Gabor Wavelet Selection and SVM Classification for Object Recognition

SHEN Lin-Lin[1] JI Zhen[1]

Abstract This paper proposes a Gabor wavelets and support vector machine (SVM)-based framework for object recognition. When discriminative features are extracted at optimized locations using selected Gabor wavelets, classifications are done via SVM. Compared to conventional Gabor feature based object recognition system, the system developed in this paper is both robust and efficient. The proposed framework has been successfully applied to two object recognition applications, i.e., object/non-object classification and face recognition. Experimental results clearly show advantages of the proposed method over other approaches.

Key words Gabor feature, support vector machine (SVM), object recognition

Feature extraction and classifier learning are essential to the performance of an object recognition system. Discriminative features and robust classifiers are always desirable to pattern recognition applications. However, discriminative features like Gabor features, either require computationally expensive feature extraction process, or have large feature dimension[1]. The large feature dimension could bring huge computation and memory cost to the following classifier training and classification, thus making robust classifiers, e.g. support vector machine (SVM), are inapplicable. This paper tries to propose a framework to design efficient and robust object recognition system.

Due to its biological similarity to human vision system, Gabor wavelets have been widely used in object recognition applications like fingerprint recognition[2], character recognition[3], etc.. One of the most important object recognition applications, face recognition, has also seen the advantages of Gabor-wavelets-based systems over many other methods in the literature. For example, the elastic bunch graph matching (EBGM) algorithm[4] has shown very competitive performance and was ranked the top performer in the face recognition technology (FERET) evaluation[5]. In a recent face verification competition (FVC2004), both of the top two methods used Gabor wavelets for feature extraction. Chung[6] used the Gabor features over a set of 12 fiducial points as input to a principal component analysis (PCA) algorithm, yielding a feature vector of 480 components. They claimed to have improved the recognition rate up to 19 % with this method, compared to that by a raw PCA. Liu[7] vectorized the Gabor responses and then applied a downsampling by a factor of 64 to reduce the computation cost of the following subspace training. Their Gabor-based enhanced Fisher linear discriminant model outperformed Gabor PCA and Gabor Fisherfaces. A more detailed survey on Gabor wavelet based face recognition methods can be found in [1].

Despite the success of Gabor-wavelet-based object recognition systems, both the feature extraction process and the huge dimension of Gabor features extracted demand large computation and memory costs, which makes them impractical for real applications[1]. For the same reason, SVM has seldom adopted Gabor wavelets for feature extraction. While subspace methods like PCA and linear discriminant analysis (LDA) could be applied for dimension reduction[7−8], they do not improve the efficiency of feature extraction process. Some of the research workers have tried to tackle this problem by: 1) downsampling the images[9]; 2) considering the Gabor responses over a reduced number of points[6]; or 3) downsampling the convolution results[7−8]. Strategies 2) and 3) have also been applied together[10]. However, downsampling methods suffer from a loss of information because of the downsampling or dimension reduction. Furthermore, the feature dimension after downsampling might still be too large for fast training of SVM. To make SVM applicable to Gabor features, Qin and He[10] reduced the size by including only the convolution results over 87 manually marked landmarks. However, locating the 87 landmarks itself was a difficult problem. Furthermore, our work[11] has also shown that facial landmarks like eyes, nose, and mouth might not be the optimal locations to extract Gabor features for face recognition.

In this paper, we propose a general object recognition framework based on SVM and the selected Gabor wavelets. The most significant positions and wavelets for extracting features are first learned using a boosting algorithm, where the optimized Gabor responses are computed and used to train a two-class-based SVM for classification. Since only the most important wavelets are used, the two-class-SVM-based system is both efficient and robust. The accuracy and efficiency of such strategy are demonstrated by two applications, i.e., face/non-face classification and face recognition.

1 Gabor wavelets and feature extraction

1.1 Gabor wavelets

Gabor wavelet, which provides the optimized resolution in both time and frequency domains for time-frequency analysis[12], was first proposed by Gabor[13] for 1D signal decomposition. The function was extended to 2D domain by Granlund[14] for 2D image analysis. Recently, we have also applied 3D Gabor wavelets to evaluate 3D image registration algorithms[15]. Gabor wavelets seem to be the optimal basis to extract local features for pattern recognition for several reasons:

1) Biological motivation: the shapes of Gabor wavelets are similar to the receptive fields of simple cells in the primary visual cortex[12].

2) Mathematical motivation: the Gabor wavelets are optimal for measuring local spatial frequencies[1].

3) Empirical motivation: Gabor wavelets have been found to yield distortion tolerant feature spaces for a number of pattern recognition tasks, including texture

Received January 28, 2008; in revised form April 28, 2008
Supported by National Natural Science Foundation of China (60572100, 60673122), Royal Society (U.K.) International Joint Projects 2006/R3-Cost Share with NSFC (60711130233), Science Foundation of Shenzhen City (CXQ2008019, 200706), and Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry (2008[890])
1. School of Computer and Software Engineering, Shenzhen University, Shenzhen 518060, P. R. China
DOI: 10.3724/SP.J.1004.2009.00350

segmentation[16−17], character recognition[3], and fingerprint recognition[2].

In the spatial domain, the 2D Gabor wavelet is a Gaussian kernel modulated by a sinusoidal plane wave. Setting the sharpness of the Gaussian in the x axis as α and y axis as β, the 2D Gabor wavelet can be defined as[18]

$$\begin{aligned}\varphi_{(f,\theta,c_x,c_y)}(x,y) &= \frac{\alpha\beta}{\pi}\exp(-(\alpha^2 x_r^2+\beta^2 y_r^2))\exp(\mathrm{j}2\pi f x_r)\\ x_r &= (x-c_x)\cos\theta+(y-c_y)\sin\theta\\ y_r &= -(x-c_x)\sin\theta+(y-c_y)\cos\theta\end{aligned}\tag{1}$$

where (c_x, c_y) is the central location of the wavelet in the image, f is the frequency of the modulating sinusoidal plane wave, $\alpha = f/\gamma$ and $\beta = f/\eta$ decides the width of the Gaussian envelop, and θ is the orientation of the major axis of the elliptical Gaussian. The size of the Gaussian envelope monotonically varies with the value of the central frequency. The higher the central frequency of the Gabor sinusoidal carrier, the smaller the area the Gaussian envelope will cover in the spatial domain. This is understandable since the high frequency signal changes faster.

A Gabor wavelet is thus determined by the following parameters: the translation (c_x, c_y), the central frequency f, the orientation θ, and the ratio between frequency and the sharpness of the Gaussian axis γ and η. In practical applications, the values of γ and η are normally fixed and a set of wavelets $\boldsymbol{\psi} = (\varphi_{n_1}, \varphi_{n_2}, \cdots, \varphi_{n_N})$ with different frequency f_u, orientation θ_v, and translations (c_x, c_y) are used. Most of the research studies reported in face recognition followed the strategies used in [4, 19], such that $\gamma = \eta = \sqrt{2}$, $f_u = F_{\max}/2^{u/2}$, $u = 0, \cdots, U-1$, and $\theta_v = (v\pi)/8$, $v = 0, \cdots, V-1$. According to the Nyquist sampling theory, a signal containing frequencies higher than half of the sampling frequency cannot be reconstructed completely. Therefore, the upper limit frequency for a 2D image is 0.5 cycles/pixel, while the lower limit is 0. However, for face images, the actual useful band is much narrower, and $F_{\max} = 0.25$ cycles/pixel has been demonstrated to be a reasonable choice[19]. The initial parameter selection aims to achieve a reasonable balance between performance and the increased computation. We have tested different combinations of U and V for face recognition in [8], the experiments led to the selection of Gabor wavelets of five scales and eight orientations for feature extraction. In general applications, (c_x, c_y) could be the coordinates of each pixel in the image. The number of wavelets available, after initial parameter selection, will thus be as large as $U \times V \times C$, where C is the number of pixels in a 2D image. Fig. 1 shows a set of 40 wavelets to be applied to extract features at a brain slice.

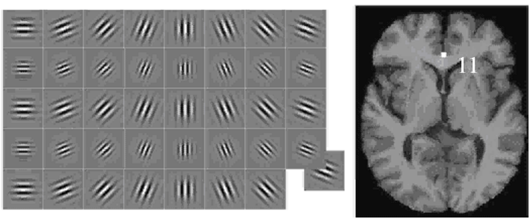

Fig. 1 A set of 40 Gabor wavelets to be applied to feature extraction

1.2 Feature extraction

Given a set of Gabor wavelets $\boldsymbol{\psi} = (\varphi_{n_1}, \varphi_{n_2}, \cdots, \varphi_{n_N})$ going through initial parameter selection, a common feature extraction approach is to construct a feature vector $G(I)$ by concatenating the inner product of an image $I(x, y)$ with each wavelet $\varphi_{n_i}(x, y)$, i.e. $G(I) = (w_{n_1}, w_{n_2}, \cdots, w_{n_N})$, where $w_{n_i} = \| \langle I, \varphi_{n_i}\rangle \|$. The feature set thus consists of the results of the local convolution of the image $I(x, y)$ with all of the initially designed Gabor wavelets φ_n, where $\boldsymbol{n} = (c_x, c_y, \theta, f)$ determines the associated parameters. As an example, taking an image of size 128×128, the Gabor feature vector will be of $128 \times 128 \times 5 \times 8 = 655\,360$ dimensions, which is incredibly large. Due to the large number of convolution operations, the computation and memory cost of feature extraction is also necessarily high. Instead of using all the Gabor wavelets, it is more sensible to select only the relevant Gabor wavelets to perform convolution with the image at appropriate positions. Two questions arise from this consideration: 1) which parameters (θ, f) should be used, and 2) at which image locations (c_x, c_y).

2 The proposed object recognition framework

As described in the last section, the feature extracted using the whole set of Gabor wavelets could bring large memory and computation cost to the following classifiers. To make the object recognition system more efficient, the most significant wavelets shall be identified and used for feature extraction.

2.1 Selecting significant Gabor wavelets for feature extraction

Based on the idea of the wavelet network, Krüeger used a set of Gabor wavelets $\boldsymbol{\psi} = (\varphi_{n_1}, \varphi_{n_2}, \cdots, \varphi_{n_N})$, namely Gabor wavelet network (GWN) to represent an image[20−22], where $\boldsymbol{n} = (c_x, c_y, \theta, \alpha, \beta)$ is a variable vector over the image, c denotes translation, (α, β) dilation, and θ orientation. Thus, each Gabor wavelet is associated with a location in the image, an orientation, and a scale parameter (α, β) (which is decided by the frequency parameter f of the Gabor wavelet, see (1) for details). The GWN for an image I is then defined by optimizing the objective function

$$E = \min_{n_i}\left\| I - \sum_i w_{n_i}\varphi_{n_i}\right\|^2 \tag{2}$$

The primary aim of Krueger′s optimization procedure is to minimize the reconstruction error. Given the optimal GWN of an image, the image can be reconstructed approximately by a linear combination of the weighted wavelets: $I = \sum_i w_i\varphi_i = \boldsymbol{W}^{\mathrm{T}}\boldsymbol{\psi}$, where $\boldsymbol{W} = (w_{n_1}, w_{n_2}, \cdots, w_{n_N})$. The weights of a GWN are directly related to Gabor wavelet responses at image locations. Large weights indicate the similarity of the corresponding wavelets and the local property of the image.

Fig. 2 shows the images reconstructed with 16, 52, 116, and 216 Gabor wavelets (left to right). The quality of the reconstruction of course depends on the number of Gabor wavelets used and can be varied to reach some desired degree of precision. As the set of Gabor wavelets $\boldsymbol{\psi} = (\varphi_{n_1}, \varphi_{n_2}, \cdots, \varphi_{n_N})$ vary from image to image, GWNs are image dependent, which means that the reconstruction coefficients are not uniform across images. Gabor expansion and GWN provide useful approaches to represent objects. However, the Gabor wavelet "basis functions" and GWNs are either non-orthogonal or image dependent. As a result, these methods either do not provide unique representations for objects or do not offer uniform representations across images. For object recognition purposes, rather than seeking the set of Gabor wavelets, which can approximate an image, we should seek Gabor wavelets that are tuned to discriminating one object from another.

Fig. 2 Original image and reconstructed images[20]

Introduced by Freund and Schapire[23], the essence of boosting algorithm is to select a number of very simple weak classifiers, which are then linearly combined into a single strong classifier. The algorithm operates as follows: for a two-class problem, m labeled training samples are given as $(\boldsymbol{x}_i, y_i)$, $i = 1, 2, \cdots, m$, where $y_i \in \{-1, 1\}$ is the class label associated with sample $\boldsymbol{x}_i \in \mathbf{R}^N$. A large number of weak classifiers $h : \mathbf{R}^N \rightarrow \{-1, 1\}$ can be generated to form a weak classifier pool for training. In each of the iterations, the space of all possible weak classifiers is searched exhaustively to find the one that contributes the least to the overall classification error. The error is then used to update the weights associated with each sample such that the wrongly classified samples have their weights increased. When a weak classifier is designed to use only a single Gabor wavelet for feature extraction, boosting is equivalent to wavelet selection.

For two-class object recognition problems, the boosting algorithm can be directly applied to the training samples for wavelet selection. However, when the task of object recognition, e.g. face recognition, is a multi-class problem, it needs to be converted to a two-class problem before the boosting algorithm could be applied. Two spaces, intra-object difference and extra-object difference spaces are defined, with intra-object space measuring dissimilarities between samples of the same object and extra-object space dissimilarities between different objects, respectively. Recall that each component of a Gabor feature is associated with a Gabor wavelet, i.e., it is obtained by performing inner product of an image with a Gabor wavelet $\|\langle I, \varphi_j \rangle\|$. The difference between two images and on this component can be represented as $\|\langle I_p, \varphi_j \rangle\| - \|\langle I_q, \varphi_j \rangle\|$. Since each weak classifier makes decision based on a single Gabor wavelet, in each of the boosting iterations, the space of all possible weak classifiers (wavelets) is searched exhaustively to find the best weak classifier (wavelet) that will produce the lowest classification error. The error is then used to update the weights such that the wrongly classified samples get more focus. Such selected Gabor wavelets should be significant for object recognition, as intra- and extra-object space discrimination is one of the major difficulties in object recognition. More details about the selection process can be found in our previous work[11].

Upon completion of T boosting iterations, T weak classifiers are selected to form the final strong classifier $H(\boldsymbol{x}) = \mathrm{sgn}(\sum_{t=1}^{T} \alpha_i h_t(\boldsymbol{x}))$. The resulting strong classifier, called boosted classifier (BC) in this paper, is a weighted linear combination of all the selected weak classifiers, with each weak classifier using certain Gabor wavelet for feature extraction. At the same time, T most significant Gabor wavelets for feature extraction can also be identified.

2.2 SVM for classification

Once the significant wavelets have been identified, they can be used to extract features for training classifiers for object recognition. Since the wavelets are selected using intra-person and extra-person space discrimination criteria, a natural choice would be the boosting algorithm learned strong classifier, namely BC. In this paper, we also tried SVM and achieved further improvement on classification accuracy with similar efficiency.

Ever since its invention, SVM has also been greatly developed and widely applied to classification and pattern recognition. One of the main reasons for the wide application of SVM is its capacity to handle nonlinearly separable data. SVM is basically a hyperplane classifier $S(\boldsymbol{x}) = \langle \boldsymbol{w}, \boldsymbol{b} \rangle + \boldsymbol{b}$ aimed at solving the two class problem[24]. For non-linearly separable data, a nonlinear mapping function $\phi : \mathbf{R}^N \rightarrow F$, $\boldsymbol{x} \rightarrow \phi(\boldsymbol{x})$ is used to map them into a higher dimension feature space where the hyperplane classifier can be applied. Using the kernel trick[25], the non-linear SVM is found to be

$$S(\boldsymbol{x}) = \mathrm{sgn}\left(\sum_k \alpha_k y_k k(\boldsymbol{x}_k, \boldsymbol{x}) + \boldsymbol{b}\right) \tag{3}$$

where $\boldsymbol{x}_k \in \mathbf{R}^N$ are the support vectors (SVs) learned by SVM, $k(\boldsymbol{x}_k, \boldsymbol{x})$ is a kernel function, e.g., a polynomial kernel, an radial basis function (RBF) kernel, etc.. The decision function of SVM is only based on the dot product of the input feature vector with the SVs, i.e., it has no requirements on the dimension of the feature vector. Theoretically, features with any dimension can be fed into SVM for training. However, in practical implementation, features with large dimension, e.g. Gabor features, could bring substantial computation and memory cost to the SVM training and classification process. In our experiments, the SVM training process did not even complete after 74 hours when a set of Gabor features of 23 040 dimension was used due to the large computation and memory costs. The dimension of features could be substantially reduced when only selected Gabor wavelets are applied. For example, our experimental results show that hundreds of features are enough to achieve very good accuracy. SVM using the selected Gabor wavelets for feature extraction, namely, optimized Gabor support vector machine (OG-SVM), is thus very efficient. Fig. 3 shows the details of the wavelet selection process and training of the proposed OG-SVM.

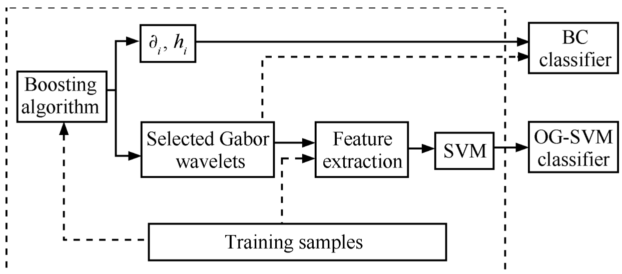

Fig. 3 Learning process of the proposed OG-SVM classifier

2.3 Object recognition

As shown in Fig. 3, once the boosting iterations and the SVM learning process are completed, T selected Gabor wavelets $\psi = (\varphi_1, \varphi_2, \cdots, \varphi_T)$ and two classifiers, i.e. BC and OG-SVM, are created. Though trained to discriminate intra-object and extra-object spaces, BC and OG-SVM could also be used for multi-class object recognition as follows: given a gallery $\{q_j\}$ of m known objects and a probe p to be identified, both classifiers will first compute the Gabor feature differences $\{\boldsymbol{x}_j = [d_1, \cdots, d_t, \cdots, d_T]\}$, $d_t = |\ \|\langle p_j, \varphi_t \rangle\| - \|\langle q, \varphi_t \rangle\|\ |$, between the probe and each of the gallery images, and then calculate an intra-object confidence score using respective decision functions:

$$\delta_j = \begin{cases} \sum\limits_{t=1}^{T} \alpha_t h_t(\boldsymbol{x}_j), & \text{BC} \\ \sum\limits_{k} \alpha_k y_k k(\boldsymbol{x}_k, \boldsymbol{x}_j) + \boldsymbol{b}, & \text{SVM} \end{cases} \tag{4}$$

The probe is then identified as an object that gives the maximum confidence score δ_j.

3 Experimental results

3.1 Object and non-object recognition

3.1.1 Database

We first apply the proposed framework to solve two class object recognition tasks, i.e., face/non-face and car/non-car classification problems. Two image sets, face image set and car image set, are used to test performance of the proposed classifier. The face image set was provided in [26] and contained 4 916 images with faces and 7 872 images without faces. Fig. 4 shows some example face and non-face images. All of the face images were of size 24×24, which were randomly split into a training set and test set containing 2 458 positive samples (faces) and 3 936 negative samples (non-faces) each. The second image set used in our experiments contained 550 images with at least a car in them and 500 images that did not contain a car[27]. The car image set was also randomly split into a training set and a test set. The training set contained 440 car images and 400 non-car images, whilst the remaining 110 car images and 100 non-car images were included in the test set. Fig. 5 shows the sample images from the car image set, which are of size 10 040. The tasks here would be training classifiers to classify images to one of the two classes, e.g. face and non-face. Classifiers were first learned using the training set, then tested using the test set.

Fig. 4 Images from the face image set

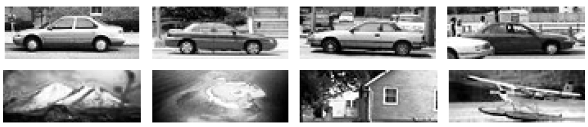

Fig. 5 Images from the car image set

3.1.2 Recognition results

The false accept rates (FAR) and false reject rates (FRR) for the trained BC classifier are shown in Figs. 6 and 7, which show the results on the test face image set and the test car image set. The best face/non-face classifier achieved 99.39 % classification rate and 1.75 % FRR with 150 selected Gabor wavelets/features, while the best car/non-car classifier achieved 97.0 % classification rate and 2.0 % FRR with only 100 wavelets/features.

Based on the results of BC, we applied 150 selected wavelets to extract features for face/non-face recognition. The performance of OG-SVM on the face image set has been shown in Table 1, together with SVM trained using the whole set of Gabor features with dimension 23 040 (G-SVM), using the raw pixels with dimension $24 \times 24 = 576$ (R-SVM) and the BC. For R-SVM, the pixel values of each sample were concatenated to a feature vector to train SVM. A Pentium 4 1.8 GHz PC and the SVM-light package[28]

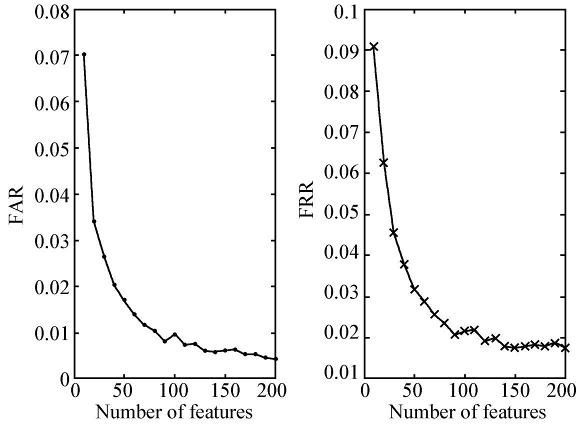

Fig. 6 FAR and FRR on the face image set

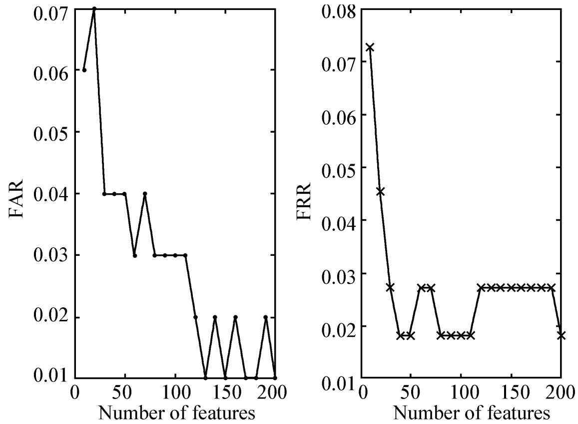

Fig. 7 FAR and FRR on the car image set

Table 1 Classification results of the face image set

	BC	SVM					
		OG-SVM		G-SVM		R-SVM	
		Linear	RBF	Linear	RBF	Linear	RBF
Feature dimension	150	150	150	23 040	23 040	576	576
Number of SVs	N/A	233	271	503	N/A	1 434	1 386
SVM training time	N/A	38 s	75 s	10 h	> 74 h	180 s	270 s
FRR (%)	1.75	1.43	1.26	1.10	N/A	10.49	4.96
FAR (%)	0.61	0.36	0.30	0.18	N/A	3.78	0.97

were used in our experiments. Compared with classifiers using Gabor features, R-SVM achieved the highest FAR and FRR, which suggests that Gabor filters are good choices to extract features for classification. However, due to the huge dimension of Gabor feature, we did not succeed in training the G-SVM using RBF kernel — the program crashed after running 74 hours, which might be caused by high memory usage and computation cost. It also took approximately 10 hours to train G-SVM with linear kernel. The training time might increase exponentially with the number of training samples. The results also suggest that the dimension of features shall be taken into consideration when designing practical SVM classifiers. In addition, the computational cost of convolving an image with 40 Gabor wavelets is very high, which makes G-SVM unsuitable for real time application. Since SVM is specially suited for classification, OG-SVM achieves lower FAR and FRR than BC. Both methods used the same 150 Gabor wavelets selected by boosting algorithm. The training of OG-SVM with RBF kernel took less than 2 min. Only 150 inner products with selected wavelets are necessary to extract Gabor features, which makes OG-SVM highly memory and computational efficient.

3.2 Face recognition

We now applied the proposed framework to a typical multi-class object recognition task, face recognition. As described in Section 2.2, the system starts with selecting significant Gabor wavelets for discriminating intra-object and extra-object spaces. Once the two-class based OG-SVM is learned, it can be applied to object recognition using the decision rule described in Section 2.3.

3.2.1 Database

The FERET database was used here to evaluate the performance of the proposed framework. The database consisted of 14 051 eight-bit grayscale images of human heads with views ranging from frontal to left and right profiles. 600 frontal face images corresponding to 200 subjects were extracted from the database for the experiments and each subject had three images of 256×384 with 256 gray levels. The images were captured at different times under different illumination conditions and contained various facial expressions. Two images of each subject were randomly chosen for training, and the remaining one was used for testing. The following procedures were applied to normalize the face images prior to the experiments:

1) Each image is rotated and scaled to align the centers of the eyes;

2) Each face image is cropped to the size of 64×64 to extract facial region;

3) Each cropped face image is normalized to zero mean and unit variance.

3.2.2 Intra-person and extra-person space discrimination

Using the method described in Section 2.1, a training set consisting of 200 intra-person difference samples and 1 600 extra-person difference samples was generated from the face database. Since both BC and OG-SVM were trained to discriminate intra-person and extra-person differences, we first evaluated their classification performances on the training set. Fig. 8 shows the classification error of BC and OG-SVM with different kernel functions, which were computed as the ratio between the number of wrongly classified difference samples and the number of training samples. One can observe from the figure that the performances of both classifiers improve when the number of features increases. However, the performance of OG-SVM is much more stable than BC. While OG-SVM with RBF kernel achieves the lowest classification error rate (0.44 %) when 140 features are used, OG-SVM with linear kernel shows similar performance.

3.2.3 Recognition performance

The classifiers were then applied to the test set (200 images, 1 image per person) for face recognition and their performances are shown in Fig. 9. Similarly, OG-SVM achieved a higher recognition rate than BC when different number of features were used. The highest identification accuracy of 92 % was achieved by OG-SVM with linear kernel when 120 Gabor features were used. The results also suggest that the difference of OG-SVM using RBF kernel and linear kernel is quite small, when the features selected by boosting algorithm are considered.

To show the efficiency and accuracy of the proposed method, we also compared its performance with other Gabor-wavelet-based approaches in Table 2. While PCA and LDA are also well known as Eigenface and Fisherface methods, details of downsample Gabor + PCA and downsample Gabor + LDA can be found in [8]. In the implementation, downsampling with rate 16 was used to reduce the dimension of extracted Gabor features before they were input to PCA, or LDA for further processing. The table shows that the proposed OG-SVM achieved a similar accuracy with downsample Gabor + LDA, but with much fewer feature dimensions and much less feature extraction costs.

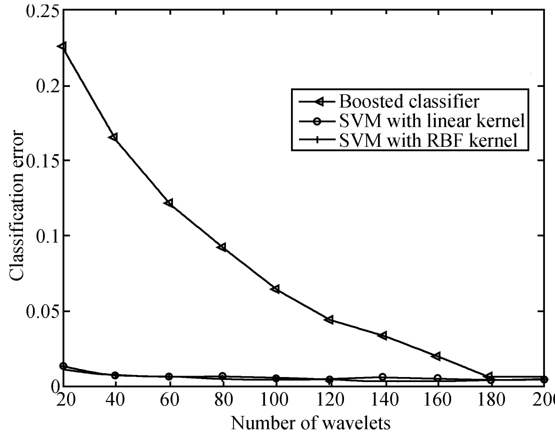

Fig. 8 Classification performances of OG-SVM and BC

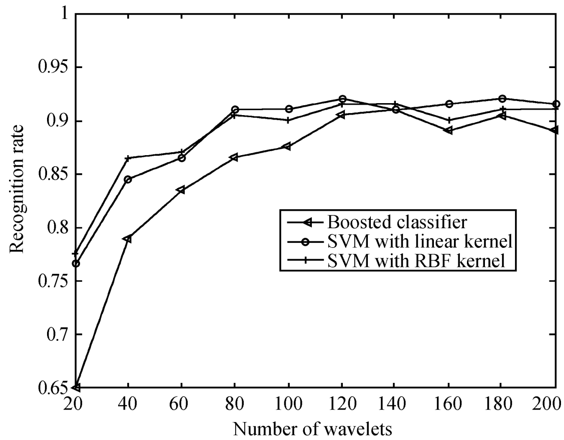

Fig. 9 Recognition performances of OG-SVM and BC

Table 2 Accuracy and efficiency of OG-SVM

Methods	Recognition rate (%)	Number of convolutions for Gabor feature extraction	Dimension of features
PCA	60	N/A	$64\times64=4\,096$
LDA	76	N/A	$64\times64=4\,096$
Downsample Gabor + PCA	80	$64 \times 64 \times 40 = 163\,840$	10 240
Downsample Gabor + LDA	92	$64 \times 64 \times 40 = 163\,840$	10 240
BC	90	120	120
OG-SVM	92	120	120

4 Discussions

We have proposed in this paper an SVM and Gabor-wavelets-based framework for object recognition. While significant Gabor wavelets were applied at selected object landmarks for feature extraction, SVM was used for classification. A boosting-based learning process was used to reduce the feature dimensions and make the Gabor feature extraction process substantially more efficient. An efficient and robust classifier, OG-SVM was then trained. While the criteria integrated in the boosting algorithm was used in the paper to select important Gabor wavelets, the feedback from SVM classification could also be further applied. However, at each iteration the associated SVM needs to be retrained and tested. When the number of candidate wavelets/features is huge, the selection process could become intractable. Intra-object and extra-object concepts have also been adopted in this paper to make boosting algorithm applicable to the multi-class problem. We are currently working on making boosting algorithm directly applicable to multi-class problems.

The proposed object recognition framework has been successfully applied to two object recognition tasks, i.e., object/non-object classification and face recognition. The results clearly show the advantages of the proposed system over other approaches. For example, the performance of OG-SVM is shown to be much more efficient than that of SVM using the whole set of Gabor features, and is much more accurate than that of SVM using the raw pixel values. When applied to face recognition, the two-class-based OG-SVM has been shown to beat several multi-class-based algorithms like PCA, LDA, and downsample Gabor + PCA. By combining optimized Gabor features with SVM, our method not only substantially reduces computation and memory cost of the feature extraction process, but also achieves very accurate recognition performance.

References

1 Shen L L, Bai L. A review on Gabor wavelets for face recognition. *Pattern Analysis and Applications*, 2006, **9**(2-3): 273−292

2 Wang W, Li J W, Huang F F, Feng H L. Design and implementation of Log-Gabor filter in fingerprint image enhancement. *Pattern Recognition Letters*, 2008, **29**(3): 301−308

3 Ding K, Liu Z B, Jin L W, Zhu X H. A comparative study of Gabor feature and gradient feature for handwritten Chinese character recognition. In: Proceedings of International Conference on Wavelet Analysis and Pattern Recognition. Washington D. C., USA: IEEE, 2007. 1182−1186

4 Wiskott L, Fellous J M, Krüger N, von der M C. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, **19**(7): 775−779

5 Phillips P J, Moon H, Rizvi S A, Rauss P J. The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, **22**(10): 1090−1104

6 Chung K C, Kee S C, Kim S R. Face recognition using principal component analysis of Gabor filter responses. In: Proceedings of International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems. Corfu, Greece: IEEE, 1999. 53−57

7 Liu C J, Wechsler H. Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition. *IEEE Transactions on Image Processing*, 2002, **11**(4): 467−476

8 Shen L L, Bai L, Fairhurst M. Gabor wavelets and general discriminant analysis for face identification and verification. *Image and Vision Computing*, 2007, **25**(5): 553−563

9 Zhang W C, Shan S Q, Gao W, Chang Y Z, Cao B, Yang P. Information fusion in face identification. In: Proceedings of the 17th International Conference on Pattern Recognition. Washington D. C., USA: IEEE, 2004. 950−953

10 Qin J, He Z S. A SVM face recognition method based on Gabor-featured key points. In: Proceedings of the 4th International Conference on Machine Learning and Cybernetics. Washington D. C., USA: IEEE, 2005. 5144−5149

11 Shen L L, Bai L. MutualBoost learning for selecting Gabor features for face recognition. *Pattern Recognition Letters*, 2006, **27**(15): 1758−1767

12 Daugman J G. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A — Optics Image Science and Vision*, 1985, **2**(7): 1160−1169

13 Gabor D. Theory of communications. *Journal of Institution of Electrical Engineers*, 1946, **93**(26): 429−457

14 Granlund G H. In search of a general picture processing operator. *Computer Graphics and Image Processing*, 1978, **8**(2): 155−173

15 Shen L L, Bai L. 3D Gabor wavelets for evaluating SPM normalization algorithm. *Medical Image Analysis*, 2008, **12**(3): 375−383

16 Bianconi F, Fernandez A. Evaluation of the effects of Gabor filter parameters on texture classification. *Pattern Recognition*, 2007, **40**(12): 3325−3335

17 Li M, Staunton R C. Optimum Gabor filter design and local binary patterns for texture segmentation. *Pattern Recognition Letters*, 2008, **29**(5): 664−672

18 Zhang B C, Shan S G, Chen X L, Gao W. Histogram of Gabor phase patterns (HGPP): a novel object representation approach for face recognition. *IEEE Transactions on Image Processing*, 2007, **16**(1): 57−68

19 Lades M, Vorbruggen J C, Buhmann J, Lange J, von der M C, Wurtz R P. Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers*, 1993, **42**(3): 300−311

20 Krüeger V, Sommer G. Gabor wavelet networks for efficient head pose estimation. *Image and Vision Computing*, 2002, **20**(9-10): 665−672

21 Krüeger V, Sommer G. Gabor Wavelet Networks for Object Representation, Technical Report CS-TR-4245, Center for Automation Research, University of Maryland, USA, 2001

22 Krüeger V, Sommer G. Wavelet networks for face processing. *Optical Society of America Journal*, 2002, **19**(6): 1112−1119

23 Freund Y, Schapire R. A short introduction to boosting. *Journal of Japanese Society for Artificial Intelligence*, 1999, **14**(5): 771−780

24 Burges C J C. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 1998, **2**(2): 121−167

25 Cristianini N, Shawe-Taylor J. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge: Cambridge University Press, 2000

26 Carbonetto P. Viola training data [Online], available: http://www.cs.ubc.ca/∼pcarbo/, April 30, 2008

27 The UIUC image database for car detection [Online], available: http://l2r.cs.uiuc.edu/cogcomp/Data/Car/, April 30, 2008

28 SVM-light package vector machine [Online], available: http://svmlight.joachims.org/, April 30, 2008

SHEN Lin-Lin Associate professor at the School of Computer and Software Engineering, Shenzhen University. He received his Ph. D. degree from University of Nottingham, UK in 2005. Before joining Shenzhen University, he had been working as a research fellow on MRI brain image processing at Medical School, University of Nottingham. His research interest covers image processing, pattern recognition, and biometrics. Corresponding author of this paper. E-mail: llshen@szu.edu.cn

JI Zhen Professor at Texas Instruments DSPs Laboratory, Shenzhen University. He received his bachelor and Ph. D. degrees from Xi′an Jiaotong University, in 1994 and 1999, respectively. In 2001, 2003, 2004, 2006, and 2007, he worked as an academic visitor in the Department of Electronics and Electrical Engineering, University of Liverpool, UK. His research interest covers DSP design, image coding, embedded system, computation intelligence, digital watermarking, and medical image processing.
E-mail: jizhen@szu.edu.cn