2010 Second International Conference on Intelligent Human-Machine Systems and Cybernetics

Human Facial Feature Localisation by Gabor Filter and Clustering

Mian Zhou*§, Hong Wei†, Xiangjun Wang‡, Pengcheng Wen‡ and Feng Liu‡

*State Key Laboratory of Precision Measuring Technology and Instruments, Tianjin University, Tianjin, China, 300204
email: zhoumian@hotmail.com
†School of Systems Engineering, University of Reading, Reading, United Kingdom, RG6 6AY
‡College of Precision Instrument and Opto-Electronics Engineering, Tianjin University, Tianjin, China, 300072
§Department of Educational Technology, Tianjin Foreign Studies University, Tianjin, China, 300072

***Abstract*—Human facial features localization is an important process of face recognition, since it helps generating face images in accordance with specified criteria, or building unique face model. This paper presents a novel method for finding facial features through Gabor filtering and k-means clustering analysis. By Gabor filtering, face images are transformed into magnitude responses. In magnitude responses, areas containing facial features demonstrate relatively strong responses. After thresholding magnitude responses, strong responses are remained, but weak responses are neglected. Points belonging to facial features are collected for the k-means clustering. Points are grouped into different clusters. Each cluster corresponds to a facial feature. By testing on the ORL face database, the method shows its accuracy and rapidness on locating facial features, such as eyes, nose, and mouth. It also displays its robustness on people who have thick beard or moustache.**

I. Introduction

As one of the most successful applications in pattern recognition, image analysis and understanding, face recognition has received significant attentions over the past 20 years. There are a large number of commercial, security and forensic applications using face recognition technologies. Face recognition techniques can be divided into two categories, one is called Appearance based approach, another is called Model based approach. In appearance based approach, the recognition is performed by analysing the statistic properties among pixel values in face images. The typical face recognition algorithm in this category is Principle Component Analysis (PCA) [1]. Model based approaches firstly set up an unique face model, secondly match face images with the model for finding parameters of model, finally perform the recognition with the parameters. This category is represented by Elastic Bunch Graph Matching (EBGM) [2] and Active Appearance Model (AAM) [3].

These two categories have different ways of processing faces, both require localising the facial features. In appearance based approaches, it is necessary to find locations of eyes, nose, and mouth. According to the locations, face images are pre-processed, *e.g.*, eyes shall be the same hight across all face images. In model-based approaches, face model consists of a set of key points, and the key points are facial features. By automatically searching features, models are set up. It follows that facial feature localisation is crucial issue in face recognition technique, and it also determines performance of the final recognition stage.

This paper presents a clustering based method for facial feature localisation. The method includes two key parts: Gabor filter and k-means algorithm. Gabor filter is used to extract facial information from face images. k-means algorithm is applied the facial information for grouping the strong responses into several clusters. Each cluster corresponds to a facial feature, so that facial features can be located rapidly and accurately.

II. Gabor Filter

2D Gabor filters were proposed by J. Daugman [4], which are widely used in Texture segmentation [5], iris recognition [6], and face recognition [7]–[10]. Gabor filters are similar to the 2D receptive fields in simple cells of the mammalian cortex [11], so that modelling the Gabor filter is analogy to simulate the mammalian visual systems. Gabor filter varies in different frequencies and different orientations, which leads to the coarse-fine analysis.

The 2d Gabor filter is defined as a five-parameter formula:

$$h(x, y, \sigma, U, V) = \frac{1}{2\pi\sigma^2} \exp\{-\frac{x^2+y^2}{2\sigma^2}\} \exp[j2\pi(Ux+Vy)] \tag{1}$$

The parameters x and y are Cartesian coordinates with respect to horizontal and vertical directions in the filter. The parameter σ is standard deviations of the Gaussian kernel along x and y coordinates respectively. The parameters U and V are the spatial frequencies of the filter in the frequency domain along x and y coordinates respectively.

In addition, the remaining three parameters σ, U and V are need to be specified. The radial frequency F is the magnitude of two spatial frequencies U and V, which is defined as

$$F = \sqrt{U^2+V^2}$$

The orientation θ of modulating Gabor is determined by the spatial frequencies U and V. The orientation is

$$\theta = \tan^{-1}\frac{U}{V}$$

978-0-7695-4151-8/10 $26.00 © 2010 IEEE
DOI 10.1109/IHMSC.2010.101

On the contrary, spatial frequencies can be calculated by $U = F \times \cos\theta$ and $V = F \times \sin\theta$.

The relationship between Gaussian deviation σ and radial frequency F is

$$\sigma = \frac{\kappa}{F} \qquad \text{where} \quad \kappa = \sqrt{2\ln 2}(\frac{2^{\varphi}+1}{2^{\varphi}-1})$$

where φ is a factor measuring half-amplitude bandwidth of frequency response along the optimal orientations in octaves. According to some physiological findings in [4], [12], the bandwidth shall be about 1 to 1.5 in octave for the best reconstruction result. In this paper, the bandwidth $\varphi = 1.0$ in octave is selected, hence $\kappa = \pi$. The deviation σ becomes

$$\sigma = \frac{\pi}{F}$$

When the orientation θ is specified, the spatial frequencies are

$$U = \frac{\pi}{\sigma}\cos\theta \quad \text{and} \quad V = \frac{\pi}{\sigma}\sin\theta \tag{2}$$

By taking Equation 2 into Equation 1, the Gabor filter becomes a formula denoted by four variables (x, y, σ, θ)

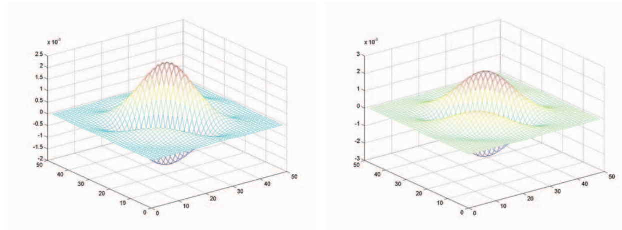
$$h(x,y,\sigma,\theta) = \frac{1}{2\pi\sigma^2}\exp\{-\frac{x^2+y^2}{2\sigma^2}\}\exp\{\frac{j2\pi^2(x\cos\theta + y\sin\theta)}{\sigma}\} \tag{3}$$

The variable σ does not indicates the standard deviation of 2d Gaussian, but only determines the size of filter. When σ increases, the size of filter template becomes bigger. The variable θ determines the orientation of filter which is used to detect grey level gradient in certain orientation. In the way, a Gabor filter can be generated by specifying σ and θ. The first exponent in Equation 3 is a 2d Gaussian kernel with the standard deviations σ. The second exponent is in a form of complex formula. The Gabor filter is sinusoidally modulated Gaussian kernel. It can be represented in pairs, often referred to as *quadrature pairs* as

$$H_{\text{even}} = \tfrac{1}{2\pi\sigma^2}\exp\{-\tfrac{x^2+y^2}{2\sigma^2}\} \times \cos\left[\tfrac{2\pi^2}{\sigma}(x\sin\theta + y\sin\theta)\right]$$
$$H_{\text{odd}} = \tfrac{1}{2\pi\sigma^2}\exp\{-\tfrac{x^2+y^2}{2\sigma^2}\} \times \sin\left[\tfrac{2\pi^2}{\sigma}(x\sin\theta + y\sin\theta)\right]$$

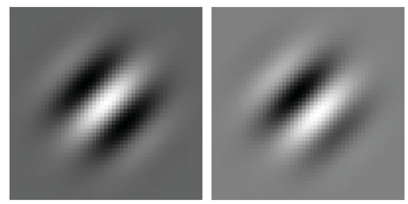
H_{even} is even pair, normally is named real part. H_{odd} is odd pair, or named imaginary part. H_{even} is a Gaussian modulated with a cosine function, so that it resides in a symmetric structure, since the standard cosine function (no dilation and no translation) is symmetric between $-\frac{\pi}{2}$ and $\frac{\pi}{2}$. H_{odd} has an antisymmetric structure, due to sine functions are increasing monotony between $-\frac{\pi}{2}$ and $\frac{\pi}{2}$. Figure 1 shows the symmetry in both real part and imaginary part of Gabor filter.

In image processing, the output of Gabor filtering is through discrete convolution by Gabor filter and digital image. The Gabor filter is a $n \times n$ template. n must be odd number, and is determined by σ. For instance, when σ is small but n is large, Gabor filter only domains a small part of template, other parts are blank. If convolving with such template, the most part of computation is meaningless. Therefore, the size of template is determined by σ dynamically. Dunn and Higgins [5] proposed that Gabor filters were truncated to a width of $6\sigma + 1$ points, which can produce filters having a wide spatial extent (e.g., when $\sigma = 8$, the size of template is 49×49). Figure 2 shows templates of a Gabor filter with $\sigma = 8$ and $\theta = \frac{\pi}{4}$



(a) The real part is symmetric (b) The imaginary part is antisymmetric

Fig. 1: The symmetry in Gabor filter



(a) The real part (b) The imaginary part

Fig. 2: The example of Gabor filter with the standard deviation $\sigma = 8$ and the orientation $\theta = \frac{\pi}{4}$.

Since Gabor filter is in a complex structure, the convolution between a Gabor filter and an image is performed by convolving the image with the real part and the imaginary part separately. The outputs of convolution are real response and imaginary response. Magnitude response are L2-Norm of real response and imaginary response, which is widely adopted in texture detection and segmentation. In this paper, magnitude responses is selected for the output of Gabor filtering. Figure 3 shows the magnitude response of Gabor filtering ($\sigma = 1, \theta = \frac{\pi}{2}$) with a face image.

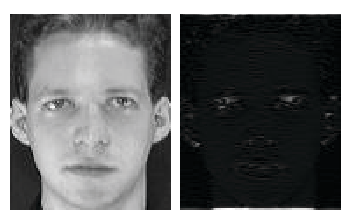


Fig. 3: The magnitude response (right) of a face image (left) convolved with the Gabor filter with the parameters $\sigma = 1$, $\theta = \pi/2$

III. K-Means Clustering Algorithm

The k-means [13] approach is adopted for clustering the pixels into couples of groups. Clustering is the process of partitioning or grouping a given set of examples into disjoint

TABLE I: The k-means clustering algorithm

1: Initialise k prototypes $(\omega_1, \omega_2, \ldots, \omega_k)$, such that each example for input is assigned to a certain prototype, which is $\omega_j = i_l \quad j \in \{1, 2, \ldots, k\} \quad l \in \{1, 2, \ldots, n\}$.
2: Each cluster C_j is represented by the prototype ω_j.
3: For each examples i_l, where $l \in \{1, 2, \ldots, n\}$, find the nearest prototype ω_{j*} with certain distance measurements by

$$\|i_l - \omega_{j*}\| \leq \|i_l - \omega_j\| \quad j \in \{1, 2, \ldots, k\} \tag{4}$$

$\|\|$ indicates the distance between the example i and the prototype ω.
4: Repeat the step 2 and 3, until the error does not change significantly. The error is calculated by

$$E = \sum_{j=1}^{k} \sum_{i_l \in C_j} \|i_l - \omega_j\|^2 \tag{5}$$

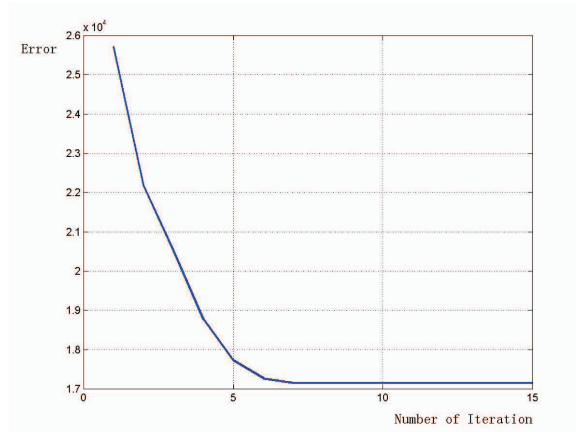


Fig. 4: The Error E converge at the 8th iteration in a k-mean clustering.

clusters. The k-means method has been shown to be effective in producing good clustering results for many practical applications among some cluster algorithms.

In k-means algorithm, it is assumed that k clusters exist. Each prototype is assigned to a cluster, so that k prototypes $(\omega_1, \omega_2, \ldots, \omega_k)$ is initialised to one of the n input examples $(i_1, i_2, \ldots, i_n)$. The principle of k-means algorithm is to assign each example to different cluster, and give the prototypes that are also the centroids of clusters.

The k-means algorithm is present as Table I. For an individual example, the best prototype shall have the minimum distance. Error E is the sum of all distance between examples and prototypes. The performance of k-means in each iteration is evaluated by the error E.

When each iteration is completed, E will be evaluated. If the errors does not change significantly between two successive iterations, the k-means clustering will be stopped. Figure 4 shows the error decreasing with the iterations. In this case, the number of examples is 398, and each example has two components. It is obvious that the k-means cluster reaches the 8th iteration, the error E does not decrease significantly. Hence, after the 8th iteration, the k-means clustering will be ceased.

Such direct k-means algorithm requires time proportional to the product of the number of examples and the number of clusters. This is computationally very expensive especially for large datasets. However, in this case, the dataset is limited to the 500 to 1000 examples. The direct k-means algorithm is still effective for the task.

IV. Facial Feature Localisation

A. ORL Face Database

In this chapter, the face database used is called ORL face database [14]. There are ten different images of each of 40 distinct person making 400 images in all. All images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement). Each image is 92×112 pixels in grey scale.

When Gabor filter is applied on face images, magnitude responses vary widely from the maximum 40.2398 to the minimum $5.4026e^{-4}$. In facial analysis, high responses are only concerned, since these responses may correspond to some crucial facial features like eyes, nose holes, lips and etc. However, lower responses may correspond to flat area in face, such as cheek and forehead. To find the key facial features, the higher responses are remained, while the lower responses are neglected.

B. Choosing the right Gabor filter

Empirically, when the size of Gabor filter template is small, more detailed local gradient variation will be captured in magnitude responses. In these responses, some facial features can be clearly recognised by human visual perception system. However, when the size of template (also the derivation of Gaussian) increases, the responses become more rough, which are difficult for human visual perception.

Among various orientations, when $\theta = \pi/2$, Gabor filtering on face images gets more responses than other orientations. This Gabor filter captures local gradient variation vertically, for examples, it likes a bar in horizontal. Also majority of facial features are laid in horizontal. Hence, Gabor filter with $\theta = \pi/2$ is adopted for locating features.

Among various standard deviations, when $\sigma = 1$, i.e., the smallest Gabor filter gives the finest responses on faces. The result of the Gabor filtering applied on the face image is shown in Figure 3. The responses correspond to the areas of two eyes, nose and mouth, and these responses resemble four brighter clusters with higher responses. The points with higher responses are the examples in the k-means clustering analysis. Each examples has two variables: one is x, and another is y.

C. Pre-processing

Before the clustering analysis is applied, the images of magnitude response are need to be pre-processed. Since the only interesting area on face is inner face area, which does not include hair and ears. It is necessary to crop the response image into a smaller one which only includes inner face area. By using a rectangle template on response image, the remaining image only contain the two eyes, nose and mouth after cropping.

Since the strong responses are only concerned for the purpose of finding facial features, a threshold scheme is applied to remove weak responses but remain strong responses. By analysing the statistic distribution of responses, the majority of points in the response images have their responses below than 4.0. Empirically, the threshold is set as 3.2 for the best results. The benefits of threshold scheme is also to reduce the number of examples in k-means clustering. k-means clustering is notorious for its long running time. The number of examples is proportional to the time of automatic analysis. Hence, decreasing the examples can speed up the clustering process. After threshold, input data for k-means is reduced to 500-1000 examples for different images, rather than $92 \times 112 = 10304$ examples.

D. Results

For performing k-means clustering, strong responses are taken as examples which include two coordinates in images. The k is specified as 5, and the k-means clustering is run on these example. The k-means algorithm applied on the face image shows some significant rates and results. The reason is that the clusters are separated distinguishably. There are some distinct clusters in different areas. And the algorithm normally runs $8 \sim 12$ iterations. It will not take long time to compute. The method automatically locates five clusters around two eyes, nostrils, and left and right part of lips. By combining left part of lips and right part of lips, the facial feature - mouth can be localised. Hence, three common facial features, i.e., eyes (right and left), nostrils, and mouth are correctly detected by the proposed method. Some results from feature localisation are shown in Figure 5. For some images, five points are around

Fig. 5: The results of automatic facial feature localisation

facial features. From Figure 5, it follows that the proposed method is robust to thick beard and mustache (2nd in the 1st row and 4th in the 2nd row).

V. Conclusion

An accurate method for locating facial features is present in this paper. The method has two key techniques, one is Gabor filter, another one is k-means clustering. Firstly, Gabor filter is used for transforming face images into magnitude responses. When some area contains facial features, the responses in the area are stronger, and verse vice. For instance, magnitude responses of cheek and forehead are almost zero. By collecting points with strong responses and neglecting points with weak responses, it can be found that facial features are presented in the form of clusters. Secondly, the k-means clustering algorithm is applied on the points to find some statistic distributions among them. Different points are assigned to different clusters. The different clusters are explained as different features on faces.

The results show that this method is capable of finding the locations of facial features - eyes, nostril and mouth on the ORL face database. In addition, the method is fast in running, since the k-means clustering only takes 8-12 iterations. The method also shows high robustness. For some people who have strong beard or moustache, it still can detect those facial features correctly.

Acknowledgement

This work is supported by Open Foundation Grant of State Key Laboratory of Precision Measuring Technology and Instruments (Tianjin University or Tsinghua University).

References

[1] M. Turk and A. Pentland, “Eigenfaces for recognition,” *Journey of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[2] L. Wiskott, J. Fellous, N. Kruger, and C. Malsburg, “Face recognition by elastic bunch graph matching,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 775–779, 1997.

[3] T. Cootes, G. Edwards, and C. Taylor, “Active appearance models,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–685, 2001.

[4] J. Daugman, “Uncertainty relation for resolution in space, spatial frequency and orientation optimized by two-dimensional visual cortical filters,” *Journey of the Optical Society of American A*, vol. 2, pp. 1160–1169, July 1985.

[5] D. Dunn and W. Higgins, “Optimal gabor filters for texture segmentation,” *IEEE Transactions on Image Processing*, vol. 4, no. 7, pp. 947–964, July 1995.

[6] J. Daugman, “High confidence visual recognition of persons by a test of statistical independence,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, pp. 1148–1160, November 1993.

[7] M. Zhou, H. Wei, and S. J. Maybank, “Gabor wavelets and AdaBoost in feature selection for face verification,” in *Proceedings of Applications of Computer Vision 2006 workshop in conjunction with ECCV2006*, May 2006, pp. 101–109.

[8] M. Zhou and H. Wei, “Face verification using gabor wavelets and AdaBoost,” in *Proceedings of 18th International Conference on Pattern Recognition*, 2006.

[9] ——, “Face feature extraction and selection by gabor wavelets and boosting,” in *Proceedings of International Conference on Image and Signal Processing*, 2009.

[10] M. Zhou, *Gabor-Boosting Face Recognition: From Machine Learning Perspective*. VDM Verlag Publishing, 2009.

[11] J. Jones and L. Palmer, “An evaluation of two-dimensional gabor filter model of simple receptive fields in cat strait cortex,” *Journal of Neurophysiology*, vol. 58, pp. 1233–1258, 1987.

[12] T. Lee, “Image representation using 2D gabor wavelets,” *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 18, pp. 959–971, 1996.

[13] D. MacKay, *Information Theory, Inference and Learning algorithms*. Cambridge University Press, 2003.

[14] F. S. Samaria and A. C. Harter, “Parameterisation of a stochastic model for human face identification,” in *IEEE 2nd Workshop on Applications of Computer Vision*, December 1994, pp. 138–142.