Kernel Enhanced Informative Gabor Features for Face Recognition

LinLin Shen and Li Bai
School of Computer Science & IT
University of Nottingham
Jubliee Campus, Nottingham, NG8 1BB, UK
{lls, bai}@cs.nott.ac.uk

Abstract

A discriminative and robust feature -- Kernel enhanced informative Gabor feature is proposed in this paper for face recognition. Mutual information is applied to select a set of informative and non-redundant Gabor features, which are then further enhanced by Kernel methods for recognition. When compared with an approach using the downsampled Gabor features, our methods introduce advantages on computation, memory cost and accuracy. The proposed method has also been fully tested on the FERET database according to the evaluation protocol, significant improvements on the test set is observed. Compared with the classical Gabor feature extraction approach using complex convolution process, our method requires less than 4ms to retrieve a few hundreds of features. Due to the substantially reduced feature dimension, only 4 seconds are required to recognize 200 face images.

1 Introduction

Motivated by the functional similarity of Gabor filters with the cells in the visual cortex of human visual system, Daugman [1] presented evidence that such visual neurons could optimize the general uncertainty relations for resolution in space, spatial frequency and orientation. From an information theoretic viewpoint, Okajima [2] derived Gabor functions as solutions for a certain mutual-information maximization problem. The work shows that the Gabor-type receptive field can extract the maximum information from local image regions. Researchers have also shown that Gabor features, when appropriately designed, are invariant against translation, rotation and scale [3]. Successful applications of Gabor filters in face recognition can be found in the FERET evaluation [4], where Elastic Bunch Graph Matching method [5] gave the best performance. More recent face verification competition 2004 [6] also demostrates the success of Gabor filters: both of the top two approaches apply Gabor filters for feature extraction.

For face recognition applications, the number of Gabor filters used to convolve face images varies with applications, but usually 40 filters (5 scales and 8 orientations) are used [5;7-9]. However, due to the large number of convolution operations, the computation cost is quite high. Even a parallel computer system has been used, it was reported in [7] that the convolution of a 128×128 pixel image with 40 Gabor filters took about 7 seconds. For global methods, the dimension of the feature vectors extracted is also incredibly large, e.g., 163,840 for image with size 64×64. To address this issue, a method is described in [10] that performed Gabor feature selection for facial landmark detection by a trial-and-error method. A sampling method is proposed in [11] to determine the "optimal" position for extracting Gabor feature. However, the selection criterion is ad hoc. Moreover, the same filters, which might not be optimal, are applied at different locations. Genetic algorithm (GA) has also been used to select Gabor features for pixel classification [12] and vehicle detection [13]. The basic approach is to create a population of randomly selected combinations of features. Each combination is considered a possible solution to the feature selection problem. However, the computation demanding of GA is very high, particularly in the case where a huge number of features are available. In addition, the GA selection is decision algorithm dependant. Recently, AdaBoost algorithm has bee used to select Haar-like features for face detection [14] and learn the most discriminative Gabor features for classification [15]. Once the learning process is finished, Gabor filters of different frequencies and orientations are applied at different locations of the image for feature extraction.

BMVC 2005 doi:10.5244/C.19.20

Despite of its success, AdaBoost algorithm selects only features that perform "individually" best, the redundancy among selected features is not considered [16]. In this paper we present a conditional mutual information [17;18] based method to select Gabor features for face recognition. A small subset of Gabor features capable of discriminating intra-person and inter-person spaces is selected using the information theory, which is then subjected to Generalized Discriminant Analysis (GDA) for face recognition. The experimental results show that 200 features are enough to achieve highly competitive accuracy. Significant computation and memory efficiency have been achieved since the number of features has been reduced from 163,840 to 200 for 64×64 images. Once the informative Gabor features are selected, Generalized Discriminant Analysis (GDA) is applied for further enhancement. Compared with the whole set of Gabor features, GDA using the selected feature achieves similar accuracy, with fewer number of features and substantially faster speed. The kernel enhanced informative Gabor features have also been tested on the full FERET database following the evaluation protocol, experimental results show that the performance of our algorithm is state of the art, but with significantly higher efficiency.

2 Gabor Feature Extraction

2.1 Gabor Wavelets

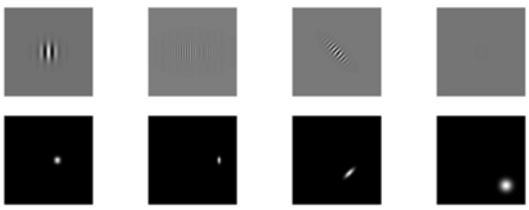
In the space domain, the 2D Gabor filter is a Gaussian kernel modulated by a sinusoidal plane wave [3]:

$$\begin{aligned} g(x,y) &= w(x,y)s(x,y) = e^{-(\alpha^2 x'^2 + \beta^2 y'^2)} e^{j2\pi f x'} \\ x' &= x\cos\theta + y\sin\theta \\ y' &= -x\sin\theta + y\cos\theta \end{aligned} \qquad (1)$$

where f (cycles/pixel) is the central frequency of the sinusoidal plane wave, θ is the anti-clockwise rotation of the Gaussian and the plane wave, α is the sharpness of the Gaussian along the major axis parallel to the wave, and β is the sharpness of the Gaussian minor axis perpendicular to the wave. To keep the ratio between frequency and sharpness constant, $\gamma = \frac{f}{\alpha}$ and $\eta = \frac{f}{\beta}$ are defined and the Gabor filters can now be rewritten as:

$$\varphi(x,y) = \frac{f^2}{\pi\gamma\eta} g(x,y) = \frac{f^2}{\pi\gamma\eta} e^{-(\alpha^2 x'^2 + \beta^2 y'^2)} e^{j2\pi f x'} \qquad (2)$$

Figure 1 shows four Gabor filters with different parameters in both spatial domain and frequency domain.



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

Figure 1 The Gabor filters with different parameter $\Pi(f,\theta,\gamma,\eta)$ in spatial domain (the 1st row) and frequency domain (the 2nd row), (a) $\Pi_a(0.1,0,1,1)$; (b) $\Pi_b(0.3,0,6,3)$ (c) $\Pi_c(0.2,\pi/4,3,1)$ (d) $\Pi_d(0.4,\pi/4,2,2)$

2.2 Gabor Feature Representation

Once Gabor filters have been designed, image features at different location, frequency and orientation can be extracted by convolving the image $I(x,y)$ with the filters:

$$O_{\Pi(f,\theta,\gamma,\eta)}(x,y) = I(x,y) * \varphi_{\Pi(f,\theta,\gamma,\eta)}(x,y) \qquad (4)$$

A number of Gabor filters at different scales and orientations are usually used. We designed a filter bank with 5 scales and 8 orientations for feature extraction [7]:

$$\gamma = \eta = \sqrt{2}\pi, f_u = f_{\max} / \sqrt{2}^u, \theta_v = \frac{v}{8}\pi, u = 0,...,4, v = 0,...,7 \quad (5)$$

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 low limit is 0. As a result, we set $f_{\max} = 0.5$. The resultant Gabor feature set thus consists of the convolution results of an input image $I(x, y)$ with all of the 40 Gabor filters:

$$S = \{O_{u,v}(x, y) : u \in \{0,...,4\}, v \in \{0,...,7\} \quad (6)$$

where $O_{u,v}(x, y) = \left| I(x, y) * \varphi_{\Pi(f_u, \theta_v, \gamma, \eta)}(x, y) \right|$. Figure 2 shows the magnitudes of Gabor representations of a face image with 5 scales and 8 orientations. A series of row vectors $\mathbf{O}_{u,v}$ could be converted out of $O_{u,v}(x, y)$ by concatenating its rows or columns, which are then concatenated together to generate a discriminative Gabor feature vector:

$$G(I) = \mathbf{O} = (\mathbf{O}_{0,0}\ \mathbf{O}_{0,1} \cdots \mathbf{O}_{4,7}) \quad (7)$$

Take an image with size 64×64 for example, the convolution result will give 64×64×5×8=163,840 features. Each Gabor feature is thus extracted by a filter with parameters f_u, θ_v at location (x, y). Since the parameters of Gabor filters are chosen empirically, we believe a lot of redundant information is included, and therefore a feature selection mechanism should be used to choose the most useful features for classification.

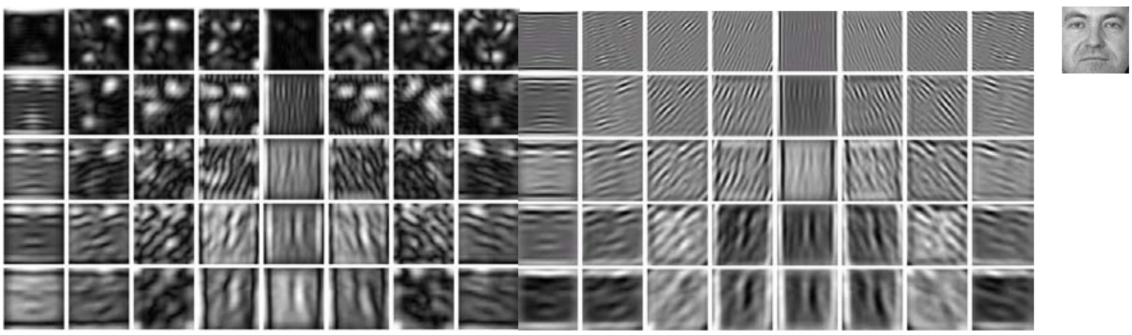


Figure 2 Convolution result - (magnitude and real part) of an image with 40 Gabor filters

2.3 The Gabor Feature Difference Space

In [19], the face recognition problem is formulated as a problem in the difference space, which model dissimilarities between two facial images. Two classes, dissimilarities between faces of the same person (intra-personal space) and dissimilarities between faces of the different people (extra-personal space) are defined. The two Gabor feature difference sets: CI (intra-personal difference) and CE (extra-personal difference) can be defined as:

$$\begin{aligned} CI &= \{\|G(I_p) - G(I_q)\|, p = q\} \\ CE &= \{\|G(I_p) - G(I_q)\|, p \neq q\} \end{aligned} \quad (8)$$

where I_p and I_q are the facial images from people p and q respectively, and $G(\cdot)$ is the Gabor feature extraction operation as defined in last section. Each of the M samples in the difference space can now be described as $g_i = [x_1 x_2 \cdots x_n \cdots x_N], i = 1,2,\cdots,M$, where N is the dimension of extracted Gabor features and $x_n = \left(\|G(I_p) - G(I_q)\|\right)_n = \left(\|\mathbf{O}_p - \mathbf{O}_q\|\right)_n$.

3 Selecting Informative Gabor Features

3.1 Entropy and Mutual Information

As a basic concept in information theory, entropy $H(X)$ is used to measure the uncertainty of a random variable (r.v.) X. If X is a discrete r.v., $H(X)$ can be defined as below:

$$H(X) = -\sum_x p(X = x)\lg(p(X = x)) \quad (9)$$

Mutual information $I(Y;X)$ is a measure of general interdependence between two random variables X and Y:

$$I(Y;X) = H(X) + H(Y) - H(X,Y) \quad (10)$$

Using Bayes rule on conditional probabilities, Eq. can be rewritten as:

$$I(Y;X) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X) \quad (11)$$

Since $H(Y)$ measures the priori uncertainty of Y and $H(Y \mid X)$ measures the conditional posteriori uncertainty of Y after X is observed, the mutual information $I(Y;X)$ measure how much the uncertainty of Y is reduced if X has been observed. It can be easily shown that if X and Y are independent, $H(X,Y) = H(X) + H(Y)$, consequently their mutual information is zero.

3.2 Conditional Mutual Information (CMI)

Given a set of candidate features $(X_1, X_2, ..., X_N)$ and sample labels Y

$K = 1$

$v(K) = \arg\max_n I(Y; X_n)$

while $K < K_{\max}$

for each candidate feature X_n

calculate CMI given each of the selected feature $X_{v(k)}, k = 1,2,...K$

$$I(Y; X_n \mid X_{v(k)}) = H(Y \mid X_{v(k)}) - H(Y \mid X_n, X_{v(k)})$$
$$= H(Y, X_{v(k)}) - H(X_{v(k)}) - H(Y, X_n, X_{v(k)}) + H(X_n, X_{v(k)})$$

end

$v(K+1) = \arg\max_n \{\min_k I(Y; X_n \mid X_{v(k)})\}$

$K = K + 1$

end

Figure 3 CMI for feature selection

In terms of information theory, the aim of feature selection is to select a small subset of features $(X_1, X_2, ..., X_N)$ that gives as much information as possible about Y, i.e. maximize $I(Y; X_1, X_2, ..., X_N)$. However, the estimation of this expression is unpractical since the number of probabilities to be decided could be as huge as 2^{N+1} even when the value of r.v. is binary. To address this issue, one approach is to use conditional mutual information (CMI) for feature fitness measurement. Given a set of features $(X_1, X_2, ..., X_K)$, CMI $I(Y; X_n \mid X_k)$ could be used to measure the information about Y carried by the feature X_n when a feature $X_k, k = 1,2,...K$ is already selected. We can justify the fitness of a candidate feature by its CMI given each feature already picked, i.e., a candidate feature is good only if it caries information about Y, and if this information has not been caught by any of the X already picked. This selection process thus takes both individual power and redundancy among selected features into consideration. As a result, the process shown in Figure 3 can be used to select a subset of $K_{\max}$ features $(X_{v(1)}, X_{v(2)}, ..., X_{v(K_{\max})})$.

3.3 Application for Gabor Feature Selection in The Difference Space

The estimation of CMI requires information about the marginal distribution $p(X_n)$, $p(Y)$ and the joint probability distribution $p(Y, X_{v(k)})$, $p(X_n, X_{v(k)})$ and $p(Y, X_n, X_{v(k)})$, which could be approximated by histograms estimation. However, it is very difficult to determine the number of histogram bins. Though Gaussian distribution could be applied as well, many of the features, as shown in the experimental section, do not show Gaussianity. To reduce the complexity and computation cost of the feature selection process, we hereby focus on random variables with binary values only, i.e., $x_n \in \{0,1\}, y \in \{0,1\}$, where x_n and y are the values of random variables X_n and Y respectively. For binary r.v., the probability could be estimated by simply counting the number of possible cases and dividing that number with the total number of training samples. For example, the possible cases will be $\{(0,0),(0,1),(1,0),(1,1)\}$ for the joint probability of two binary r.v. $p(Y, X_{v(k)})$.

The concept of intra-personal, extra-personal space and Gabor feature differences between two facial images are used in this paper, see section 2.3 for details. Each sample $g_i = [x_1 x_2 \cdots x_n \cdots x_N]$ in the difference space is now associated with a binary label: $y = 0$ for an intra-personal difference, while $y = 1$ for an extra-personal difference. Each feature of the sample in the difference space is also converted to binary value as below, i.e., if the difference is less than a threshold, the difference is set as 0, otherwise it is set as 1.

$$x_n = \begin{cases} 0, & if \quad x_n < t_n \\ 1, & if \quad x_n \geq t_n \end{cases} \tag{12}$$

Since we are only interested in the selection of features, the threshold t_n is simply determined by the centre of intra-personal samples mean and extra-personal samples mean.

$$t_n = \frac{1}{2}\left(\frac{1}{m}\sum_{p=1}^{m}\left(\left(g_p\right)_n \middle| y_p = 1\right) + \frac{1}{l}\sum_{q=1}^{l}\left(\left(g_q\right)_n \middle| y_p = 0\right)\right) \tag{13}$$

where m and l is the number of intra and extra personal difference samples, respectively. Once a set of labelled training samples is given as $\{(g_1, y_1), (g_2, y_2), \ldots (g_M, y_M)\}$, where $y_i \in \{0,1\}$ is the class label (intra-personal or extra-personal) associated with example $g_i = [x_1 x_2 \cdots x_N], x_n \in \{0,1\}$, the iterative process described in last section can be used to select the informative Gabor features. The Gabor features thus selected are carrying important information about predicting whether the sample is an intra-personal difference, or an extra-personal difference. Based on the fact that face recognition is actually to find the most similar match with the least difference, the selected features will also be very important for recognition.

4 Training Samples Generation

For a training set with L facial images captured for each of the D persons, $D\binom{L}{2}$ samples could be generated for intra-personal difference class while $\binom{DL}{2} - D\binom{L}{2}$ samples are available for extra-personal difference class. There are always much more extra-personal samples than intra-personal samples for face recognition problems. Take a database with 400 images from 200 subjects for example, 200 intra-personal image pairs and $\binom{400}{2} - 200 = 79{,}800$ extra-personal image pairs are available. To achieve a balance between the numbers of training samples from the two classes, a random subset of the extra-personal samples could be produced. However, we also want to make the subset be representative of the whole set as much as possible. To achieve this trade off, we proposed the procedure shown in Figure 4 to generate M extra-personal samples using U Gabor filters:Instead of using only M pairs, our method randomly generates M samples from $M \times U$ extra-personal image pairs. As a result, without increasing the number of extra-personal samples to bias the feature selection process, the training samples thus generated are more representative.

```
For j = 1,2,···M
    For i = 1,2,···U
        Randomly generate an image pair from different person
        Calculate the Gabor feature difference f_i corresponding to filter i using the image pair
    End
    Combine the U feature differences into an extra-personal sample, g_j = [f_1, f_2, ···, f_i, ··· f_U]
End
```

Figure 4 Training samples generation algorithm

5 Kernel Enhancement for Recognition

Once the most informative Gabor features are selected, different classification strateges could be used for face recognition, e.g., Principal Component Analysis (PCA) or Linear Discriminant Analysis (LDA) can be further applied for enhancement before the nearest neighbour (NN) classifier is used for classification. Recently, kernel methods have been successfully applied to solve pattern recognition problems because of their capacity in handling nonlinear data. By mapping sample data to a higher dimensional feature space, effectively a nonlinear problem defined in the original image space is turned into a linear problem in the feature space [20]. Support Vector Machine (SVM) is a successful example of using Kernel trick for classification. However, SVM is basically designed for two classes problem and it has been shown in [21] that non-linear Kernel subspace methods perform better than SVM for face recognition. As a result, we use in this work Generalized Discrimniant Analysis (GDA) [22] for further feature enhancement and KNN classifier for recognition. GDA subspace is firstly constructed from the training image set and each image in the gallery set is projected onto the subspace. To classify an input image, the selected Gabor features are extracted and then projected to the GDA subspace. The similarity between any two facial images can then be determined by the normalized correlation distance of the projected vectors. Details of applying GDA for face recognition can be found in [9].

6 Experimental Results

We firstly analyse the performance of our algorithm using a subset of FERET database, which is a standard testbed for face recognition technologies [4]. 600 frontal face images corresponding to 200 subjects are extracted from the database for the experiments - each subject has three images of size 256×384 with 256 gray levels. The images were captured at different photo sessions so that they display different illumination and facial expressions. Two images of each subject are randomly chosen for training, and the remaining one is used for testing. The following procedures were applied to normalize the face images prior to the experiments:

- The centres of the eyes of each image are manually marked,
- Each image is rotated and scaled to align the centres of the eyes,
- Each face image is cropped to the size of 64×64 to extract facial region
- Each cropped face image is normalized to zero mean and unit variance

Figure 5 Sample images used in experiments

Figure 5 shows the sample images from the database. The first two rows are the example training images while the third row shows the example test images.

6.1 Selected Gabor Features

The randomly selected 400 face images (2 images each subject) are used to learn the most important Gabor feature for intra-personal and extra-personal face space discrimination. As a result, 200 intra-personal face difference samples and 1,600 extra-personal face difference samples using the method as described in section 4 are randomly generated for feature selection. Figure 6 shows the first six selected Gabor features and locations of the first 200 Gabor features on a typical face image in the database. It is interesting to see that most of the selected Gabor features are located around the prominent facial features such as eye brows, eyes, noses and chins, which indicates that these regions are more robust against the variance of expression and illumination. This result is agreeable with the fact that the eye and eyebrow regions remain relatively stable when the person's expression changes. Figure 7 shows the distribution of selected filters in different scales and orientations. As shown in Figure 7, filters centred at low frequency band are selected much more frequently than those at high frequency band.

On the other hand, majority of the discriminative Gabor features are with orientation around $3\pi/8$, $\pi/2$ and $5\pi/8$.

(a) (b) (c) (d) (e) (f) (g)

Figure 6 First six selected Gabor features (a)-(f); and the 200 selected feature points (g)

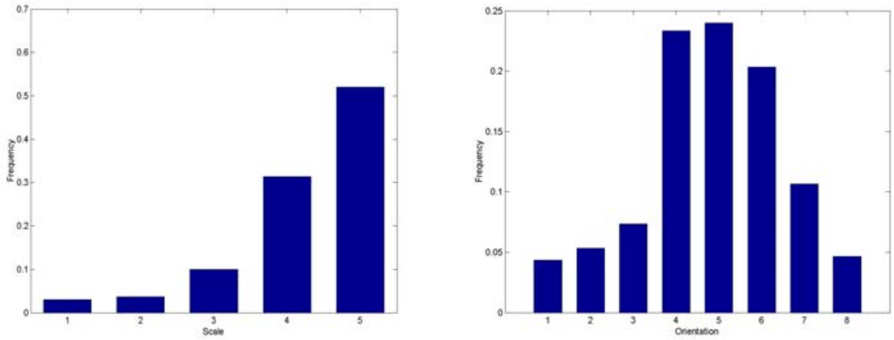


Figure 7 Distribution of selected filters in scale and orientation

6.2 Recognition Performance on the Subset of FERET database

Once the informative Gabor features (InfoGabor) are selected, we are now able to apply them directly for face recognition. Normalized correlation distance measure and 1-NN classifier are used. For comparison, we have also implemented the Adaboost algorithm to select Gabor features for face recognition (BoostedGabor), using exactly the same training set. During boosting, exhaustive search is performed in the Gabor feature difference space as defined in (7). By picking up at each iteration the feature with the lowest weighted classification error, AdaBoost algorithm selects one by one those features that are significant for classification. Details of the learning process can be found in [15]. The performance shown in Figure 8 proves the advantage of InfoGabor over BoostedGabor. The performance drop using 120 features could be caused by the variance between test images and training images -- some features significant to discriminate training images might not be the appropriate ones for test images. A more representative training set could alleviate this problem. As shown in the figure, InfoGabor achieved as high as 95% recognition rate with 200 features.

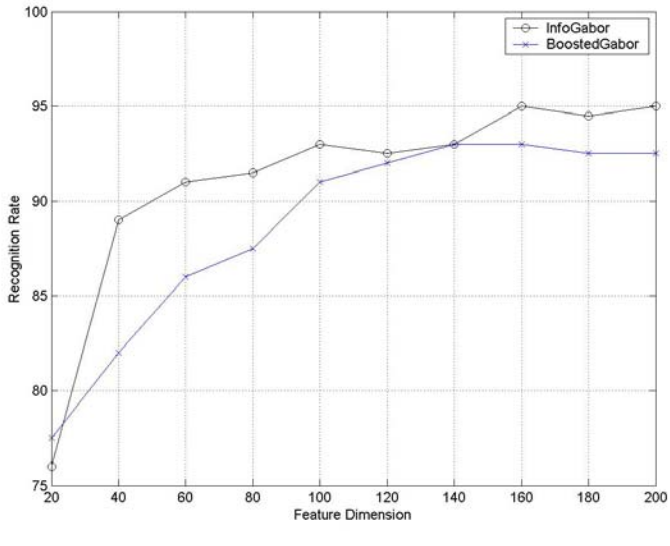


Figure 8 Recognition performance

In the following experiments, we perform GDA on the selected Gabor features (InfoGabor-GDA) for further feature enhancement. To show the robustness and efficiency of the proposed methods, we also perform GDA on the whole Gabor feature set (Gabor-GDA) for comparison purposes. To make the subspace learning process tractable, downsampling is adopted to reduce feature dimension to a certain level [9]. GDA with RBF kernel are used in our experiments. Normalized correlation distance measure and the nearest neighbour classifier are adopted for both methods. Due to the enhancement of kernel methods, the InfoGabor-GDA further improve the accuracy from 95% to 98%, which is even higher than the method of Gabor-GDA. The comparison shows that some important Gabor features may have been missing during the dowsampling process for Gabor-GDA, while many features remained are, on the other hand, redundant. With non-redundant and informative Gabor features, the proposed method achieves better accuracy with significantly less computation. We also compare the computation and memory costs of Gabor-GDA and InfoGabor-GDA in Table 1. This shows that InfoGabor-GDA requires significantly less computation and memory cost than Gabor-GDA, e.g., the number of convolutions to extract Gabor features is reduced from 163,840 to 200. Although Fast Fourier Transform (FFT) could be used here to circumvent the convolution process, the feature extraction still takes about 1.5 seconds in our C implementation while the 200 convolutions takes only less than 4ms. For Gabor-GDA with downsample rate = 16, the feature dimension is reduced to 10,240, which is still 50 times of the dimension of InfoGabor-GDA. As a result, InfoGabor-GDA is much faster in training and testing. While it takes Gabor-GDA 275 seconds to construct the GDA subspace using the 400 training images, it takes InfoGabor-GDA only about 6 seconds. InfoGabor-GDA also achieves substantial recognition efficiency - only 4 seconds are required to recognize the 200 test images. The computation time is recorded in Matlab 6.1, with a P4-1.8GHz PC.

	Number of Convolutions to Extract Gabor Feature	**Dimension of Gabor Features before GDA**	**Training Time**	**Test Time**	**Recognition Rate**
Gabor-GDA	64×64×40 = 16,3840	10,240	275 sec.	263 sec.	97%
InfoGabor-GDA	200	200	6 sec.	4 sec.	98%

Table 1 Comparative of Gabor-GDA and InfoGabor-GDA

6.3 Recognition Performance on the Full Set of FERET Database

After showing the comparative results with a state of the art Gabor feature based algorithm, we are now testing our InfoGabor-GDA algorithm on the whole FERET database. According to the evaluation protocol, a gallery of 1196 frontal face images and 4 different prob sets are used for testing. The numbers of images in different prob sets are listed at Table 2, with example images shown in Figure 9. Fb and Fc prob sets are used for assessment of the effect of facial expression and illumination changes respectively, and there is only a few seconds between the capture of the gallery-probe pairs. DupI and Dup II consist of images taken on different days with their corresponding gallery images, and particularly, there is at least one year between the acquisition of the probe image in Dup II and the corresponding gallery image. A training set consists of 736 images, is used to select the most informative Gabor features and construct the GDA subspace. Note that the same set was released to researchers to develop their algorithms during FERET evaluation. As a result, 592 intra-personal and 2000 extra-personal samples are produced to select 300 Gabor features using the sample generation algorithm and information theory. During development phase, the training set is randomly divided into a gallery set and a test set to decide the dimension for GDA for optimal performance. The same parameters developed are used through the testing process.

Prob Set	**Gallery**	**Prob set size**	**Gallery size**	**Variations**
Fb	Fa	1195	1196	Expression
Fc	Fa	194	1196	Illumination and Camera
Dup I	Fa	722	1196	Time gap < 1 week
Dup II	Fa	234	1196	Time gap > 1 year

Table 2 List of different prob sets

Performance results of the proposed algorithm are shown in Table 3, together with that of the main approaches participating FERET evaluation [4], and an approach extract Gabor features from variable feature points for recognition [23]. The results show that our method achieves the best result on sets Fb, Fc and Dup II, due to the robustness of selected Gabor features against variation of expression and capture time. Particularly, the performance of our methods is significantly better than all of other methods on Dup II. The Elastic Bunch Graph Matching (EBGM) method, based on the Dynamic Link Architecture, perform a little better than our methods on Dup I. However, the method requires intensive computation complexity for both Gabor feature extraction and graph matching. Compared with their approach, our method is much faster in efficiency.

Figure 9 Examples of different probe images

Method	Fb	Fc	Dup I	Dup II
PCA	83.4%	18.2%	40.8%	17.0%
PCA + Bayesian	94.8%	32.0%	57.6%	35.0%
LDA	96.1%	58.8%	47.2%	20.9%
Elastic Bunch Graph Matching	95.0%	82.0%	**59.1%**	52.1%
Variable Gabor Features [24]	96.3%	69.6%	58.3%	47.4%
InfoGabor-GDA	**96.9%**	**85.57%**	56.5%	**65.38%**

Table 3 FERET evaluation results for various face recognition algorithms

7 Conclusions

Mutual information theory has been successfully applied to select informative Gabor features for face recognition. To simplify the computation cost and algorithm complexity, the intra-personal and extra-personal difference spaces are used. In this space, the value of each random variable is binary. The Gabor features thus selected are non-redundant of each other, while carrying important information about the identity of face images. Being applied Generalized Dsicriminant Analysis, the selected features are further enhanced in the non-linear Kernel space. Our algorithm has been fully tested using extensive database. Compared with an approach using the downsampled Gabor features, our method shows advantage over both accuracy and efficiency. The results on the full FERET database following the evaluation protocol also show that our algorithm achieves better performance on 3 test data sets than the top method in the competition – the elastic graph matching algorithm. However, our algorithm has advantage in computation cost and efficiency since no graph matching process is needed. In addition, our method achieves significantly better performance on the most difficult test set Dup II.

When the feature selection process in this paper address the r.v. with binary values only, they could certainly be extended to the case of continuous variable. The distribution of the feature could either be represented using Gaussian model, or discretized using histogram. When the r.v.s with multiple values are used, the feature selection process will require much more computation cost and complexity.

The number of features to be selected is currently decided by experiments. A more reasonable method is to use the value of information gain for reference. If the gain of including a new feature is less than a threshold, we can say that the inclusion of new feature does not bring any more useful information, which will terminate the selection process. We are currently working on how to determine the threshold.

References

[1] J. G. Daugman, "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*, vol. 2, no. 7, pp. 1160-1169, 1985.

[2] K. Okajima, "Two-dimensional Gabor-type receptive field as derived by mutual information maximization," *Neural Networks*, vol. 11, no. 3, pp. 441-447, 1998.

[3] V. Kyrki, J. K. Kamarainen, and H. Kalviainen, "Simple Gabor feature space for invariant object recognition," *Pattern Recognition Letters*, vol. 25, no. 3, pp. 311-318, 2004.

[4] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *Ieee Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090-1104, 2000.

[5] L. Wiskott, J. M. Fellous, N. Kruger, and C. vonderMalsburg, "Face recognition by elastic bunch graph matching," *Ieee Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775-779, 1997.

[6] K. Messer, J. Kittler, M. Sadeghi, M. Hamouz, A. Kostin, F. Cardinaux, S. Marcel, S. Bengio, C. Sanderson, N. Poh, Y. Rondriguez, J. Czyz, L. Vandendorpe, C. McCool, S. Lowther, S. Sridharan, V. Chandran, R. P. Palacios, E. Vidal, L. Bai, L. Shen, Y. Wang, Y. H. Chiang, H. C. Liu, Y. P. Huang, A. Heinrichs, M. Miiller, A. Tewes, C. v. d. Malsburg, R. Wiirtz, Z. G. Wang, F. Xue, Y. Ma, Q. Yang, C. Fang, X. Q. Ding, S. Lucey, R. Goss, and H. Schneiderman, "Face authentication test on the BANCA database," in *Proc. of International Conference on Pattern Recognition* Cambridge, UK: 2004.

[7] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. Vandermalsburg, R. P. Wurtz, and W. Konen, "Distortion invariant object recognition in the Dynamic Link Architecture," *Ieee Transactions on Computers*, vol. 42, no. 3, pp. 300-311, 1993.

[8] C. J. Liu and H. Wechsler, "Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition," *Ieee Transactions on Image Processing*, vol. 11, no. 4, pp. 467-476, 2002.

[9] L. Shen and L. Bai, "Face recognition based on Gabor features using kernel methods," in *Proc. of the 6th IEEE Conference on Face and Gesture Recognition* Korea: 2004, pp. 170-175.

[10] I. R. Fasel, M. S. Barlett, and J. R. Movellan, "A comparison of Gabor filter methods for automatic detection of facial landmarks," in *Proceedings of the 5th IEEE Int. Conf. on Automatic Face and Gesture Recognition,* 2002.

[11] D. H. Liu, K. M. Lam, and L. S. Shen, "Optimal sampling of Gabor features for face recognition," *Pattern Recognition Letters*, vol. 25, no. 2, pp. 267-276, 2004.

[12] N. W. Campbell and B. T. Thomas, "Automatic selection of Gabor filters for pixel classification," in *Sixth International Conference on Image Processing and its Applications* 1997, pp. 761-765.

[13] Z. Sun, G. Bebis, and R. Miller, "Evaluationary Gabor filter optimization with application to vehicle detection," in *Proc. of the 3rd IEEE International Conference on Data Mining* Florida, USA: 2003.

[14] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition* Kauai, Hawaii: 2001.

[15] L. Shen and L. Bai, "AdaBoost Gabor Feature Selection for Classification," in *Proc. of Image and Vision Computing NewZealand* 2004, pp. 77-83.

[16] S. Z. Li and Z. Q. Zhang, "FloatBoost learning and statistical face detection," *Ieee Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1112-1123, 2004.

[17] G. D. Tourassi, E. D. Frederick, M. K. Markey, and C. E. Floyd, "Application of the mutual information criterion for feature selection in computer-aided diagnosis," *Medical Physics*, vol. 28, no. 12, pp. 2394-2402, 2001.

[18] F. Fleuret, "Fast Binary Feature Selection with Conditional Mutual Information," *Journal of Machine Learning Research*, vol. 5, pp. 1531-1555, 2004.

[19] P. J. Phillips, "Support vector machines applied to face recognition," in *Proceedings of the 1998 conference on Advances in neural information processing systems II* MIT press, 1999, pp. 803-809.

[20] B. Scholkopf, S. Mika, C. J. C. Burges, P. Knirsch, K. R. Muller, G. Ratsch, and A. J. Smola, "Input space versus feature space in kernel-based methods," *Ieee Transactions on Neural Networks*, vol. 10, no. 5, pp. 1000-1017, 1999.

[21] M. Yang, "Kernel Eigenfaces vs. Kernel Fisherfaces: Face recognition using Kernel methods," in *Proc. of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition* Washington, D.C.: 2002, pp. 205-211.

[22] G. Baudat and F. E. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural Computation*, vol. 12, no. 10, pp. 2385-2404, 2000.

[23] B. Kepenekci, "Face recognition using Gabor wavelet transform." M.S. thesis, Dept. of electrical and electronic engineering, The Middle Ease Technical University, 2001.

[24] B. Kepenekci, F. B. Tek, and G. B. Akar, "Occluded face recognition based on gabor wavelets," in *Proc. of the IEEE International Conference on Image Processing* 2002, pp. 293-296.