FACE RECOGNITION USING BAGGING KNN

Dr. Hossein Ebrahimpour[1], Dr. Abbas Kouzani[2]
[1] The University of Kashan, Faculty of Engineering, Kashan, Iran
[2] Deakin University, Geelong, Victoria 3217, Australia

Abstract. In this paper a novel ensemble based techniques for face recognition is presented. In ensemble learning a group of methods are employed and their results are combined to form the final results of the system. Gaining the higher accuracy rate is the main advantage of this system. Two of the most successful wrapping classification methods are bagging and boosting. In this paper we used the K nearest neighbors (kNN) as the main classification technique and Bagging as the wrapping classification methods. The results of these setting for the ORL face database are reported.

Keywords: Face recognition, Ensemble based learning, k nearest neighbors, bagging; boosting; Adaboost.

1 INTRODUCTION

Face recognition is one of the most challenging tasks in the field of pattern recognition. Besides of the many variations in the images, the training data set in the face recognition system is only a very small portion of the possible cases. This fact makes the face recognition a very difficult task for most of the current classifiers. In this paper we present a novel technique of using an ensemble based classifier method for face recognition. In Section 2 a brief introduction to face recognition will be given, section 3 explain the structure of the classier and discusses the parameters that will effect on the performance of the classier. Section 4 introduces the ensemble technique used in this paper. The experimental results are given in the section 5 and the summary and references are given in the section 6 and 7.

2 FACE RECOGNITION

In this context, face recognition refers to the automatic method of identification of an individual based on the information contained in a gray scale image. Many techniques have been developed to solve this problem, such as Principal Component Analysis (PCA) [1], Dynamic Link Matching [2] and Face Recognition using Neural Networks [3].

Most of the practical face recognition systems need a face detection stage to detect the location of the face within a source image. Face recognition systems also normalize the size and orientation of the face to achieve more robustness.

The normalization methods use the location of the significant facial feature such as eyes, nose or mouth. The importance of robust facial feature detection for both detection and recognition has resulted in the development of a variety of different facial feature detection algorithms [4], [5], [6], [7], [8], [9].

The geometric feature based approaches [10], [11], [12], [13] are the earliest approaches to face recognition and detection. These approaches were focused on detecting individual features such as eyes, ears, head outline and mouth, and measuring different properties such as eyebrow thickness and their vertical position or nose position and width, in a feature vector that is used to represent a face. To recognize a face, first feature vectors of the test image and the images in the database are obtained. Second, a similarity measure between these vectors, most often a minimum distance criterion, is used to determine the identity of the face.

Principal component analysis (PCA) [14], is a simple statistical dimensionality reducing technique that has perhaps become the most popular and widely used method for representation and recognition of human faces. PCA, via the KL transform [15] can extract most statistically significant information for a set of images as a set of eigenfaces when which can be used both to recognize and reconstruct face images.

Recently some ensemble based face recognition systems are introduced. In 2003 Martinez and Fuentes [16] developed a system based on homogeneous ensembles with manipulation of input features called face recognition using unlabeled data. Experimental results on the UMIST Face database show that using unlabeled data improves accuracy when a small set was appended to the training set. Three different learning algorithms: k-nearest neighbor, artificial neural networks and locally weighted linear regression were used. They reported an accuracy of 92.07% as their best results using locally weighted linear regression.

Another ensemble based face recognition system, called ensemble linear discriminant analysis (EnLDA), is proposed by Kong et al [17] in 2006. In their system a Boosting LDA and a Random Sub-feature LDA schemes are incorporated together to construct the total weak-LDA classifier ensemble. The performances of this system over two face databases (ORL and Yale face databases) are reported in the form of two graphs.

In this paper we proposed a novel ensemble based face recognition system which is based on K nearest neighbour classifier and bagging. The performance of 97.5 % accuracy is achieved over the ORL face database.

3 K NEAREST NEIGHBOR CLASSIFIER (KNN)

The K nearest neighbor (kNN) classifier is an extension of the simple nearest neighbor (NN) classifier system. The nearest neighbor classifier works based on a simple nonparametric decision. Each query image I_q is examined based on the distance of its features from the features of other images in the training database. The nearest neighbor is the image which has the minimum distance from the query image in the feature space. The distance between two features can be measured based on one of the distance functions such as, city block distance d_1, Euclidean distance d_2 or cosine distance d_{cos}:

$$d_1(x, y) = \sum_{i=1}^{N} | x_i - y_i | \tag{1}$$

$$d_2(x, y) = \sqrt{\sum_{i=1}^{N} | x_i - y_i |} \tag{2}$$

$$d_{cos}(x, y) = 1 - \frac{\vec{x} . \vec{y}}{|x|.|y|} \tag{3}$$

K nearest neighbor algorithm uses K closest samples to the query image. Each of these samples belongs to a known class Ci. The query image Iq is categorized to the class CM which has the majority of occurrences among the K samples. The performance of the kNN classifiers highly related to value of the k, the number of the samples and their topological distribution over the feature space. Appendix A describes this through a series of experiments. Many approaches are introduced to improve the performance of the kNN systems using wavelet techniques [17], Cluster-Based Trees [18] and Tolerant rough sets [19] and so on. In this paper we show ensemble based techniques can be used to improve the performance of the system.

.

4 ENSEMBLE OF CLASSIFIERS

An Ensemble classifier wrapping is a method of improving the accuracy of a group of classifiers by combining their results using one of the voting methods. A typical learning system is consisting of feature detection unit and a decision making unit (classifier). Classifier examines all the training data against a decision function. A learning algorithm sets the parameters of classifier based on the training data to result a certain accuracy rate. Then the system is used to predict the result of a testing data set. It is shown [20] in the most cases, an ensemble (committee) of classifiers can produce a better prediction than the individual classifiers. There are different classifier wrappers such as Mixture-of-experts model [21], Adaptive Boosting (Adaboost) [22] and Bagging [23].

4.1 BAGGING

Bagging or “bootstrap aggregation”, is the first effective meta-algorithm method of ensemble learning. This method is a special case of the model averaging, which was designed for decision tree models, but it can be used with other type of model for classification.

Bagging method uses multiple versions of a training set to train a different model of classifier. Each version of the training set can be generated by sampling with replacement. The outputs of the models are combined by voting to create a single output. The Bagging algorithm can be summarized as followed [24]:

Algorithm 1 The Bagging algorithm

1. Input: Training data set S with correct labels •*i* i=1,...,C representing C classes, Number of Iterations T, Weak learning function G(.), Percent F to create bootstrapped training data.

2. Initialize:

for all t = 1, . . . ,T

2.1 Take a bootstrapped replica St by randomly drawing F percent of S.

2.2 Call G(.) with St and receive the hypothesis (classifier) ht .

2.3 Add ht to the ensemble, E.

End.

3. Test: Simple Majority Voting – Given unlabeled instance x

3.1. Evaluate the ensemble E= $\{h_1, \ldots, h_t\}$ on x.

3.2. Let $V_{t,j}$ be the vote given to class ω_j by classifier ht.

$$V_{t,j} = \begin{cases} 1, & \text{if} \quad h_t \text{ picks class } \omega_j \\ 0, & \text{otherwise} \end{cases}$$

3.3. Total vote received by each class is calculate as

$$V_j = \sum_{t=1}^{T} v_{t,j},\ j = 1,\ldots,C$$

```
3.4. Select the class that
receives the highest total vote
as the final classification.
```

end.

5 RESULTS

In order to demonstrate the performance and capabilities of this system, ORL face database is used to test the system. In this database, there are 40 subjects and each subject consists of 10 images of the size of 112x 92 pixels with variations in scale, orientation and facial expressions as it illustrated in Fig 1. The first image of each subject is used as the test image while the other 9 images are used as training set.

Each 112x92 image in the database is reshaped to a 10304x1 feature vector. To reduce the size of the feature space, we used SVM feature reduction method and reduce the size of feature vector from 10304 to 15 features.

These feature sets are feed to an Ensemble based K nearest neighbour (kNN) classifier with k=3 and Bagging wrappers. The Accuracy rate of 97.5% is achieved, which means 39 of the 40 subjects are recognized correctly and only one of the subjects (subject number 35 as shown in Fig. 2) was recognized incorrectly.

in order to study and compare the results of this system with other methods a series of experiments have been conducted. Table 1 shows the result of our system and two other techniques (SVM and maximum entropy models). Each method tested once without classifier wrapping and once with ensemble classifier wrapping. Two different classifier wrappers, AdaBoost and Bagging, are used in these experiments. kNN is also tested with 3 different values for k (k={1,3,5}). This results still shows the best accuracy rates belongs to kNN (k=3) +Bagging with 97.5% accuracy and only 2.5% error rate.

6 SUMMARY

In this paper we proposed a novel method for face recognition using ensemble based K nearest neighbour classifier. This system uses Bagging as Ensemble classifier wrapper. Based on the results of the system on the ORL face database the accuracy rate of 97.5% is achieved.

REFERENCES

[1] M.A. Turk and A.P. Pentland. " Face Recognition Using Eigenfaces '. IEEE Computer Society- Conference on Computer Vision and Pattern Recognition. Pages 586-591 .June 1991.

[2] L. Wiskott. J.M. Fellous. N. Kruger and C. von der Malsburg. Face Recognition by Elastic Bunch Graph Matching ". IEEE Trans. Pattern Anal-vsis and Machine.

[3] S. Y. Kung, M. Fang, S. P. Liou, M. Y. Chiu, and J. S. Taur, “Decision-based neural network for face recognition system,” in *Proc. ICIP’96*, vol. I, Washington, D.C., Oct. 1995, pp.430–437.

[4] A. Alattar and S. Rajala, “Facial features localization in frontal view head and shoulders images,” IEEE international Conference on Acoustics, Speech and signal Processing, vol. 6, pp. 3557-3560, 1999.

[5] G. Chow and X. Li, “Towards a system for automatic facial feature detection," Pattern Recognition, vol. 26, no. 12, pp. 1739-1755, 1993.

[6] B. Manjunath, R. Chellappa, and C. D. Malsburg, “A feature based approach to face recognition,” Proceedings of IEEE Computer Society. Conference on Computer Vision and Pattern Recognition, pp. 373-378, 1992.

[7] M. Michaelis, R. Herpers, L. Witta, and G. Sommer, “Hierarchical filtering scheme for the detection of facial key points," International Conference on Acostics, Speech, and Signal Processing, vol. 4, pp. 2541-2544, 1997.

[8] L. Stringa, “Eyes detection for face recognition," Applied Artificial Intelligence, vol. 7, pp. 365-382, 1993.

[9] L. Yuille, P. Hallinan, and D. Cohen, “Feature extraction from faces using deformable templates," International Journal of Computer Vision, vol. 8, no. 2, pp. 99-111, 1992.

[10] A. J. Goldstein, L. Harmon, and A. Lesk, “Identification of human faces,” Proceedings of the IEEE, pp. 748-760, 1971.

[11] L. D. Harmon, M. K. Khan, R. Lasch, and P. F. Ramig, “Machine identification of human faces,” Pattern Recognition, pp. 97-110, 1981.

[12] T. Kanade, “Picture Processing by Computer Complex and Recognition of Human Faces”. PhD thesis, Kyoto University, 1973.

[13] M. D. KELLY, “Visual identification of people by computer,” Technical report AI-130, Stanford AI Project, Stanford, CA., 1970

[14] K. Fukunaga, “Introduction to statistical pattern recognition,” Academic Press, 2nd ed, 1990

[15] M. Turk and A. Pentland, “Eigenfaces for recognition” Journal of Cognitive Neuroscience, vol. 3, pp. 71-86, 1991

[16] C. Martínez and O. Fuentes, “ Face Recognition using Unlabeled Data”, Computation y Systems - Journal of Computer Science Research, Vol. 7, No.2, pp 123-129, 2003.

[17] Hui Kong, Xuchun Li, Jian-Gang Wang and Chandra Kambhamettu, "Ensemble Learning for Face Recognition". The International Conference on Biometrics Authentification (ICBA), Hong Kong, 2006

[18] Yu-Long Qiao; Jeng-Shyang Pan; Sheng-He Sun, "Improved K nearest neighbor classification algorithm," Circuits and Systems, 2004. Proceedings. The 2004 IEEE Asia-Pacific Conference on , vol.2, no.pp. 1101- 1104 vol.2, 6-9 Dec. 2004

[19] Bin Zhang, Sargur N. Srihari, "Fast k-Nearest Neighbor Classification Using Cluster-Based Trees," IEEE Transactions on Pattern Analysis and Machine Intelligence ,vol. 26, no. 4, pp. 525-528, April, 2004.

[20] Yongguang Bao; Xiaoyong Du; Ishii, N., “Improving performance of the k-nearest neighbor classifier by tolerant rough sets,” Cooperative Database Systems for Advanced Applications, 2001. CODAS 2001. The Proceedings of the Third International Symposium on , vol., no.pp.167-171, 2001

[21] Dietterich T.G. “Machine-learning research: Four current directions “(1997) AI Magazine, 18 (4), pp. 97-136.

[22] R.A. Jacobs, M.I. Jordan, S.J. Nowlan, and G.E. Hinton, “Adaptive mixtures of local experts,” Neural Computation, vol. 3, pp. 79–87, 1991.

[23] Y. Freund and R.E. Schapire, “Decision-theoretic generalization of on-line learning and an application to boosting,” Journal of Computer and System Sciences, vol. 55, no. 1, pp. 119–139, 1997.

[24] BREIMAN, Leo, “Bagging predictors,” Machine Learning 24(2):pp123-140, 1996.

[25] R. Polikar, “Ensemble Based Systems in Decision Making,” IEEE Circuits and Systems Magazine, vol.6, no.3, pp. 21-45, 2006

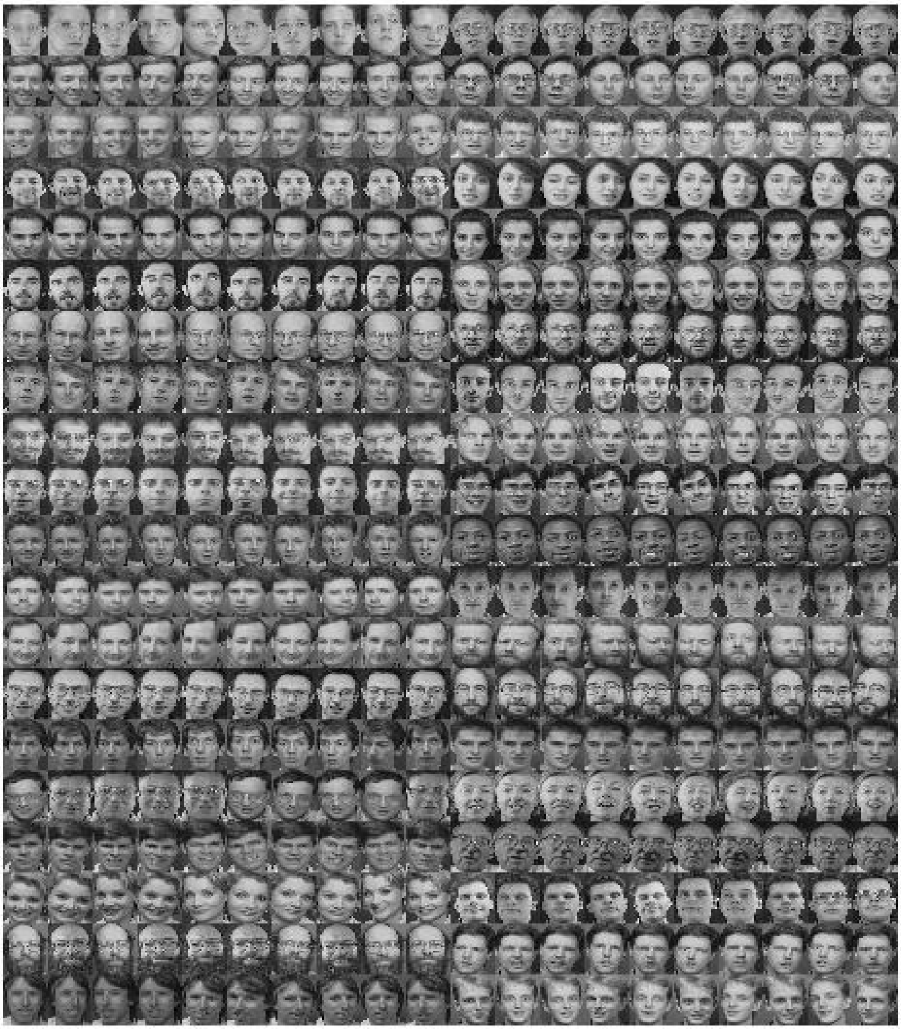

Figure 1- The ORL face database.

Figure 2- Ten images of subject number 35 in ORL face database, the first image is used as the test image and the other 9 images are used as the train set.

APPENDIX: THE PERFORMANCE OF THE KNN CLASSIFIERS

To examine the performance of the kNN classifiers and its relation to value of the k, the number of the samples and their topological distribution over the feature space a series of tests has been conducted. Table 1 shows the results of 90 experiments over a 2D feature space. For each experiment a random distribution of the samples over the N= {2, 20} number of clusters is generated. The number of clusters N is an argument of the complexity of the distributions of samples. The number of samples in this experiment varies between 10, 100 and 1000 samples. The value of k is selected from the set S={1,3,5,15,51}. Fig. 3 illustrates the distribution of 10,100 and 1000 random samples over a 2D feature space and the result of a kNN classifier. The Error rate in this example is 1.87%. As the Table.1 shows the performance of the system is rapidly increased if a large number of the samples are used. However in the most cases, increase in the number of the data is impossible. To improve the performance of the kNN systems some techniques such as wavelet techniques [4], Cluster-Based Trees [5] and Tolerant rough sets [6] and Ensemble-based techniques can be employed.

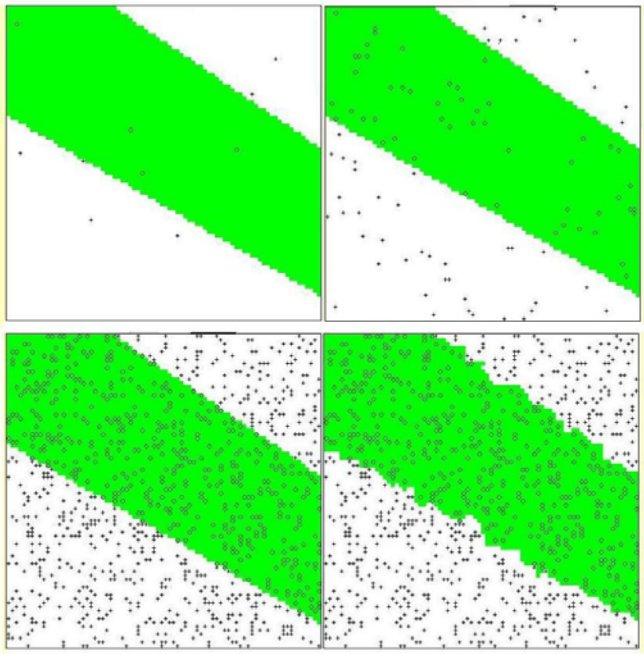

Figure 3 - A(top left), B(top right) and C(bottom left) illustrate the distribution of 10,100 and 1000 random samples over a 2D feature space. The result of a kNN classifier is shown on D (bottom right) Error = 1.87%

Table 1. Error rates for SVM, kNN and maximum entropy models (IIS). Each method tested with and without Ensemble wrapping (AdaBoost and Bagging).

		Error Rate%		
		No Wrapping	+AdaBoost	+Bagging
SVM		7.5	97.5	7.5
maximum entropy models (IIS)		7.5	5	7.5
kNN	K=1	2.5	97.5	5
	K=3	5	5	2.5
	K=5	12.5	7.5	12.5

Table 2. Results of 90 experiments over a 2D feature space.

Number of clusters	Number of samples	K	Error%		
			Test 1	Test 2	Test 3
2	10	1	5.00	7.80	11.56
		3	11.56	10.45	6.70
		5	11.56	26.73	5.80
		15	11.56	26.73	66.09
		51	88.43	73.26	66.09
	100	1	5.53	6.07	5.65
		3	5.03	6.95	7.40
		5	4.84	6.59	7.53
		15	5.96	9.90	7.35
		51	21.68	37.40	24.43
	1000	1	1.57	1.71	1.70
		3	1.65	1.60	2.07
		5	1.68	1.87	2.14
		15	2.7	1.84	2.20
		51	5.84	3.54	3.78
20	10	1	43.96	56.64	38.18
		3	42.84	59.46	47.42
		5	43.21	50.43	47.45
		15	61.26	51.12	50.67
		51	61.26	51.12	50.67
	100	1	26.21	28.68	32.06
		3	28.96	36.59	38.54
		5	32.53	39.20	44.85
		15	38.00	43.23	50.34
		51	38.84	48.92	52.29
	1000	1	10.06	11.06	12.43
		3	12.18	12.23	13.93
		5	12.96	13.04	15.25
		15	15.34	17.07	19.07
		51	21.14	25.35	31.60