9-2009

Multi-view ear recognition based on moving least square pose interpolation

Heng LIU
Southwest University of Science and Technology

David ZHANG
Shanghai Jiaotong University

Zhiyuan ZHANG
Singapore Management University, zhiyuanzhang@smu.edu.sg

Multi-view Ear Recognition Based on Moving Least Square Pose Interpolation

Heng Liu[1,2], David Zhang[3], and Zhiyuan Zhang[4]

[1] School of Information Engineering, Southwest University of Science and Technology, Mianyang, 621010, P.R. China
[2] Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai, 200240, P.R. China
hengliusky@gmail.com
[3] Department of Computing, The Hong Kong Polytechnic University, Hong Kong, P.R. China
csdzhang@comp.polyu.edu.hk
[4] Bio-computing Research center, Harbin Institute of Technology, Shenzhen Graduate School
zzy380@163.com

Abstract. Based on moving least square, a multi-view ear pose interpolation and corresponding recognition approach is proposed. This work firstly analyzes the shape characteristics of actual trace caused by ear pose varying in feature space. Then according to training samples pose projection, we manage to recover the complete multi-view ear pose manifold by using moving least square pose interpolation. The constructed multi-view ear pose manifolds can be easily utilized to recognize ear images captured under different views based on finding the minimal projection distance to the manifolds. The experimental results and some comparisons show the new method is superior to manifold learning method and B-Spline based recognition method.

1 Introduction

Since human ear holds many similar characteristics as human face - in the head, a stable structure of appearance, ease of data acquisition, it is natural to consider applying ear recognition under occasions where face recognition has been to shown to work. So far, some works have shown ear owns some desirable properties for biometrics such as universality, uniqueness, and permanence [1]. Ear biometrics is expected to be a creditable technique for human identification, and it actually attracts much attention in recently. Compared with some popular biometrics such as face, fingerprint and iris, ear biometrics holds many excellences. Firstly, ear biometrics is not affected by facial expressions and cosmetics. And the appearance of the auricle (external ear) is nearly unaffected by aging (from 7 to 70). Secondly, ear data can be easily collected by a non-invasive way. In addition, on the ear itself, ear owns abundant and unique shape structure features and special pose information which make itself play a primary role in human profile image recognition. Therefore, in the field

D.-S. Huang et al. (Eds.): ICIC 2009, LNAI 5755, pp. 1085–1095, 2009.

of human appearance-oriented biometrics, ear recognition may work as a useful and necessary biometrics means.

Recent years there have appeared some biometrics techniques [2-5] which utilize 2D intensity images for ear recognition. D.J. Hurley et al. [2] apply force field energy function to acquire 2D ear features. Michał Choraś [3] has tried to extract abundant ear geometrical features from ear edge images. M. Burge and W. Burger [4] constructs Voronoi neighborhood graph model for ear curves matching. Bustard et al. [5] utilize SIFT feature points detection and matching for 2D ear registration and recognition. Moreno et al. [6] apply three kinds of neural networks to do 2D ear recognition experiments. K. Chang et al. [7] combine and compare ear and face images in appearance based biometrics.

Actually, all above mentioned works only use front view images for ear recognition and do not address ear multiple views recognition problems. As we know, the imaging of ear shape is always affected by the rotation of human head. When subject's ear is captured in different views between training samples and test samples, these front view based methods will result in matching failure. On the other hand, Ping Yan et al. [8] and Hui Chen et al. [9] propose 3D ear recognition techniques respectively. Although they both report good 3D recognition performance respectively, the cost of 3D ear data capturing and the time consuming for the whole recognition procedure are very high, which hints that the 3D approach will be inefficient in practice. That is also to say, the current 3D ear recognition approach has limited capability of solving the problem of ear pose variation.

However, among these existing 2D front view and 3D range image recognition methods, there is a margin way: multi-view ear recognition, which not only extends 2D ear front view method but also has partly similar traits of 3D range technique. Multiple views of ear carry abundant shape features in different sides which can boost up recognition performance. Naturally, we turn to explore the alternative way multi-view recognition against ear pose variation. In very recently, Zhao-xia Xie et al. [10] try to apply the LLE manifold learning method for multi-view ear recognition. Due to the requirement of large learning samples and noise sensitivity, the practical recognition performance of the method is not good. In addition, Zhiyuan Zhang et al. [11] propose a B-Spline based pose interpolation strategy for multi-view ear recognition. They report good performance using pose interpolation techniques.

As moving least square (MLS) has better numerical approximation traits and with weight function MLS is more flexible not only for global fit but also local approximation than B-Spline does, thus in this work, we consider to take moving least square technique to interpolate multiple ear poses in discriminative feature projection space formed by null space kernel discriminate analysis (NKDA), and then propose the corresponding multi-view ear recognition approach.

The rest of this paper is organized as follows. Section 2 illustrates the basic motivation and consideration of our multi-view ear pose interpolation based recognition strategy. And the multi-view ear dataset used in the work is introduced in the Section also. In Section 3, the concrete processing of NKDA multi-view ear feature extraction and moving least square pose interpolation, and the corresponding recognition approach are presented. Section 4 provides our recognition experiments, and some comparisons with B-Spline based interpolation and manifold learning recognition methods. Finally, we make short conclusions in Section 5.

2 The Framework of MLS Pose Interpolation Based Recognition

2.1 Recognition Framework

According to human vision system (HVS), human vision can capture the differences of different ears and even can take advantage of the trace continuity when ear pose varies. In addition, recent works [12-14] on face recognition across large pose changing have shown that face varieties of subjects, with lighting fixed, can be regarded as a 2D manifold which contain and represent face pose variation and facial expression transformation. Stimulated by these thinking, since ear do not hold expression changing, we may get the hints: ear pose varying will form a one dimension manifold - curve in certain feature space; and different ears with pose changing lie on different manifolds. The illustration of these suggestions is shown in Fig.1.

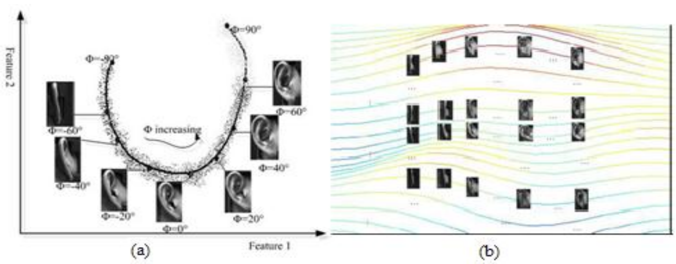

Fig. 1. Ear pose manifold characteristics illustration. (a) Ear pose trajectory in feature space. (b) Different ears lying on different pose manifolds.

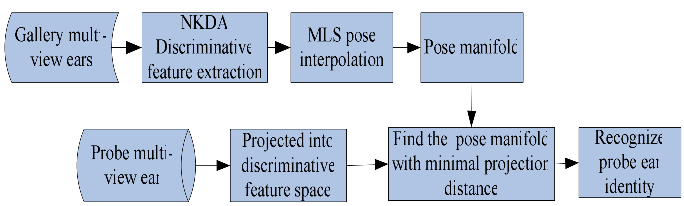

Fig. 2. Block diagram of the proposed MLS pose interpolation based multi-view ear recognition approach

From these hints and the figure, we can deduct that if we can use pose interpolation technique to construct the complete pose manifold, multi-view ear recognition can be solved. Based on this, our multi-view ear recognition approach can be formed with two main parts: NKDA feature extraction and moving least square pose interpolation. Fig.2 shows the framework of our MLS pose interpolation based recognition approach.

Considering if multiple pose interpolation can be built upon certain discriminative feature space, the good multi-view ear recognition performance could be expected. Thus following such thingking , we take NKDA method to extract discriminative features from gallery multi-view ears and thus construct the discriminative feature space for pose interpolation. Obviously, every ear training sample can be projected onto this feature space at different locations. Then we can adopt MLS interpolation method to produce new multiple poses of certain projected training ear. In this way, since the good virtue of MLS, new poses of each subject ear can be accurately and smoothly obtained which are denoted by new projected points in the discriminative feature space. Once pose interpolation for all subject ears is completed, the pose manifold of every subject ear is constructed. In test stage, one probe ear is projected onto the discriminative feature space firstly. Then, we calculate the project distance of probe point to all existing MLS pose manifolds. And the subject of certain pose manifold which has the minimal distance to the probe point in this space is recognized as its identity.

2.2 Multi-view Ear Dataset

Our multi-view ear dataset are obtained by using high-resolution camera capturing with moving on a half-circle orbit with angle indication. In our case, we capture ear eight views ($-60^o, -50^o, -40^o, -20^o, 0^o, +20^o, +40^o, +60^o$) in two sessions (one month interval) for multi-view ear recognition. Totally, we sampled 60 individuals' right ear and led to acquire 480 sampled images (eight views per individual). The pixel size of every sampled image is 1280×1024. The original sampled images were segmented in semi-supervised way and saved as multi-view ear data set. Thus, there are 60 ear individuals, and every ear individual has eight multiple views ($-60^o, -50^o, -40^o, -20^o, 0^o, +20^o, +40^o, +60^o$) in our data set. The size of each final image is 134×255. For example, Fig. 3 shows ten ear individual different views and the views of each individual turn from -60^o into 60^o gradually from left to right. From Fig. 3, we see, multi-views ear preserve abundant shape features in different pose.

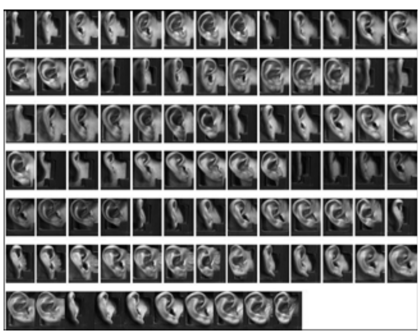

Fig. 3. Ten ears' multiple pose samples

3 Methodology

Multi-view ear location regression is analyzed firstly. Then null space kernel discriminate analysis (NKDA) is used to extract multiple pose non-linear features.

Finally, MLS is utilized to interpolate pose for constructing pose manifold explicitly in NKDA discriminative feature space for multi-view ear recognition.

3.1 Multi-view Ear Pose Location Regression Analysis

As far as ear is concerned, without expression (different from the face), the variability is simpler – only pose variation can take place. In our case, ear multiple pose images are acquired by rotating camera horizontally, which means there is only one freedom degree in motion, i.e., our pose variation is one dimension intrinsic manifold.

Suppose multi-view ear samples have been projected to certain feature space of which the dimension is d , x_i is one arbitrary projected pose sample in this space, e.g., $x_i \in IR^d$, and y_i is the true pose location of this sample in the feature space. From a statistical learning point of view, ear pose variation and the corresponding formed manifold (one dimension curve) can be regarded as one dimension pose location regression problem, i.e., finding one pose location regression function with constrains on smoothness and low variance to minimize the penalized residual sum of squares (RSS) via least squares method:

$$\arg\min_{f} RSS(f,\lambda) = \sum_{i=1}^{N} \left(y_i - f(x_i)\right)^2 + \lambda \int \left(f''(t)\right)^2 dt \tag{1}$$

where $f(x)$ denotes pose location regression function, N is the number of samples. The penalty term $\int \left(f''(t)\right)^2 dt$, weighted by smoothing parameter λ , forces smoothness on the function whereas the data fidelity measure term $\sum_{i=1}^{N} \left(y_i - f(x_i)\right)^2$ would make it rough so as to mimic the data. Remarkably, it has been shown that (1) has an explicit unique solution which is a natural cubic spline with knots at each $x_i, i = 1 \ldots N$ [15]. This approach is also called "spline smoothing" procedure.

With certain basis expansion, the regression pose location function $f(x)$ can be denoted by

$$f(x_i) = \sum_{m=1}^{M} \alpha_m p_m(x_i) = p^T(x_i)\alpha(x_i) \tag{2}$$

where $f(x_i)$ is a pose location approximation function for pose x_i . $p(x) = [p_1(x), p_2(x), \cdots, p_M(x)]^T$ is the k order complete monomial basis function, $\alpha(x) = [\alpha_1(x), \alpha_2(x), \cdots, \alpha_M(x)]^T$ is the coefficient of the basis function or parameter of the regression model, and M is the number of basis functions. For example, for two variables linear and quadratic monomial basis functions are, respectively,

$$p(x) = [1, x, y]^T, M = 3$$
$$p(x) = [1, x, y, x^2, xy, y^2]^T, M = 6 \tag{3}$$

Obviously, the work [11] takes the cubic spline, i.e. B-Spline, as pose location approximation basis function. However, for B-Spline, the number of support knots of the basis function and the calculation form of coefficient $\alpha(x)$ are fixed, this may lead pose approximation or interpolation not to be good. Thus, considering the coefficient $\alpha(x)$ are determined by a weighted least squares method minimizing the error $J(\alpha)$ between the experimental and approximated values of the objective function

$$J(\alpha) = \sum_{i=1}^{N} w(\|x_i - x\|)(p^T(x_i - x)\alpha - f(x_i))^2 \tag{4}$$

where N is the number of performed experiments and x_i is the experimental designs. The weights w_i insure the continuity and the locality of the approximation and are defined $w_i > 0$, decreasing within a fixed region around the point i called domain of support of x_i and vanish outside.

Min (J) gives

$$\alpha(x) = A^{-1} B f(x) \tag{5}$$

where

$$A(x) = \sum_{i=1}^{N} w(x - x_i) p(x_i) p(x_i)^T, B(x) = [w(x - x_1)p(x_1), \cdots, w(x - x_N)p(x_N)],$$

finally, the approximate pose location function can be gotten as

$$f^h(x) = p^T(x) A^{-1}(x) B(x) \cdot f(x) \tag{6}$$

The more detail about MLS approximation procedure can be found in [16]. The weight function $w(x)$ plays a crucial role by influencing the way that the coefficients α depend on the location of the design point x. The precision of MLS approximation or interpolation in a large extent depends on weight function chosen.

Thus, in this way, we choose proper weight function and basis function to do MLS multi-view ear pose interpolation.

3.2 NKDA Multi-view Ear Discriminative Feature Extraction

Since multi-view ears are severely non-linear, in this work, null space kernel discriminate analysis (NKDA) is adopted to extract the non-linear ear discriminative features across multiple views. The details of NKDA can be found in [17]. We make a short introduction on the basic rules to derivate NKDA. Suppose we have a set of

$n\ d-$ dimension samples $x_1, x_2, \ldots x_n$ belonging to c classes of ear. The optimal projection direction W will be obtained from the following:

$$W = \arg\max_{W} \frac{W^T S_B W}{W^T S_W W} \tag{7}$$

where S_W is within-class scatter matrix and S_B is between-class scatter matrix. This is the basic rule of LDA. It will be easily proved that if S_W is a non-singular matrix then the optimal projection vectors W are the eigenvectors of $S_W^{-1} S_B$. Unfortunately, in small sample size problem in which the numbers of samples is much smaller than the dimension of samples space which makes S_W is always singular. Seeing that, one substitute- null space based LDA (NLDA) [18] was proposed. In this kind of method, the optimal projection W should satisfy:

$$W^T S_W W = 0, W^T S_B W = I \tag{8}$$

This equation means that the optimal discriminatory vectors must exist in the null space of S_W. Based on kernel function, NLDA can be extended to NKDA which will hold stronger potential both on non-linear feature extraction and class discriminating. Assuming the number of samples is N, the number of total classes is c, the dimension of input space is , the kernel function is $K(x, y)$, and the output non-linear dimension reduction mapping is Γ, the main step of NKDA can be summarized as:

Step1. Computing Kennel mapping matrix on every training sample: $K(x_i, x_j), i = 1 \ldots N, j = 1 \ldots N$.

Step2. Calculating every class mean and within-class scatter matrix: $m_j = \sum_{i \in C_j} K(x_i) / N_j$, $K_w = \sum_{j=1}^{c} \sum_{i \in C_j} (K(x_i) - m_j)(K(x_i) - m_j)^T$.

Step3. Extract the null space P of K_w such that $P^T K_w P = 0$. P is usually in $(N-1) \times (c-1)$.

Step4. The output mapping on the sample set is: $\Gamma(x) = (P^T K) \cdot (x) = P^T \cdot K(x)$.

We apply NKDA to extract non-linear features of multi-view ears. There are three popular kernel functions, and the RBF (Gauss) kernel is adopted in our experiments as:

$$K(x, y) = (\phi(x) \cdot \phi(y)) = \exp\left(\frac{-\|x - y\|^2}{\sigma^2} \right) \tag{9}$$

where σ is set to1000 in our experiments.

3.3 MLS Pose Interpolation Based Multi-view Ear Recognition

In most cases, existent ear pose are not consecutive and they need to be interpolated to form a continuous and smooth pose curve (see the illustration in Fig.1). Among all forms of numerical interpolation, thanks to MLS good characteristics, we use MLS method for pose interpolation to construct pose manifold. As we have discussed, in MLS approximation, if we take one order basis function with cubic spline weight function, MLS pose interpolation equals B-Spline based pose interpolation. In our approach, we take two order monomial basis function and gauss weight function for MLS pose interpolation. The gauss weight function is taken as:

$$w_i(x)=\begin{cases}\dfrac{e^{-r^2\beta^2}-e^{-\beta^2}}{1-e^{-\beta^2}} & 0\le r\le 1\\ 0 & r>1\end{cases} \tag{10}$$

where $r=d_I/d_{mI}$, $d_I=\|x-x_I\|$ is the distance from the point x_I to its' neighbor point x , $d_{mI}=\kappa\times c_I$ is the radius of influencing region at the knot point x_I , κ is the influencing region radius factor, greater than one), β is called weight factors. c_I represents the density feature of points distribution around the knot point x_I , β reflects the contribution degree for the neighbor points around the knot point x_I to get the weights. In our case, we find, as for ears' pose variation in $\left[-40^\circ,+40^\circ\right]$, if c_I is set to be the distance from x_I to its' closest neighbor with $\beta=2$, and for other ear poses, if c_I is set to be the distance from x_I to the second closest neighbor distance with $\beta=4$, the multi-view recognition performance is better

Then after pose interpolation, the category of unknown ear can be determined by ransacking all constructed ears pose manifolds. Suppose the pose manifold C_p for each individual ear p is known, then the recognition of ear multiple pose can be formally defined as: for a probe ear view I , the identity p^* can be determined by ransacking all pose manifolds to find the manifold C_p with minimal "distance" to I , i.e.,

$$p^*=\arg\min_p d_f\left(I,C_p\right) \tag{11}$$

where d_f represents L^2 -Hausdorff distance between image I and C_p in feature space. Let $x\in C_p$ denotes a point on a manifold C_p and x^* is the point on C_p which is closest to I (i.e. at minimal L^2 distance), then $d_f\left(I,C_p\right)=d\left(I,x^*\right)$ where $d(\cdot,\cdot)$ denotes the corresponding L^2 distance. This procedure can be illustrated in Fig. 4.

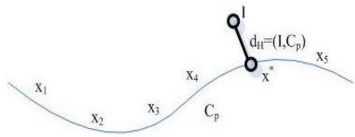

Fig. 4. Constructed pose manifold for multi-view ear recognition

4 Experiments and Comparisons

Our experiments are done based on our multi-view ear data set and ear samples are chosen randomly. In this data set, every ear has 8 multi-view samples. Firstly, we apply MLS pose interpolation technique in NKDA feature extraction space to construct pose manifold for multi-view ear identification. In order to make compare the performance with the work [11], we adopt the same training and test strategy, i.e., five views ($-60^o,-50^o,-20^o,+20^o,+60^o$) are treated as training samples and the left views ($-40^o,+0^o,+40^o$) are taken as probe samples. Since using Gauss weighted function for MLS pose interpolation, we show our method recognition performance with different parameters in Tabel.1.

Table 1. MLS pose interpolation based multi-view ear recognition with different parameters. (Adaptive c_I and β mean in pose range [-40°, $+40^\circ$] c_I takes the 1st closest point distance with $\beta = 4$, elsewhere c_I takes the 2rd closest point distance with $\beta = 2$)

Parameters	Object numbers	Recognition rate
c_I takes 1st, β=4	30, 60	91.7%, 96.7%
c_I takes 2rd, β=2	30, 60	90.0%, 95.0%
Adaptive c_I and β	30, 60	93.3%, 98.3%

For performance comparisons, we adopt different methods including manifold learning method [19-20], B-Spline based method, and our MLS pose interpolation based method for 60 subjects multi-view ears recognition. All experimental conditions are consistent to the previous experiments. The contrast results are shown in Table.2.

Table 2. Recognition contrast with manifold learning and B-Spline based methods

Methods	Dimension embedded	Neighbors	Recognition rate
LLE	22	35	65.0%
LPP	59	18	81.7%
B-Spline	59	3	96.7%
MLS pose interpolation based	59	Adaptive	98.3%

In addition, MLS pose interpolation based method does not need at least four training pose samples or knots, which is very different from B-Spline based recognition method. If we adjust d_{mI} and β of Gauss weight function to proper values for MLS interpolation, we can need the less available training views for multi-view ear recognition. Table.3 shows such experimental results with 60 subjects' ears. In the experiments, the left views excluding training views are taken as test views.

Table 3. Recogniton performance when taking different available training views

Available training views	Methods	Recognition rate
$(-50^\circ, -40^\circ, 0^\circ, +40^\circ)$	MLS interpolation based	85%
	B-Spline based	70.0%
$(-40^\circ, 0^\circ, +40^\circ)$	MLS interpolation based	73.3%
	B-Spline based	Unable to work

5 Conclusions

In this work, we investigate multi-view ear recognition methods beyond existing work in 2D ear biometrics. We take NKDA to extract multi-view ear non-linear features and do MLS pose interpolation in discriminative feature space to construct pose manifold. Adopting gauss weight function for MLS, we can get a 98.3% rank-one recognition rate against large pose variations in our multi-view ear data set. Experiments and comparisons have shown our MLS pose interpolation based method is very efficient for solving multi-view ear recognition problem. In the future, we will pay more attention on applying our MLS pose interpolation based method for multi-view face recognition.

Acknowledgement

This work is partially support by the Key Research Foundation of the Education Bureau of Sichuan Province (Grant no. 08ZA013), the CERG fund from the HKSAR Government, the central fund from Hong Kong Polytechnic University, and the Scientific Research Starting Foundation for Doctoral Scientist of Southwest University of Science and Technology.

References

1. Iannarelli, A.: Ear Identification. Paramont Publishing Company (1989)
2. Hurley, D., Nixon, M., Carter, J.: Force Field Feature Extraction for Ear Biometrics. Computer Vision and Image Understanding 98, 491–512 (2005)
3. Choras, M.: Ear Biometrics Based on Geometric Feature Extraction. Electronic Letters on Computer Vision and Image Analysis 5(3), 84–95 (2005)

4. Burge, M., Burger, W.: Ear Biometrics in Computer Vision. In: Proceedings of 15th International Conf. of Pattern Recognition, vol. 2, pp. 822–826 (2000)
5. Bustard, J.D., Nixon, M.: Robust 2D Ear Registration and Recognition Based on SIFT Point Matching. In: Proceedings of 2nd IEEE International Conference on Biometrics: Theory, Applications and Systems, pp. 1–6 (2008)
6. Moreno, B., Sanchez, A., Velez, J.: On the Use of Outer Ear Images for Personal Identification in Security Applications. In: Proceedings of IEEE Int'l Carnaham Conf. Security Technology, pp. 469–476 (1999)
7. Chang, K., Bowyer, K., Sarkar, S., Victor, B.: Comparison and Combination of Ear and Face Images in Appearance-based Biometrics. IEEE Trans. Pattern Analysis and Machine Intelligence 25, 1160–1165 (2003)
8. Yan, P., Bowyer, K.W.: Biometric Recognition Using 3D Ear Shape. IEEE Trans. Pattern Analysis and Machine Intelligence 29, 1297–1308 (2007)
9. Chen, H., Bhanu, B.: Human Ear Recognition in 3D. IEEE Trans. Pattern Analysis and Machine Intelligence 29, 718–737 (2007)
10. Xie, Z.X., Mu, Z.C.: Improved Locally Linear Embedding and Its Application on Multi-Pose Ear Recognition. In: Proceedings of 2007 International Conference on Wavelet Analysis and Pattern Recognition, pp. 1367–1371 (2007)
11. Zhang, Z.Y., Liu, H.: Multi-view Ear Recognition Based On B-Spline Pose Manifold Construction. In: Proceedings of the 7th World Congress on Intelligent Control and Automation, Chongqing, China, pp. 2416–2421 (2008)
12. Prince, S.J.S., Elder, J.H.: Tied Factor Analysis for Face Recognition Across Large Pose Changes. In: Proc. British Machine Vision Conference, vol. 3, pp. 889–898 (2006)
13. Li, Y.M., Gong, S.G., Liddell, H.: Constructing Facial Identity Surfaces for Recognition. International Journal of Computer Vision 53, 71–92 (2003)
14. Lee, K.C., Ho, J., Yang, M.H., Kriegman, D.: Video-Based Face Recognition Using Probabilistic Appearance Manifolds. In: Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. I, pp. 313–320 (2003)
15. Green, P., Silverman, B.: Nonparametric Regression and Generalized Linear Models. Chapman and Hall, Glasgow (1994)
16. Lancaster, P., Salkauskas, K.: Surfaces Generated by Moving Least Squares Methods. Mathematics of Computation 37, 141–158 (1981)
17. Liu, W., Wang, Y.H., Li, S.Z., Tan, T.N.: Null Space-based Kernel Fisher Discriminate Analysis for Face Recognition. In: Proc. Sixth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 369–374 (2004)
18. Chen, L.F., Liao, H.Y.M., Lin, J.C., Ko, M.T., Yu, G.J.: A New Debased Face Recognition System Which Can Solve the Small Sample Size Problem. Pattern Recognition 33, 1713–1726 (2000)
19. Roweis, S.T., Saul, L.K.: Nonlinear Dimensionality Reduction by Locally Linear Embedding. Science 290 (2000)
20. He, X.F., Yan, S.C., Hu, Y.X., Niyogi, P., Zhang, H.J.: Face Recognition Using Laplacianfaces. IEEE Trans. Pattern Analysis and Machine Intelligence 27, 328–340 (2005)