MVA2000 IAPR Workshop on Machine Vision Applications, Nov. 28-30, 2000, The University of Tokyo, Japan

8—28

Performance Analysis of Gabor Responses in Face Recognition

Menaka Rajapakse and Yan Guo
RWCP* Multi-Modal Functions KRDL† Lab
21 Heng Mui Keng Terrace, Singapore 119613
menaka@krdl.org.sg

Abstract

In this paper, we investigate the behavior of Gabor responses at automatically located facial feature points for face recognition. In our approach, a set of feature points on the facial landmarks is selected, where features of interest are eyes, nose, mouth and left and right contours of the face. As a preprocessing step all images in the database are normalized, eye positions in the normalized images are set at preset fixed coordinates, and integral projection is used for automatically locating facial landmarks such as nose and mouth areas in face images. Finally, Gabor responses at respective feature locations are extracted. The extracted features are analyzed for recognition performance using a neural network classifier with backpropagation learning, where input to the network is a similarity vector corresponding to feature points of two faces.

1 Introduction

Vision is one of the most powerful assets inherited by human beings who has the ability to perceive information as soon as an object is encountered. With the help of information we gather from birth we build up models to represent objects familiar to us. As we grow up, our minds start exploring the models which represent the objects in the world. Even though the research on investigating human vision system is immense, the development of an automated vision system rivaling human capabilities seems far-fetched. This may be due to the complexity of human vision system and the difficulty in simulating its physical behavior. From many psychologists point of view, face identification mechanism is hard-wired in the brain. Brain cells which perform feature extraction are known to exist at low level areas of the visual cortex and these cells are believed to learn these features from the environment adaptively. Suitability of mimicking simple cells using Gabor wavelets has been investigated [5] [6] and established as a promising approach to model receptive fields. Further, Gabor responses are known to be robust because of their low sensitivity to local distortions, translations and rotations.

Visual features in faces such as hair, face outline, eyes, and mouth are said to be important in perceiving and remembering faces. Manual localization of these features has been practised by many of the applications detailed in literature. Here we use a set of feature points representing a face in the facial feature space. In our approach, we take the horizontal integral projections of vertical gradient of images for the detection of fiducial point locations within the face image. The method described here consists of three main phases namely, facial landmark detection, Gabor response extraction at these points, and classification for recognition. Classification was carried out using a multi-layer perceptron neural network (MLP) with backpropagation (BP) learning because this type of classifier has been proven as effective mapping tools for a wide variety of problems of practical interest. A well known characteristic of these networks is their ability to adapt and learn complex mappings. In our approach we use MLP to classify same faces and different faces presenting feature similarities as inputs to the network.

2 Feature point selection

A set of feature points corresponding to facial landmarks is selected. Still images used in the experiment are normalized to have similar grey levels and same size by setting the eye positions fixed at preset coordinates and the vertical distance of the eye positioning is fixed for all the images. As the eye area seems to

*Real World Computing Partnership
†Kent Ridge Digital Lab

be more invariant to facial expressions and gestures we selected more points enclosing eyes compared to the number of points selected in the other areas such as nose and mouth. Three points across the base of the nose were sected, another two points on the left and right contours and three points on the mouth covering the corners of the mouth and the middle point between them were also chosen.

Figure 1: Illustration of selected feature points

3 Facial Landmarks Detection

Integral projection has been successfully applied to extract features for human face recognition [1] [2]. Projections are extremely effective in determining the positions of features provided that the window in which they act is suitably located to avoid misleading interferences [2]. Given an image $I(x,y)$, the vertical and horizontal integral projections in an area given by a window $[x_1,x_2] \times [y_1,y_2]$ is defined respectively as:

$$V(x) = \sum_{y=y_1}^{y_2} I(x,y)$$

and

$$H(y) = \sum_{x=x_1}^{x_2} I(x,y)$$

Left and right contours of a face can be detected using horizontal projection while the vertical projection can be used to detect hair-line, nose base and mouth. As the initial step, the vertical locations of nose and mouth in the image with respect to known eye positions are roughly estimated. A refined estimate is then calculated by applying horizontal projection to the edge map (vertical gradient) of the image. The peaks evaluated from the horizontal projection are used to determine the nose base and mouth locations. The retrieved locations are then rated based on their prominence and locations within the estimated range. Figure 2 illustrates examples of the original

Figure 2: (a)Original image (b)Horizontal edge map

image and corresponding horizontal edge map used in the experiment.

Once the vertical locations of the nose base and mouth are located, the estimations of feature points corresponding to the detected locations are evaluated. The evaluation of these points is carried out with respect to the center of eye positions which lie on a line connecting the preset eye coordinates.

4 Gabor Responses

Multi-scale, complex Gabor kernels at different orientations are known to mimic receptive fields of simple cells in the visual cortex. Also, Gabor responses in an image are robust against brightness and contrast variations, and able to highlight low level features. The phase components of the complex valued Gabor function used in retrieving multi-scale features can be written as:

$$\psi_k(\mathrm{z}) = \frac{k^2}{\sigma^2}\exp(-\frac{k^2\mathrm{z}^2}{2\sigma^2})[\exp(ik\mathrm{z}) - \exp(-\frac{\sigma^2}{2})] \quad (1)$$

where value of the Gabor kernels $\psi_k(\cdot)$ is given by image coordinates and take the form of a plane wave restricted by a Gaussian envelope. For our experiment, we selected spatial frequencies $k = \frac{\pi}{2}, \frac{\pi}{4}, \frac{\pi}{8}$ with six levels of orientations with angles spaced equally at intervals of $\frac{\pi}{6}$ from 0 to π and a value of σ equal to π.

5 Face Representation

The Gabor response at each feature point is evaluated by convolving the feature value with both odd and even Gabor kernels. The amplitude of the complex Gabor response was used for all calculations due to its less sensitivity to displacements. Feature vectors, each consisting of 18 Gabor response amplitudes are generated for all feature points of the face images in the database.

A face is represented by a set of feature vectors corresponding to outputs obtained by convolving the feature points with a set of Gabor wavelets. A face F in the database can be represented as $F = \{\overrightarrow{f_1}, \overrightarrow{f_2} \cdots, \overrightarrow{f_I}\}$ where I=number of feature points and the feature vector corresponds to each feature is written as $\overrightarrow{f_i} = (f_{i1} f_{i2} \cdots f_{i18})_{i=1...I}$

6 Neural Network Classifier

For our analysis, we used a two layer MLP with backpropagation learning algorithm to recognize faces in a face database. The basic idea behind this classification is to decide whether the two faces presented to the network are of the same person or different persons.

The implementation of the neural net classifier used is as follows: Let the weight between the input layer unit i and the hidden layer unit j be $v_{ij}, i = 1, 2, ..n, j = 1, 2, ...h$ and between hidden layer unit j and output unit k be $w_{kj}, k = 1, 2, ..., m$. Let the n dimensional input training pattern p be denoted as $\mathbf{s}^p$, $p = 1, 2, ...P$, and the output of the hidden layer unit j for input pattern $\mathbf{s}^p$ is denoted as $y_j{}^p$, $j = 1, 2, ...h$. Likewise, the output from unit k of the output layer for input pattern $\mathbf{s}^p$ is $z_k{}^p$, while the desired output is denoted as $t_k{}^p, k = 1, 2, ..., m$.

Let $H_j = \sum_i v_{ij} s_i$ $j = 1, 2, ...h$ and $O_k = \sum_j w_{kj} y_j$ $k = 1, 2, ...m$ be the net inputs to the hidden layer unit j and output layer unit k, respectively. Then the outputs computed by the unit j of the hidden layer and the unit k of the output layer are given by $y_j = g(H_j)$ $j = 1, 2, ..., h$ $z_k = g(O_k)$ $k = 1, 2, ..., m$ where g is the non-linear sigmoidal activation function given by

$$g(s) = \frac{1}{1 + e^{-1.0*s}} \tag{2}$$

The adaptation rules for the output and hidden layers are

$$\Delta w_{kj} = \eta \delta_k y_j \tag{3}$$

where $\delta_k = (t_k - z_k) g'(O_k) y_j$, η is the learning coefficient and

$$\Delta v_{ji} = \eta \delta_j s_i \tag{4}$$

where $\delta_j = g'(H_j) \sum_k \delta_k w_{kj}$ respectively.

A momentum term was added to the gradient expression in order to improve the rate of convergence. This was accomplished by adding a fraction of the previous weight change to the current weight change. The update rule given in [9] is as below.

$$\Delta w_j(t+1) = -\eta \frac{\partial E}{\partial w_j(t)} + \alpha \Delta w_j(t) \tag{5}$$

where α is the momentum coefficient. The value of α should be positive and less than one.

Input to the MLP network is a similarity vector $\overrightarrow{S}$ corresponding to similarities between chosen feature points of two faces F and F' in the database. Let $\overrightarrow{S} = (s_1 s_2 .... s_k)$ where k = number of feature points and $s_i = \frac{\overrightarrow{f_i} \cdot \overrightarrow{f_i'}}{|\overrightarrow{f_i}||\overrightarrow{f_i'}|}$.

Inputs to the network are normalized such that the similarity values at input lie between zero and one. Hidden layer is fully interconnected to the input and output layers and the output layer consists of a single node corresponding to the decision responses; whether the two faces are of the same person or different persons.

During the training process the network was trained with pattern vectors of positive sample pairs (faces of the same person) and with negative samples (faces of two different persons) generated by shuffling the images in the database. During testing, a set of images independent of the training image set was used to establish system performance after training. The probability of misclassification was evaluated by dividing the number of misclassified faces by the total number of faces tested. We have also explored the desired number of hidden units to achieve a lower equal error rate(EER).

7 Recognition and Results

The present system was tested on 1180 images of XM2VTSDB face database from the University of Surrey. The database consists of four images per person taken at a time interval of one month apart. Similar lighting conditions and homogeneous backgrounds have been used during image acquisition. The set of images is composed of frontal and near frontal images with varying facial expressions. The original image size is 726 x 576 pixels and the database contains images of caucasian and asian males and females. The images were cropped and normalized to yield pixel size 150x200 and then scaled down to 128x128 pixels to use in the experiment. Once the similarity vectors for all feature points between faces in the training set were calculated, they were used to train the MLP network. During testing the feature(similarity) vectors of same person and different persons were presented to the network in order to get performance measures of false rejection and false acceptance ratios respectively. The equal error rate (EER) calculated was 4.75% which was an improvement compared to the EER gained by using the average inner product alone 5.4% as the final confidence measure [10].

8 Conclusion

We implemented a simple face recognition algorithm based on Gabor wavelets taking a vector of nineteen features representing landmark features of human faces. Metric variations of locations of selected fiducial points were compensated by taking horizontal integral projection of vertical gradient (horizontal edge map) of an image. It is also worthy note that the

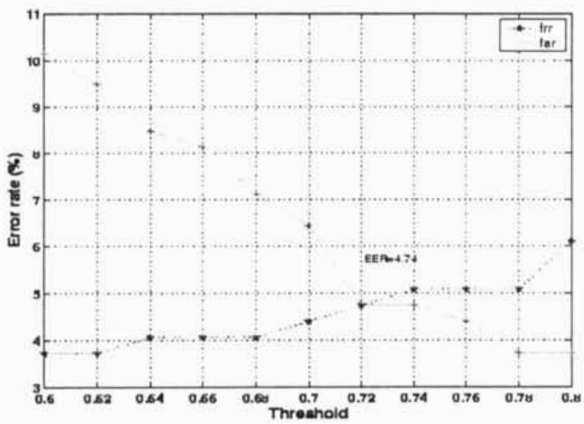

Figure 3: Verification Performance - Image size 128x128.

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

Figure 4: Examples of false rejected faces: (a)-test image (b),(c),(d)-stored images

detection of facial features has been made more robust by incorporating constraints on the geometry of the face in terms of relative feature locations. These constraints were used to guide the location of matches and restrict the regions over which integral projection was computed. Also, localization of the mouth and nose was made easier by anchoring the search of these locations with respect to preset eye positions. The system misclassified a few faces for the reasons we assume being due to changes in appearance (presence or absence of glasses/beards/mustache, large pose expression variations) causing a face to populate itself more than one area in the face space. By increasing the number of feature points and allowing a tolerance for the facial landmark point shifts the system may able to produce better results. We also came to a conclusion that the optimum number of hidden units required to acquire the above mentioned recognition rate is between three to ten hidden units empirically.

References

[1] Kanade, T. "Picture processing by computer complex and recognition of human faces", *Technical report, Kyoto University, Dept. of Information Science* 1973.

[2] Brunelli, R. , Poggio, T. 1992 "Face Recognition through Geometrical Features" *Proceedings of ECCV* 1992, pp 792-800.

[3] Qi Bin Sun,Chian Prong Lam, and Jian Kang Wu. "A Practical Automatic face Recognition System", *Face Recognition, From Theory to Applications* 1998.

[4] Manjunath, B.S, et al. 1992 "A Feature Based Approach to Face Recognition " *Proc. IEEE Computer Vision and Pattern Recognition* 1992, pp 373-378.

[5] Daugman, J.G, 1998 " Complete discrete 2-D Gabor Transforms by neural networks for image analysis and compression" *IEEE Trans. Acoustic Speech Signal Processing* 36, 7 1998, pp 1169-1179.

[6] Lee T.S, 1997 "Image Representation Using 2D Gabor wavelets",*IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 18, no. 10, 1996.

[7] Huang, W, et al., 1998 "A robust approach to face and eye detection from images with cluttered background",*Proc. IEEE 14th International Conference on Pattern Recognition,* vol. 1,1998, pp.110-113.

[8] Wiskott, L, et al., 1997 " Face Recognition by Elastic Bunch Graph matching", *Proc. IEEE International Conference on Image Processing,* Vol. 1 1997 pp. 129-132.

[9] Rumelhart, D, E., et al., 1986 " Learning Internal Representations by Error Propagation ", *Parallel Distributed Processing: Explorations in the Microstructures of Cognition,* Vol. 1 MIT Press, Cambridge, MA.

[10] Rajapakse, M, et al., 2000 " Efficient Gabor feature based Approach for Face Recognition", *Proc. World Multi Conference on Systemics, Cybernetics and Informatics,* Vol. 5, Part 1 2000 pp. 265-269.