

Face Detection Using Skin Color and Gabor Wavelet Representation

Resmana Lim^{1,2)}, Marcel J.T. Reinders¹⁾ and Thiang²⁾

¹⁾Information and Communication Theory Group
Faculty of Information Technology and Systems, Delft University of Technology
P.O. Box 5031, 2600 GA Delft, The Netherlands
{R.Lim; M.J.T.Reinders}@its.tudelft.nl

²⁾Electrical Engineering Department, Petra Christian University
Siwalankerto 121-131, Surabaya 60236, Indonesia
{resmana;thiang}@petra.ac.id

Abstract-This paper proposes a method that finds human faces as well as its landmark points on the face in an image. Salient facial landmark detection is one of the main task in a face recognition system since it enables face normalization and therefore leads to a size and orientation invariant face recognition system. The proposed approach first uses color information to detect face candidate regions and then uses a deformable Gabor wavelet graph matching 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 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.

Keywords - facial landmark detection, skin color detection, graph matching, genetic algorithm, Gabor filtering.

I. INTRODUCTION

Face detection and detecting facial landmarks (such as position of eyes, nose, mouth, etc.) play an important role in face recognition systems. In practical face recognition system, these subsystems determine the quality of the recognition rate because they are used to normalize the recorded image(s). Once normalized, face images can quite reliably be recognized using well-known techniques like principal component analysis [1] [2]. See also [3] for more information on face recognition.

Cai and Goshtasby [4] proposed an algorithm for extracting human faces from color images. They used the CIE *Lab* color representation and in particular the *a* and *b* components to separate skin from non-skin regions. After candidate regions are generated in this way, they locate faces using an intensity template matching. Intensity template matching, however, is not so robust since it is quite sensitive to variation in lighting and contrast and in particular to scale and orientation variations.

This paper focuses on the robust and accurate detection of landmark points on the face. The approach first uses color information to detect face candidate regions and then uses a deformable graph matching to locate facial landmark points in these candidate regions. The method is made robust against lighting variations and variations between people by representing the landmark points using Gabor filter responses. The choice of Gabor filter responses is biologically motivated because they model the response properties of human visual cortical cells [5]. The essence of the success of Gabor filters is that they remove most of the variability in images due to variation in lighting and contrast. At the same time they are robust against small shifts and deformations. The Gabor filter representation in fact increases the dimensions of the feature space such that salient features can effectively be discriminated in the extended feature space because the tailored feature detectors are supposed to be as statistically independent from each other as possible [6].

To make our method robust against size and orientation variations, the facial landmarks are located using an affine graph matching technique. In this approach a graph that models the face is matched against the probed image under different affine transformations.

The face is modeled as a graph with its nodes representing the landmark points of interest and edges representing the geometry of the face. Each node models the corresponding landmark point by the expected local Gabor filter responses at that position. The ensemble of all such local filter responses at a certain position is called a *jet*. Hence, a node represents a landmark point by its *jet* representation. The geometry of the face model (edges) needs to be somewhat flexible to account for the unknown scale and rotations of the face in the probed image. This flexibility is incorporated by allowing affine transformation of the face model graph during the detection of the landmark points. Our approach is related to the work in [7] where the graph matching process is performed in a coarse-to-fine approach with the degrees of freedom of the face graph incorporated progressively. In progressive order these are translation, scale, and finally in [7] they also incorporate local distortions. The progressive matching process that is proposed in [7] is, however, still computationally expensive due to the large search space. The main contribution of this paper are 1) the use of color to preselect region of interest and 2) the introduction of an alternative matching process by using a Genetic Algorithm (GA) to optimize the matching criteria. Finding faces candidate regions based on color drastically speeds up the detection process in general. The use of a GA considerably speeds up the detection of the facial landmarks. The search space we currently use (but is not restricted to it) allows for translation, scale and rotation of the face model graph.

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 Gabor feature extraction and face graph representation are presented in section 3. The graph matching process based on GAs is discussed in section 4. In section 5, 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 in detecting skin regions in an image is to reduce the search space for the faces drastically. Note that since this part of the process only generates candidate regions, a higher false positive is allowed whereas a false negative detection should be avoided. From previous work [4] [8], it becomes clear that skin-color is a relatively good feature to detect such face candidate regions in the image.

It has been found that for a variety of different skin types, skin color is clustered within a rather well-defined region in chromaticity space [4] and HSV space [8]. The HSV color space has been effectively used to segment color images in many applications [8]. 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. Although skin colors of different people appear to vary over a wide range, they differ less in color than in brightness.

The identification of the facial region is determined by utilizing a priori knowledge of the skin distribution 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 $N(\mathbf{m}, \Sigma)$, where \mathbf{m} represent the mean value and Σ the 2D covariance matrix. These parameters can be estimated from the manually labeled skin area in the training images. Having such a posterior distribution of skin color, we can now obtain the skin-likelihood image from the original image. 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 very unlikely to be face regions. Here, candidate regions having an area lower than predefined area are removed. The flow diagram of the skin region extraction scheme then looks as shown in Fig 1.

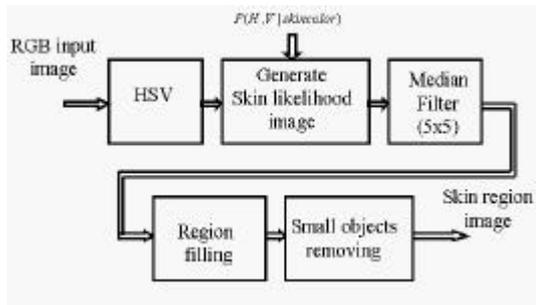


Fig 1. Flow Diagram of Skin Region Extraction



(d)

Fig 2. a) original color image b) segmented object c) after introducing median filter d) detected face candidate regions



(a)



(b)



(c)

Figures 2a-d illustrates the process of skin regions extraction described above. From these figures it can be noted that not all detected regions contain faces. Some regions represent the hands and other exposed parts of the body. In the next step, we will show how the facial landmarks are used to select the faces from the face candidate regions. In short the candidate face regions are matching against a face model and if the match is close enough we will confirm the hypothesis that the candidate region is the region of a face. Otherwise, we will reject the hypothesis.

III. FACE DETECTION IN CANDIDATE FACE REGIONS

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 landmarks points are. Hereto, we evaluate every face region by matching it against face model graph. To test the hypothesis that the region shows a face, we apply an affine graph matching procedure between the region and an affine transformed face model graph. During the matching the similarity between the transformed model and the image is maximized 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. To judge whether the candidate face region constitutes a face we finally can threshold these fit values. The matching procedure itself is based on a Genetic search algorithm (GA) in order to efficiently and robustly evaluate the search space. Only after it has been decided that the candidate face region constitutes a face, the search process continues to find facial landmarks in the face region again by maximizing the similarity between the face model graph and the face region image.

The next section elaborates on face model graph and the matching process.

3.1. Gabor Feature Extraction and Face Representation

Each facial landmark is represented by the expected local Gabor filter responses in the image. Before elaborating on the face model graph we first give some details on how to calculate the Gabor filter responses. A 2-D Gabor filter is obtained by modulating a 2-D sine wave (at particular frequencies and orientations) with a Gaussian envelope. We follow the notation in [9]. The 2-D Gabor filter kernel is defined by

$$f(x, y, \mathbf{q}_k, I) = \exp \left[-\frac{1}{2} \left\{ \frac{(x \cos \mathbf{q}_k + y \sin \mathbf{q}_k)^2}{s_x^2} + \frac{(-x \sin \mathbf{q}_k + y \cos \mathbf{q}_k)^2}{s_y^2} \right\} \right] \cdot \exp \left\{ \frac{2\pi(x \cos \mathbf{q}_k + y \sin \mathbf{q}_k)}{I} \right\} \quad (1)$$

where s_x and s_y are the standard deviations of the Gaussian envelope along the x and y -dimensions, respectively. I and \mathbf{q}_k are the wavelength and orientation of the 2-D sine wave, respectively. The spread of the Gaussian envelope is defined in terms of the sine wavelength I . Here, we set $s_x = s_y = \lambda/2$. A rotation of the $x - y$ plane by angle \mathbf{q}_k results in a Gabor filter at orientation \mathbf{q}_k . \mathbf{q}_k is defined by

$$\mathbf{q}_k = \frac{P}{n}(k-1) \quad k=1,2,\dots,n \quad (2)$$

where n denotes the number of orientations that are taken into account. Throughout this paper we use eight different orientations ($n=8$) and four different wavelengths ($I=3,5,7,10$).

A single Gabor filter response is obtained by convolving one of the filter kernels (a specific I, \mathbf{q}_k) in Eq. 1 with the image. For sampling point (x, y) , this response, denoted as $g(\cdot)$, is defined as:

$$g(x, y, \mathbf{q}_k, I) = \sum_{u=-(N-x)}^{N-x-1} \sum_{v=-(N-y)}^{N-y-1} I(x+u, y+v) f(u, v, \mathbf{q}_k, I) \quad (3)$$

where $I(x, y)$ denotes a $N \times N$ grayscale image.

When we apply *all* Gabor filters at multiple wavelengths (I 's) and orientations (\mathbf{q} 's) at a specific point (x, y) , we get multiple filter responses for that point. 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

$$J_j = a_j \exp(i\mathbf{f}_j) \quad j=1,\dots,m \quad (4)$$

where a_j is the magnitude, \mathbf{f}_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 grey value. In this paper, the following landmark points are used to represent the face: center of the eyes, nose and mouth, see also figure 3. 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. 1. Then the nodes (p_1, \dots, p_4) represent the separate facial landmark points by describing their corresponding Gabor jet responses. The edges in the graph between the landmark points (e_1, e_2, e_3) then represent the topographical information about the interrelationships between the landmark points by describing the distances between them.

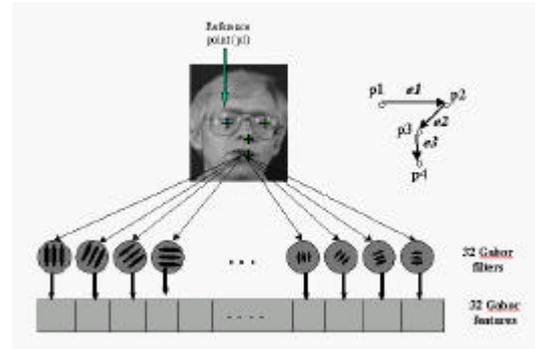


Fig 3. Face model consisting of four landmark points (p_1, \dots, p_4) represented by their Gabor filter bank responses.

We used an averaged face image taken from the MIT face database as a model image [1]. The reference face model graph is generated manually from the model face image. We manually mark the four landmark points (p_1, p_2, p_3 and p_4) in the model image (i.e. eyes, nose and mouth). Then, the Gabor features are computed to provide jets of the node. After that the edges vector (e_1, e_2 and e_3) between the marked landmark points are computed and the corresponding distances are stored.

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 [7]) i.e.

$$\max_{\forall J} S(J, J') \quad (5)$$

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 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 [7] as follows:

$$S(J, J') = \frac{\sum_j a_j a_j'}{\sqrt{\sum_j a_j^2 \sum_j a_j'^2}} - \frac{b}{E} \sum_l \frac{(\vec{e}_l^{J'} - \vec{e}_l^J)^2}{(\vec{e}_l^J)^2} \quad (6)$$

Where edges label $l=1, \dots, E$; b determines the relative importance of topographical/metric structure and \vec{e}_l are the distance vectors used as labels at edges l . Note that we only take the amplitudes of the jet responses into account as it turned out in [7] to be a sufficient representation. In the next section, we present our approach to perform the maximization procedure i.e. the actual matching of the graph in a cost-effective way.

IV. GRAPH MATCHING BY GENETIC ALGORITHM

In the previous section, we showed that finding the landmark points in an unknown probe image can be achieved by solving Eq. 5 for all possible affine transformation of the face model graph. This maximization can easily be performed by an exhaustive search over all possible combination of affine parameters (here a five dimensional parameter space). To speed up the exhaustive search procedure as well as to be able to incorporate more variability of the face model graph (increasing the dimension of the search space even further) we propose a search strategy in this section that overcomes the exhaustive searching over the affine parameter set while maintaining the invariance against translation, scale and orientation. The strategy is based upon the evolutionary optimization strategy called genetic algorithm (GA). A GA method samples the facial landscape to detect

salient matches between the probe image and the transformed face model graph. It 'remembers' only those salient areas where the match has high values (i.e. similarity), which are used as a starting point for sampling new areas until the algorithm converges.

The GA approach encodes each point in the parameter (or solution) space into a binary bit string called a chromosome, and each point is associated with a 'fitness' value that is usually equal to the objective function evaluated at that point. In our case, the objective function is equal to the similarity function between the transformed face model graph and the corresponding image graph (Eq. 6) without considering metric similarity by setting $b=0$. For our problem, the solution is represented by five parameters. These are 1,2) x and y the position of the probe graph (we have taken the x and y coordinate of the left eye as reference point), 3,4) the scaling factor in x and y directions, and 5) rotation angle. Note that the positions of all landmark points other than the left eye can be determined from the left eye⁽⁶⁾ position using the face model graph in conjunction with the scale and orientation parameters.

Each parameter is coded into an 8-bit binary string. The GA keeps a set of solutions as a population (or gene pool), that is then evolved repeatedly towards a better overall fitness value. In each generation, the GA constructs a new population using genetic operators such as crossover and mutation. Only population members that have high fitness values participate in crossover (mating) operations because they are more likely to survive. After a number of generations, the population contains members with better fitness values.

After the GA searching is converged, it provides us with values for the left-eye/reference position, scaling factor and rotation angle. These values in conjunction with the stored face model graph's structure results in the positions of the landmark points in the probe image that we were interested in.

For achieving even better results, the method has been further tuned by introducing local distortions on each of the nodes of the graph by giving more flexibility to the topographical structure of the graph. Now the metric similarity is taken into account by setting $b=0.4$ of Eq. (6).

V. EXPERIMENT RESULTS

The proposed method is tested on various color images. Figures 4a-c demonstrates the performance of

our facial landmarks detection. The approach first uses color information to preselect face candidate regions and a GA search is performed next to locate the facial landmarks.

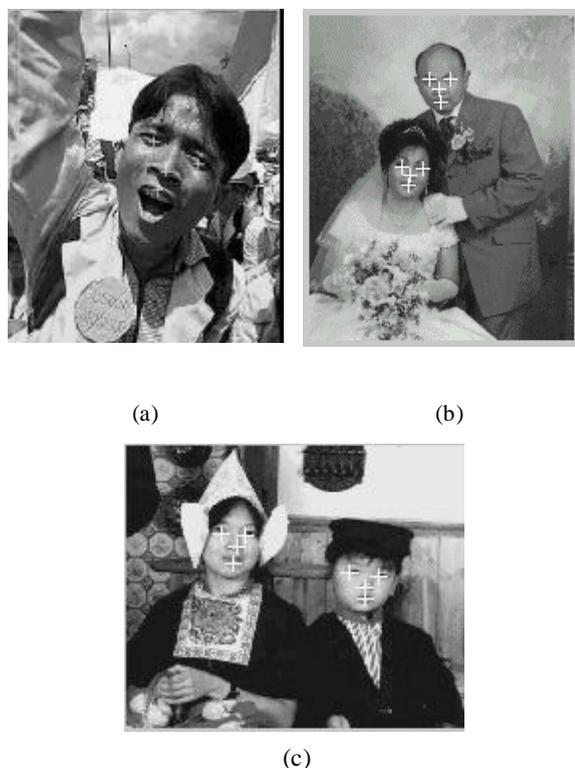


Figure 4. Facial Landmark localization

For the 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. For every facial landmarks localization we ran 30 generations and final fitness is evaluated. The threshold value of 0.7 is used to judge whether the candidate face region constitutes a face or not. If the fitness value exceed the threshold value, another GA of 15 generations was ran to perform facial landmarks localization. Finally we introduced pseudo-random local distortions on each of the nodes of the graph by giving more flexibility to the topographical structure of the graph.

The proposed detection process takes 45 generations to converge in which 80 individuals are developed at each generation, thus during the matching

process a total of 3600 operations is needed. This number is much lower than the operation needed when employing the exhaustive matching which needs 255^5 iterations to cope with all possible translation, scaling and rotation of the transformed model face graph.

VI. CONCLUSION

We have proposed a detection scheme for locating facial landmarks based on color information and graph matching using a genetic algorithm as optimization strategy. 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. 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] Moghaddam, B. and Pentland, A., Face Recognition Using View-Based and Modular Eigenspaces, In *Automatic Systems for the Identification and Inspection of Humans, SPIE*, volume 2257, 1994.
- [2] Turk, M. and Pentland, A., Eigenfaces for recognition, *J. of Cognitive Neuroscience*, vol.3, pp.71-86, 1991.
- [3] Jain, L.C. et al. (eds.), *Intelligent Biometric Techniques in Fingerprint and Face Recognition*, CRC Press, NJ, 1999.
- [4] Cai, J. and Goshtasby, A., Detecting Human Faces in Color Images, *Image and Vision Computing* 18, pp. 63-75, 1999.
- [5] Jones, J. and Palmer, L., An Evaluation of the Two dimensional Gabor Filter Model of Simple Receptive Fields in Cat Striate Cortex, *Journal Neurophysiology*, vol. 58, pp.1233-1258, 1987.
- [6] Bell, A.J. and Sejnowski, T.J., The Independent Components of Natural Scenes Are Edge Filters, *Vision Research*, vol. 7, no. 23, pp. 3,327-3,338, 1997.
- [7] Wiskott, L., Fellous, J.M., Kruger, N., and von der Malsburg, C., Face Recognition by Elastic Bunch Graph Matching, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), pp.775-779, 1997.
- [8] Herodotou, N., Plataniotis, K.N., and Venetsanopoulos, A.N., Automatic location and tracking of the facial region in color video sequence, *Signal Processing: Image Communication*, 14, pp. 359-365, 1999.
- [9] Hamamoto, Y., A Gabor Filter-based Method for Fingerprint Identification, *Intelligent Biometric Techniques in Fingerprint and Face Recognition*, eds. L.C. Jain et al., CRC Press, NJ, pp.137-151, 1999.
- [10] Resmana Lim, and M.J.T. Reinders, "Facial Landmark Detection using a Gabor Filter Representation and a Genetic Search Algorithm", proceeding of ASCI 2000 conference.