and face detection", In ICPR'96, pp.601-605, 1996.

[13] T.Leung, N.Burl and P.Perona, "Finding face in cluttered scenes using labeled random graph matching", In ICCV'95, pp.637-644, 1995.

[14] R.Vaillant C.Monrocq and Y.Le Cun, "Original approach for the localization of objects in images", IEEE Proc. Vis. Image Signal Processing, 141(4), August 1994.

[15] Y.Dai Y.Nakano, "Face-texture model-based on sgld and its application in face detection in a color scene", PR, 29(6):pp.1007-1017, June 1996.

[16] E.Osuna, R.Freund and F.Girosi, "Training support vector machines: An application to face detection", Proc. CVPR'97, pp.130-136, 1997.

[17] A.Lanitis, C.J. Taylor, T.F.Cootes and T.Ahmed, "Automatic interpretation of human faces and hand gestures using flexible models", Proc. Intl. Workshop on Automatic Face- and Gesture-Recognition, pp.98-103, 1995.

[18] [85] H. Wu, Q. Chen, and M. Yachida, "Face detection from color images using a fuzzy pattern matching method," *IEEE Trans.* Pattern Analysis and Machine Intelligence, vol. 21, pp. 557–563, June 1999.

[19] H. Wechsler, P. Phillips, V. Bruce, F. Soulie, and T. Huang, Eds., *Face Recognition: From Theory to Applications*, Springer-Verlag, 1998.

[20] R. Chellappa, C.L. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey," *Proc. IEEE*, vol. 83, pp. 705–740, May 1995.

[21] W. Zhao, R. Chellappa, A. Rosenfeld, and P.J. Phillips, "Face recognition : A literature survey," *CVL Technical Report, Center for Automation Research, University of Maryland at College Park*, Oct. 2000, <ftp://ftp.cfar.umd.edu/TRs/CVL-Reports-2000/TR4167-zhao.ps.gz>.

[22] *Faceblind.org*, <http://www.faceblind.org/research/index.php>.

[23] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Contentbased image retrieval at the end of the early years," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349–1380, Dec. 2000.

[24] V. Bruce and M. Burton, Eds., *Processing Images of Faces*, Ablex publishing, Norwood, NJ, 1992.

[25] J. Weng and D.L. Swets, "Face recognition," in *Biometrics: Personal Identification in Networked Society*, A.K. Jain, R. Bolle, and S. Pankanti, Eds., pp. 67–86, Kluwer Academic, Boston, MA, 1999.

[26] H.D. Ellis, M.A. Jeeves, F. Newcombe, and A. Young, Eds., *Aspects of Face Processing*, Martinus Nijho® Publishers, Dordrecht, Netherlands, 1985.

[27] O.A. Uwechue and A.S. Pandya, Eds., *Human Face Recognition Using Thirdorder Synthetic Neural Networks*, Kluwer Academic Publishers, Norwell, MA, 1997.

[28] P.L. Hallinan, G.G. Gordon, A.L. Yuille, P. Giblin, and D. Mumford, *Two- and Three- Dimensional Patterns of the Face*, A.K. Peters, Natick, MA, 1999.

[29] S. Gong, S.J. McKenna, and A. Psarrou, *Dynamic Vision: From Images to Face Recognition*, Imperial College Press, London, 1999.

[30] C.A. Murthy and S.K. Pal, *Histogram thresholding by minimizing gray level fuzziness, {Information Science}*, vol. 60, no. 1/2 , pp. 107-135, 1992.

[31] R. F´eraud, O.J. Bernier, J.-E. Viallet, and M. Collobert, "A fast and accurate face detection based on neural network," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 1, pp. 42–53, Jan. 2001.

. . . .