MVA2009 IAPR Conference on Machine Vision Applications, May 20-22, 2009, Yokohama, JAPAN

On Face Recognition using Hierarchical Self-organized Gabor Features

Saleh Aly
Kyushu University, Japan
aly@limu.is.kyushu-u.ac.jp

Naoyuki Tsuruta
Fukuoka University, Japan
tsuruta@tl.fukuoka-u.ac.jp

Rin-ichiro Taniguchi
Kyushu University, Japan
rin@limu.is.kyushu-u.ac.jp

Abstract

Gabor-based face representation has achieve enormous success in face recognition. However, one drawback of Gabor-based face representation is the huge amount of data that must be stored. Due to the nonlinear structure of the data obtained from Gabor response, classical linear projection methods like principal component analysis failed to reduce this large amount of data. As a way to solve this problem, a nonlinear projection method is exploited. A set of hierarchical self-organizing maps is employed to capture the nonlinearity of the data and to represent it in a new reduced feature space. Experimental results on ORL face database prove the validity of our proposed feature extraction method

1 Introduction

Important practical applications of automatic face recognition have made it a very popular research area in the last three decades [1]. An automated vision system that performs the functions of face detection, verification and recognition will find countless applications, such as airport security and access control, building surveillance and monitoring, human-computer intelligent interaction. Even though humans can detect and identify faces in scene with a little or no effort, building an automated system that accomplishes such objectives is, however, very challenging [1]. The challenges are even more profound when one considers the large variations in the visual stimulus due to illumination conditions, viewing directions or pose, and facial expression.

Solving face recognition problem can be mainly divided into two subtasks, feature extraction and feature classification. Among the varieties of feature extraction methods, Gabor wavelets [2] considered one of the most successful feature extractors due to its robustness against illumination, rotation, and scale variations. There are different pattern recognition algorithms employed Gabor filters as the primary feature extraction method. These algorithms can be classified into two algorithms, analytic [3, 4] and global [5, 6] methods. Analytic methods extract feature from a set of selected facial landmarks, while global methods extract features at each pixel then concatenate these features to construct one global vector for the whole face. Lades et al. [3] demonstrated the use of Gabor wavelets for face recognition using Dynamic Link Architecture (DLA) framework. The DLA starts by computing the Gabor jets, and then it performs a flexible template comparison between the resulting image decompositions using graph matching. Wiskott et al. [4] have expanded on DLA when they developed a Gabor wavelet based elastic bunch graph matching method to label and recognize human faces. Recently, Gabor filters have also been applied in global form for face recognition. For example, ICA [5], and LDA [6] are used to reduce the huge dimensionality of the Gabor feature vector.

This paper introduces a hierarchical self-organized Gabor features (HSOGF) method for face recognition. HSOGF is used as a feature extractor to represent face image in a new feature space, while simple classifiers like KNN or linear-SVM are used for classification. The novelty of the HSOGF method comes from (i) constructing a feature map from the response of Gabor filters at each position in the image using two-layer self-organizing map (SOM). (ii) use subset of Gabor filters instead of the commonly used 40 filters reduce the computational complexity while keeping the recognition accuracy at the same level. The feasibility of the HSOGF method has been successfully tested on face recognition using ORL database.

The remainder of this paper is organized as follows: section 2 derives a Gabor feature vector representation. Section 3 describes the structure and training algorithm of hierarchical self-organized map neural network. The architecture of the proposed face recognition system is described in Section 4. Section 5 assess the performance of SOGF method on the face recognition task by applying ORL database. Finally we conclude our paper and discuss promising directions for future work in Section 6.

2 Gabor wavelet

Gabor filters [2] are used for image analysis because of their biological relevance and computational properties. The Gabor wavelets, whose kernels are similar to the $2D$ receptive field profiles of the mammalian cortical simple cells, exhibit strong characteristics of spatial locality and orientation selectivity, and are optimally localized in the space and frequency domain. In this paper, the following form of a $2D$ Gabor filter function, in the continues spatial domain has been employed:

$$\psi(x, y, \theta, \lambda, \gamma) = e^{-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}} \cos(2\pi \frac{x'}{\lambda}) \qquad (1)$$

$$x' = x\cos\theta + y\sin\theta, y' = -x\sin\theta + y\cos\theta \qquad (2)$$

Where θ specifies the orientation of the wavelet, the orientation of the wavelets dictates the angle of edges or bars for which the wavelet will respond. λ specifies the wavelength of the cosine wave or inversely the frequency of the wavelet. Wavelets with a large wavelength (low frequency) will respond to gradual changes in intensity of the image, while wavelet with short wavelength (high frequency) will respond to sharp edges and bars. The radius of Gaussian specified by σ, the Gaussian size determines the amount of the image that affects the convolution; this parameter is usually proportional to the wavelength, such that wavelets of different size and frequency are scaled versions of each other i.e. $\sigma = c\lambda$. Finally, γ specifies the aspect ratio of the Gaussian.

2.1 Filter parameter setting

Filter parameters are inspired by the work in [8] because it gives biologically plausible Gabor filters which perform well for filtering tasks related with object recognition. We arranged the filters to form a pyramid of scales, $\lambda \in \{2, 2\sqrt{2}, 4, 4\sqrt{2}, 8\}$, the filter span a range of sizes from 3×3 to 11×11 pixels in steps of two pixels. The orientation parameter is sampled into 8 different orientations over the interval 0 to π, i.e. $\theta \in \{0, \frac{\pi}{8}, \frac{2\pi}{8}, \frac{3\pi}{8}, \frac{4\pi}{8}, \frac{5\pi}{8}, \frac{6\pi}{8}, \frac{7\pi}{8}\}$. The radius of Gaussian function is set such that the wavelets of different size and frequency are scaled versions of each other, i.e. $\sigma = \lambda$. The aspect ratio parameter included such that the wavelets could also approximate some biological models, the wavelets used in this paper have circular Gaussian, i.e. $\gamma = 1$.

In order to tolerate for shift variations, Serr et. al. [8] propose a biologically motivated MAX filter to reduce the effect of feature shifting in the Gabor response image. A MAX filter is employed by performing maximum operation over a local area in the Gabor-filtered image. In our experiment the size of the block used by MAX filters equal 2×2 pixels. In which the dimension of the output image reduced by half.

2.2 Normalized Gabor feature vector

The Gabor wavelet representation of an image is the convolution of the image with a family of Gabor kernels as defined by Eq. (1). Let $I(x, y)$ be the gray level distribution of an image, the convolution output of image I and a Gabor kernel $\psi(x, y, \theta, \lambda, \gamma)$ normalized to unit length in order to reduce the effect of lighting variations, the following equation define a normalized Gabor feature vector:

$$O(x, y, \theta, \lambda, \gamma) = \frac{I(x, y) * \psi(x, y, \theta, \lambda, \gamma)}{\|O\|} \quad (3)$$

Where $*$ denotes the convolution operator. Applying the convolution theorem, we can derive the convolution output from Eq. (3) via the Fast Fourier Transform (FFT), which reduce the computational complexity by a significant factor.

The distribution of Gabor filter response at specific feature point is investigated using principal component analysis algorithm. The distribution of Gabor features at the center of eye for 13 different viewpoints (ranged from full frontal, full left-profile, and fully right profile views) and 21 illumination conditions is visualized in Fig. 1 using one subject from CMU-PIE face database [9]. The figure shows the distribution of features projected in a plane spanned by the two greatest eigenvectors. It is clear that the distribution exhibit a nonlinear

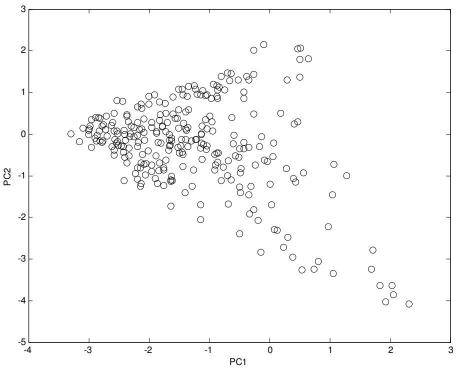

Figure 1: Distribution of Gabor filter responses around eye using the two greatest eigenvectors

manifold in the Gabor space, in this sense linear projection methods like principal component analysis will fail to model these variations. Therefore, we propose a hierarchical set of SOM cells, each cell consists of two-layer of SOM maps which are employed to model this curved distribution.

3 Hypercolumn model neural network

Before we explain the structure and training algorithm of Hypercolumn model (HCM) [10], it is necessary to explain the characteristics of self-organizing map and the two-layer hierarchical SOM.

3.1 Self-organizing map

SOM network [7] is a biologically motivated neural network, which emulate the ability of the brain to map the outside world into cortex, where nearby stimuli are coded on nearby cortical areas. Kohonen has proposed a simple algorithm for the formation of such mapping. A pattern is projected from an input space to a position in the map- where the information is coded as the location of the activated neuron. The SOM is unlike most classification or clustering techniques in that it provides a topological ordering of data; Similarity in the input space is preserved in the output space. The topological preservation of the SOM process makes it especially useful in the classification of data, which includes a large number of classes. The SOM codebook has the following two characteristics. First, the probability distribution function (PDF) of the codebook is a good approximation for the PDF of the training data. Second, the topographic order of the training data is preserved in the codebook, even if the dimensionality of the SOM is smaller than that of training data. The second characteristic means that similar features are mapped to nearby positions in the feature map. These characteristics make SOM an appropriates choice to learn the distribution of Gabor features. However, for curved data distribution, one SOM model can not capture the whole structure of the data, therefore two-layer SOM is employed to model this complex structure.

3.2 Hierarchical self-organizing map (HSOM)

The HSOM [11] is a two-layer SOM network connected like any feedforward network: every unit in the

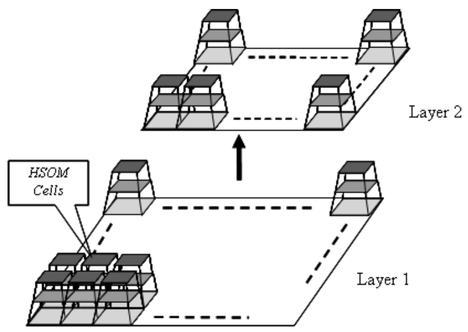

Figure 2: HCM Network structure

sending layer is connected to every unit in the receiving layer. However, using SOM layers has a special characteristic that lead to much more efficient implementation using the simple learning algorithm of Kohonen in both layers. Neurons in the second layer of the network are fed by the index of the winner neuron in the first layer as input data. The training algorithm for the second layer is the same as that in the first layer. These two-layers differ from each other in the number of neurons. The number of neurons in the second layer is smaller than that in the first layer. The reduction in the number of neurons helps HSOM network to integrate features extracted in the first layer. Employing this network structure to solve pattern recognition problem with high dimensional feature space like face recognition require a replication of this structure at each region in the input image, this network structure called Hypercolumn model which will explained in the next section.

3.3 HCM structure and training algorithm

The HCM network shown in Fig.2 is derived from the structure of Neocognitron (NC) [12] by replacing each C-cells and the lower directly connected S-cells with a two-layer HSOM network. The first map in HSOM stands for feature extraction map (FEM), while the second one is used for feature integration (FIM). The number of HSOM in X and Y directions are decided by the width and height of the input Gabor-filtered image respectively. FEM in the first layer receives its input from the response of all Gabor filters at different scale and orientation at specific point in the input image. Feature integration layer is an SOM map, whose input is the index of the winner in the lower feature extraction layer. In the feature integration layer, therefore, all distorted patterns are mapped to the same neuron, since the number of neurons in the feature integration layer is smaller than the number of neurons in the feature extraction layer. After that, the index of the winner in the feature integration layer is presented to the next feature extraction layer. These local features are pilled up hierarchically in the higher layers. This hierarchical structure allows the network to increase the invariance of the features.

The HCM training algorithm uses the unsupervised learning algorithm of the competitive neural networks to construct its feature maps. The learning process is applied layer-by-layer starting from the bottom layer, where the normal learning algorithm of the HSOM is used to train each unit in the map. All HSOMs in the same layer can be trained in parallel. After presenting the input pattern to the HCM network, the training algorithm find the best matching winner from the first feature extraction map and update the weights of this winner and its neighbors. This process is repeated for each FEM in the first layer. Similarly, FIM layer repeat the same algorithm using index of the winner neuron in the first layer as input data. In the recognition phase, the competition also applied layer by layer starting from the bottom layer until the top layer. The network uses the same competition algorithm used by the HSOM to find the winner. Then the winner position indexes of all HSOMs in the first layer are used as the input for the second layer.

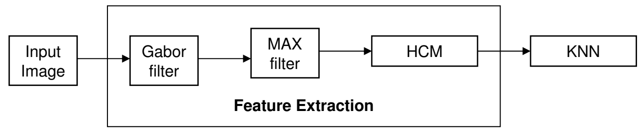

Figure 3: Architecture of proposed face recognition system

4 Proposed face recognition system

In this paper, we introduce a hierarchical self-organized Gabor features (HSOGF) method for face recognition whose system architecture is shown in Fig. 3. Gabor filters are used as a feature extractor to represent face image and HCM neural network used to project Gabor features in a learned feature space.

The HSOGF method first derives a Gabor feature vector from a set of Gabor wavelet representation of face images. MAX filter is applied for the image response, in which the filtered image gives the same response in spite of small shifts in the local features. Finally, the dimensionality of the feature vector is reduced using a set of HSOMs, furthermore the outlier features are removed due to the neighborhood learning capabilities of SOM.

The rationale behind integrating Gabor wavelet and SOM is two-fold. On the one hand, the Gabor transformed face images exhibit strong characteristics of spatial locality, scale and orientation selectivity. These images can thus produce salient local features that are most suitable for face recognition. On the other hand, SOM would further reduce the redundancy and represent Gabor features in a set of topological ordered nodes. In which the features are encoded only by the position of the best matching nodes.

5 Experimental results

5.1 ORL face database

The ORL database [13] contains ten different images of each of 40 distinct subjects. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open / closed eyes, smiling / not smiling) and facial details (glasses / no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement).

Table 1: Recognition accuracy (%) for each Gabor filter using ORL face database at various frequency and orientation

λ/θ	0	$\frac{\pi}{8}$	$\frac{2\pi}{8}$	$\frac{3\pi}{8}$	$\frac{4\pi}{8}$	$\frac{5\pi}{8}$	$\frac{6\pi}{8}$	$\frac{7\pi}{8}$
2	58	61	65	72	75	73	68	56
$2\sqrt{2}$	57	63	71	80	77	69	66	60
4	62	67	77	78	77	70	67	62
$4\sqrt{2}$	57	71	77	**80**	**80**	71	68	63
8	57	68	75	**77**	**75**	76	65	62

5.2 The influence of each Gabor filter over the classification accuracy

In this experiment, we evaluate the influence of each Gabor feature in the classification accuracy. Table 1 illustrates the independent contributions of the frequency and orientation parameters of each Gabor filter on the recognition performance. In this experiment, only Gabor features without any projection method are combined with KNN classifier to evaluate the discriminative ability of each filter. It could be observed that good results are obtained at low frequencies than high frequencies i.e. $\lambda \in \{4\sqrt{2}, 8\}$. For the orientation parameter, vertical frequencies (i.e. horizontal bar features) are the most suitable for recognition i.e. $\theta \in \{\frac{3\pi}{8}, \frac{4\pi}{8}\}$. In this sense, face images can be represented using only those subset of filters which significantly reduce the computational complexity without any degrading in the accuracy.

5.3 Classification accuracy using a subset of Gabor filters and HCM neural network

The aim of this experiment is to investigate the performance of the proposed method as an invariant feature descriptor for face images. The above best four filters are used to extract local features from face image, we use one layer HCM neural network consists of 24 HSOM cells, each FEM and FIM map in the cell has 10×8 , 5×4 neurons respectively. All SOMs were trained in sequential-mode using the extracted Gabor features, only 10 updates were performed. The initial weights of all neurons were set to the greatest eigenvectors of the training data, and the neighborhood widths of neurons converged exponentially to 0.1 with the increase of time.

ORL face database is divided into two halves; one-half used as a training data and the other half for test. The training data for our system consists of 200 images, the first five images for each person, the other 200 images are used for blind test. All images in the database are scaled to the size 48×48 pixels, and photometrically normalized with histogram equalization and the image is normalized so that it has zero mean and unit standard deviation, no geometric normalization is performed. In order to classify extracted features, the performance of two simple classifiers such as KNN (K=1) and SVM with linear kernel are compared. Both KNN and SVM classifiers give the same recognition rate (95%), which indicate the robustness of our feature extraction method. For comparisons, recognition accuracies using Eigenfaces [14] and Fisherfaces [15] algorithms combined with KNN classifier, give 83% and 89% respectively.

6 Conclusion and future works

In this paper we addressed the problem of face recognition using Gabor features. Experimental results show the effectiveness of the combination of Gabor features and hierarchical self-organizing map. Using subset of Gabor features reduce the computational cost without any degradation in the performance of the system. The Hierarchical self-organizing map successfully model the nonlinear structure of Gabor responses. In the future, we will investigate the capability of the proposed method to tackle pose and illumination variations of face images.

References

[1] W. Zhao, R. Chellappa P. Phillips, and A. Rosenfeld, Face recognition: a literature survey, Acm Computing Surveys vol. 35, no. 4, pp.399-459, 2003.

[2] D. Field, What is the Goal of Sensory Coding? Neural Computation Vol 6, pp.559-601, 1994.

[3] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C.von der Malsburg, R. P. Wurtz, and W. Konen, Distortion Invariant Object Recognition in the Dynamic Link Architecture, IEEE Trans. Comput., vol. 42, pp.300-311, 1993.

[4] L. Wiskott, Fellous j.M., N. Kruger, and C. von der Malsburg, Face recognition by elastic bunch graph matching, IEEE Trans. PAMI, vol. 19, no. 7, pp. 775-779, 1997.

[5] C. Liu and H. Wechsler, Independent component analysis of gabor features for face recognition, Neural Networks, IEEE Transactions on, 14 (4) (2003) 919-928,

[6] C. Liu and H. Wechsler , Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition, IEEE Transactions on Image Processing, 11 (2002) 467-476.

[7] T. Kohonen, Self-Organizing Maps, Springer, Berlin, 1997.

[8] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, T. Poggio, Robust object recognition with cortex-like mechanisms, IEEE Trans. Pattern Analysis and Machine Intelligence, 29 (3) (2007), 411-426.

[9] T. Sim, S. Baker, M. Bsat, The CMU pose, illumination, and expression database, IEEE Trans. PAMI, 25 (12) (2003) 1615-1618.

[10] N. Tsuruta, R. Taniguchi and M. Amamiya, Hyper column Model: A Combination Model Hierarchical Self-Organizing Maps and Neocognitron for Image Recognition, System and Computer in Japan, vol. 31, no. 2, pp. 49-61, 2000.

[11] J. Lampinen and E. Oja, Clustering Properties of Hierarchical Self Organizing Maps, Journal of Math. Imaging and Vision, vol. 2, no 2-3, pp. 261-272, 1992.

[12] K. Fukushima, Neocognitron, Biological Cybernetics, vol. 36, no. 4, pp. 193-202, 1980.

[13] F. Sammaria, and A. Harter Parameterization of a stochastic model for human face identification In 2nd IEEE workshop on applications of computer vision, Sarasota, FL, 1994.

[14] M. Turk, A. Pentland, : Face recognition using eigenfaces, J Cog Neurosci, vol.19, no.7, pp.71-86, 1991.

[15] Belhumeur P, Hespanha J, Kriegman D (1997) Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Trans PAMI 19:711-720