Face Detection Based on Similarity-based Clustering Algorithm

Wen-Hui Lin [1], Yui-Lang Chen [2], [1]Jhen-Chih Liao, [1]Kuo-Lung Wu

[1]*Department of Information Management*
Kun Shan University
Tainan, Taiwan
[2] *Department of computer Science and Information Engineering*
National Chung Cheng University
Chia-Yi, Taiwan
[1]*linwh@mail.ksut.edu.tw,* [2] *vtmtcic@hotmail.com*

ABSTRACT

There are many methods proposed for human face detection, but the size of filtering mask for detecting face is still a difficult problem now. In this paper, an adaptive face detection method is proposed to precisely detect faces in image with a variety of face size and contaminated by the noise such as the non-face skin-color objects, arms, object's color similar to skin-color, wearing clothes and background object, and some of these faces overlapped. The adaptive face detection method is composed by four steps. The first, based on skin-color classification in color space system, the skin-color pixel features include both its position and color information are extracted and used to classify the skin-color pixel to generate candidate skin-color region frames by a similarity-based clustering method (SCM). By considering the face regions property the region frames can be merged, partitioned and removed to get the optimum face frame. Then a frame integration algorithm is used for merging frames if they belong to the same face. Next, a frame segmentation algorithm can be used to partition different faces in the same region and generate the optimum of face boundaries. Finally, after performing the three algorithms above, the candidate face regions will be found by rejecting most framed regions if the ratio of height to weight is over than 2.3. The detection of the face regions in a color image can be effectively achieved by performing an appearance-based method with spectral histograms as representation and support vector machines (SVMs) as classifiers.
Keyword: face detection, skin-color classification, a similarity-based clustering method (SCM).

1: INTRODUCTION

The task of finding suitable face regions in an image with a variety of human's face size is difficulty to implement. Garcia and G. Tziritas [8] used quantized skin color regions to iteratively merge on the set of homogeneous skin-color regions in color quantized images, but the method doesn't give reasonable rules and sensible definitions. Moreover, the performance of their results will give too many framed regions, which include a few face regions and many non-face regions, for face detection. Saad et al. [13] proposed a simply method of faces detection in a color natural scene using skin-color classification and template matching, but a serious weakness that face regions would be missed if they are connected with the other skin-color areas like arms, human wear clothes and background. Waring and Liu [14] present a face detection method using spectral histograms and support vector machines. Their discrimination function is designed on a 21×21 image window, it implicitly requires that all faces lie roughly within a 21×21 window. So a three-level Gaussian pyramid they suggest is built by down sampling the test image by a factor of 1.1. However, about this approach, the face regions would also be lost if the size of these face regions are bigger or smaller over than an approximate 21×21 image window. Hence, in this paper, an adaptive face detection method is proposed to precisely detect faces in image with a variety of face size and contaminated by the noise such as the non-face skin-color objects, arms, object's color similar to skin-color, wearing clothes and background object, and some of these faces overlapped.

2: COLOR MODELS FOR SKIN-COLOR CLASSIFICATION

In face detection, preprocess based on skin-color classified can effectively provide probable of face locations in color images. However, a Bayesian approach will give more effective and precision in setting skin-color distribution from ample training data to skin-color classify.

2.1: YCbCr COLOR SPACE

About YCbCr color space [2], it has been defined in response to increasing demands for digital algorithms in handling video information, and has since become a widely used model in a digital video. The luminance component Y has an excursion of 219 and an offset of +16. And the chrominance components Cb and Cr have excursions of +112 and offset of +128, producing a range from 16 to 240 inclusively. However, in YCbCr color space, the Bayes decision rule for minimum cost [3] can be used to classify sample I into skin color class (ω_1 or 1) and non-skin color class (ω_2 or 0). The Bayes decision rule for minimum cost is expressed as

$$p(I\,|\,\omega_1)/p(I\,|\,\omega_2) > T \quad \Rightarrow \quad X \in \omega_1 \qquad (1)$$

$$p(I\,|\,\omega_1)/p(I\,|\,\omega_2) < T \quad \Rightarrow \quad X \in \omega_2 \qquad (2)$$

and

$$p(i\,|\,\omega_1) = Cs_i / Ts \qquad i \in I \qquad (3)$$

$$p(i\,|\,\omega_2) = Cn_i / Tn \qquad i \in I \qquad (4)$$

where T is a threshold refers to [1]. Cs_i represents the number of skin color at position of Cb and Cr, and Ts is the total number of skin color from all samples in the color space. Correspondingly, Cn_i represents the number of non-skin color at position of Cb and Cr, and Tn is the total number of non-skin color from all samples in the color space.

In addition to classify skin-color by this Bayesian approach, there were equations already defined the bounding planes in many researches in YCbCr color space [7-8]. So, it is also a convenient for researchers to skin-color classify without consuming most of time in training skin-color map.

2.2: HSV COLOR SPACE

Beside YCbCr color space, HSV color space is also a main concern in skin-color classification. The HSV (hue, saturation, and value) model [4] is commonly used in computer graphics applications. It also known as HSB (hue, saturation, brightness) and defined a color space in terms of three constituent components. Hue is the color type, such as red, blue, or yellow, and ranges [0, 360]. Saturation is the vibrancy of the color, intensity of a specific hue. It is based on the color's purity. A highly saturated hue has a vivid, intense color, while a less saturated hue appears more muted and grey. With no saturation at all, the hue becomes a shade of grey. Value is the brightness of the color, similar as luminance component Y of the YCbCr color space.

For skin-color classification, HSV color space is also a main concern. The skin-color region in HSV color space can be searched out as above approach. Or also there are equations already defined the bounding planes in HSV color space [8-9]. The skin-color pixels, be classified in an image, is then denoted as $\{d_i\}_{i=1}^{Ts}$ for Similarity-based Clustering Algorithm.

3: SKIN-COLOR FEATURES EXTRACTION

The feature set for a skin-color pixel is selected as $X = \{position, color\}$, where the position and the color respectively indicate as, $\{v,h\}$ and $\{\{S \times \cos H, S \times \sin H\}, \{Cb, Cr\}, Y\,or\,V\}$. The $\{v,h\}$ subset of the feature is the vertical and horizontal coordinate component of a skin-color pixel. The $\{S \times \cos H, S \times \sin H\}$ is the Euclidian distance vector in cylinder of HSV color space. The $\{Cb, Cr\}$ components in YCbCr color space are also selected as features, due to obtain sufficiency information for classification. The feature vector composed by seven components $\{\{v,h\}, \{S \times \cos H, S \times \sin H\}, \{Cb, Cr\}, Y\}$ is selected to represent the property of a skin-color pixel in a skin-color image. For getting higher speed in the feature classified algorithm and reduce noise, it is necessary to preprocess the size of original image, $M \times N$, by downsizing uniformly with a block 8×8 then the resulted image size become to $M/8 \times N/8$. The feature vector for a skin-color pixel in this resulted image must be recalculated. Setting a threshold value τ_1 for deciding each block whether it is a skin-color or not when a block downsize into a pixel? Let bn denote the number of all blocks and the amount of skin-color pixels in m'th block as bs_m. Then the feature vector for a downsized skin-color pixel, 8×8 block in original skin-color image, can be recalculated by the following process:

Skin-color Features Extraction
Begin
 $j = 0$;
 For $m = 1$ to bn
  If ($count(bs_m) > \tau_1$)
  {
   $j = j + 1$;
   $x_j = \{\{v,h\,|\,\text{the center position in } m\text{'th block}\},$
     $\{\{E[S \times \cos H], E[S \times \sin H]\},$
     $\{E[Cb], E[Cr]\},$
     $E[Y\,or\,V]\} \in bs_m\}$;
  }
  $n = j$;
End

4: THE ADAPTIVE FACE DETECTION METHOD

The adaptive face detection approach that consists of feature classification algorithm, frame integration algorithm and frame segmentation algorithm is developed to effectively detect the candidate face regions

in an image with variety of human's faces and counteract noise.

4.1: SIMILARITY-BASED CLUSTERING ALGORITHM FOR SKIN-COLOR PIXELS

For face detection, the skin-color pixels is classified by applying the extracted pixel feature vector to a similarity-based clustering method (SCM) [5] to find skin-color regions frame. Before implementing clustering, the features must be respectively standardized because the essence and measured units of these variances are diverse to mutually comparison. Let x_j denote the j'th skin-color pixel feature

$$x_j = \left\{ v_j, h_j, (S \times \cos H)_j, (S \times \sin H)_j, Cb_j, Cr_j, Y_j \right\}, (5)$$

and x_{jk} represents the k'th element of the j'th feature vector and the standardized feature elements x_{jk}^s are noted as

$$x_{jk}^s = (x_{jk} - \bar{x}_k) / \sigma_k \quad \forall j = 1, \cdots, n \text{ and } k = 1, \cdots, 7. (6)$$

where $\bar{x}_k$ and σ_k are the mean and the standard deviation of k'th element, and n is represents the total number of clustering skin-color pixels.

After standardization, there are five elements in the color components, those contain $\{S \times \cos H\}$, $\{S \times \sin H\}$, $\{Cb\}$, $\{Cr\}$, and $\{Y\}$ and each of them is respectively standardized to 1 unit. In the position components, only two elements $\{v\}$ and $\{h\}$ are standardized to one unit respectively. Thus the total quantities of color and position elements are not equal. Owing to balance the influence between both position and color component, we have to adjust position element value as follow:

$$y_j^t = \left\{ 2.5 \times \{v_j, h_j\}^s, (S \times \cos H)_j^s, (S \times \sin H)_j^s, Cb_j^s, Cr_j^s, Y_j^s \right\} (7)$$

In order to facility the representation of the features vector, y_j will be redenoted as

$$y_j = \left\{ v_j'^s, h_j'^s, (S \times \cos H)_j^s, (S \times \sin H)_j^s, Cb_j^s, Cr_j^s, Y_j^s \right\}^T$$
$$v' = 2.5v; \ h' = 2.5h, \quad \forall j = 1, \cdots, n. \quad (8)$$

For accuracy and efficiency, the principal component of features y_j can be extracted by a generally method, Karhunen-Loeve Transform [6]:

$$\Lambda = \Phi^T \Sigma \Phi, \tag{9}$$

and the covariance Σ will be computed by

$$\Sigma = \sum_{j=1}^{n} y_j \cdot y_j^T. \tag{10}$$

Then the principal component vector y_j' of the feature vector y_j can be obtained by projecting it onto an orthogonal space.

$$y_j' = \Phi^T \cdot y_j. \tag{11}$$

When refining the extracted skin-color pixel feature vector, a similarity-based clustering method based on a total similarity objective function $J_s(z)$ related to the approximate Gaussian density shape estimation is used to effectively classify skin-color pixel features to find out candidate face regions frame in an image.

$$J_s(z) = \sum_{i=1}^{c} \sum_{j=1}^{n} f\left(S\left(y_j', z_i\right)\right), \tag{12}$$

where $\{z_i\}_{i=1}^{n}$ are cluster centers and $f(.)$ is a monotone increasing function. The similarity relation $S\left(y_j', z_i\right)$ is set up with

$$S\left(y_j', z_i\right) = \exp\left(-\beta^{-1} \cdot \left\|y_j' - z_i\right\|^2\right), \tag{13}$$

where β as a sample variance is the normalized term.

By clustering algorithm, all face would be roughly framed and the performance was shown by a sample result as Fig. 1. And in order to get the better result, the optimum of framed boundaries, candidate face regions would be further treated by follow algorithms.

4.2: FRAME INTEGRATION ALGORITHM FOR MERGING REGIONS

After performance by clustering algorithm, many of faces in image were precisely framed for fit regions, but there is an exception if the factual face area was bigger to a certain extent. It would cause the factual face area to be multi-framed at the same face because the location was a main distance about clustering element. A sample result is shown in Fig. 2 (a). Hence, we simply propose a frame
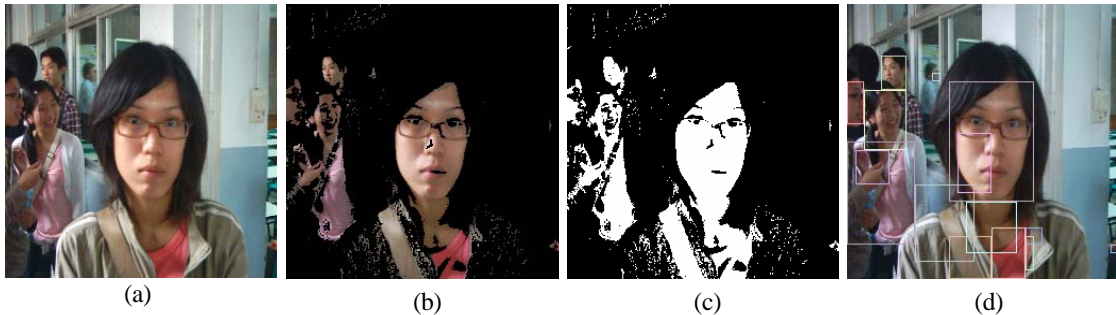

(a)　　　　(b)　　　　(c)　　　　(d)

Fig 1. A sample test by a similarity-based clustering algorithm. (a) The original image; (b) The skin color image based on union of YCbCr and HSV color spaces; (c) The Bitmap; (d) Result after the algorithm.

Fig 2. A sample test for candidate face regions. (a) The result by clustering algorithm; (b) The performance after frame integration algorithm; (c) Processing by frame segmentation algorithm; (d) The final result after forsaking the unlikely regions.

integration algorithm for regions union if these regions belong to the same face. The values of initial setting are our experience from a variety of images. Fig. 2 (b) shows the sample result after the frame integration algorithm.
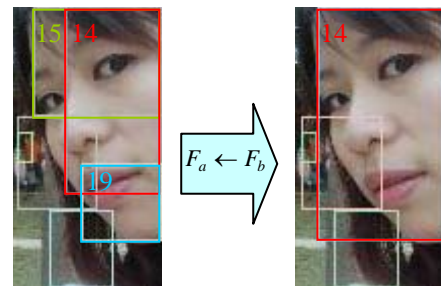
Frame Integration Algorithm

Initial: Set $\tau_2 = 1/2$, $\tau_3 = 1/5$, $\varsigma_1 = 1.5$ for follow decision rules;

Rule 1: Exists area overlap between adjacency frames.

Rule 2: Number of skin pixels in each framed should be greater than τ_2 .

Rule 3: The length of the edge of the overlapped are of the two overlapped frames must larger than τ_3 .

Rule 4: If the above rules satisfied, the crucial rule which according to the statistic of empirical rule based on bell shaped, Fig. 2.5, is mainly aimed at deviation of color relation, sets of $\{Y\}$, $\{Cb\}$, $\{Cr\}$, $\{S \times \cos H\}$, and $\{S \times \sin H\}$. First, concerning about brightness set $\{Y\}$ between F_a and F_b, the mean and standard deviation of F_a and F_b must be computed, respectively denote as μ_{ya} , σ_{ya} and μ_{yb} . About F_a, value of μ_{yb} should be limited between $\mu_{ya} \pm \varsigma_1 \cdot \sigma_{ya}$.
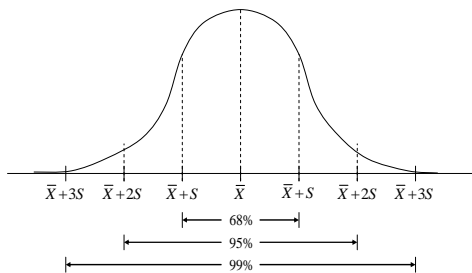

Fig 3. The statistic of empirical rule about bell shaped, where $\overline{X}$ and S are represent the mean value and standard deviation.

And then the other restrictions in color sets of $\{Cb\}$, $\{Cr\}$, $\{S \times \cos H\}$ and $\{S \times \sin H\}$ are same as above, $\{Y\}$.

Rule 5: If through all rules above were accepted, we can use a union list, as Fig. 4, to record $F_a \leftarrow F_b$, which represent $F_b \in F_a$.

Rule 6: Combining regions according to the union list.

F_a	3	8	12	15	19	20	24	26	30
F_b	5	5	17	14	14	16	5	17	16

Fig 4. A test of the union list.

4.3: FRAME SEGMENTATION ALGORITHM

By way of the frame integration algorithm, most of the regions which belong to the same face would be effectively united. Then, in this section, there is an algorithm which is used to frame segmentation with candidate face regions, that maybe there are faces in a same frame caused by step of clustered or integrated approach, to give an optimum of boundaries. The analysis in this algorithm, describes as follows, mainly by the bitmap and a sample result is shown in Fig. 2 (c).

Frame Segmentation Algorithm

Preprocess: Generally, there were many non-skin color pixels (as eyes, eyebrows, nose, mouth, and too light or dark location) in the face region. These would make to partition of actual face region into independent framed areas by frame segmentation rules follow. Hence, we suggest a simple method to fill up these pieces of non-skin color pixels. For main of horizontal fractions, we can use a column mask (5×1 pixels in our experiment) to filter with the bitmap. In filtering, the step first is to search position of both upper and lower pixels, which are belong to ω_1 denote as p_1 and p_2. Then, the second step is setting ω_1 for pixels between p_1 and p_2. In the same way, for main of vertical fractions, we can use a row mask to filter with the bitmap. And then to iterate until the bitmap is no changed, the performance of a sample result is shown in Fig. 5.

Step 1: About vertical segmentation, in order to get a threshold of suitable value in each framed region, we also use a statistic of empirical rule, which is mainly aimed at deviation about each number of skin-color pixels of column in a framed region by the bitmap. So, the step first is to compute the mean and standard deviation, respectively denote as $\mu_v^{(R_a)}$ and $\sigma_v^{(R_a)}$, by each number of skin-color pixels of column $col_i^{(R_a)}$ in a'th framed region R_a. However, a notion is not to underestimate for $\sigma_v^{(R_a)}$, let set

$$\sigma_v^{(R_a)} = \left\{ \sum_{i=1}^{l_v} (col_i - \mu_v)^2 \Big/ (l_v - 1) \right\}^{1/2} \qquad (14)$$

and $l_v^{(R_a)}$ is the length of row pixels in the framed region. Then, the threshold of suitable value $\tau_v^{(R_a)}$ is given by

$$\tau_v^{(R_a)} = \mu_v^{(R_a)} - \sigma_v^{(R_a)} . \qquad (15)$$

Step 2: To consider that if $\tau_v^{(R_a)}$ becomes a smaller or greater of unreasonable value. About segmentation, the framed regions would be bound to segment when any $col_i^{(R_a)}$ is equal to zero. And the framed regions would not be segmented if $col_i^{(R_a)}$ is greater enough. Therefore, let set

$$\tau_v^{(R_a)} = \max(\tau_v^{(R_a)}, 1) \qquad (16)$$

and

$$\tau_v^{(R_a)} = \min(\tau_v^{(R_a)}, \varsigma_2 \cdot l_v) , \qquad (17)$$

where ς_2 is a fraction by l_v (our experiment set $\varsigma_2 = 1/3$).

Step 3: The framed region $\{R_a\}_{a=1}^{c^*}$, where c^* is the number of framed regions, would be segmented if position of $col_i^{(R_a)}$ smaller than $\tau_v^{(R_a)}$. Equally, we can use the same steps above for horizontal segmentation.

Step 4: Iterating between steps 1 and 3 until all framed regions are non-alteration. The optimum of candidate face regions will be found as Fig. 2 (c).

4.4: UNLIKELY FACE REGIONS DETECTION

After algorithms above for detecting of candidate face regions, the final process is to forsake unlikely regions, similar as smaller area or a wide gap about the ratio of height to width over than 2.3 to 1. A sample result is shown in Fig. 2 (d).

5: EXPERIMENTAL RESULTS

Generally, some time color light due to the departure of color of testing images. So to get more efficient results of face detection, in our experiment, we can eliminate the disturbance of color lights by "color balance suppose" [9-10] before we detect skin-color pixels use skin color classification algorithms.

The performance of the proposed method has demonstrated in Fig.2. The size of these testing images is 256×256 pixels. After processing by algorithms above, about forsaking the unlikely regions, we only use some threshold rules based on shape, similar as too small area or a wide gap about the ratio of width and height, in our experiment. But there was a fundamental relationship [13] between the number of connected object components and the number of object holes in a candidate face region called the Euler number, defined by [16], could be use to reject some survival of regions. Due to a skin region is defined as a closed region in a candidate face region, which can have at least one hole, represent eye, mouth, or etc., inside it because they are not skin-color pixels. They appear as holes inside the region, but other skin regions such as arms or legs have no holes inside them. So if the
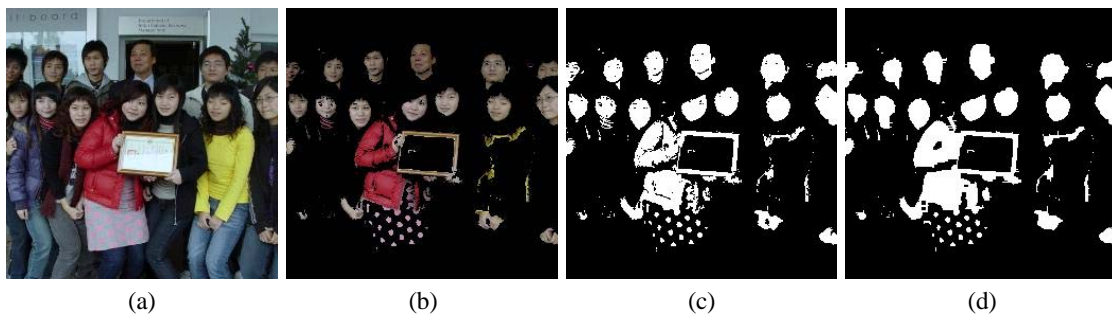

| (a) | (b) | (c) | (d) |

Fig 5. A sample result by preprocess of frame segmentation algorithm . (a) The original image; (b) The skin-color image based on union of YCbCr and HSV color spaces; (c) The Bitmap from original image;

candidate face region with no holes, it also can be rejected.

Finally, recognition of the face regions in a color image is performed by an appearance-based method using spectral histograms as representation and support vector machines (SVMs) as classifiers [14].

6: CONCLUSTION AND DISCUSSION

The uniformly down-size pre-process is first performed. For this procedure, there is a problem about precision and efficiency; the larger compression ratio can gain more efficiency for clustering algorithm but also gain the worse precision. Hence, if concerning the precision, we can downsize by using a micro-block with 4×4 pixels or lower. About the threshold τ_1 based on the block with 8×8 pixels for our experiment, the value is selected as 12 to give more faulting tolerance.

By our proposed adaptive face detection method in this paper, the suitable size of facial regions will be effectively detected in images even if the images contain skin-color objects that are not human faces, such as hands and background object, or any object near faces or these faces are overlap. The result of detected face region is more suitable in comparison with the methods proposed by J.R. Casas et al. [17] and B. Mohabbati et al. [18]. For avoiding the influence by the clustering factor, an unsupervised possibilistic clustering method [19] can be used to obtain optimal clustering result. However, most captured faces are usually distorted by a linear transformation, which typically includes scaling, rotation, shearing, and translation transformations of an object called affine transform. The affine distorted will critically affect most of decision methods. Thus, we can use some robust method [11-13] to make restored for these distorted faces. Moreover, a reliable method [14] using spectral histograms [15] and support machines (SVMs) in this paper can provide an accurate detection without considering affine distorted.

REFERENCES

[1] D. Chai and A. Bouzerdoum, "A Bayesian Approach to Skin Color Classification in YCbCr Color Space," IEEE Proceedings, TENCON 2000, vol. 2, pp. 421-424, 2000.

[2] S. K. Singh, D. S. Chauhan, M. Vatsa, R. Singh, "A Robust Skin Color Based Face Detection Algorithm," Tamkang Journal, Science and Engineering, vol. 6, no. 4, pp. 227-234, 2003.

[3] K. Fukunaga, "Introduction to Statistical Pattern Recognition, Boston: Academic Press," 2nd edition, 1990.

[4] HSV color space, http://en.wikipedia.org/wiki/HSV_color_space.

[5] M. S. Yang and K. L. Wu, "A Similarity-Based Robust Clustering Method," IEEE Trans. Pattern Analysis and Machine Intelligence, April 2004, vol. 26, no. 4, pp. 434-448.

[6] Jolliffe, I.T., "Principal Component Analysis," Springer-Verlag, New York, 1986.

[7] D. Chai and K. N. Ngan, "Face segmentation using skin-color map in videophone applications," IEEE Trans. Circuits and Syst. for Video Technol., vol. 9, pp. 551-564, June 1999.

[8] C. Garcia and G. Tziritas, "Face Detection Using Quantized Skin Color Regions Merging and Wavelet Packet Analysis," IEEE Trans. Multimedia, vol. 1, pp. 264-277, Sept. 1999.

[9] K. Wang and Q. Ruan, "Eliminate the influence of vary light conditions in face detection," In Proc. Eighth IEEE Int. Conf. On Signal Processing, vol. 2, pp. 942-945, Aug. 2004.

[10] B. Mensor, M. Bruning, "Segmentation of Human Faces in Color Image Using Connected Operator," In Proc. Seventh IEE Int. Conf. On Image Processing and its Applications IPA'99, 1999.

[11] W. H. Lin, S.W. Pan and C.W. Chang, "A Geometric Restoration Method for Objects under Affine Transforms," Computer Graphics Workshop, 2003.

[12] W. H. Lin, Y. L. Chen, "Affine Face Clustering and Recognition Based on Wavelet Features," Int. Conf. of Information Management (ICIM), 2005.

[13] E. M. Saad, M. M. Hadhoud, M. I. Moawad, M. El-Halawany, and A. M. Abbas, "Detection of Faces in A Color Natural Scene Using Skin Color Classification and Template Matching," In Proc. Twenty-Second National Conf. On Radio Science, pp. 301-308, March 2005.

[14] C. A. Waring and X. Liu, "Face Detection Using Spectral Histograms and SVMs," IEEE Trans. Man and Cybernetics, vol. 35, pp. 467-476, June 2005.

[15] X. Liu and D. Wang, "A Spectral Histogram Model for Texton Modeling and Texture Discrimination," Vision Res., vol. 42, pp. 2617–2634, 2002.

[16] William K. Pratt, "Digital Image Processing," A wiley-interscience publication (second edition), 1991.

[17] J.R. Casas, A.P. Sitjes, and P.P. Folch, "Mutual feedback scheme for face detection and tracking aimed at density estimation in demonstrations," IEE Proc. Vision, Image and Signal Processing, vol. 152, pp. 334-346, June 2005.

[18] B. Mohabbati and S. Kasaei, "Face Localization and Versatile Tracking in Wavelet Domain," Information and Communication Technologies (ICTTA), vol. 1, pp. 1552-1556, April 2006.

[19] M.S. Yang and K.L. Wu, "Unsupervised possibilistic clustering," Pattern Recognition, 39(1), pp. 5-21, 2006.