

A Classifier Ensemble for Face Recognition Using Gabor Wavelet Features

Hamid Parvin, Nasser Mozayani, and Akram Beigi

School of Computer Engineering, Iran University of Science and Technology
(IUST), Tehran, Iran

{parvin,mozayani,beigi}@iust.ac.ir

Abstract. Gabor wavelet-based methods have been proven that are useful in many problems including face detection. It has been shown that these features tackle well facing into image recognition. In image identification, while there is a number of human faces in a repository of employees, it is aimed to identify the face of an arrived employee is which one? So the application of gabor wavelet-based features is reasonable. We propose a weighted majority average voting classifier ensemble to handle the problem. We show that the proposed mechanism works well in an employees' repository of our laboratory.

Keywords: Classifier Ensemble, Gabor Wavelet Features, Face Recognition, Image Processing.

1 Introduction

Gabor wavelet-based methods have been successfully employed in many computer-vision problems, such as fingerprint enhancement and texture segmentation [10, 11]. Also similar to the human visual system, Gabor wavelets represent the characteristics of the spatial localities and the orientation selectivity, and are locally optimal in the space and frequency domains [12]. Therefore, Gabor Wavelets are proper the choice for image decomposition and representation when the goal is to derive local and discriminating features [13].

Combinational Classifiers are so versatile in the fields of artificial intelligence. It has been proved that a single classifier is not able to learn all the problems because of three reasons:

1. Problem may inherently be multifunctional.
2. From other side, it is possible that a problem is well-defined for a base classifier which its recognition is very hard problem.
3. And finally, because of the instability of some base classifiers like Artificial Neural Networks, Decision Trees, and Bayesian Classifier and so on, the usage of Combinational Classifiers can be inevitable.

There are several methods to combine a number of classifiers in the field of image processing. Some of the most important are sum/mean and product methods, ordering (like max or min) methods and voting methods. There is a good coverage over their comparisons and evaluations in the [1], [2], [3] and [4]. In [5] and [6] it is shown that

the product method can be considered as the best approach when the classifiers have correlation in their outputs. Also it is proved that in the case of outliers, the rank methods are the best choice [4]. For a more detailed study of combining classifiers, the reader is referred to [7].

Applications of combinational classifiers to improve the performance of classification have had significant interest in image processing recently. Sameer Singh and Maneesha Singh [8] have proposed a new knowledge-based predictive approach based on estimating the Mahalanobis distance between the test sample and the corresponding probability distribution function from training data that selectively triggers classifiers. They also have shown the superior performance of their method over the traditional challenging methods empirically.

This paper aims at producing an ensemble-based classification of face recognition by use of gabor features with different frequencies. The face images are first gave to the gabor feature extractor with different frequencies, then the features of all trainset union with the test data are compared with each other in each frequency. This results in a similarity vector per each frequency. The similarity vectors are finally combined to vote to which image the test data is belonged.

2 Voting Classifier Ensemble

Classifier ensemble works well in classification because different classifiers with the different characteristics and methodologies can complement each other and cover their internal weaknesses. If a number of different classifiers votes as an ensemble, the overall error rate will decrease significantly rather using each of them individually.

One of the oldest and the most common policy in classifier ensembles is majority voting. In this approach as it is obvious, each classifier of the ensemble is tested for an input instance and the output of each classifier is considered as its vote. The class is the winner which the most of the classifiers vote for it. The correct class is the one most which is often chosen by different classifiers. If all the classifiers indicate different classes, then the one with the highest overall outputs is selected to be the correct class.

Let us assume that E is the ensemble of n classifiers $\{e_1, e_2, e_3 \dots e_n\}$. Also assume that there are m classes in the case. Next, assume applying the ensemble over data sample d results in a binary D matrix like equation 1.

$$D = \begin{bmatrix} d_{1\ 1} & d_{1\ 2} & \cdot & d_{1\ n} \\ \cdot & \cdot & \cdot & \cdot \\ d_{m-1\ 1} & d_{m-1\ 2} & \cdot & d_{m-1\ n} \\ d_{m\ 1} & d_{m\ 2} & \cdot & d_{m\ n} \end{bmatrix} \tag{1}$$

where $d_{i,j}$ is equal to one if the classifier j votes that data sample belongs to class i . Otherwise it is equal to zero. Now the ensemble decides the data sample to belong class b according to equation 2.

$$b = \arg \max_i \left| \sum_{j=1}^n d_{ij} \right| \quad (2)$$

Another method to combine a number of classifiers which employs d_{ij} as confidence of classifier j for belonging the test data sample to class i is called majority average voting. The majority average voting uses equation 2 as majority voting.

Weighted majority vote is another approach of voting; in this method members' votes have different worths. Unlike the previous versions of voting this is not like democracy. For example if a classifier has 99% recognition ratio, it is more worthy to use its vote with a more effect than the vote of another classifier with 80% accuracy rate. Therefore in weighted majority vote approach, every vote is multiplied by its worth. Kuncheva [7] has shown that this worth can optimally be a function of accuracy.

To sum up assume that the classifiers existing in the ensemble E have accuracies $\{p_1, p_2, p_3 \dots p_n\}$ respectively. According to Kuncheva [7] the worth of them are $\{w_1, w_2, w_3 \dots w_n\}$ respectively where

$$w_i = \log \frac{p_i}{1-p_i} \quad (3)$$

Weighted majority vote mechanism decides the data sample to belong class b according to equation 2.

$$b = \arg \max_i \left| \sum_{j=1}^n w_j * d_{ij} \right| \quad (4)$$

Similarly another method of combining which again employs d_{ij} as confidence of classifier j for belonging the test data sample to class i is called weighted majority average voting. Weighted majority average voting uses equation 4 as weighted majority voting.

3 Feature Extraction

Gabor filter can capture salient visual properties such as the spatial localization, the orientation selectivity, and the spatial frequency characteristics. The Gabor responses describe a small patch of gray values in an image $I(x)$ around a given pixel $x=(x,y)^T$. It is based on a wavelet transformation, given by the equation 5.

$$R_i(x) = \int I(x') \psi_i(x-x') dx' \quad (5)$$

which $\psi_i(x)$ is a convolution of image with a family of Gabor kernels like equation 6.

$$\psi_i(x) = \frac{\|k_i\|^2}{\sigma^2} e^{-\frac{\|k_i\|^2 \|x\|^2}{2\sigma^2}} \left[e^{jk_i x} - e^{-\frac{\sigma^2}{2}} \right] \quad (6)$$

where

$$k_i = \begin{pmatrix} k_{ix} \\ k_{iy} \end{pmatrix} = \begin{pmatrix} k_v \cos \theta_\mu \\ k_v \sin \theta_\mu \end{pmatrix} \quad (7)$$

Each $\Psi_i(x)$ is a plane wave characterized by the vector k_i enveloped by a Gaussian function, where σ is the standard deviation of this Gaussian. The center frequency of f^{th} filter is given by the characteristic wave vector k_i having a scale and orientation given by (k_v, θ_μ) . Convolving the input image with a number of complex Gabor filters with 5 spatial frequencies ($\nu = 0, \dots, 4$) and 8 orientations ($\mu = 0, \dots, 7$) will capture the whole frequency spectrums, both amplitude and phase as illustrated in [9].

According to equation 5, each image I^q of face train dataset is mapped to 40 images $I^q_{v,\mu}$, where $\nu \in \{0, \dots, 4\}$ and $\mu \in \{0, \dots, 7\}$. Test Image H is also mapped to $H'_{v,\mu}$. Now orientation matching between each train image and the test image is gained using equation 8.

$$r_{q,H,f} = n_{q,H,f} - m_{q,H,f} \quad (8)$$

where $m_{q,f}$ and $n_{q,f}$ are extracted from equation 9.

$$[m_{q,H,f}, n_{q,H,f}] = \arg \max_{m,n} \left(\text{mean}_{x \in \text{Center}} \left(\left| I^q_{f,m}(x) - H'_{f,n}(x) \right| \right) \right) \quad (9)$$

where *Center* is a 9×9 square in the middle of the image, e.g. for image with size 80×40 , it is $\{36, \dots, 44\} \times \{16, \dots, 24\}$. Now the orientation matched image denoted by $OMI^q_{v,\mu'}$ is defined as equation 10.

$$OMI^q_{v,H,\mu'} = I^q_{v,\mu} \quad (10)$$

where

$$\mu' = (\mu + r_{q,H,f}) \bmod 8 \quad (11)$$

Now we define the similarity vector \mathbf{sim}^f whose i th element indicates the similarity between i th train image and the test image, according to equation 12.

$$\mathbf{sim}_i^f(H) = \frac{1}{8} \sum_{m=0}^7 \text{mean}_{x \in C} \left(\left| OMI^q_{f,H,m}(x) - H'_{f,m}(x) \right| \right) \quad (12)$$

4 Employed Classification

Let assume that there exist n training images and one test image. Also assume that the training images are indexed as number one to n respectively and the test image indexed as number $n+1$. The goal is to understand to which training image the test

image is similar. The Gabor Wavelet features of r_l frequency and eight directions are first extracted from images number one to $n+l$. Then the similarities between each of train images and test image are evaluated according to equation 12, as discussed in the previous section. It is obvious that in order to become these similarities comparable they must be normalized in such a way that the sum of similarities vector of test image becomes unit. So they are normalized in range $[0,1]$. After calculating each of these similarities between each two training and test images, a similarity vector named sim^{r_l} which is a vector with n elements, is obtained. It is important to note that the $\mathit{sim}^{r_l}_i$ means the similarity between images number i and test image.

As the reader can guess, the problem mentioned here, is an n class problem. sim^{r_l} can be also served as a simple classifier C_{r_l} which uses image number 1 to n as its train dataset. So as to the index of the maximum value in the vector can be considered as class label of test image.

Considering sim^{r_l} , $r_l \in \{0, \dots, 4\}$ there are five classifiers to classify the test image. Now the majority-votes ensemble is employed to classify the test image. Assume that the accuracy of classifier C_{r_l} is denoted by p_{r_l} , the weight vector \mathbf{w} can straightforwardly be calculated in the weighted-majority-votes ensemble.

5 Parameters of Classification

In the experiments, there exist 2×300 training images. Here there are 300 real classes, 2 images per each class denoted by TI_i and VI_i where $i \in \{1, \dots, 300\}$. Indeed one image of class i is denoted by TI_i and the other by VI_i . 300 fixed images i.e. TI_i , are selected as training dataset. Running the algorithm 299 times, each time one of VI_i is considered as test image and the other 299 images as validation dataset. In z th running of algorithm image VI_z is selected as test image and images VI_j where $j \in \{1, \dots, 300\} - \{VI_z\}$ are considered as validation dataset. Now we obtain 5 classifiers C_{r_l} , $r_l \in \{0, \dots, 4\}$ based on sim^{r_l} . To calculate the accuracies of C_{r_l} the mentioned validation dataset is used as following. The similarities between each pairs of images denoted by TI_i and VI_j , where $i \in \{1, \dots, 300\}$ and $j \in \{1, \dots, 300\} - \{VI_z\}$ are evaluated employing equation 13.

$$\mathit{SIMILARITY}_{i,j}^{r_l} = \mathit{sim}_i^{r_l}(I^j) \quad (13)$$

It is obvious that these similarities must also again be normalized in order to become them comparable. So they are again normalized in range $[0,1]$ as mentioned before. After calculating each of these similarities between each two of training datasets, a similarity matrix named $\mathit{SIMILARITY}^{r_l}$ which is an $n \times n$ matrix, is obtained. It is important to note that the $\mathit{SIMILARITY}^{r_l}_{i,j}$ means the similarity between image number i of training dataset and image number j of validation dataset and the VI_z^{th} column of that matrix is invalid.

Now the accuracy of classifier C_{r_l} , on the training data, is the number of training data that correctly assigned to its correct class, divided to n . In other words, the number of the columns which its maximum value is over matrix diagonal, divided to n can be considered as the accuracy of this classifier as stated in equation 14. Although it is obvious that diagonal elements of this matrix must be the largest in their columns, it is not true in many cases.

$$p_{r1} = \frac{1}{299} \sum_{j \in \{1, \dots, 300\} - \{V_{i_2}\}} \text{isequal}(\arg \max_i (\text{SIMILARITY}_{i,j}^{r1}), j) \quad (14)$$

where $\text{isequal}(x,y)$ is defined as equation 15.

$$\text{isequal}(x, y) = \begin{cases} 1 & x == y \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

6 Experimental Study

Experimental results are reported over 300 pairs of images. Each pair of images belongs to an employee (personnel) of our laboratory. All the images have the same resolution. All of them are first equalized using equalizing their histograms.

Live-one-out technique is used to test ensemble classifier over these images. Also features of 5 different frequencies and 8 orientations are extracted. So, there are forty similarity matrices. 599 images, except in weighted majority voting, are used as training set because there is no longer need to validation set. It is worthy to mention that the best classifier using only one of the similarity matrix, has just 76.63% recognition ratio. While recognition ratio of classifier mentioned above has 90.17% recognition ratio with majority voting, by use of the average voting as final results the 89.32% recognition ratio is achieved. But the combinational proposed approach has 92.67% recognition ratio. The Table 1 summarizes the results.

Table 1. Face recognition ratios of different methods

Best C^f	MV(C^f)	MAV(C^f)	WMAV
76.63	89.32	90.17	92.67

7 Conclusions

In this paper, new face identification algorithm is proposed. In the proposed algorithm first gabor wavelet features with different frequencies are extracted maximizing over different orientations. Defining one classifier per each frequency an ensemble is obtained. The ensemble uses weighted majority average voting as the consensus function. It is shown that the proposed mechanism works well in an employees' repository of laboratory.

References

1. Ross, A., Jain, A.K.: Information fusion in biometrics. *Pattern Recognition Letters* 24(13), 2115–2125 (2003)
2. Pekalska, E., Duin, R., Skurichina, M.: A discussion on the classifier projection space for classifier combining. In: Roli, F., Kittler, J. (eds.) *MCS 2002. LNCS*, vol. 2364, pp. 137–148. Springer, Heidelberg (2002)

3. Kittler, J., Li, Y.P., Matas, J., Sánchez, M.U.R.: Combining evidence in multimodal personal identity recognition systems. *Audio- and Video-Based Biometric Authentication* (1997)
4. Kittler, J., Matas, G., Jonsson, K., Sánchez, M.: Combining evidence in personal identity verification systems. *Pattern Recognition Letters* 18(9), 845–852 (1997)
5. Bilmes, J., Kirchhoff, K.: Directed graphical models of classifier combination: Application to phone recognition. *Spoken Language Processing* (2000)
6. Tax, D.M.J., Breukelen, M.V., Duin, R.P.W., Kittler, J.: Combining multiple classifiers by averaging or by multiplying. *Pattern Recognition* 33, 1475–1485 (2000)
7. Kuncheva, L.I.: *Combining Pattern Classifiers, Methods and Algorithms*. Wiley, New York (2005)
8. Singh, S., Singh, M.: A dynamic classifier selection and combination approach to image region labeling. *Signal Processing: Image Communication* 20, 219–231 (2005)
9. Štruc, V., Pavešić, N.: Gabor-Based Kernel Partial-Least-Squares Discrimination Features for Face Recognition. *INFORMATICA* 20(1), 115–138 (2009)
10. Hong, L., Jain, A., Pankanti, S., Bolle, R.: Fingerprint enhancement. In: *IEEE WACV*, Sarasota, pp. 202–207 (1996)
11. Bashar, M.K., Matsumoto, T., Ohnishi, N.: Wavelet transform-based locally orderless images for texture segmentation. *Pattern Recognition Letters* 24(15), 2633–2650 (2003)
12. Žibert, J., Mihelič, F.: Zhao–Atlas–Marks representation of speech signals. *Electrotechnical Review* 69(3-4), 159–164 (2002)
13. Liu, C.: Gabor-based kernel PCA with fractional power polynomial models for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26(5), 572–581 (2004)