

# Weighted Voting Classifier Ensemble for Image Processing



Zahra Rezaei, Mohsen Parvin, Sajad Parvin, Alinejad-Rokny Hamid  
Science and Technology Branch  
Islamic Azad University  
Iran  
{rezaei, parvin}@iust.ac.ir, {s.parvin, alinejad}@gmail.com

**ABSTRACT:** Feature extraction for object identification in image processing systems has gained value and significance in the last few years. Realizing its value we have postulated the feature of images as the Gabor response and we deployed a defined processing. In this system we used points for noting the images. In this process we generated matrixes to aggregate them using ensemble mechanisms. Then we used the weighted approach to measure the ensemble which proved to produce acceptable results in our testing experiments.

**Keywords:** Classifier Ensemble, Feature extraction, Face identification, Image classification

**Received:** 31 July 2011, Revised 14 September 2011, Accepted 19 September 2011

© 2012 DLINE. All rights reserved

## 1. Introduction

Feature extraction for object representation performs an important role in automatic object detection systems. Previous methods have used many representations for object feature extraction, such as raw pixel intensities [16], [17] and [21], rectangle features [18], [19] and [20], and local binary pattern [22]. Gabor-wavelet based feature extraction 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-wavelet features represent the characteristics of the spatial localities and the orientation selectivity, and are locally optimal in the space and frequency domains [12]. Therefore, Gabor-wavelet features are the proper choice for image decomposition and representation when the goal is to derive local and discriminating features [13].

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 around a given pixel. It is obtained based on a wavelet transformation. To obtain a Gabor response from a typical image 3 inputs must be chosen: (a) the pixel around that the Gabor response is to be extracted denoted by  $x$ , (b) spatial frequency value denoted by  $k$  and (c) orientation frequency value denoted by  $q$ . We can call each Gabor response to a tuple  $(x, k, \theta)$  in a typical image a Gabor wavelet-feature.

Usage of recognition systems has found many applications in almost all fields. However, Most of classification algorithms have obtained good performance for specific problems; they have not enough robustness and generality for other problems. Ensemble

of multiple classifiers can be considered as a general solution method for pattern recognition problems. It has been shown that combination of classifiers can usually operate better than a single classifier provided that its components are independent or they have diverse outputs. It is shown that the necessary diversity of an ensemble can be achieved by manipulating of dataset features. Parvin et al. have proposed some methods of creating the necessary diversity for an ensemble success [14] and [15].

As it is said 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.

Applications of combinational classifiers to improve the performance of classification have had significant interest in image processing recently. Singh and 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.

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].

This paper aims at producing an ensemble-based classification of face recognition by use of Gabor-wavelet features with different orientation and spatial frequencies. The face images are first gave to the Gabor feature extractor with different orientation and spatial frequencies, and then the features of all trainset union with the test data are compared with each other in each orientation frequency. This results in a similarity matrix per each orientation frequency. The similarity matrices are finally combined to vote to which training image the test data belongs.

## 2. Weighted Voting Classifier Ensemble

An ensemble learns classification better than a single classifier because different single classifiers with the different characteristics and methodologies can complement each other and cover their internal weaknesses. If a number of different classifiers vote 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.

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

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.

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

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 method uses equation 4 as weighted majority voting.

### 3. Overview of Proposed Method

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 around a given pixel. It is obtained based on a wavelet transformation. To obtain a Gabor response form a typical image 3 inputs must be chosen: (a) the pixel around that the Gabor response is to be extracted denoted by  $x$ , (b) spatial frequency value denoted by  $k$  and (c) orientation frequency value denoted by  $\theta$ . We can call each Gabor response to a tuple  $(x, k, \theta)$  in a typical image a Gabor wavelet-feature.

It has been proven that Gabor wavelet-feature based recognition methods are useful in many problems including face detection. It has been shown that these features can tackle the image recognition problem well. 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 due to the ability of Gabor-wavelet feature in well-encoding and well-representing the characteristics of an image the application of Gabor-wavelet based features in the case of face identification is reasonable. Each image is an employee face in our benchmark. So we can use a specific spatial localization for all images without lacking generality of the problem. So a pre-processing phase is necessary to get rid of the high possible domain for the pixel  $x$  around that the Gabor response is to be extracted. As it is presented in Figure 1, we first cut the images so as to all marginal non-facial pixel be removed in pre-processing phase. Then we rescale all modified images in a fixed size. After that we can select the middle point of all rescaled images as the pixel  $x$  around that the Gabor response is to be extracted.

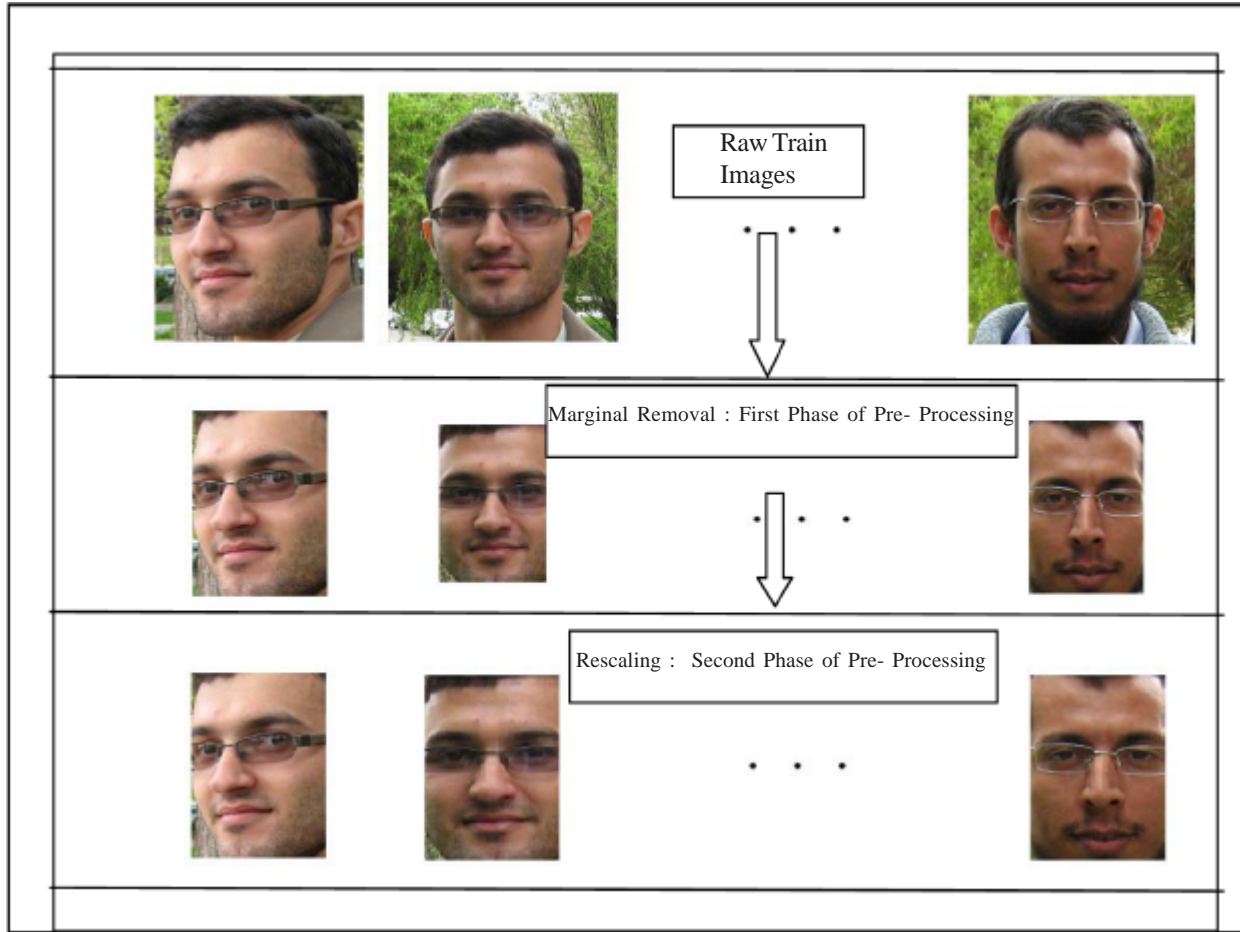


Figure 1. Pre-Processing Phase

#### 4. 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.

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

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  $\sigma$  is the standard deviation of this Gaussian. The center frequency of  $i^{th}$  filter is given by the characteristic wave vector  $k_i$  having a scale and orientation given by  $(k_v, \theta_\mu)$ . Convolution of the input image with a number of complex Gabor filters with 5 spatial frequencies ( $v = 0, \dots, 4$ ) and 8 orientations ( $\mu = 0, \dots, 7$ ) will capture the whole frequency spectrum, both amplitude and phase as illustrated in [9].

According to equation 5, each image  $I^q$  of face train dataset is mapped to 40 features  $I^q_{v,\mu}$ , where  $v \in \{0, \dots, 4\}$  and  $\mu \in \{0, \dots, 7\}$ . Test image  $H$  is also mapped to  $H_{v,\mu}$ . Now we define the similarity vector  $sim^f$  whose  $i$ th element indicates the similarity between  $i$ th train image and the test image,  $H$ . Similarity between the train images  $I^q$  and the test image  $H$  in a fixed orientation frequency  $f$  is defined according to equation 8.

$$sim_q^f(H) = \frac{1}{5} \sum_{v=0}^4 \underset{x \in C}{mean}(|I_{v,f}^q(x) - H_{v,f}(x)|) \quad (8)$$

where  $C$  is a  $9 \times 9$  square in the middle of the image, e.g. for image with size  $80 \times 40$ , it is  $\{36, \dots, 44\} \times \{16, \dots, 24\}$ . Indeed we try to compute for a very few points in the middle of the images rather than all of them. So we have really 3240 ( $9 \times 9 \times 8 \times 5$ ) features.

### 5. Employed Classification

Let assume that there exists  $n$  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  $r_l$  orientation frequency and five orientation frequency are first extracted from images number one to  $n+1$ . Then the similarities between each of train images and test image are evaluated according to equation 8, 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 the similarity 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  $sim^{r_l}_i$  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 images number 1 to  $n$  as its train dataset. It acts very similar to 1-NN classifier where it assigns the index of the maximum value in the vector to class label of test image.

Considering  $sim^{r_l}$ ,  $r_l \in \{0, \dots, 7\}$  there are eight 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  $w$  can straightforwardly be calculated in the weighted-majority-votes ensemble.

### 6. Parameters of Classification

Parameter  $k_v$  is set to one of the values  $\{0.2, 0.4, 0.6, 0.8, 1\}$ . In the experiments, there exist  $2 \times 300$  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 300 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 8 classifiers  $C_{r_l}$ ,  $r_l \in \{0, \dots, 7\}$  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 9.

$$SIMILARITY_{i,j}^{r_l} = sim_i^{r_l}(I^j) \quad (9)$$

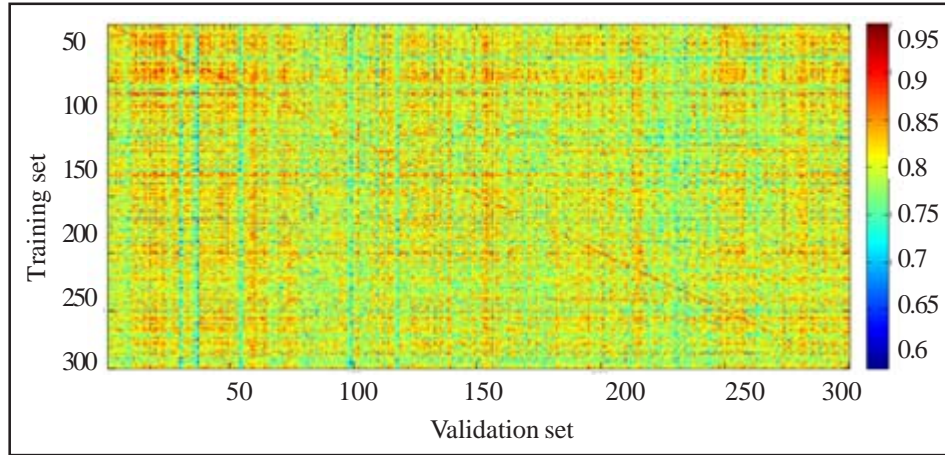


Figure 2. Similarity matrix in frequency 5

To show the effectiveness of the similarity matrix, the Figure 2 shows the matrix SIMILARITY<sup>5</sup>. In the Figure 2, all validation set are considered in the matrix.

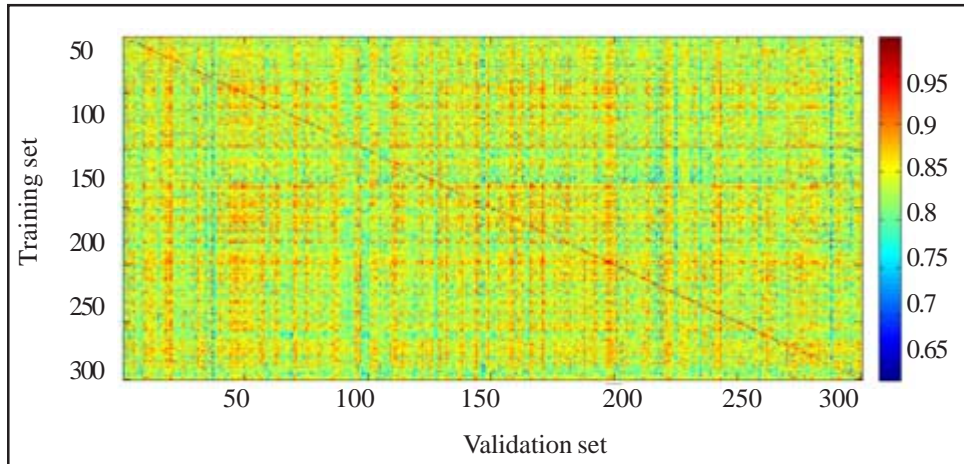


Figure 3. Similarity matrix in frequency 10

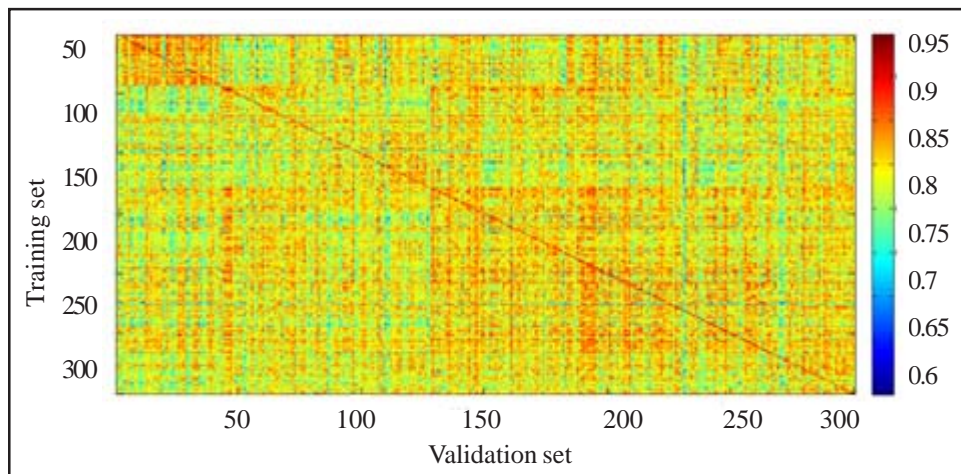


Figure 4. Similarity matrix in frequency 20

Figure 3, Figure 4, Figure 5, Figure 6, Figure 7 and Figure 8 also show the matrix  $SIMILARITY^{10}$ ,  $SIMILARITY^{15}$ ,  $SIMILARITY^{20}$ ,  $SIMILARITY^{27}$ ,  $SIMILARITY^{36}$  and  $SIMILARITY^{50}$  respectively.

As is obvious the best frequency is 27. As we increase the frequency, after the 27 the quality of classification decreases.

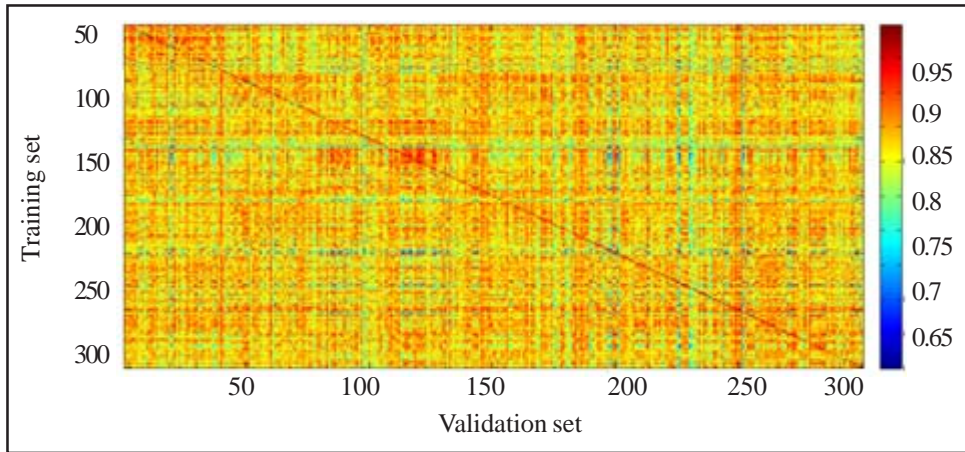


Figure 5. Similarity matrix in frequency 27

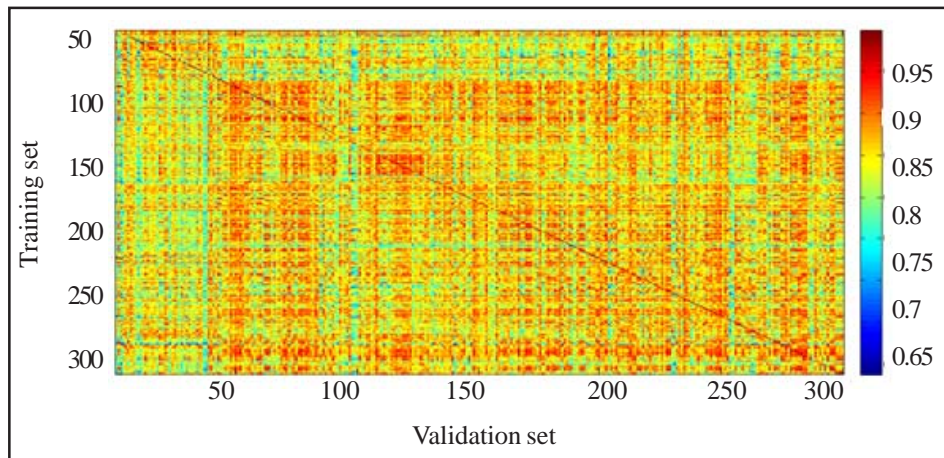


Figure 6. Similarity matrix in frequency 36

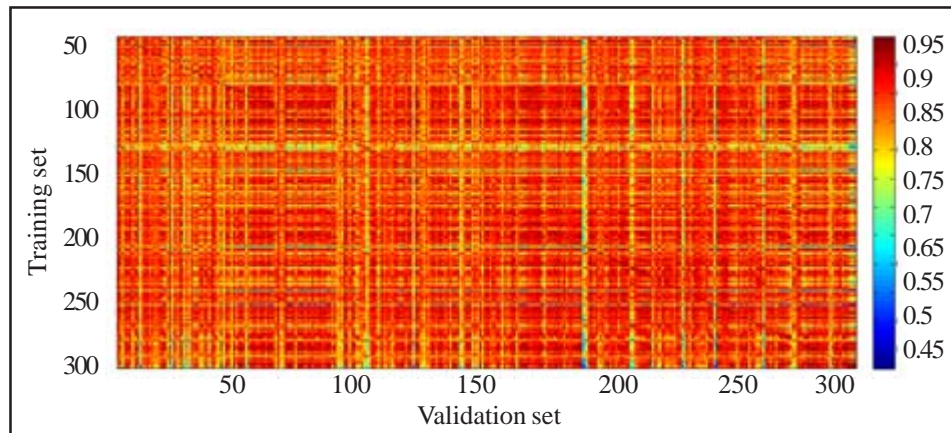


Figure 7. Similarity matrix in frequency 50

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  $SIMILARITY^{r1}$  which is an  $n \times n$  matrix, is obtained. It is important to note that the  $SIMILARITY^{r1}_{i,j}$  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  $C_{r1}$ , 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 10. Although it is obvious that diagonal elements of this matrix must be the largest in their columns, it is not true in many cases.

$$P_{r1} = \frac{1}{299} \sum_{j \in \{1, \dots, 300\} - \{VI_z\}} isequal(\arg \max_i (SIMILARITY^{r1}_{i,j}), j) \tag{10}$$

where  $isequal(x,y)$  is defined as equation 11.

$$isequal(x, y) = \begin{cases} 1 & x == y \\ 0 & otherwise \end{cases} \tag{11}$$

## 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 scales 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.

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

Table 1. Face recognition ratios of different methods

## 7. Conclusion and Discussion

In this paper, new face identification algorithm is proposed. We first extract a large number of different features from each employee face image. Each feature is a Gabor response of image with a different tuple  $(x, k, \theta)$ . We use a pre-processing whereby we can use a fixed point  $x$  for all images without missing of the generality. Eight orientation frequency values are selected for  $\theta$  parameter. Five spatial frequency values are also selected for domain of  $k$  parameter. So we reach a  $k \times \theta$  Gabor-wavelet based feature space. We compute a similarity matrix per different values  $\theta$  parameter. By considering each of these matrices as a classifier we finally use an ensemble mechanism to aggregate them into final classification.

To validate the employed face identification algorithm we use live-one-out technique. We turn to a weighted majority average voting classifier ensemble to handle the problem. It is shown that the proposed mechanism works well in an employees' repository of laboratory containing 600 face images from 300 different individuals.



## References

- [1] Ross, A., Jain, A. K. (2005). Information fusion in biometrics. *Pattern Recognition Letters*, 24 (13) 2115-2125.
- [2] Pekalska, E., Duin, R., Skurichina, M. (2002). A discussion on the classifier projection space for classifier combining, *Multiple Classifier Systems*, p. 137–148.
- [3] Kittler, J., Li, Y. P, Matas, J., Sanchez, M. U. R. (1997). Combining evidence in multimodal personal identity recognition systems, *Audio- and Video-Based Biometric Authentication*.
- [4] Kittler, J., Matas, G., Jonsson, K. Sanchez, M. (1997). Combining evidence in personal identity verification systems, *Pattern Recognition Letters*, 18 (9) 845–852.
- [5] Bilmes, J., Kirchhoff, K. (2000). Directed graphical models of classifier combination: Application to phone recognition, *Spoken Language Processing*.
- [6] Tax, D. M. J., Breukelen, M. V. R. P. W., Duin, Kittler, J. (2000). Combining multiple classifiers by averaging or by multiplying, *Pattern Recognition*, 33, 1475 – 1485.
- [7] Kuncheva, L. I. (2005). *Combining Pattern Classifiers, Methods and Algorithms*, New York: Wiley.
- [8] Singh, S., Singh, M. (2005). A dynamic classifier selection and combination approach to image region labeling, *Signal Processing: Image Communication*, 20, 219 – 231.
- [9] Štruc, V., Pavesi, C. N. (2009). Gabor-Based Kernel Partial-Least-Squares Discrimination Features for Face Recognition, *INFORMATICA*, 20 (1)115–138.
- [10] Hong, L., Jain, A., Pankanti, S., Bolle, R. (1996). Fingerprint enhancement, *IEEE WACV*, p. 202–207.
- [11] Bashar, M. K., Matsumoto, T., Ohnishi, N. (2003). Wavelet transform-based locally orderless images for texture segmentation”, *Pattern Recognition Letters*, 24 (15) 2633–2650.
- [12] Zibert, J., Mihelic, F. (2002). Zhao–Atlas–Marks representation of speech signals, *Electrotechnical Review*, 69(3) 159–164.
- [13] Liu, C. (2004). Gabor-based kernel PCA with fractional power polynomial models for face recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26 (5) 572–581.
- [14] Parvin, H., Alizadeh, H., Minaei-Bidgoli, B. (2009). A New Method for Constructing Classifier Ensembles, *JDCTA* 3(2) 62-66.
- [15] Parvin, H., Alizadeh, H., Minaei-Bidgoli, B. (2009). Using Clustering for Generating Diversity in Classifier Ensemble, *JDCTA*, 3 (1) 51-57.
- [16] Rowley, H. A., Baluja, S., Kanade, T. (1998). Neural network-based face detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, V. 20(1), p. 29–38.
- [17] Garcia, C., Delakis, M. (2004). Convolutional face finder: a neural architecture for fast and robust face detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(11) 1408–1423.
- [18] Viola, P., Jones, M. (2001). Rapid object detection using a boosted cascade of simple features, *IEEE Conference on Computer Vision and Pattern Recognition*, p. 511–518.
- [19] Li, S. Z., Zhu, L., Zhang, Z. Q., Blake, A., Zhang, H. J., Shum, H. (2002). Statistical learning of multi-view face detection, *European Conference on Computer Vision*, 4, 67–81 .
- [20] Chen, X. R., Yuille, A. (2004). Detecting and reading text in natural scenes, *IEEE International Conference on Computer Vision and Pattern Recognition*, p. 366–373.
- [21] Kim, K. I., Jung, K., Kim, J. H. (2003). Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12) 1631–1639.
- [22] Hadid, A., Pietikainen, M., Ahonen. (2004). A discriminative feature space for detecting and recognizing faces, *Conference on Computer Vision and Pattern Recognition*, p. 797–804.
- [23] Parvin, H., Mozayani, N., Beigi, A. (2011). A Classifier Ensemble for Face Recognition Using Gabor Wavelet Features, *International Conference of Computational Intelligence in Security for Information Systems*, p. 301-307.