Facial Landmarks Localization based on Fuzzy and Gabor Wavelet Graph Matching

Resmana Lim [1,2)]
[1)] Electrical Engineering Department
Petra Christian University
Siwalankerto 121-131, Surabaya 60236, Indonesia
resmana@petra.ac.id

Marcel J.T. Reinders
[2)] Information and Communication Theory Group
Faculty of Information Technology and Systems
Delft University of Technology
The Netherlands
{ R.Lim; M.J.T.Reinders@its.tudelft.nl }

Abstract

This paper proposes a method that automatically finds human faces as well as its landmark points in color images based on a fuzzy analysis. The proposed approach first uses color information to detect face candidate regions and then uses a fuzzy analysis of the color, shape, symmetry and interior facial features. A deformable Gabor wavelet graph matching is used to locate the facial landmark points describing the face. The latter allows for size and orientation variation since the search for landmark points allows for affine transformations as well as local deformations of the Gabor wavelet graph. The search is performed using a genetic algorithm that is essential because it effectively searches the solution space. Results based on the proposed method are included to verify the effectiveness of the proposed approach.

I. INTRODUCTION

Face detection and detecting facial landmarks (such as position of eyes, nose, mouth, etc.) play an important role in face recognition systems. A number of Fuzzy approaches [1] [2] have been proposed to detect faces in color images. Fuzzy model of color, face shape and skin-hair pattern was used in [1] to detect faces, but it ignore all the details about interior facial features during the face detection, and hence this method might give some false positives under some condition. Fuzzy analysis of the face color, shape, symmetry and likelihood of face position is proposed in [2] resulting face detection in color images/sequences. These fuzzy methods provide good face detection but they do not provide facial landmarks localization. This paper proposes to improve the work of [2] by combining a deformable Gabor wavelet graph matching [3] [4] into the face detection system. The novelty of this paper lies in the combination of techniques that have been used for fuzzy face detection.

The approach first uses color information to detect face candidate regions and then uses a fuzzy analysis of color, shape, symmetry and a deformable graph matching to locate facial landmark points in these candidate regions. The graph matching between face model feature and probe image is optimized by a genetic algorithm (GA). The method is made robust against lighting variations and variations between people by representing the landmark points using Gabor filter responses.

The feasibility of our methodology for face detection has been deployed using color images with single and multiple faces of different ethnicities. The experiments show promising results under relatively wide conditions of the probe images.

The remainder of this paper is organized as follows. First, the approach to find face candidate regions based on skin color is presented in section 2. Then, the Fuzzy face characteristic analysis and the deformable Gabor graph matching are presented in section 3. In section 4, results on various images are presented. Finally, a conclusion and directions for future work are briefly covered in the last section.

II. EXTRACTION OF FACE CANDIDATE REGIONS

In the first step, skin regions are separated from non-skin regions based on color information. The main objective of detecting skin regions in an image is to reduce the search space for faces drastically. The HSV color space has been effectively used to segment color images in many applications. It turns out that it is also well suited to segment skin regions from non-skin regions. The skin color distribution of different people was found to be clustered in a relatively small area of the Hue-Saturated (HS) color space in which the Hue component appears to be the most significant feature.

The identification of candidate facial regions is determined by utilizing a priori knowledge about the skin distribution of skin colors in the HS color space. A number of skin samples extracted from 20 color images were used to determine the color distribution of human skin in the HS color space. Our samples were taken from persons of different ethnicities.

Before determining the HSV values, the skin samples were first filtered using a low-pass (mean) filter to reduce the effect of noise in the samples. The distribution of the HS values of the skin samples is represented by a 2-D Gaussian model $N(\mu, \Sigma)$, where μ represent the mean value and Σ the 2D covariance matrix. These parameters can be estimated from the manually labeled skin areas in the training images.

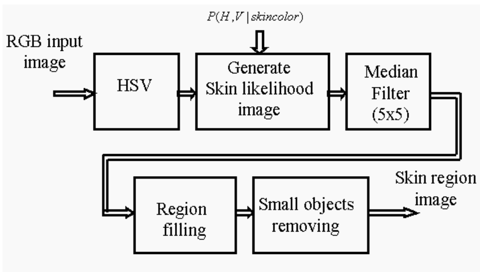

Fig. 1. Flow diagram of skin region extraction.

Having such a posterior distribution of skin color, we can now obtain the skin-likelihood image from the original image by using the Bayes rule. High values in such a skin-likelihood image thus indicate the presence of a skin region whereas low values don't. Consequently, the face candidate regions can be extracted by proper thresholding of this image. After candidate face regions are generated, the regions are post-processed to smoothen object silhouettes, and also to eliminate any isolated misclassified pixels that may appear as impulsive-type noise. Subsequently, the holes in the face are filled. Finally, small isolated regions that remain after this step are removed because they are unlikely to be face regions. Here, candidate regions having an area lower than a predefined area are removed. The flow diagram of the skin region extraction scheme then looks as shown in Fig. 1.

III. Fuzzy Analysis Of Shape, Color & Facial Features

After the face candidate regions are extracted we need to identify for each region whether it is a face or not and if so what the position of the facial landmark points are. Hereto, a number of expected facial characteristics such as color, symmetry, shapes, and facial landmarks feature similarity to a face reference model are used in the selection process. In order to quantify these values of each characteristic, Fuzzy membership functions are constructed. The value of a particular membership function provides an indication of the *fitness (similarity of fit)* of the object under consideration with the corresponding feature. Finally an overall fitness value can be derived for each object by combining the measure obtained from the individual primitives.

In our facial landmarks localization system, more specifically we consider the following 3 primitives as suggested in [2]:

- Deviation from the average hue value of the different skin-type categories.
- Face aspect ration. Given the geometry and shape of the human face, it is reasonable to expect that the ratio of height to width falls within a specific range.
- Vertical orientation. We assume that only reasonable rotation of the head are allowed in the image plane. This corresponds to a certain deviation of the facial symmetry axis from the vertical direction.

In our paper, we propose to include a similarity measure of interior facial feature between the candidate object and the model face. Hereto, we evaluate every face candidate region by matching it against a face model graph. We apply an affine graph matching procedure between the region and the face model graph. During the matching, the similarity between the transformed model and the image is maximized (by using GA) over the set of all affine transformed models in order to cope with a translated, scaled and rotated face in the probe image. For each candidate region the value of the maximal fit between the model and the region is found in this way. This fitness value is used as a primitive to the Fuzzy decision system.

Fuzzy Membership Function

We construct a number of membership function models and empirically evaluated. Here we use a trapezoidal function model for each primitive that attain the maximum value only over a limited range of input values. The membership function assume any value in the interval [0,1]. The estimates for the four membership functions are obtained by a collection of physical measurements of each primitive from a database of facial images [2].

The first primitive is the hue characteristic of the face region. As reported in [2], this function is built using the discrete universe of discourse [-20° ,50°] (i.e., $-20^{\circ} = 340^{\circ}$). The lower bound of the average hue observed in the image database is approximately 8^{o} (African-American distribution). Whereas the upper bound average value is around 30^{o} (Asian distribution). Then, the membership function associated with the first primitive is defined as follows:

$$\mu(x) = \begin{cases} \dfrac{(x+20)}{28}, if & -20^{o} \le x \le 8^{o} \\ 1 \qquad\;\;, if & 8^{o} \le x \le 30^{o} \\ \dfrac{(50-x)}{20}, if & 30^{o} \le x \le 50^{o} \end{cases} \tag{1}$$

The second primitive is face aspect ratio. We use the membership function as reported in [2] that the aspect ration (height/width) of the human face has a nominal value approximately 1.5 :

$$\mu(x) = \begin{cases} \dfrac{(x-0.75)}{0.5}, if & -0.75 \le x \le 1.25 \\ 1 \qquad\quad\;, if & 1.25 \le x \le 1.75 \\ \dfrac{(2.25-x)}{0.5}, if & 1.75 \le x \le 2.25 \\ 0 \qquad\quad\;, otherwise & \end{cases} \tag{2}$$

The vertical orientation of the face in the image is used as the third primitive. The orientation of the face (i.e. deviation of the facial symmetry from the vertical axis) is more likely to be aligned toward the vertical. A reasonable threshold selection of 30 ^{o} can be made for valid head rotation as suggested in [2]. Thus, the membership function for the primitive is describe as follows:

$$\mu(x)=\begin{cases}1 & ,if \quad 0^{o}\le x\le 30^{o}\\ \dfrac{(90-x)}{60} & ,if \quad 30^{o}\le x\le 90^{o}\end{cases} \tag{3}$$

In addition to the above 3 primitives, we propose to use a similarity measure of interior facial landmarks between face model and the object under consideration. The proposed primitive is important to be included in our knowledge-based system because it allows low false positive and also provides us locations of the facial landmarks. In the next subsection, we will show how to compute the similarity measure. In short, the candidate face regions are matched against a transformed face model using an optimization procedure of GA. The maximum fitness value from the matching process was used as the last fuzzy primitive. We found in experiment that the approximate fitness value of 0.7 is appropriate to judge whether the candidate face region constitutes a face. Thus, the membership function of the last primitive will be:

$$\mu(x)=\begin{cases}\dfrac{x}{0.7} & ,if \quad 0\le x\le 0.7\\ 1 & ,if \quad 0.7\le x\le 1\end{cases} \tag{4}$$

The individual membership function described above are combined to form an overall decision. In our paper, we use the fuzzy aggregators as expressed in Eq. 5 to form the overall function as also used in [2]. The overall fuzzy membership function, which assumes the form of a weighted product as follows:

$$\mu_c(x)=\left(\left(\min_{j=1}^{m}\mu_j\right)\left(\max_{j=1}^{m}\mu_j\right)\right)^{0.5} \tag{5}$$

To judge whether the candidate face region constitutes a face, we finally threshold the overall membership function value by a predefined threshold value.

Gabor Feature Extraction and Face Representation

The 2-D Gabor filter kernel is defined by

$$f(x,y,\theta_k,\lambda)=\exp\left[-\frac{1}{2}\left\{\frac{(x\cos\theta_k+y\sin\theta_k)^2}{\sigma_x^2}+\frac{(-x\sin\theta_k+y\cos\theta_k)^2}{\sigma_y^2}\right\}\right]\cdot \tag{6}$$
$$\exp\left\{\frac{2\pi(x\cos\theta_k+y\sin\theta_k)}{\lambda}i\right\}$$

where σ_x and σ_y are the standard deviations of the Gaussian envelope along the x and y-dimensions, respectively. λ and θ_k are the wavelength and orientation of the 2-D sine wave, respectively. A rotation of the $x-y$ plane by angle θ_k results in a Gabor filter at orientation θ_k. A single Gabor filter response is obtained by convolving one of the filter kernels (a specific λ,θ_k) in Eq. 6 with the image. For sampling point (x,y), this response, denoted as $g(.)$, is defined as:

$$g(x,y,\theta_k,\lambda)=\sum_{u=-(N-x)}^{N-x-1}\sum_{v=-(N-y)}^{N-y-1}I(x+u,y+v)\,f(u,v,\theta_k,\lambda) \tag{7}$$

where $I(x,y)$ denotes a NxN grayscale image. Throughout this paper we consider eight orientations and four wavelengths resulting in 32 filter responses. This multi-valued (32) vector is denoted as the Gabor *jet* representation of that point *(x,y)*. A jet J is thus defined as the set $\{J_j\}$ of 32 complex coefficients obtained from one image point, and can be written as

$$J_j = a_j \exp(i\phi_j) \quad j=1,..,m \tag{8}$$

where a_j is the magnitude, ϕ_j is the phase of the Gabor features/coefficients and m is number of Gabor jets/landmark points.

Each facial landmark point can thus be represented by such a Gabor jet instead of just its gray value. In this paper, the following landmark points are used to represent the face: centers both eyes, nose and mouth, see also Fig. 2. Using the Gabor representation, the face is thus modeled by four (*m=4)* jets each consisting of 32 complex numbers. This representation can also be represented as a graph, see also Fig. 2. Then the nodes (p1,..,p4) represent the separate facial landmark points by describing their corresponding Gabor jet responses. The edges in the graph between the landmark points (*e1*,*e2*,*e3*) then represent the topographical information about the interrelationships between the landmark points by describing the distances between them.

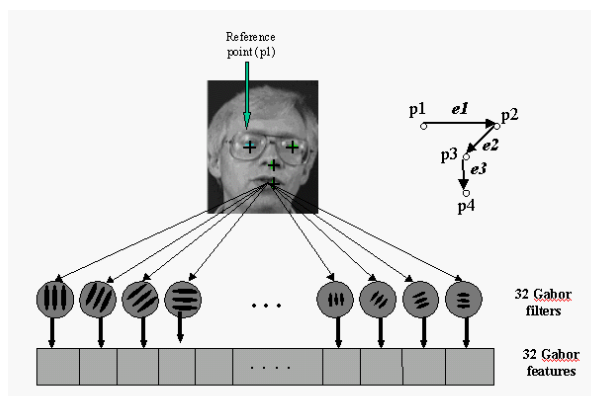

Fig. 2. Face model representation.

To find the landmark points in an unknown probe face image, we apply an affine graph matching procedure on the probe image that maximizes the Gabor magnitude similarity between the overlaid affine transformed face model graph with the corresponding graph representation of the probe image (also reported in [5]), i.e.

$$\max_{\forall J} S\ (J,J') \tag{9}$$

where J ranges over the set of affine transformed face model jets and J' are the corresponding jet graphs of the probe face image. The similarity is maximized by GA over the set of affine transformed face models in order to cope with a translated, scaled and rotated face in the probe image.

For the definition of the similarity function $S(J, J')$ we use the definition of [5] as follows:

$$S(J,J') = \frac{\sum_j a_j a_j'}{\sqrt{\sum_j a_j^2 \sum_j a_j'^2}} - \frac{\beta}{E}\sum_l \frac{(\vec{e}_l^{\,J'} - \vec{e}_l^{\,J})^2}{(\vec{e}_l^{\,J})^2} \qquad (10)$$

where β determines the relative importance of topographical/metric structure and $\vec{e}_l^{\,J}$ are the edge labels of graph J denoting distances between the landmark points; and E denoting a number of edges of the graph (i.e. $E=3$). Note that we only take the amplitudes of the jet responses into account as it turned out in [5] to be a sufficient representation.

IV. EXPERIMENT RESULTS

The proposed method is tested on various color images. Figure 3 demonstrates the performance of our facial landmarks detection. The color information is used to detect face candidate regions. On all images we used a 5x5 mean filter to smoothen the skin likelihood images and all objects smaller than 110 pixels are removed. The fuzzy analysis of color, shape, symmetry and deformable graph matching was used to locate facial landmarks from the set of candidate objects. An object was classified as a facial region if its overall membership function, μ_c , exceeded a predefined threshold of 0.75.

For the deformable graph matching by using GA, we used a constant population size of 80 individuals at each generation, the crossover rate and mutation rate are 1.0 and 0.01, respectively. The search space for each of the parameters was initialized around the position of the candidate face region that we are evaluating. This is done by placing the center of the face graph model at the center of the candidate region. The Scaling factor and orientation of the graph model are set to the estimated size and orientation of the candidate region. To localize the facial landmarks, we ran 45 generations and final fitness is evaluated. We used an average of frontal view face images taken from the MIT face database to create a face graph model.

Figure 3 shows some results of the facial landmark points localization. Currently only visual inspection of the results is performed. The algorithm was able to localize facial landmarks for frontal pose faces with moderate rotation and tilting. Facial landmarks of faces with large view angle can not be correctly localized. We found that our face model graph could not adapt the faces with large pose angle.

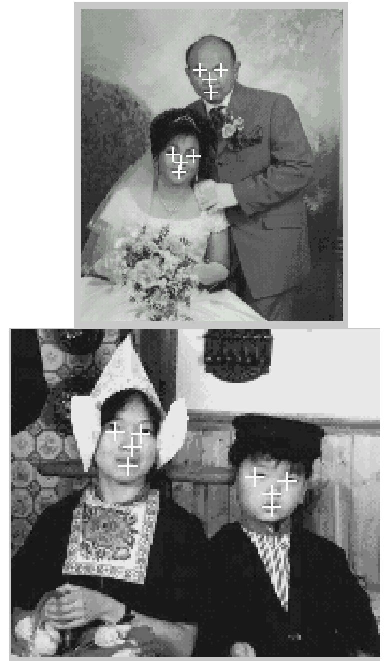

Fig. 3. Face landmarks localization.

V. CONCLUSION

We have proposed a detection scheme for locating facial landmarks based fuzzy analysis of color, shape, symmetry and interior facial feature similarity to a reference model. The performance of the proposed method was demonstrated on various color images containing single and multiple faces. The results are quite promising for frontal pose faces with moderate rotation and tilting. The future work should be dedicated to adapt the modeling such that it can cope with views that do not show faces frontally. From the results of the experiment, we conclude that the proposed method has a good prospect and should be considered in the design of face recognition systems.

REFERENCES

[1] Haiyuan Wu, Qian Chen, and M. Yachida, “Face Detection From Color Images Using a Fuzzy Pattern Matching Method”, *IEEE Trans. On PAMI*, Vol. 21, pp 557-563, 1999.

[2] N. Herodotou, K.N. Plataniotis, and A.N. Venetsanopoulos, “Automatic location and tracking of the facial region in color video sequence”, *Signal Processing: Image Comunication*, Vol. 14, pp. 359-365, 1999.

[3] R. Lim, and M.J.T. Reinders, “Facial Landmark Detection using a Gabor Filter Representation and a Genetic Search Algorithm”, *proceeding of ASCI 2000 conference*, Lommel Belgium, 2000.

[4] R. Lim, and M.J.T. Reinders, “Face Detection Using Skin Color and Gabor Wavelet Graph Matching”, *proceeding of ASCI 2001 conference*, Heijen the Netherlands, 2001.

[5] Wiskott, L., Fellous, J.M., Kruger, N., and von der Malsburg, C., “Face Recognition by Elastic Bunch Graph Matching”, *IEEE Transactions on PAMI*, Vol. 19, pp.775-779, 1997.