Ensemble-Based Discriminant Manifold Learning For Face Recognition

Junping Zhang[1 2], Li He[1], and Zhi-Hua Zhou[3]

[1] Shanghai Key Laboratory of Intelligent Information Processing
Department of Computer Science and Engineering
Fudan University, Shanghai 200433, China
jpzhang@fudan.edu.cn; demonstrate@163.com
[2] The Key Laboratory of Complex Systems and Intelligence Science
Institute of Automation, Chinese Academy of Sciences, Beijing, 100080, China
[3] National Laboratory for Novel Software Technology
Nanjing University, Nanjing 210093, China
zhouzh@nju.edu.cn

Abstract. The locally linear embedding (LLE) algorithm can be used to discover a low-dimensional subspace from face manifolds. However, it does not mean that a good accuracy can be obtained when classifiers work under the subspace. Based on the proposed ULLELDA (Unified LLE and linear discriminant analysis) algorithm, an ensemble version of the ULLELDA (En-ULLELDA) is proposed by perturbing the neighbor factors of the LLE algorithm. Here many component learners are generated, each of which produces a single face subspace through some neighborhood parameter of the ULLELDA algorithm and is trained by a classifier. The classification results of these component learners are then combined through majority voting to produce the final prediction. Experiments on several face databases show the promising of the En-ULLELDA algorithm.

1 Introduction

During the past decades, much research on face recognition has been done by computer scientists [1]. Many assumptions are generated for building a high-accuracy recognition system. One assumption is that a facial image can be regarded as single or multiple arrays of pixel values. The other one is that face images are empirically assumed to points lying on manifolds which are embedded in the high-dimensional observation space [2, 3]. As a result, many manifold learning approaches are proposed for discovering the intrinsic face dimensions, such as expression and pose, and so on [4, 5]. However, when the manifold learning-based approaches are employed for discriminant learning, the accuracy of classifiers easily suffer. A possible reason is that the distances among the smallest d eigenvalues or eigenvectors which are obtained by the LLE algorithm are so small and close [6] that the face subspace is ill-posed and instability. Furthermore, the choose of neighbor factors influences the accuracy of classifiers.

To overcome the mentioned two disadvantages, we propose an ensemble ULLELDA (En-ULLEDA) algorithm. First, the previous proposed ULLELDA algorithm with a single-value neighborhood parameter is performed for obtaining a single subspace [7]. Second, a set of neighborhood parameters are employed for generating a collection of subspaces, each of which is used for discriminant learning with a specific classifier. Finally, the classification is performed by majority voting. Experiments on several face databases show the promising of the En-ULLELDA algorithm when compared with the ULLELDA algorithm.

The rest of the paper is organized as follows. In Section 2 we propose the En-ULLELDA algorithm. In Section 3 the experimental results are reported. Finally, In section 4 we conclude the paper with some discusses.

2 Ensemble of ULLELDA (En-ULLELDA)

For better understanding the proposed En-ULLELDA algorithm, we first give a brief introduction on the ULLELDA algorithm. The objective of the previous proposed ULLELDA algorithm is to refine the classification ability of manifold learning. The basic procedure of the algorithm is demonstrated as follows:

Step 1 Approximate manifold around sample $\boldsymbol{x}_i$ with a linear hyperplane passing through its neighbors $\{\boldsymbol{x}_j, j \in \mathcal{N}(\boldsymbol{x}_i)\}$. Therefore, define

$$\psi(\boldsymbol{W}) = \sum_{i=1}^{N} \|\boldsymbol{x}_i - \sum_{j \in \mathcal{N}(\boldsymbol{x}_i)} \boldsymbol{W}_{i,j} \boldsymbol{x}_j\|^2 \tag{1}$$

Subject to constraint $\sum_{j \in \mathcal{N}(\boldsymbol{x}_i)} \boldsymbol{W}_{i,j} = 1$, and $\boldsymbol{W}_{i,j} = 0, j \notin \mathcal{N}(x_i)$, all the weights $\boldsymbol{W}$ in Eq. 1 are calculated in the least square sense [4]. Where $\mathcal{N}(\boldsymbol{x}_i)$ denotes the index set of neighbor samples of x_i.

Step 2 The LLE algorithm assumes that the weight relationship of sample and its neighborhood samples is invariant when data are mapped into a low-dimensional subspace. Let $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N)$. Therefore, define

$$\varphi(\boldsymbol{Y}) = \sum_{i=1}^{N} \|\boldsymbol{y}_i - \sum_{j \in \mathcal{N}(\boldsymbol{x}_i)} \boldsymbol{W}_{i,j} \boldsymbol{y}_j\|^2 \tag{2}$$

where $\boldsymbol{W}$ can minimize Eq. 1 and satisfies the constraints, and

$$\sum_i \boldsymbol{y}_i = 0 \tag{3}$$

$$\frac{1}{N} \sum_i \boldsymbol{y}_i \boldsymbol{y}_i^\top = \boldsymbol{I} \tag{4}$$

The optimal solution of $\boldsymbol{Y}^*$ in Eq. 2 can be transformed into computing the smallest or bottom $d+1$ eigenvectors of matrix $(\boldsymbol{I} - \boldsymbol{W})^T (\boldsymbol{I} - \boldsymbol{W})$ except that the zero eigenvalue needs to be discarded.

Step 3 The d-dimensional data $\boldsymbol{Y}$ are further mapped into a p-dimensional discriminant subspace through the LDA algorithm (linear discriminant analysis) [8]. Therefore, we have

$$\boldsymbol{Z} = \boldsymbol{D}\boldsymbol{Y} \tag{5}$$

Where $\boldsymbol{Z} = (\boldsymbol{z}_i \in \mathrm{I\!R}^p, i = 1, \cdots, N)$, and the size of matrix $\boldsymbol{D}$ is p rows with d columns which is calculated based on the LDA algorithm [8].

Step 4 To project out-of-the-samples into the discriminant subspace, the mapping idea of the LLE algorithm is employed [4]. A main difference between the LLE and the ULLELDA algorithms is that in the latter one, the out-of-the-samples are directly projected into a discriminant subspace without the computation of the LLE algorithm.

Step 5 A specific classifier is employed in the subspace obtained by the ULLELDA algorithm for discriminant learning.

The previous experiments on several face databases show that the combination of the proposed ULLELDA algorithm and some specific classifier has better accuracy than the combination of the traditional PCA (principal component analysis) algorithm and classifier [7]. However, two potential instabilities influence the performance of the ULLELDA algorithm:

1. Spectral decomposition in the LLE algorithm may generate different intrinsic low-dimensional subspaces even if the neighborhood parameters are fixed. The reason is that the differences among the principal eigenvalues obtained by the LLE algorithm are so small and close that the sequences among the principal eigenvalues are easily alternated.
2. The neighbor parameter in the LLE algorithm plays a trade-off role between global measure and local one. If the size is very large, the LLE algorithm is approximately equivalent to the classical linear dimensionality reduction approach. And if it is very small, the LLE algorithm cannot achieve an effective dimensionality reduction.

While some refinements on how to select a suitable neighborhood size had been proposed [9] for unsupervised manifold learning, it is still difficult for supervised learning to choose an optimal parameter because data are noisy and the mentioned two instabilities are dependent each other. As a result, choosing the neighbor parameter for supervised manifold learning still depends on user's experience. These problems motivate us to further improve the proposed ULLELDA algorithm through ensemble learning.

Krogh and Vedelsby [10] have derived a famous equation $\boldsymbol{E} = \overline{\boldsymbol{E}} - \overline{\boldsymbol{A}}$ in the case of regression, where $\boldsymbol{E}$ is the generalization error of an ensemble, while $\overline{\boldsymbol{E}}$ and $\overline{\boldsymbol{A}}$ are the average generalization error and average ambiguity of the component learners, respectively. The ambiguity was defined as the variance of the component predictions around the ensemble prediction, which measures the disagreement among the component learners [10]. This equation discloses that the more accurate and the more diverse the component learners are, the better the ensemble is. However, measuring diversity is not straightforward because there is no generally accepted formal definition, and so it remains a trick at present to generate accurate but diverse component learners. Several known tricks include perturbing the training data, perturbing input attributes and learning parameters.

As for the ULLELDA algorithm, it is clear that when neighbor factor is perturbed, a set of different discriminant subspaces which result in diversity will be generated. When a test sample is classified under the subspaces, the accuracy may be different from one subspace to the other one. Therefore, we can use majority voting to model ensemble ULLELDA algorithm (En-ULLELDA). Let ULLELDA_K be the ULLELDA algorithm with some neighbor factor K (Here K denotes the number of $\mathcal{N}(\cdot)$), and base classifier be BC, then the classification criterion of the En-ULLELDA algorithm is written as follows:

$$C(\boldsymbol{v}_i) = \arg\max_{l}\{K|(\text{ULLELDA}_K(v_i), \text{BC}) = l\} \tag{6}$$

Where $C(\boldsymbol{v}_i)$ denotes the label of test sample $\boldsymbol{v}_i$, and $(\text{ULLELDA}_K(\boldsymbol{v}_i), \text{BC})$ denotes a component learner with neighbor factor K and some specific base classifier. It is noticeable that a base classifier is explicit contained by the proposed En-ULLELDA framework.

A pseudo-code of the En-ULLELDA algorithm is illustrated as follows:

Table 1. The En-ULLELDA algorithm

Input:
  Data set X; Learner L; Neighborhood Parameters Set $K = \{k_1, k_2, \cdots, k_m\}$,
  Stepsize= $k_i - k_j, i - j = 1$
  Reduced Dimension d and Reduced Discriminant Dimension d', Trials T
Procedure:
1. Normalization of X
2. for $t = 1$ to T {
3. Let $Enerror$ be an empty set
4. for $k = k_1$: Stepsize: k_m
5. {
6. Generate training set $X_{training}$ and test set X_{test} with random partition
7. Based on LLE, k and d,
   Calculate the corresponding one $Y_{training}$ of training set $X_{training}$
8. Based on LDA and d',
   Calculate the corresponding one $Z_{training}$ of $Y_{training}$
9. Project test set X_{test} into the subspace of $Z_{training}$ with ULLELDA,
   and obtain the corresponding one Z_{test}
10. Compute error rate $e(Z_{test}, t, k, L)$ of Z_{test} based on $Z_{training}$ and learner L
11. Storage classification labels $Lab(test, k, t, L)$ of test samples.
12. }
13. $Enerror = \{Enerror; \text{MajorityVoting}(Lab(test, K, t, L))\}$
14. }
15. Output:
  The average error and standard deviation of $Enerror, t = 1, \cdots, T$

3 Experiments

To test the proposed En-ULLELDA algorithm, three face databases (the ORL database (40 subjects and 10 images per person) [11], the UMIST database (575 multi-view facial images with 20 subjects) [12], and the Yale database (15 individuals and 11 images per person) [13]) are used. By combining the ORL database and the UMIST database with the Yale database, a large database Henceforth the OUY face database) including 75 individuals and 1140 images is built. In the paper, the intensity of each pixel is characterized by one dimension, and the size of an image is $112 * 92$ pixels which form a 10304 dimensional vector. All the dimensions are standardized to the range $[0, 1]$. The training samples and test samples are randomly separated without overlapping. All the reported results are the average of 100 repetitions.

For overcoming the curse of dimensionality, the reduced dimensions of data based on the LLE algorithm are set to be 150. For the 2nd mapping (LDA-based reduction) of the ULLELDA algorithm, the reduced dimension is generally no more than $L-1$ (Where L means the number of classes). Otherwise eigenvalues and eigenvectors will appear complex numbers. Actually, we only keep the real part of complex values when the 2th reduced dimensions are higher than $L-1$. When the ULLELDA algorithm is used, the neighbor parameter need to be predefined. With broad experiments, let the neighbor size K be 40 for ORL, UMIST, and be 15 for OUY databases.

Finally, four base classifiers (1-nearest neighbor algorithm (NN) and nearest feature line (NFL) algorithm [14], the nearest mean algorithm (M) and the Nearest-Manifold-based (NM) algorithm) are employed to test classification performance based on the face subspace(s). Here the NFL algorithm denotes that classification is achieved by searching the nearest projection distance from sample to line segments of each class. And the NM algorithm is to calculate the minimum projection distance from each unknown sample to hyperplanes of different classes where each hyperplane is made up of three prototypes of the same class [15].

A set of comparative experiments between the En-ULLELDA algorithm and the ULLELDA algorithm are performed on three face databases. It is noticeable that due to the fact that the number of face images each class is different in the OUY database, we divide the database into training /test set based on the ratio of the number of training samples to the number of samples of the same class. And the ratio is equal to 0.4 when we investigate the influence of neighbor factor K. The reported results are shown in Fig. 1 to Fig. 3. In these figures, the ranges of neighbor factors are shown in the horizontal axis of Fig. 1 to Fig. 3. The vertical axis denotes the ratio of the error rate of the ULLELDA algorithm against that of the En-ULLELDA algorithm. Also, in the title of subplots on the OUY database, the abbreviation "ROUY" denotes that the training samples is sampled based on the mentioned ratio. If a value in vertical direction is greater than 1, it means that the En-ULLELDA algorithm has better recognition performance than the ULLELDA with single neighbor parameter. Furthermore, the experimental results obtained by the En-ULLELDA algorithm with base classifiers are shown in the titles of subplots. By analyzing the experimental results, it is not

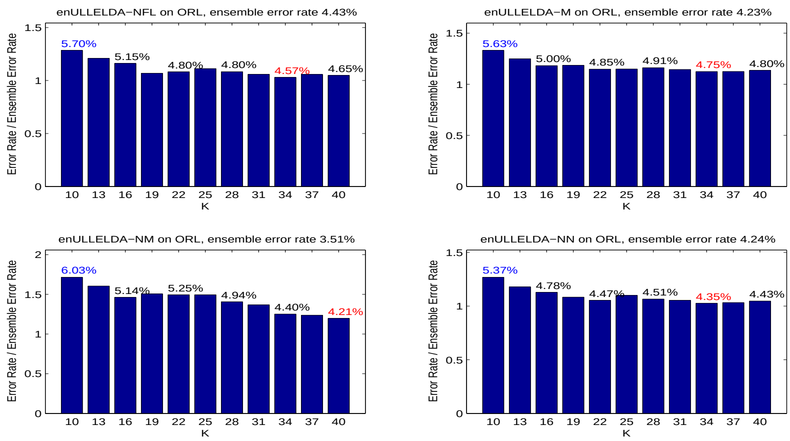

Fig. 1. The En-ULLELDA and The ULLELDA algorithms on The ORL Database

difficult to see that as K varies, the error rates have remarkable difference. For example, when $K = 10$, the error rate based on the ULLELDA+NM is 6.32%; and when $K = 40$, the error rate is 4.21% in the ORL database. Due to the fact that in real applications, it is almost always impossible to prior know which K value is the best, therefore it is not easy for the ULLELDA algorithm to get its lowest error rate. From the figures we can also see that the lowest error rates are obtained by the proposed En-ULLELDA algorithm. So, the En-ULLELDA algorithm is a better choice. In summary, the results argue that 1) the selection of K is a crucial factor to the performance of face recognition. 2) The accuracy of the En-ULLELDA algorithm is superior to that of the ULLELDA algorithm in all of the mentioned face databases.

For testing the generalization performance of the En-ULLELDA algorithm, the influence of the number of training samples is also studied. The results are displayed as in Fig. 4 through Fig. 6. It is worth noting that for better visualization, we only draw some main results achieved by the ULLELDA algorithm. In these figures, all the dashed lines represent the error rates of the ULLELDA algorithm with some specific neighbor factors. Each factor is in the range of $[10, 40]$. Also, each bar is error rate of the En-ULLELDA algorithm with fixed number of training samples. Meanwhile, in the top of each bar, "$x + x\%$" denotes the error rate plus standard deviation of the En-ULLELDA algorithm. From experimental results it can be seen that the En-ULLELDA with base learners obtains better recognition performance than the ULLELDA algorithm with single neighbor factor. For example, when the En-ULLEDA with NM algorithm is used for ORL face database, the lowest error rate and standard deviation are 0.88% and

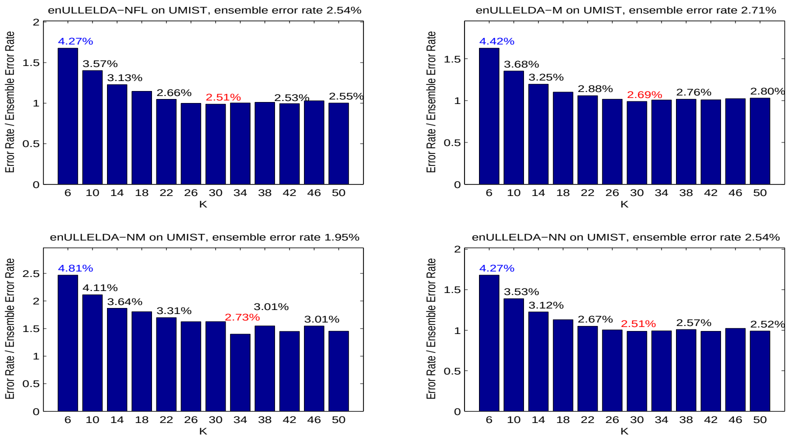

Fig. 2. The En-ULLELDA and The ULLELDA algorithms on The UMIST Database

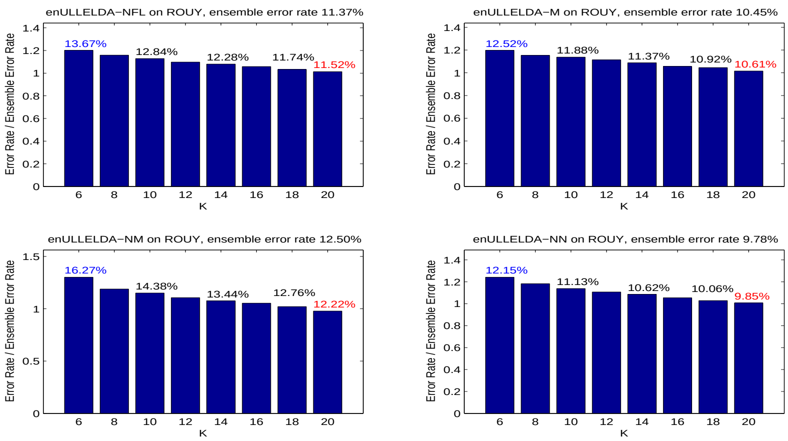

Fig. 3. The En-ULLELDA and The ULLELDA algorithms on The OUY face Database

1.39%, respectively (where training samples= 9). From these figures it can be seen that compared with the combination of the ULLELDA algorithm and three base classifiers, the En-ULLELDA algorithm has better recognition ability. It show again that the En-ULLELDA algorithm is a refinement of the ULLELDA algorithm in enhancing the recognition ability of base classifiers.

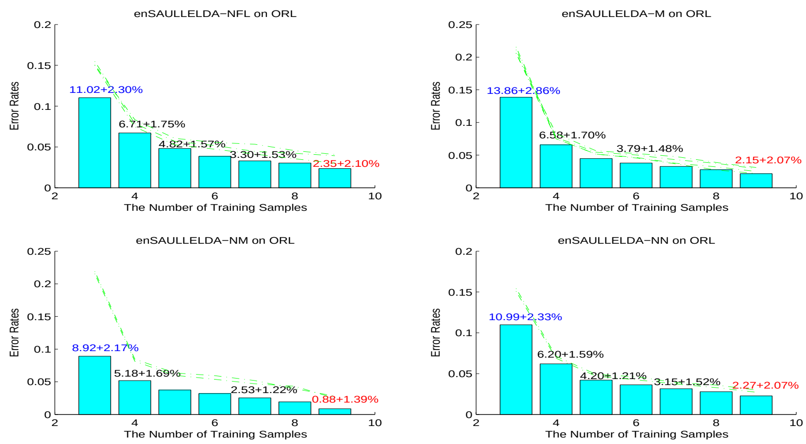

Fig. 4. The Influence of Training Samples on The ORL Database

4 Conclusions

In this paper, we study ensemble-based discriminant manifold learning in face subspaces and propose the En-ULLELDA algorithm to improve the recognition ability of base classifiers and reduce the influences of choosing neighbor factors and small eigenvalues.

By perturbing the neighbor factor of the LLE algorithm and introducing majority voting, a group of discriminant subspaces with base classifiers are integrated for classifying face images. Experiments show that the classification performance of the En-ULLELDA algorithm is better than that of the combination of the ULLELDA algorithm and base classifier.

In the future, we will consider the combination of ULLELDA with other ensemble learning methods, such as boosting algorithms and bagging algorithms, to further enhance the separability of the common discriminant subspace and decrease the computational complexity of the proposed En-ULLELDA algorithm.

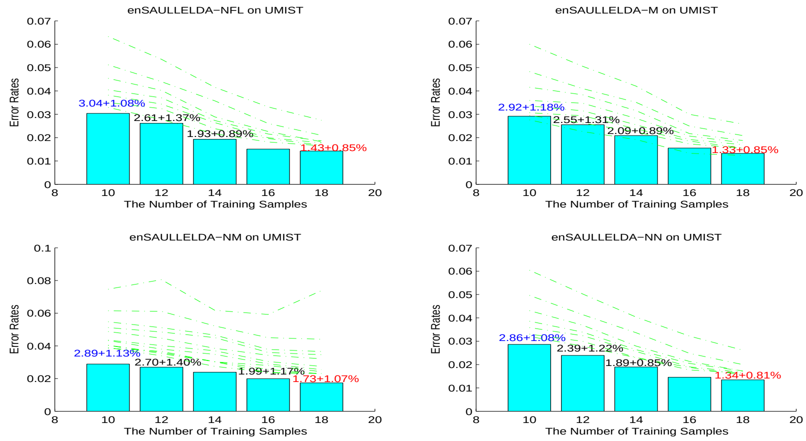

Fig. 5. The Influence of Training Samples on The UMIST Database

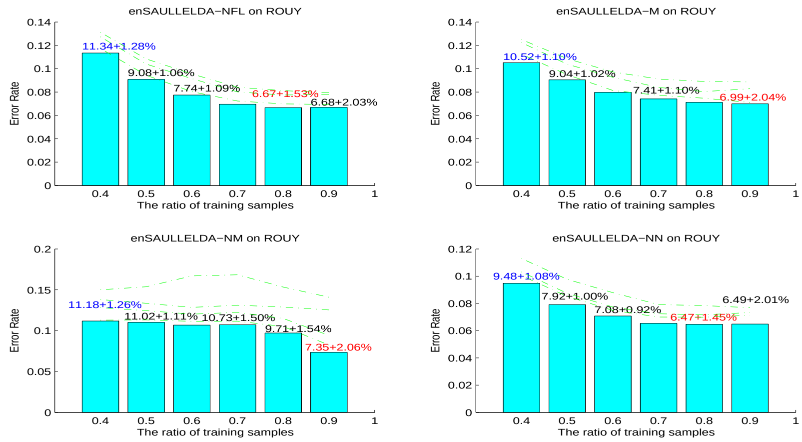

Fig. 6. The Influence of The Training Samples on The OUY Database

Acknowledgements

This work is sponsored by NSF of China (60505002), the Fok Ying Tung Education Foundation (91067) and the Excellent Young Teachers Program of MOE of China.

References

1. Samal, A., Iyengar, P. A.: Automatic Recognition and Analysis of Human Faces and Facial Expressions: A Survey. Pattern Recognition. 25 (1992) 65-77
2. Nayar, S. K., Nene, S. A., Murase, H.: Subspace Methods for Robot Vision. Technical Report CUCS-06-95, Columbia University, New York (1995)
3. Seung H. S., Daniel, D. L.: The Manifold Ways of Perception. Science 290 (2000) 2268-2269
4. Roweis S. T., Lawrance, K. S.: Nonlinear Dimensionality Reduction by Locally Linear Embedding. Science 290 (2000) 2323-2326
5. Tenenbaum, J. B., Silva, de, Langford, J. C.: A Global Geometric Framework for Nonlinear Dimensionality Reduction. Science 290 (2000)2319-2323
6. Sha, F., Saul, L. K.: Analysis and Extension of Spectral Methods for Nonlinear Dimensionality Reduction. In: Proceedings of the 22nd International Conference on Machine Learning, Bonn, Germany, 15 (2005) 721–728
7. Zhang, J., Shen, H., Zhou, Z. –H.: Unified Locally Linear Embedding and Linear Discriminant Analysis Algorithm (ULLELDA) for Face Recognition. In: Stan Z. Li, Jianhuang Lai, Tieniu Tan, Guocan Feng, Yunhong Wang (eds.): Advances in Biometric Personal Authentication. Lecture Notes in Computer Science, Vol. 3338. Springer-Verlag, Berlin Heidelberg New York (2004) 209-307
8. Swets D. L., Weng, J.: Using Discriminant Eigenfeatures for Image Retrieval. IEEE Trans. on PAMI. 18 (1996) 831-836
9. Kouropteva, O., Okun, O., PietikÄinen, M.: Selection of the Optimal Parameter Value for the Locally Linear Embedding Algorithm. In: Wang L, Halgamuge SK, Yao X, (eds.): Computational Intelligence for the E-Age, Singapore, (2002) 359-363
10. Zhou Z.–H., Yu, Y.: Ensembling local learners through multimodal perturbation. IEEE Trans. on SMCB, 35(2005) 725-735
11. Samaria, F. S.: Face Recognition Using Hidden Markov Models. PhD thesis, University of Cambridge, (1994)
12. Daniel, B., Graham, Allinson, N. M.: em Characterizing virtual Eigensignatures for General Purpose Face Recognition. In: H. Wechsler, P. J. Phillips, V. Bruce, F. Fogelman-Soulie and T. S. Huang (eds.): Face Recognition: From Theory to Applications. NATO ASI Series F, Computer and Systems Sciences, 163 (1998) 446-456
13. Bellhumer, P. N., Hespanha, J., Kriegman, D.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Trans. on PAMI. 17(1997) 711–720
14. Li, S. Z., Chan K. L., Wang, C. L.: Performance Evaluation of the Nearest Feature Line Method in Image Classification and Retrieval. IEEE Trans. on PAMI, 22 (2000) 1335-1339
15. Zhang, J., Li, S. Z., Wang, J.: Nearest Manifold Approach for Face Recognition. The 6th IEEE International Conference on Automatic Face and Gesture Recognition (2004) 223-228