Boosting Local Feature Based Classifiers for Face Recognition

LeiZhang[1], Stan Z. Li[2], ZhiYiQu[1],Xiangsheng Huang[3]
[1]Lanzhou University, Lanzhou ,China
[2]Microsoft Research Asia,China
[3]Institute of Automation ,Chinese Academy of Sciences
leizh02@st.lzu.edu.cn,szli@microsoft.com,quzy@lzu.edu.cn,
xiangshenghuang@hotmail.com

Abstract

In this paper, we present a method for face recognition using boosted Gabor feature based classifiers. Weak classifiers are constructed based on both magnitude and phase features derived from Gabor filters [6]. The multi-class problem is transformed into a two-class one of intra- and extra-class classification using intra-personal and extra-personal difference images, as in [12]. A cascade of strong classifiers are learned using bootstrapped negative examples, similar to the way in face detection framework [17]. The combination of classifiers based on two different types of features produces better results than using either type. Experiments on FERET database show good results comparable to the best one reported in literature [14].

1. Introduction

Face recognition has attracted much attention due to its potential values for applications as well as theoretical challenges. However, after decades of effort, the problem still remains difficult. It has been observed that in face recognition the variations between the images of the same face due to illumination and viewing direction changes are almost always larger than those due to change in face identity [13].

As a typical pattern recognition problem, face recognition has to deal with two main issues: (i) what features to use to represent a face, and (ii) how to classify a new face image based on the chosen representation. Up until now, many representation approaches have been introduced, including eigenface (PCA) [16], FisherFace (LDA) [2], independent component analysis(ICA)[1], and Gabor wavelet features [18].while regarding classification methods, nearest neighbor[16], convolutional neural networks [8], nearest feature line [9], Bayesian classification [11] and AdaBoost method[4]have been widely used. Among various representations, multi-scale and multi-orientation Gabor features have attracted much attention and achieved great success in face recognition. On the other hand, AdaBoost method, introduced by Freund and Schapire [4], which can fulfil both feature selection and object classification at the same time, has achieved great success in face detection [17] and other applications.

Gabor wavelets, whose kernels can satisfy both the wavelet theory and the neurophysiological constraints for simple cells at the same time, exhibit desirable characteristics of spatial locality and orientation selectivity. The biological relevance and computational properties of Gabor wavelets for image analysis have been described in [5]. The Gabor wavelet representation facilitates recognition without correspondence (hence, no need for manual annotations) because it captures the local structure corresponding to spatial frequency (scale), spatial localization, and orientation selectivity [3]. However, Gabor features are overcomplete and high dimensional representation of face images. Straightforward implementation is both computationally expensive and exhibits a lack of efficiency.

AdaBoost method, which is among the top performed classification approaches and adopted in many applications, provides a simple yet effective stagewise learning approach for feature selection and nonlinear classification. Adaboosted cascade framework,which is also widely used in many applications especially in face detection occasions [17], is a divide-and-solve strategy, that can make training and testing process much easier and faster. Moreover, it is an efficient way to treat asymmetric problems and hardly converging training processes. But as a classification method, its performance relies largely on the representation of samples. In general, face images span a nonlinear manifold in the image space. To obtain good performance, it needs a proper nonlinear representation which can be transformed into another space in which different faces (persons) are linearly separable. Up to now, such a feature representation is not available.

Combining different kinds of feature is an efficient way to improve classifier's performance, because using several feature streams is a perennially successful approach to improve performance in speech recognition, as stated in [15].

We argue that some nonlinear non-convex parts of the manifold in one image representation are linear and convex in another feature representation, as long as the two kinds of feature are not totally dependent with each other, that is, they can reveal complementary information of face images. Motivated by this idea, we make full use of Gabor features, to build two different feature streams: Gabor magnitude features and phase features. Later, we will also discuss it in conditional mutual information's point of view to analyze the beneficial of using the two kinds of Gabor features. Since either Gabor magnitude features or phase features are redundant, and face recognition is a multi-class problem. We adopt a method as introduced in [12], to turn multi-class problem into two-class problem by using intra-personal and extra-personal variation in Gabor feature space. Then we adopt AdaBoost Cascade framework to select the most discriminant Gabor features and also build a strong classifier by using one kind of feature at early stages, then the other kind of feature to solve the unsolved problem at later stages.

The remaining part of this paper is organized as follows: In section 2, the two kinds of Gabor feature face representation approach is introduced. In section 3, the cascade boosting learning for two kinds of feature selection and classifier construction are described. And the experiment results using the FERET database and FERET evaluation protocol [14] and analysis are shown in section 4. In section 5, the last section, we present the conclusion and future work.

2. Gabor Features face image representation

The representation of faces using Gabor feature has been extensively and successfully used in face recognition [18]. and Significant improvements in the face recognition rate have also been reported in literature. Gabor features (later called Gaborface) exhibit desirable characteristics of spatial locality and orientation selectively, and are optimally localized in the space and frequency domains. The Gabor kernels can be defined as follows [19]:

$$\Psi_{\mu,\nu} = \frac{k_{\mu,\nu}^2}{\sigma^2} exp(-\frac{k_{\mu,\nu}^2 z^2}{2\sigma^2})[exp(ik_{\mu,\nu}z) - exp(-\frac{\sigma^2}{2})] \quad (1)$$

where μ and ν define the orientation and scale of the Gabor kernels respectively, $z = (x, y)$, and the wave vector $k_{\mu,\nu}$ is defined as follows:

$$k_{\mu,\nu} = k_\nu e^{i\phi_\mu} \quad (2)$$

where $k_\nu = k_{max}/f^\nu, k_{max} = \pi/2, f = \sqrt{2}$, $\phi_\mu = 2\pi\mu/8$. The Gabor kernels in equ.(1) are all self-similar since they can be generated from one filter, the mother wavelet, by scaling and rotating via the wave vector $k_{\mu,\nu}$. Each kernel is a product of a Gaussian envelope and a complex plane wave, while the first term in the square brackets in equ. (1) determines the oscillatory part of the kernel and the second term compensates for the DC value. The effect of the DC term becomes negligible when the parameter σ, which determines the ratio of the Gaussian window width to wavelength, has sufficiently large values. Hence, a bank of Gabor filters is generated by a set of various scales and rotations. More scales or rotations could increase the dependencies of neighbor samples while less scales or rotations could decrease the accuracy.

We use Gabor kernels at five scales $\nu \in \{0, 1, 2, 3, 4\}$ and eight orientations $\mu \in \{0, 1, 2, 3, 4, 5, 6, 7\}$, with the parameter $\sigma = 2\pi$ [7]. The numbers of scales and directions are selected in a way that makes the features extracted suitable to represent the characteristics of spatial locality and orientation selectivity. The Gaborfaces are computed by convoluting face images with corresponding Gabor filters. For every image pixel we have two Gabor parts, real part and imaginary part. We transform the two parts into two kinds of Gabor features, Gabor magnitude features and Gabor phase features.

Gabor Real part:

$$Re(\Psi_{\mu,\nu}) = \frac{k_{\mu,\nu}^2}{\sigma^2} exp(-\frac{k_{\mu,\nu}^2 z^2}{2\sigma^2})[cos(k_{\mu,\nu}z) - exp(-\frac{\sigma^2}{2})] \quad (3)$$

Gabor Imaginary part:

$$Im(\Psi_{\mu,\nu}) = \frac{k_{\mu,\nu}^2}{\sigma^2} exp(-\frac{k_{\mu,\nu}^2 z^2}{2\sigma^2}) sin(k_{\mu,\nu}z) \quad (4)$$

The magnitude features are formed by:

$$\sqrt{Re(\Psi_{\mu,\nu})^2 + Im(\Psi_{\mu,\nu})^2} \quad (5)$$

The phase features are formed by:

$$tan^{-1}(Im(\Psi_{\mu,\nu})/Re(\Psi_{\mu,\nu})) \quad (6)$$

Because we naturally obtain Gabor features in real and imaginary part, to form two kinds of Gabor features doesn't require much extra computation than to just form magnitude features. But in most applications they only use Gabor magnitude features, and discard Gabor phase features. We found that the two kinds of feature can provide complementary information although they are dependent. As stated in [15], conditional mutual information (CMI), which estimates the amount of information that one feature space contains about the other, is a useful tool in deciding different feature spaces combination, and they also suggest that CMI between the outputs of independent classifiers based on each feature spaces should help predict which stream can be combined most beneficially. Moreover, this CMI between the outputs can be estimated by the differences in classifier outputs. The more differences in classifier outputs, the lower the CMI between them. Our experiments show that classifiers based on different kind of Gabor feature, make different error recognition. We hope that we can gain more from combining them.

Then we use the obtained Gaborfaces to form the intra-personal and extra-personal space as introduced in paper [12]: the variation in different images of the same individual to form the intra-personal space, and the variation in different images of different individuals to form the extra-personal space. That is, if you have two images from one person, subtract the two gaborfaces from the two images, you get one sample of the intra-personal space, if the two images are from two different persons, subtract the two gaborfaces form the two images, you get one sample of the extra-personal space, this naturally converts face recognition, a multi-class problem into two-class problem. The Intra-personal image and extra-personal image represented by Gabor magnitude features are illuminated in Figure1.

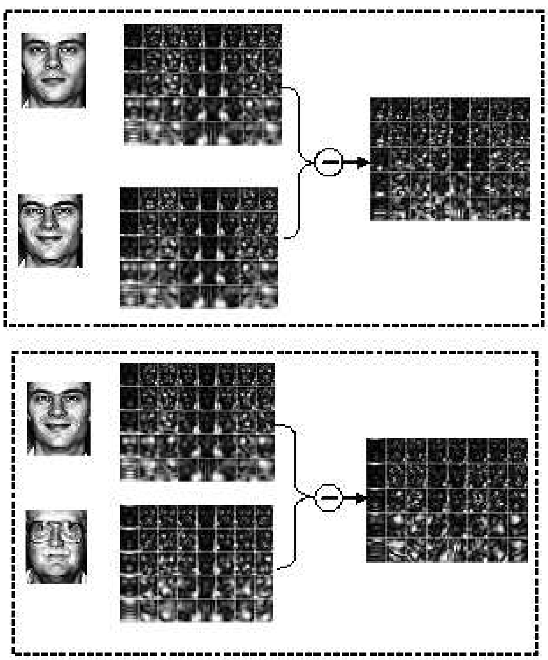

Figure 1: Intra-personal image and extra-personal image represented by Gabor magnitude features, the upper Figure demonstrates Intra-personal image Gabor magnitude representation, the lower one demonstrates extra-personal image Gabor magnitude representation

3. Feature Selection and Classifier Learning

Both Gabor magnitude features and phase features are redundant, subspace techniques have been used to reduce the dimensionality [10]. Here we propose to use Adaboost with cascade structure to select most significant Gabor features from a large Gabor feature set. AdaBoost method, which can do feature selection and build a strong classifier at the same time, provides a simple yet effective stagewise learning approach for feature selection and nonlinear classification. It learns a sequence of easily learnable weak classifiers, each of which needs only slightly better than random guessing, and boosts them into a single strong classifier by a linear combination of them. The weak classifiers, each derived based on some simple, coarse estimates, need not be optimal. Yet, the AdaBoost learning procedure provides an optimal approach to combine them into a strong classifier.

Therefore, AdaBoost is adapted to solve the following three fundamental problems in one boosting procedure: (1) learning effective features from a large feature set, (2) constructing weak classifiers each of which is based on one of the selected features, and (3) boosting the weak classifiers into a stronger classifier.

AdaBoost Algorithm

Input: n training examples$(x_1, y_1), \ldots, (x_n, y_n)$ with $y_i \in \{+1, -1\}$ is the class label for the sample x_i;where $i = 1, \ldots, n$.

Initialize: weights $w_{1i} = 1/2l$ or $1/2m$ for $y_i = 1$ or -1 with $l + m = n$, where $i = 1, \ldots, n$.

Do for t=1, . . . , T:

1. Train one hypothesis h_j for each feature $j$with w_t,and error $e_j = Pr_i^{w_t}\ [h_j(x_i) \neq y_i]$.

2. Choose

$$h_t(x) = h_{k^*}(x) \tag{7}$$

such that $\forall j \neq k$,if $e_k < e_j$.Let $e_t = e_k$.

3. Update: $w_{t+1,i} = w_{t,i}\beta_t^{e_i}$. where $e_i = 1$ or 0 for example x_i classified correctly or incorrectly respectively, and $\beta_t = e_t/(1 - e_t)$.

4.Normalize the weights so that they are a distribution, $w_{t+1,i} \longleftarrow w_{t+1,i}/\sum_{j=1}^{n} w_{t+1,j}$.

Output: the final hypothesis,

$$h_f(x) = \begin{cases} 1 & if \ \sum\limits_{t=1}^{T} \alpha_t h_t(x) > \sum\limits_{t=1}^{T} \alpha_t \\ 0 & otherwise \end{cases} \tag{8}$$

where $\alpha_t = log\frac{1}{\beta_t}$

The AdaBoost learning procedure is aimed to derive α_t and $h_t(x)$.

Our AdaBoost algorithm uses cascade structure, which has been successfully applied in many applications, like face detection.According to this algorithm, the training set of positive samples (intra-personal) are fixed, while new negative samples (extra-personal) are re-sampled in each stage. This algorithm works as following:

1.**Initialize**: S = a random sample set with size n.

2.**For**: $t = 1$ to T

a) Use Adaboost to train a strong classifier h_{ft}, based on S;

b) Generate a new sample set S with size n in which all the negative samples are misclassified by previous classifiers, while positive samples is the same as previous stages;

3.**Output**: $g(h_{f1}, \ldots, h_{fT})$=$h_{f1} * \ldots * h_{fT}$

In many cases, training a single strong classifier without using cascade structure rarely converges, which leads to many undesirable weak classifiers learned and low efficiency. Moreover, cascade structure splits problem into many subproblems, and solves them separately. Making the training process much easier, and also making the testing process much faster. Of course, it also has minor drawbacks: every weak classifier normally does not just help solve one sub-problem. If we use cascade structure, they may learn more features than non-cascade structure does when the given problem is easy, using non-cascade structure could result fast convergence. Despite this, cascade is still a very useful structure, especially in solving hard problems like face recognition.

In our application, we use cascade structure to train a strong classifier. It is a natural way to solve the asymmetric problem, since positive samples are relatively small, and negative sample pool is extremely large– this fits the cascade structure. Moreover, this structure can combine the two kinds of feature easily, naturally, and efficiently, by using one kind of feature at early stages, then the other kind of features to solve the unsolved problem at later stages. Our experiment results exhibit efficiency of the proposed approach.

4. Experiments

We tested the proposed method on the FERET $fafb$ face database, and this training set is also from the training set of the FERET database, which includes 1002 images of 429 subjects. All images are cropped to 42 pixels high by 36 pixels wide and rectified according to the manually located eye positions supplied with the FERET data. All normalized images use histogram normalization for preprocessing. The cropped and preprocessed images are illuminated in Figure 2. The training set yields 795 intra-personal image pairs and $500,706$ extra-personal image pairs. In order to make the obtained classifier more robust, we generated one virtual sample for each person by geometric transforms, including shift, rotation transformations. The shift is done randomly by 1 or by 2 pixels in the four directions along the coordinates. Rotations are done randomly in $\pm 5°$, and these geometric transformations are a random factor which affects a random image from every person's image folder. Then we have 1797 intra-face image pairs. At any given time, all 1797 intra-personal pairs and 4000 extra-personal pairs are used for training. A new set of 4000 extra-personal pairs, which is misclassified by early stages, is generated from the full training set at the beginning of the next layer, until almost all extra-face pairs are correctly classified.

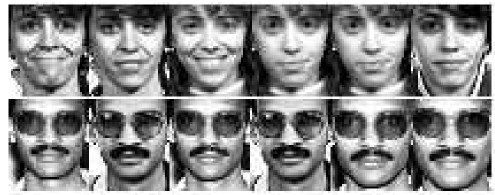

Figure 2: some samples of preprocessed images

To test the efficiency of our proposed method, several comparative experiments were tested on the probe set fb with the gallery fa of the FERET database. There are 1196 images in fa, 1195 images in fb, and all the subjects have exactly one image in both fa and fb. (1) Build several Adaboost cascade classifiers by selecting weak classifiers from Gabor magnitude feature space; build several Adaboost cascade classifiers by selecting weak classifiers from Gabor phase feature space, then compare the outputs of the two kinds of classifiers. We found that: the intersections of the two error sets just 45% of the Unions; the best performance of each kind of classifiers is demonstrated in Figure 3; (2) Test our proposed method on this database, using an Adaboost cascade classifier which has several earlier stages that are formed by selecting weak classifiers from one kind of Gabor feature space, and the remaining stages by selecting weak classifiers from the other Gabor feature space to solve the unconquered samples. The result is also demonstrated in Figure 3. and the Rank$-N$ results of these programs are illustrated in Figure 4.

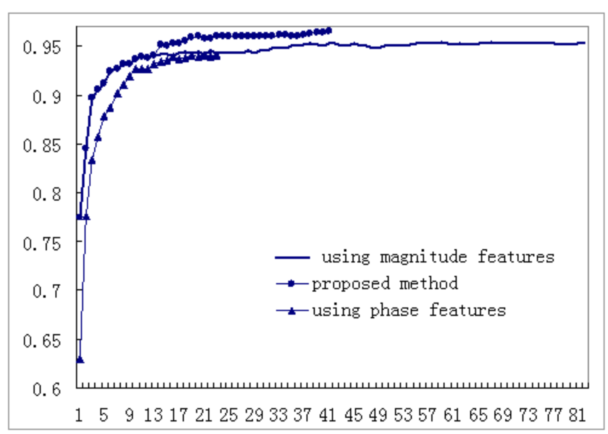

Figure 3: Face recognition performances with respect to the number of stages

Representation and similarity metric for Algorithms Evaluated

Algorithm	Representation	Similarity meature
U.of So.CA	Dynamic Link Architecture(Gabor Jets)	Elastic graph matching
U.of MD 96	Fisher discriminate	L2
U.of MD 97	Fisher discriminate	weighted L2
Baseline	PCA(eigenface)	L1
Baseline	Correlation	Angle

Table 1: Comparable algorithms on FERET ($fafb$)

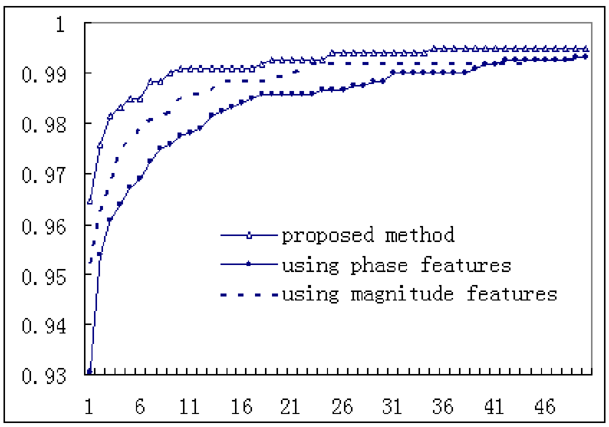

Figure 4: Rank-N recognition rate

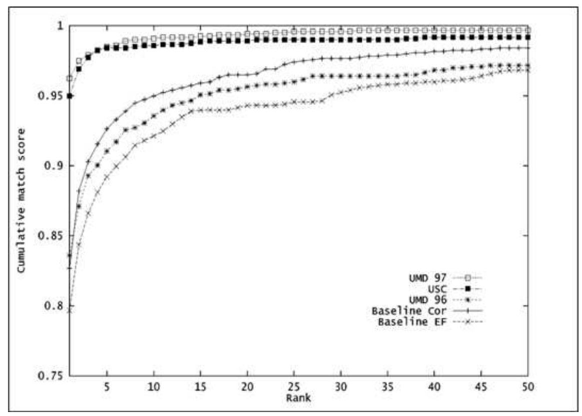

Figure 5: [14]on FERET $fafb$ sets

As Figure 3 shown: the best Rank 1 performance of classifier built on Gabor phase features is 93.56% using 3990 features, and that of classifier built on Gabor magnitude features is 95.4% using 4242 features; While our proposed method just uses 2672 features to achieve a recognition rate of 96.5% and also has the best rank-N performance among the three comparative algorithms which are demonstrated in Figure 4. our approach has also achieved the upper bound recognition performance shown in Figure 5, the algorithms evaluated in this figure are listed in Table1,[14]three of which are named after the universities that developed the procedure.

While most recently reported results only use subsets of FERET $fafb$ sets, not much comparable results are available.

From the outputs of the two kinds of classifiers that are based on different feature sets, we can tell that the two kinds of feature can provide complementary information, thus they have low CMI ,which should favor their combination performance in theory. This is also proven by the combined classifier's efficiency in our experiment.

5. Conclusion and Future Work

In this paper, we applied two kinds of Gabor feature for face image representation with a suggested efficient combination approach in face recognition; and also presented a face recognition system by AdaBoost cascade structure. Experimental results on FERET $(fafb)$ database has proven the effectiveness of our new approach. In addition, for a practical system, using two kinds of Gabor features will not add much computation time, compared to using just one kind of them, but could obviously improve the classifier's performance. While the problem of how to combine them and when to combine them could contribute most to the recognition system is still open, and this will be the focus in our future research.

References

[1] M. S. Bartlett, H. M. Lades, and T. J. Sejnowski. Independent component representations for face recognition. *Proceedings*

of the SPIE, Conference on Human Vision and Electronic Imaging III, 3299:528–539, 1998.

[2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. “Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, July 1997.

[3] B.Schiele and J.L.Crowley. Recognition without correspondence using multidimensional receptive field histograms. *International Journal of Computer Vision*, 36:31–52, 2000.

[4] Y. Freund and R. Schapire. “A decision-theoretic generalization of on-line learning and an application to boosting”. *Journal of Computer and System Sciences*, 55(1):119–139, August 1997.

[5] J.G.Daugman. Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research*, 20:847–856, 1980.

[6] J.G.Daugman. Quadrature-phase simple-cell pairs are appropriately described in complex analytic from. *Journal Opt.Soc.Amer*, 1992.

[7] M. Lades, J. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, W. R.P., and W. Konen. Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers*, 42:300–311, 1993.

[8] S. Lawrence, C. L. Giles, A. Tsoi, and A. Back. “Face recognition: A convolutional neural network approach”. *IEEE Transactions on Neural Networks*, 8(1):98–113, 1997.

[9] S. Z. Li and J. Lu. “Face recognition using the nearest feature line method”. *IEEE Transactions on Neural Networks*, 10(2):439–443, March 1999.

[10] 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(4):467–476, 2002.

[11] B. Moghaddam. Principal manifolds and bayesian subspaces for visual recognition. In *International Conference on Computer Vision (ICCV’99)*, pages 1131–1136, September 1999.

[12] B. Moghaddam and A. Pentland. Beyond euclidean eigenspaces:bayesian matching for visian recognition.

[13] Y. Moses, Y. Adini, and S. Ullman. “Face recognition: The problem of compensating for changes in illumination direction”. In *Proceedings of the European Conference on Computer Vision*, volume A, pages 286–296, 1994.

[14] 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*, 22(10):1090–1104, 2000.

[15] D. P.W.Ellis and J. A.Bilmes. Using mutual information to design feature combinations. *Int. Conference on Spoken Language Processing*, 3:79–82, 2000.

[16] M. A. Turk and A. P. Pentland. “Eigenfaces for recognition”. *Journal of Cognitive Neuroscience*, 3(1):71–86, March 1991.

[17] P. Viola and M. Jones. “Robust real time object detection”. In *IEEE ICCV Workshop on Statistical and Computational Theories of Vision*, Vancouver, Canada, July 13 2001.

[18] L. Wiskott, J. Fellous, N. Kruger, and C. V. malsburg. ”face recognition by elastic bunch graph matching”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):775–779, 1997.

[19] W. S. Yambor. Analysis of pca-based and fisher discriminant-based image recognition algorithms. M.s., Colorado State University, July 2000. Technical Report CS-00-103, Computer Science.