Similarity Rank Correlation for Face Recognition Under Unenrolled Pose

Marco K. Müller, Alexander Heinrichs, Andreas H.J. Tewes,
Achim Schäfer, and Rolf P. Würtz

Institut für Neuroinformatik, Ruhr-Universität, D–44780 Bochum, Germany

Abstract. Face recognition systems have to deal with the problem that not all variations of all persons can be enrolled. Rather, the variations of most persons must be modeled. Explicit modeling of different poses is awkward and time consuming. Here, we present a subsystem that builds a model of pose variation by keeping a model database of persons in both poses, additionally to the gallery of clients known in only one pose. An identification or verification decision for probe images is made on the basis of the rank order of similarities with the model database. Identification achieves up to 100% recognition rate on 300 pairs of testing images with 45 degrees pose variation within the CAS-PEAL database, the equal error rate for verification reaches 0.5%.

1 Introduction

Despite significant progress automatic face recognition still suffers from the multitude of variations a face undergoes when images are taken. This article aims at improving face recognition in the presence of head pose variation. Rather than creating a 3D model explicitly like in [1] or by introducing a manifold using many views of each subject [2,3] an approach is presented where the only preprocessing necessary is to find facial landmarks in the image followed by pose estimation. This is realised by bunch graph matching, but more sophisticated methods like EFOM matching [4,5] are available.

Contrary to approaches with explicit models for this variation, it is not necessary to collect many images showing different head poses in order to enroll a subject into the database. To generalize to different poses, images showing subjects in a variety of different head poses are additionally collected. These model images are taken offline in a preprocessing step.

As explicit modeling of pose variation is awkward and computationally expensive, a simpler approach is studied, which relies on the assumption that similar people in one pose will also be similar in another one. Then, the pose variation can be learned by a list of persons known in both poses. Recognizing a client is now achieved by comparing him with all model identities showing the same head pose, and the ranking of similarities creates a similarity function, which can be transferred to the new pose. Thus, the expensive process of enrollment for different head poses has therefore been shifted from the recognition task itself to a preprocessing step. This similarity is a special case of *rank correlation*,

S.-W. Lee and S.Z. Li (Eds.): ICB 2007, LNCS 4642, pp. 67–76, 2007.

one example being Spearman's rank order correlation coefficient [6]. This sort of statistics is used for the evaluation of biometric systems [7] and for face matching in [8]. Here, we apply it to pose generalization.

2 Method

We describe a person's identity by a list containing identity numbers from a large number of reference persons (the model database M). These lists describing a new identity are orderd according to the similarity values to our model database, they build a rank list. To model pose variation, the model database must contain one image per pose for every model identity.

Beside this model there is a gallery T_G of enrolled subjects which contains only one pose. During recognition or verification of an image of an unknown person in an unenrolled but estimated pose these ranks are calculated for the corresponding view. All these persons make up the testing probes T_P, the combination of both is the testing database T. For two poses, the model database M is structured analogous into M_G and M_P (see fig. 1). A similarity function measuring the correlation of the rank lists then measures how similar the new images of persons with different poses are.

The advantage of such a model is that no views need to be generated to deal with pose variation. New images are only compared to model images of the same kind (pose), and the between-poses comparison is done by correlation of two rank lists containing the person's identity.

2.1 Landmark Finding

The main data structure used in the algorithm is the *model graph* with nodes on facial landmarks that are labeled with Gabor jets [9]. A set of such model graphs for different faces constitutes a bunch graph [10]. The method proposed here needs galleries of known people in different poses. These galleries have been created with an automatic bunch graph generator, which works as follows.

Starting with a couple of manually labeled graphs, which are processed to a bunchgraph, the system finds the best position for the bunchgraph on the first image. If the similarity of the graph with the jets at that point crosses a threshold, matching has worked fine and the found landmarks are a potential face graph. At the beginning of the process this threshold is high, so that only a few potential graphs pass the test. To counter the problem that the similarity rises statistically with the size of the bunchgraph, another test checks the result for consistency, meaning that each node of the bunch graph should be most similar to the landmark it is placed on. Combination of these two tests leads to a stable system.

The correctly matched graph is then used to improve the bunch graph. For this purpose, a *Growing Neural Gas* (GNG) [11] is used as a clustering method.

It generates representatives of the shown data and has the very helpful property, that the optimal number of representatives need not be given by the user, but is determined according to the distribution of the data. At the beginning, a GNG is created for every node of the bunchgraph. The neurons of these GNGs are initialized on the values of the jets at the according node. The nodes of every correctly matched graph in the later process are used as input for the respective GNG. After processing of the input, the values of the neurons of the GNG are assigned to the jets of the bunchgraph.

Using the GNG delivers two main advantages compared to the simpler approach of enhancing the bunchgraph by the correctly matched graph directly. First, the bunchgraph grows more slowly and reaches a stable size once the number of representatives at each node is enough for a good representation of the data. The size of the bunchgraph determines the speed of the matching. Second, the influence of a failed matching on the bunchgraph is minimal. Since the corresponding jets are only used as an input to the GNG, the GNG will in fact move a little bit to the wrong direction, but this can be compensated fast by new and correct input.

Whenever a correctly matched graph is stored and used as input for the GNG, the corresponding image is removed from the list of images to be processed. In this way, the whole set of images is treated. After this first run, the remaining images are processed again. As the increased bunch graph covers the face space better, faces on images with incorrect matches in first run now may be matched correctly. These runs are repeated until no face is matched correctly in a whole run. Then both threshold values of the tests are lowered a little bit and the next run is started. This is repeated, until every image of the database is processed. A comparison of the position of the eye nodes of the found graphs with ground truth data showed that all faces were located with adequate accuracy. A simpler version of this is described in [12].

2.2 Rank Correlation Function

Given a testing set $T = (T_G, T_P)$ consisting of N_G gallery images ($0 \leq t_G < N_G$) and N_P probe images($0 \leq t_P < N_P$). A model set $M = (M_G, M_P)$ of N_M identities is given providing both kinds of images as in gallery ($0 \leq m_G < N_M$) and probe ($0 \leq m_P < N_M$). From the model graphs only those N_N nodes are selected that are visible in both poses, corresponding nodes have the same node index n ($1 \leq n \leq N_N$). For compact notation, in the following X is to be substituted by either "G" or "P". Jets are denoted by $J(\cdot, \cdot, \cdot)$ and depend on node number, identity number, and image set.

The similarities of each test image to all corresponding model images can be calculated nodewise with the jet similarity $S(J_1, J_2)$ as in [9,10]. On the resulting vectors of similarities

$$X^S(n, t_X, m_X) = S(J(n, m_X, M_X), J(n, t_X, T_X)) \tag{1}$$

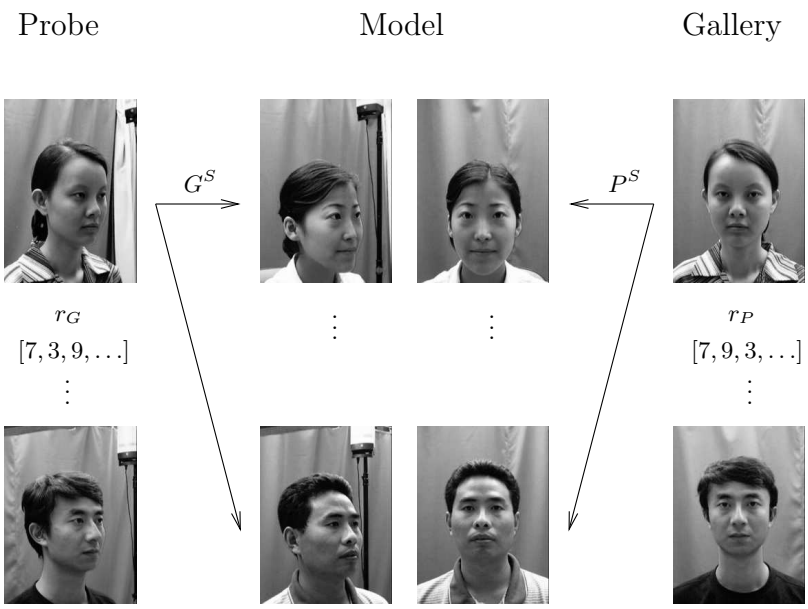


Fig. 1. For each test image the jet similarities X^S to all corresponding model jets can be calculated. Sorting the model indices according to these similarities results in a field $r_X(n, t_X)$ which contains for all test images nodewise the most similar model images. For one node these ranklist may be like the example in the figure.

the ranking $r_X(n, t_X; m_X)$ is defined by the sorting permutation such that

$$m_X > m'_X \Rightarrow X^S(n, t_X; r_X(n, t_X; m_X)) \leq X^S(n, t_X; r_X(n, t_X; m'_X)) \,. \tag{2}$$

Similarity between two test images is then given by:

$$S(t_G, t_P) = \frac{1}{F} \cdot \sum_n \sum_{m_G, m_P} \frac{\delta\big(r_G(n, t_G; m_G), r_P(n, t_P; m_P)\big)}{\sqrt{m_G + m_P + 1}} \,, \tag{3}$$

the normalization factor F is defined as the maximal possible similarity:

$$F = N_N \cdot \sum_{i=1}^{N_M} \frac{1}{\sqrt{2i-1}} \,. \tag{4}$$

To give an example, let $r_g(n, t_G)$ be $[1, 3, 2]$. This means (for node n) the most similar model images to the test image t_G are number 1, then 3 and then 2. If r_p is $[1, 2, 3]$ the similarity for this node would be

$$S = \frac{\frac{1}{\sqrt{1}} + \frac{1}{\sqrt{4}} + \frac{1}{\sqrt{4}}}{\frac{1}{\sqrt{1}} + \frac{1}{\sqrt{3}} + \frac{1}{\sqrt{5}}} = \frac{2}{2.02} = 0.99. \tag{5}$$

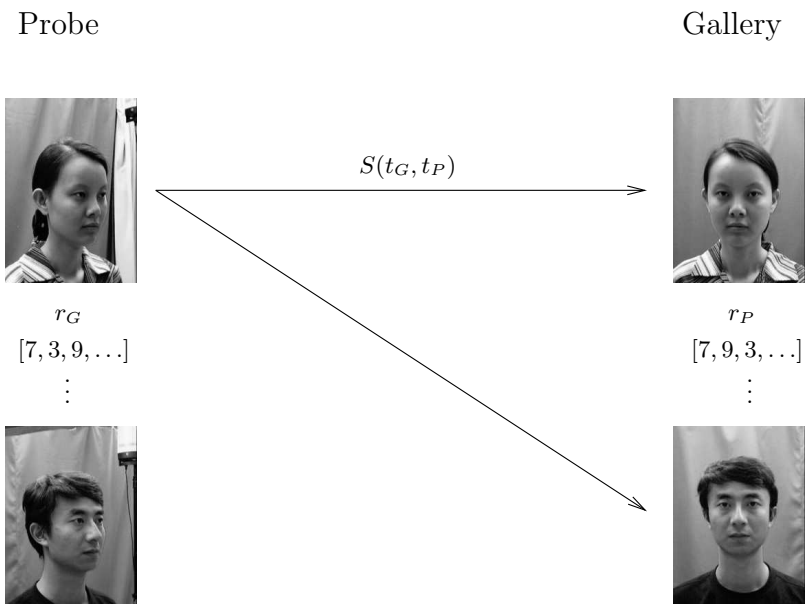


Fig. 2. Once the rank lists have been calculated, they are describing the person's identity. For each probe image showing a person in a very different pose than the image in the gallery the similarity to all gallery images can be calculated by the similarity function $S(t_G, t_P)$

3 Experiments

3.1 Data

Three poses of the CAS-PEAL face Database [13] have been used: PM+45, PM+00 and PM−45. The first N_M identities build the model database, the last N_T identities the testing database. In most experiments $N_M = N_T = 500$ has been used, variations are shown in table 2. Twelve of the remaining identities have been labeled manually to find the landmarks in the model and testing databases, as described in section 2.1.

3.2 Identification Results

Recognition rates are shown in the right half of table 1. For comparison, the results of direct jet comparison (which is not appropriate for large pose variation), are shown on the left half of that table.

Table 2 shows the recognition rates for the case album PM+00 – probe PM−45 for different numbers of identities in model and testing database. In general, recognition rates increase if the number of model identities increases or the number of testing images decreases. If the model size is equal to the test size, the results are even better with a higher number of identities, at least up to 500

(a) PM+45 (b) PM+00 (c) PM−45

Fig. 3. Examples of used CAS-PEAL poses

Table 1. Recognition rates for 500 CAS-PEAL identities for the three poses PM+45, PM+00 and PM−45. The left table shows the results of simple jet comparison, the right one the result of our system with 500 different identities in the model database.

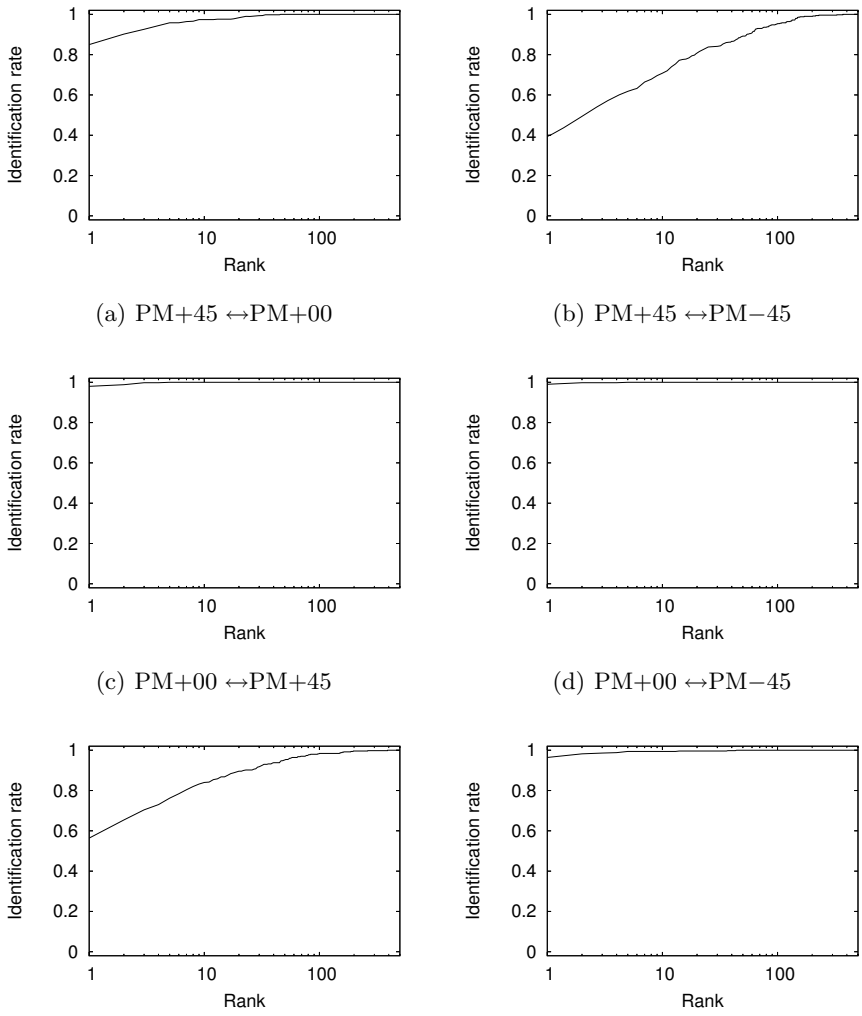
Gallery	Probe		
	PM+45	PM+00	PM−45
PM+45	100.0	8.0	2.6
PM+00	33.0	100.0	54.2
PM−45	3.2	24.4	100.0

Gallery	Probe		
	PM+45	PM+00	PM−45
PM+45	100.0	85.0	39.4
PM+00	98.0	100.0	99.0
PM−45	56.4	96.4	100.0

Table 2. Recognition rates for PM+00 – PM−45 for different numbers of model and test identities. The model database consists of the first N_M identities, the testing database of the N_T last ones of the 1000 used images.

Model	Test								
	100	200	300	400	500	600	700	800	900
100	96.0	95.5	89.7	88.5	83.8	81.2	79.6	78.3	75.9
200	99.0	97.5	96.7	95.3	93.4	92.0	90.4	89.6	—
300	99.0	97.5	97.3	96.8	96.8	96.5	96.1	—	—
400	100.0	98.0	100.0	98.8	98.6	98.3	—	—	—
500	99.0	98.5	99.0	99.0	99.0	—	—	—	—
600	99.0	98.0	99.7	99.5	—	—	—	—	—
700	100.0	99.0	99.3	—	—	—	—	—	—
800	100.0	99.5	—	—	—	—	—	—	—
900	98.0	—	—	—	—	—	—	—	—

identities. Figure 4 shows the cumulative match scores for the closed identification scenario we used. Our method performs quite well for pose differences of 45°, pose variations of 90° are difficult, because the number of visible landmarks in both images decreases. Experiments with node independent rank lists have been made, but combination of node dependent with node independent rank

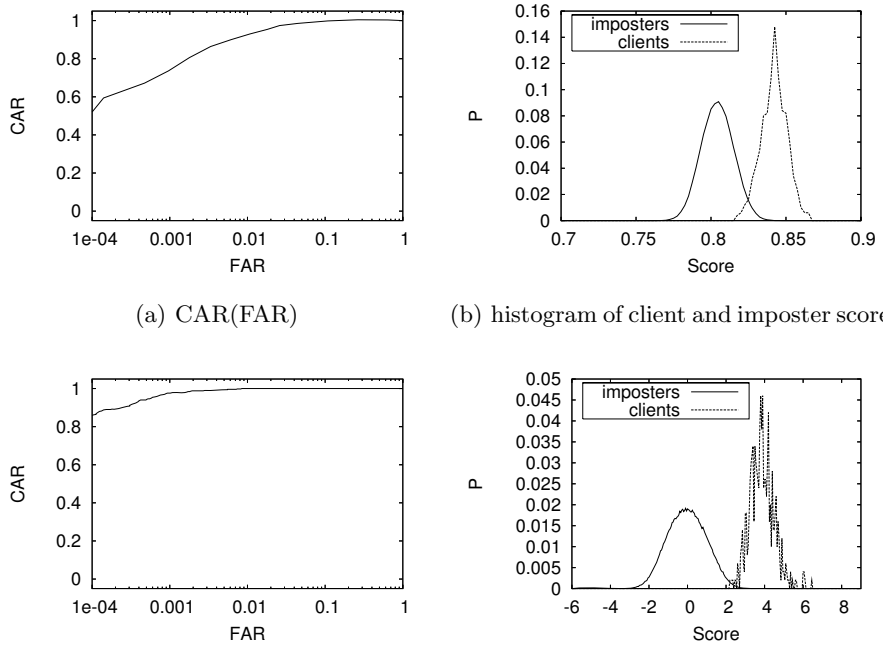


(a) PM+45 ↔PM+00 (b) PM+45 ↔PM−45

(c) PM+00 ↔PM+45 (d) PM+00 ↔PM−45

Fig. 4. Identification performance (cumulative match scores) for different poses in gallery (left of ↔) and probe (right of ↔) set. The model database consists of 500 identities, the testing database of 500 different ones.

similarities lead to better recognition rates for even bigger pose differences than 90°. What can be observed, as already in Table 1, is that recognition is generally better if the gallery consists of frontal pose. In our experiments, the rank similarity function as defined in (3) performed best. For PM+00↔PM−45 it reached 99% of recognition rate. The Spearman rank-order correlation coefficient lead to only 93.4%.



(a) CAR(FAR) (b) histogram of client and imposter scores

(c) CAR(FAR) (d) histogram of client and imposter scores

Fig. 5. Verification performance for PM+00 – PM−45 for 500 model and 500 test identities. a) and b) are without normalization and have an EER of 2.6%. c) and d) are the results with normalization according to (7); the EER is between 0.4% and 0.6%.

3.3 Verification Results

To test our method for verification tasks each of the 500 gallery identities is compared to all probe identities. This means there are 500 clients and 500·499 imposters. This has been done for the case album PM+00 – probe PM−45. Figure 5 shows the results. Figure 5 a) shows Correct Acceptance Rate (CAR) over False Acceptance Rate (FAR), Figure 5 b) the probability distributions for clients and imposters to attain a certain score. The Equal Error Rate (EER) is 2.6%.

To improve the EER in verification experiments, one can use the distribution of similarities to all gallery images to normalize the similiarity values of each probe image. For that, each probe image is tested against all gallery identities, and the resulting 500 similarities for each probe image are normalized in the following way:

$$\bar{S}(t_P) = \frac{1}{N_G} \sum_{t_G=0}^{N_G-1} S(t_G, t_P) \tag{6}$$

$$S_n(t_G, t_P) = \frac{S(t_G, t_P) - \bar{S}(t_P)}{\sqrt{\frac{1}{N_G} \cdot \sum_{t_G=0}^{N_G-1} \left(S(t_G, t_P) - \bar{S}(t_P)\right)^2}} \tag{7}$$

Fig 5 shows the resulting improvement for CAR and EER, the latter lies between 0.4 and 0.6%.

4 Discussion

We have presented a module for a face recognition system which can recognize or verify persons in a pose which is very different from the one enrolled. We have demonstrated the efficiency on 45° pose variation. An interesting side result is that the frontal pose is the best to be used as a gallery. For a complete system, the recognition step must be preceded by a rough pose estimation and look-up in the respective pose model. As the database used has considerable scatter in the actual pose angles it can be concluded that our method is robust enough to expect that the full range of poses can be covered with relatively few examples.

Acknowledgments

We gratefully acknowledge funding from Deutsche Forschungsgemeinschaft (WU 314/2-2 and MA 697/5-1), the European Union, the European Regional Development Fund, and the land of Northrhine-Westphalia in the program "Zukunftswettbewerb Ruhrgebiet" and the "Ruhrpakt". Portions of the research in this paper use the CAS-PEAL face database collected under the sponsorship of the Chinese National Hi-Tech Program and ISVISION Tech. Co. Ltd.

References

1. Blanz, V., Vetter, T.: Face recognition based on fitting a 3d morphable model. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(9), 1063–1074 (2003)
2. Saul, L.K., Roweis, S.T.: Think Globally, Fit Locally: Unsupervised Learning of Low Dimensional Manifolds. Journal of Machine Learning Research 4, 119–155 (2003)
3. Okada, K., von der Malsburg, C.: Pose-invariant face recognition with parametric linear subspaces. In: Proceedings of the International Conference on Automatic Face and Gesture Recognition, Washington D.C., pp. 71–76 (2002)
4. Tewes, A.: A Flexible Object Model for Encoding and Matching Human Faces. Shaker Verlag, Ithaca, NY, USA (2006)
5. Tewes, A., Würtz, R.P., von der Malsburg, C.: A flexible object model for recognising and synthesising facial expressions. In: Kanade, T., Ratha, N., Jain, A. (eds.) Proceedings of the International Conference on Audio- and Video-based Biometric Person Authentication. LNCS, pp. 81–90. Springer, Heidelberg (2005)
6. Press, W., Flannery, B., Teukolsky, S., Vetterling, W.: Numerical Recipes in C — The Art of Scientific Programmming. Cambridge University Press, Cambridge (1988)
7. Andrew, L., Rukhin, A.O.: Nonparametric measures of dependence for biometric data studies. Journal of Statistical Planning and Inference 131, 1–18 (2005)

8. Ayinde, O., Yang, Y.H: Face recognition approach based on rank correlation of gabor-filtered images. Pattern Recognition 35(6), 1275–1289 (2002)
9. Lades, M., Vorbrüggen, J.C., Buhmann, J., Lange, J., von der Malsburg, C., Würtz, R.P., Konen, W.: Distortion invariant object recognition in the dynamic link architecture. IEEE Transactions on Computers 42(3), 300–311 (1993)
10. Wiskott, L., Fellous, J.M., Krüger, N., von der Malsburg, C.: Face recognition by elastic bunch graph matching. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7), 775–779 (1997)
11. Fritzke, B.: A growing neural gas network learns topologies. In: Tesauro, G., Touretzky, D.S., Leen, T.K. (eds.) Advances NIPS 7, pp. 625–632. MIT Press, Cambridge, MA (1995)
12. Heinrichs, A., Müller, M.K., Tewes, A.H., Würtz, R.P.: Graphs with principal components of Gabor wavelet features for improved face recognition. In: Cristóbal, G., Javidi, B., Vallmitjana, S. (eds.) Information Optics: 5th International Workshop on Information Optics; WIO'06, American Institute of Physics, pp. 243–252 (2006)
13. Gao, W., Cao, B., Shan, S., Zhou, D., Zhang, X., Zhao, D.: The CAS-PEAL large-scale Chinese face database and baseline evaluations. Technical Report JDL-TR-04-FR-001, Joint Research & Development Laboratory for Face Recognition, Chinese Academy of Sciences (2004)