

Facial Landmark Detection using Affine Graph Matching and a Genetic Search Algorithm

Resmana Lim ¹⁾ and Marcel J.T. Reinders ²⁾

¹⁾Electrical Engineering Department, Faculty of Industrial Technology, Petra Christian University
email: resmana@petra.ac.id

²⁾Information and Communication Theory Group, Faculty of Information Technology and Systems
Delft University of Technology, The Netherlands

Abstract

This paper proposes a method that finds landmark points on the face, which is one of the main tasks in a face recognition system. Salient facial landmark detection is important because it enables face normalization and leads to size and orientation invariant face recognition. The presented approach is based on an affine graph matching technique and uses a genetic algorithm to perform the search. The feasibility of our methodology for detection tasks related to face landmark point detection has been deployed using the ORL face image database. Experiments show satisfactory results under relatively wide conditions. The GA searching approach is essential because it effectively searches the solution space.

Keywords: facial landmark detection, face recognition, affine graph matching, genetic algorithm, Gabor filtering.

Abstrak

Paper ini membicarakan sebuah metode pencarian titik-titik penting pada citra wajah, yang mana ini adalah langkah penting dalam system pengenalan wajah. Titik-titik penting wajah (seperti mata, hidung dan mulut) adalah sangat penting untuk proses normalisasi wajah sehingga dihasilkan system pengenalan wajah yang handal terhadap variasi ukuran dan orientasi. Pendekatan yang digunakan disini menggunakan teknik affine graph matching dan algoritma genetika untuk pencarian titik-titik penting wajah. Sistem diuji menggunakan database wajah dari ORL. Percobaan menunjukan hasil yang cukup menggembirakan. Pencarian dengan algoritma genetika sangat penting disini sebab dapat mencari solusi secara efektif.

Kata kunci : deteksi titik-titik penting wajah, pengenalan wajah, affine graph matching, algoritma genetika, Gabor filter.

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. This paper focuses on the robust and accurate detection of landmark points on the face. The presented approach uses Gabor filter responses to effectively represent the landmark points. The choice of Gabor filter responses is biologically

motivated because they model the response properties of human visual cortical cells [4]. Further, feature extraction based on Gabor filtering has been deployed successfully for texture analysis [5], character recognition [6], fingerprint recognition [7], as well as face recognition [8] [9]. 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 [10]. Recent work [11] has shown that a Gabor approach for local feature extraction outperformed PCA (Principal Component

Note: Discussion is expected before November, 1st 2003. The proper discussion will be published in "Jurnal Teknik Elektro" Volume 4 Number 1 March 2004.

Analysis), FLD (Fisher's Linear Discriminant) and LFA (Local Feature Analysis).

Our approach in finding facial landmarks is based on affine graph matching [8] [9]. This is done by matching a graph that models the face against the probed image. The face model is represented 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 filter responses 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 [9] 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 [9] they also incorporate local distortions. During the matching process, the model graph is thus translated, scaled, and rotated in order to compensate for the unknown position of the face in the probed image. The progressive matching process that is proposed in [9] is, however, still computationally expensive due to the large search space. The main contribution of this paper is that we introduce an alternative matching process by using a Genetic Algorithm (GA) to optimize the matching criteria. 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 has been tested using the ORL face image database.

The use of a GA in feature detection is not new. For example in [12] a GA has been used for eye detection by fitting image distributions (mean, entropy and standard deviation) between the training and probe image. Here, however, we combine the GA search strategy with the Gabor representation that has a more optimal description of the face landmark points [11].

The remainder of this paper is organized as follows. The Gabor feature extraction and face graph representation are presented in section 2. The graph matching process based on GAs is discussed in section 3. In section 4, our experiments on the ORL face database are presented. Finally, a conclusion and directions for future work are briefly covered in the last section.

Gabor Feature Extraction and Face Representation

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 [7]. The 2-D Gabor filter kernel is defined by

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

where s_x and s_y are the standard deviations of the Gaussian envelope along the x and y -dimensions, respectively. I and 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 q_k results in a Gabor filter at orientation q_k . q_k is defined by

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

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, q_k) in Eq. 1 with the image. For sampling point (x, y) , this response, denoted as $g(\cdot)$, is defined as:

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

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

When we apply *all* Gabor filters at multiple wavelengths (I)'s and orientations (q_k)'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 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 1. 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. Then the nodes (p_1,\dots,p_4) represent the landmark points by describing their corresponding Gabor jet responses. The edges in the graph (e_1,e_2,e_3) then can represent the topographical information about the inter relationships between the landmark points by describing the distances between them. Because we use four landmark points on the face, the complete jet set produces 128 Gabor filter responses per face graph.

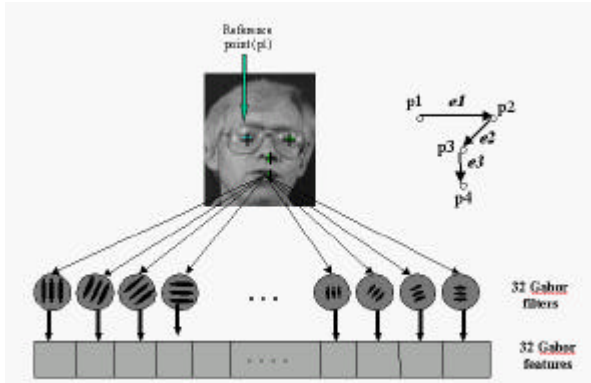


Figure 1. Face model consisting of four landmark points (p_1,\dots,p_4) represented by their Gabor filter bank responses.

The reference face model graph is generated manually from a number of training faces taken from the database. First, we manually mark the four landmark points (p_1,p_2,p_3 and p_4) in the training image (i.e. eyes, nose and mouth). Then, the Gabor features are computed to provide jets of the node. After that the edges (e_1,e_2 and e_3) between the marked landmark points are computed and the corresponding distances are stored. This procedure is repeated for all training images and the final face model graph is then

computed as the average of these manually sampled training face models, i.e. for each node the average jet is calculated and for every edge the average distance is calculated. The number of training faces that should be used depends on the variability of faces one wants to represent. Namely, if one wants to detect faces of many different races, facial expression, age, etc., the set of training faces must contain those variations also in order to cope with this variability. In general, the set of training faces contained in the face model should be as different as possible to increase the generalization capacity of the method.

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 overlayed affine transformed face model graph with the corresponding graph representation of the probe image (also reported in [9]) 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 [9] as follows:

$$S(J, J') = \frac{\sum_j a_j a_j'}{\sqrt{\sum_j a_j^2 \sum_j a_j'^2}} \quad (6)$$

Note that we only take the amplitudes of the jet responses into account as it turned out in [9] 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.

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 exhaustive search over all possible combination of affine parameters (here a four dimensional parameter space). To overcome this huge exhaustive search and also 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 method of a 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). It then uses these high fitness areas 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 the 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). For our problem, the solution is represented by only four parameters. These are 1,2) the position of the probe graph (we have taken the x and y coordinate of the left eye as reference point), 3) the scaling factor, and 4) rotation angle. Note that the positions of all landmark points other than the left eye can be determined from the left eye position using the face model graph in conjunction with the scale and orientation parameters.

Each parameter is coded into an 8-bit binary string (Figure 2). 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 (Figure 3). 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. Note that the GA improves performance

by upgrading entire populations rather than individual members. In addition, we choose to keep only the best two population members unchanged when creating a new population. In this way, we ensure that the fitness value is monotonically increasing (elitism principle).

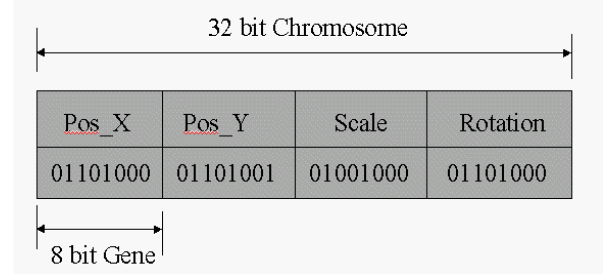


Figure 2. Chromosomes format.

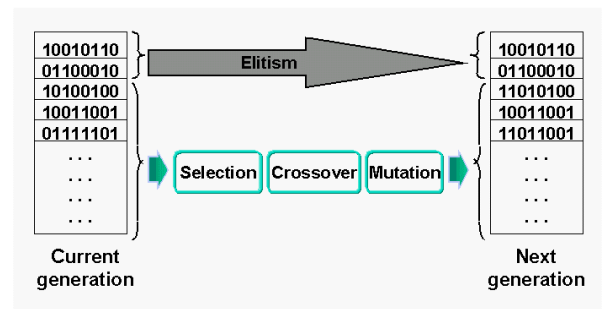


Figure 3. Producing the next generation in GAs.

Based on the above concepts, a simple GA algorithm for maximizing the similarity between the given probe image and the transformed face model graph (Eq. 5) is described as follows [13], also illustrated in Figure 3:

1. Initialization

Initialize a population with randomly generated individuals i.e. random values for the four parameters: position (x,y) of the reference point, scaling factor and rotation angle. Evaluate the fitness value of each individual by evaluating the similarity function (Eq. 6) between the probe image graph denoted by the parameter of the individual and the transformed face model graph.

2. Generate New Solution

- Keep the best 2 individuals.
- Select two members from the population with probabilities proportional to their fitness values.
- Apply crossover with the probability equal to the predefined crossover rate.
- Apply mutation with the probability equal to the predefined mutation rate.
- Repeat (b) to (e) until enough members are generated to form the next generation.

3. Evaluate Solution

Repeat step 2 and 3 until a stopping criterion is met. The stopping criteria is met if the fitness does not improved within four generations or if the number of generation exceeds a certain pre-defined number (here we have used 20).

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.

Experiment Results

We have used the ORL database that contains a set of faces taken between April 1992 and April 1994 at the Olivetti Research Laboratory in Cambridge, UK [14]. There are 10 different images of 40 distinct subjects. For some of the subjects, the images were taken at different times. There are variations in facial expression (open/closed eyes, smiling/non-smiling), and facial details (glasses/no-glasses). All the images were taken against a dark homogeneous background with the subjects in an up-right, frontal position, with tolerance for some tilting and rotation of up to about 20 degrees. There is some variation in scale of up to about 10%. The images are grayscale with a resolution of 112 x 92. In addition, we have used scaled version (75%) of these images to demonstrate the scaling invariance of the proposed landmark detector.

For the GA we used a constant population size of 20 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 additionally constraint by the following boundaries:

- left-eye reference point position $X=[1 \dots 55]$, $Y=[1 \dots 70]$
- scaling factor = [60% .. 110%]
- rotation angle = $[-20^\circ \dots +20^\circ]$

From the set of 40 subjects we have used the face of five randomly chosen subject to train the face model graph.

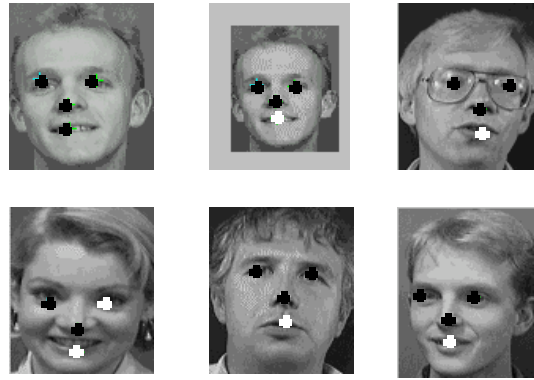


Figure 4 shows some examples of the results of the detected facial landmarks. Figure 5 shows the fitness of the population for one of the experiments.

Figure 4. Found facial landmark points by the proposed algorithm.

We have tested the algorithm over 60 probed face images taken from 20 subjects (different from the training subjects). The average displacement between the detected points and a manually detected reference set was 7.2 pixels. This displacement was relatively high because the face model was not representative enough to cope with all testing subjects. The high mismatch occurred for images of the bearded subjects or for subjects that have skin color that deviates from those of subjects used to create the face model. In the case when the subjects' images are included on the forming of face model graph, the displacement error decreased to 5.8 pixels on average. The displacement can also be reduced by incorporating local distortions into the model face graph.

The proposed detection process takes 20 generations to converge in which 20 individuals are developed at each generation, thus during the matching process a total of 400 operations is needed (20 generations times 20 individuals). This number is much lower than the operation needed when employing the exhaustive matching which needs 255^4 iterations to cope with all possible translation, scaling and rotation of the transformed model face graph.

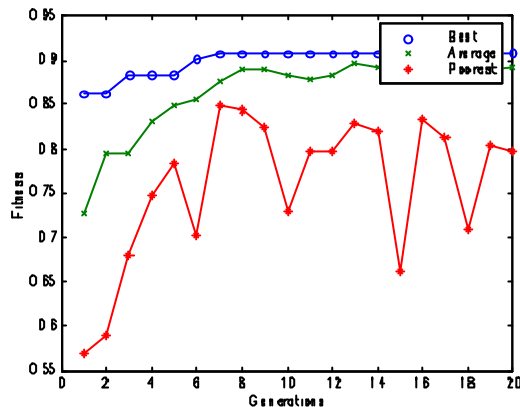


Figure 5. Fitness function of the best, poorest, and average population members as function of the generation.

Conclusion

We have proposed a detection scheme for locating facial landmarks based on graph matching using a genetic algorithm as optimization strategy. The performance of the proposed method was demonstrated on the ORL face database. The results are quite promising for frontal pose faces with moderate rotation and tilting. The proposed method should be further tuned by introducing local distortions on each of the nodes of the graph for achieving even better results by giving more flexibility to the topographical structure of the graph. 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] 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.
- [5] Tunner, M.R., Texture Discrimination by Gabor Functions, Biological Cybernetics, Vol. 55, pp.71-82, 1986.
- [6] Hammamoto, Y., Uchimura, S., Masamizu, K. and Tomita, S., Recognition of Handprinted Chinese Characters using Gabor Features, in Proc. of the 3rd Intl. Conf. Document Analysis and Recognition, pp.819-823, 1995.
- [7] 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.
- [8] Lades, L., J. Vorbruggen, Buhmann J., Lange, J., von der Malsburg, C., and Wurtz, R., Distortion Invariant Object Recognition in the Dynamic Link Architecture, IEEE Trans. Computers, Vol. 42, pp.300-311, 1993, 1993.
- [9] 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.
- [10] 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.
- [11] Donato, G., Bartlett, M.S., Hager, J.C., Ekman, P., and Sejnowski, T.J., Classifying Facial Actions, IEEE Transactions on Pattern Analysis and Machine Intelligence, 21(10), pp.974-989, 1999.
- [12] Huang, J., Liu, C., and Wechsler, H., Eye Detection and Face Recognition using Evolutionary Computation, in Face Recognition: from Theory to Applications, eds. Harry Wechsler et al. Springer-Verlag, pp.348-377, 1997.
- [13] Jang, J.S.R., Sun, C.T., and Mizutani, E., Neuro-Fuzzy and Soft Computing: A Computational approach to Learning and Machine Intelligence, Prentice-Hall, Inc. NJ, 1997.
- [14] ORL website, <http://www.cam-orl.co.uk/facedatabase.html>