Performance Evaluation and Prediction for 3D Ear Recognition

Hui Chen, Bir Bhanu, and Rong Wang

Center for Research in Intelligent Systems
University of California, Riverside, California 92521, USA
{hchen,bhanu,rwang}@vislab.ucr.edu

Abstract. Existing ear recognition approaches do not give theoretical or experimental performance prediction. Therefore, the discriminating power of ear biometric for human identification cannot be evaluated. This paper addresses two interrelated problems: (a) proposes an integrated local descriptor for representation to recognize human ears in 3D. Comparing local surface descriptors between a test and a model image, an initial correspondence of local surface patches is established and then filtered using simple geometric constraints. The performance of the proposed ear recognition system is evaluated on a real range image database of 52 subjects. (b) A binomial model is also presented to predict the ear recognition performance. Match and non-matched distances obtained from the database of 52 subjects are used to estimate the distributions. By modeling cumulative match characteristic (CMC) curve as a binomial distribution, the ear recognition performance can be predicted on a larger gallery.

1 Introduction

Ear is a viable new class of biometrics since the ear has desirable properties such as universality, uniqueness and permanence [1]. For example, ear is rich in features; it is a stable structure which does not change with the age; it doesn't change its shape with facial expressions, cosmetics and hair styles. Although it has above advantages over other biometrics, it has received little attention compared to other popular biometrics such as face, fingerprint and gait.

In recent years, some approaches have been developed for the ear recognition from 2D images. Burge and Burger [2] proposed an adjacency graph, which is built from the Voronoi diagram of the ear's edge segments, to describe the ear. Ear recognition is done by subgraph matching. Hurley et al. [3] applied force field transform to ear images and wells and channels are shown to be invariant to affine transformations. Chang et al. [4] used principal component analysis (PCA) to ear images. All of these works for ear recognition have used 2D intensity images and therefore the performance of their systems is greatly affected by imaging conditions such as lighting and shadow. However currently available range sensors can directly provide us 3D geometric information which is insensitive to above imaging problems. Therefore, it is desirable to design a human ear recognition system from 3D side face images obtained at a distance. In fact, different methods to design biometrics system based on 3D data have been addressed [5–10].

T. Kanade, A. Jain, and N.K. Ratha (Eds.): AVBPA 2005, LNCS 3546, pp. 748–757, 2005.

However, no existing ear recognition approaches gives theoretical or experimental performance prediction. Evaluation and prediction of the performance of biometrics system to identify individuals is always considered in real world applications. Researchers have built mathematical models to evaluate and predict the performance on biometrics such as face, fingerprint, iris and gait. Bhanu et al. [11] develop a binomial model to predict fingerprint recognition performance. Tan et al. [12] present a two-point model and a three-point model to estimate the error rate for the point based fingerprint recognition. Johnson et al. [13] build a CMC model that is based on the feature space to predict the gait identification performance. Wayman [14] derives equations for the general biometric identification system. Daugman [15] analyzes the statistical variability of iris recognition using a binomial model. Johnson et al. [16] model a CMC curve to estimate recognition performance for larger galleries. Grother et al. [17] introduce the joint density function of the match and non-match scores to predict open- and closed-set identification performance.

In this paper, we first introduce an integrated local surface descriptor for 3D ear representation. A local surface descriptor is defined by a centroid, its surface type and 2D histogram. The 2D histogram consists of shape indexes, calculated from principal curvatures, and angles between the normal of reference point and that of its neighbors. The local surface descriptors are calculated only for the feature points which are defined as the local minimum and maximum of shape indexes. By comparison of local surface descriptors between a test and a model image, correspondences of local surface patches are established and then filtered using simple geometric constraints. The initial transformation is estimated based on the corresponding surface patches and applied to randomly selected locations of model ears. Iterative closest point (ICP) algorithm [18] iteratively refines the transformation to bring model ears and test ear into best alignment. The root mean square (RMS) registration error is used as the matching error criterion.

Next, a binomial model is presented to predict the proposed ear recognition performance. We calculate the RMS registration errors between 3D ears in the probe set with 3D ears in the gallery. RMS errors are used as matching distances to estimate the distribution of match and non-match distances. Then the *cumulative match characteristic* (CMC) curve is modeled by a binomial distribution and the probability that the match score is within rank r can be calculated. Using this model we can predict ear recognition performance for a large gallery.

2 Technical Approach

2.1 Feature Extraction

In our approach, feature points are defined as local minimum and maximum of shape indexes, which can be calculated from principal curvatures. In order to estimate curvatures, we fit a quadratic surface $f(x, y) = ax^2 + by^2 + cxy + dx + ey + f$ to a local window and use the least square method to estimate the parameters of the quadratic surface, and then use differential geometry to calculate the surface normal, Gaussian and mean curvatures and principal curvatures [19].

Shape index (S_i), a quantitative measure of the shape of a surface at a point p, is defined by (1) where k_1 and k_2 are maximum and minimum principal curvatures respectively. With this definition, all shapes are mapped into the interval $[0, 1]$ [20].

$$S_i(p) = \frac{1}{2} - \frac{1}{\pi} tan^{-1} \frac{k_1(p) + k_2(p)}{k_1(p) - k_2(p)} \tag{1}$$

Within a $w \times w$ window, the center point is marked as a feature point if its shape index is higher or lower than those of its neighbors.

2.2 Local Surface Patches

We define a "local surface patch" as the region consisting of a feature point P and its neighbors N. The neighbors should satisfy these two conditions,

$$\begin{aligned} N = \{pixels \quad & N, ||N - P|| \leq \epsilon_1\} \\ & and\ acos(n_p \bullet n_n < A), \end{aligned} \tag{2}$$

where n_p and n_n are the surface normal vectors at point P and N. The two parameters ϵ_1 and A are important since they determine how the local surface patch is resistant to clutter and occlusion. Johnson [21] discussed the choices for the two parameters. For every local surface patch, we compute the shape indexes and normal angles between point P and its neighbors. Then we can form a 2D histogram. One axis of this histogram is the shape index which is in the range [0,1]; the other is the dot product of surface normal vectors at P and N which is in the range [-1,1]. In order to reduce the effect of the noise, we use bilinear interpolation when we calculate the 2D histogram.

We also compute the centroid of local surface patches. We classify surface shape of the local surface patch into three types: concave, saddle and convex based on shape index value of the feature point. The shape index range and its corresponding surface type are listed in Table 1 [22]. Note that a feature point and the centroid of a patch may not coincide.

Table 1. Surface type Tp based on the shape index

Type tag (Tp)	S_i range	Surface type
0	[0,5/16)	Concave
1	[5/16,11/16)	Saddle
2	[11/16,1]	Convex

In summary, every local surface patch is described by a 2D histogram, surface type and the centroid. The local surface patch encodes the geometric information of a local surface.

2.3 Off-Line Model Building

Considering the uncertainty of location of a feature point, we repeat the above process to calculate descriptor of local surface patches for neighbors of feature point P and save these descriptions into the model database. For each model object, we repeat the same process to build the model database.

2.4 Surface Matching

Comparing Local Surface Patches. Given a test range image, we repeat the above steps and get local surface patches. Considering the inaccuracy of feature points' location, we also extract local surface patches from neighbors of feature points. Then we compare them with all of the local surface patches saved in the model database. This comparison is based on the surface type and histogram dissimilarity. Since histogram can be thought of as an approximation of probability distributed function, we use statistical method to assess the dissimilarity between two probability distributions. The $\chi^2 - divergence$ is among the most prominent divergence used in statistics to assess the dissimilarity between two probability density functions. We use it to measure the dissimilarity between two observed histograms Q and V, which is defined by (3) [23].

$$\chi^2(Q,V) = \sum_i \frac{(q_i - v_i)^2}{q_i + v_i} \tag{3}$$

Figure 1 shows an experimental validation that the local surface patch has the discriminative power to distinguish shapes. We do experiments under three cases. 1) a local surface patch (Lsp1) generated for an ear is compared to another local surface patch (Lsp2) corresponding to the same physical area of the same ear imaged at different viewpoints; a low dissimilarity exists. 2) The Lsp1 is compared to Lsp3 which lies in different area of the same ear; the dissimilarity is high. 3) The Lsp1 is compared to Lsp4 which lies in the similar area as the Lsp 1 but not the same ear, there exists a higher dissimilarity than the first case. The experimental results suggest that the local surface patch can be used for differentiation between ears. Table 2 lists the comparison results.

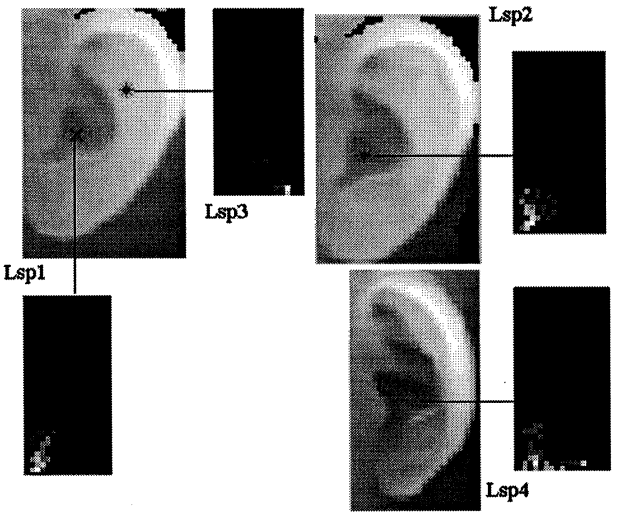


Fig. 1. Experimental validation of discriminatory power of Local Surface Patches

Table 2. Comparison results of local surface patches

Surface Type	Lsp1	Lsp2	Lsp3	Lsp4
	Tp=0	Tp=0	Tp=2	Tp=1
$\chi^2 - divergence$	$\chi^2(Lsp1, Lsp2)$	$\chi^2(Lsp1, Lsp3)$	$\chi^2(Lsp1, Lsp4)$	
	0.479	1.99	0.984	

Grouping Corresponding Pairs of Local Surface Patch. For every local surface patch from the test ear, we choose the local surface patch from the database with minimum dissimilarity and the same surface type as the possible corresponding patch. We filter the possible corresponding pairs based on the geometric constraints given below.

$$d_{C_1,C_2} = |d_{S_1,S_2} - d_{M_1,M_2}| < \epsilon_2, \tag{4}$$

where d_{S_1,S_2} and d_{M_1,M_2} are Euclidean distance between centroids of two surface patches. For two correspondences $C_1 = \{S_1, M_1\}$ and $C_2 = \{S_2, M_2\}$ where S means test surface patch and M means model surface patch, they should satisfy (4) if they are consistent corresponding pairs. Thus, we use geometric constraints to partition the potential corresponding pairs into different groups. The largest group would more likely to be the true corresponding pair.

Given a list of corresponding pairs $L = \{C_1, C_2, \ldots, C_n\}$, the grouping procedure for every pair in the list is as follows: Initialize each pair of a group. For every group, add other pairs to it if they satisfy (4). Repeat the same procedure for every group. Select the group which has the largest size.

Aligning Model Ears with Test Ears. We estimate the initial rigid transformation based on the corresponding local surface patches using quaternion representation [24]. Starting with the initial transformation, Iterative closest point (ICP) algorithm [18] is run to refine the transformation by minimizing the distance between the control points of the model ear and their closest points of the test ear. Since the ear is assumed to be in the center of the image, the control points are selected around the center of the image. Every time the control points are randomly selected from model ears and ICP is applied to those points. We repeat the same procedure several times and choose the minimum RMS error as the final result.

2.5 Prediction Model

The mathematical prediction model is based on the distribution of match and non-match scores [11]. We use $ms(x)$ and $ns(x)$ to denote the distributions of match and non-match scores. If the similarity score is higher, the match is closer. The error occurs when any given match score is less than any of the non-match scores. The probability that the non-match score is greater than or equal to the match score x is $NS(x)$ computed by (5).

$$NS(x) = \int_x^\infty ns(t)dt \tag{5}$$

The probability that the match score x has rank r exactly, is given by the binomial probability distribution:

$$C_{r-1}^{N-1}(1-NS(x))^{N-r}NS(x)^{r-1} \tag{6}$$

By integrating over all the match scores, we get

$$\int_{-\infty}^{\infty} C_{r-1}^{N-1}(1-NS(x))^{N-r}NS(x)^{r-1}ms(x)dx \tag{7}$$

In theory the match scores can be any value within $(-\infty, \infty)$. Therefore the probability that the match score is within rank r, which is definition of a CMC curve, is

$$P(N,r) = \sum_{i=1}^{r}\int_{-\infty}^{\infty} C_{i-1}^{N-1}(1-NS(x))^{N-i}NS(x)^{i-1}ms(x)dx \tag{8}$$

In above equations N is the size of large population whose performance needs to be estimated. Here we assume that the match score and non-match score are independent and the match and non-match score distributions are the same for all the persons. The small size gallery is used to estimate distributions of $ms(x)$ and $ns(x)$.

For the ear recognition case, every 3D ear in the probe set is matched to every 3D ear in the gallery and the RMS registration error is calculated using the procedure described in Section 2.4. The RMS registration error is used as matching error criterion. In our case, the matching error is smaller, the match is closer. In order to use the above prediction model, we modify the equations accordingly.

3 Experimental Results

3.1 Data Acquisition

We use real range data acquired using Minolta Vivid 300. The range image contains 200×200 grid points and each grid point has a 3D coordinate (x, y, z). There are 52 subjects in our database and every subject has two left side face range images taken at different viewpoints. The ears are manually extracted from side face images. The data is split into model and test sets. Each set has 52 ears. The extracted model ears and corresponding test ears are shown in Figure 2 and Figure 3 respectively.

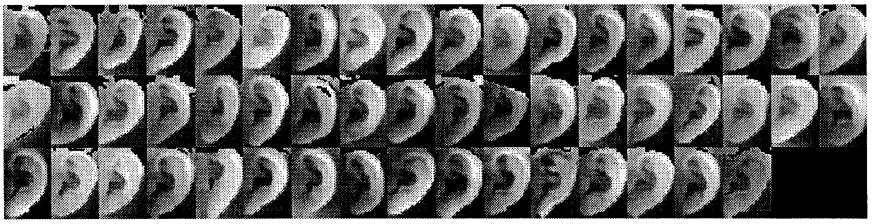


Fig. 2. 3D model ears shown as gray scale images

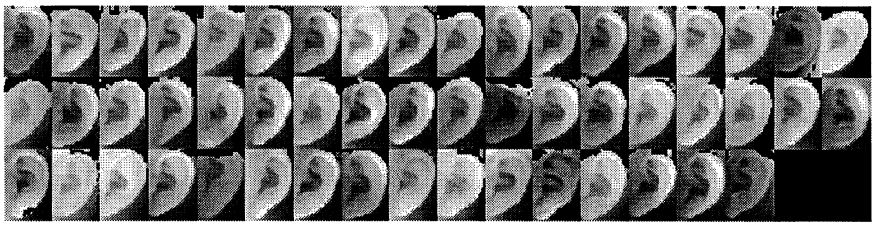
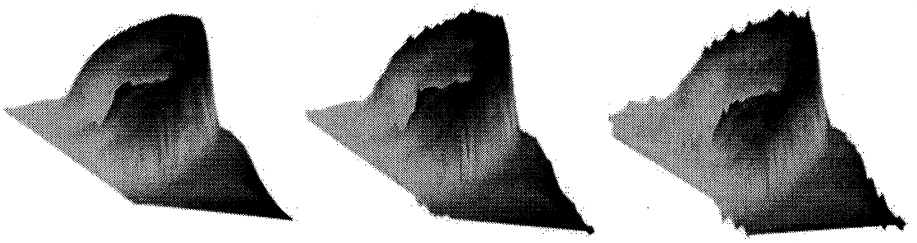


Fig. 3. 3D test ears shown as gray scale images



(a) No added noise (b) $\sigma = 0.6mm$ (c) $\sigma = 1.0mm$

Fig. 4. Test scans corrupted with Gaussian noise

3.2 Performance Evaluation

To test the proposed system's performance, we add Gaussian noise to the test scans along the viewing direction (Z-axis). The standard deviation of Gaussian noise we add depends on the mesh resolution of test scans. However the mesh resolution is not well defined. We use the Johnson's definition [21] "Mesh resolution is defined as the median of all edge lengths in a mesh". Given a test range image, we triangulate it and get a triangular mesh. Then we calculate the median of all edge lengths in the mesh. The average median calculated from test scans is about $1.25mm$. We add Gaussian noise with $\sigma = 0.6mm$ and $\sigma = 1.0mm$ to test scans. Therefore, we have three probe sets: one probe set has no added Gaussian noise; the second probe set has Gaussian noise $N(0, \sigma = 0.6mm)$; the third probe set has Gaussian noise $N(0, \sigma = 1.0mm)$. Examples of one test scan corrupted with Gaussian noise are shown in Figure 4.

The CMC curve is used to evaluate the system's recognition performance. The rank-1, rank-2 and rank-3 recognition rates for three probe sets are listed in Table 3. The verification performance results are given by the receiver operating characteristic (ROC)

Table 3. Recognition results for three probe sets

Probe set	Rank-1	Rank-2	Rank-3
No added noise	90.4%	96.2%	96.2%
$N(0, \sigma = 0.6mm)$	76.9%	86.5%	86.5%
$N(0, \sigma = 1.0mm)$	44.2%	61.5%	67.3%

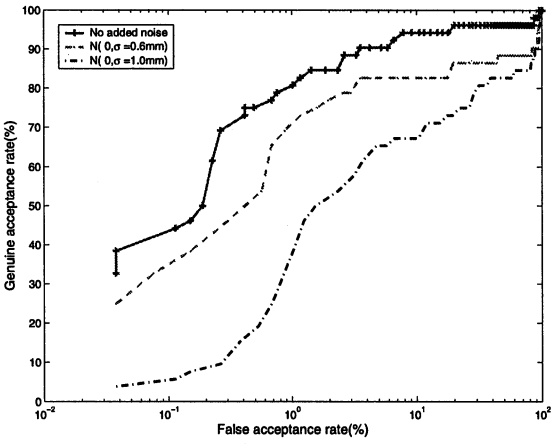


Fig. 5. ROC curve for three probe sets

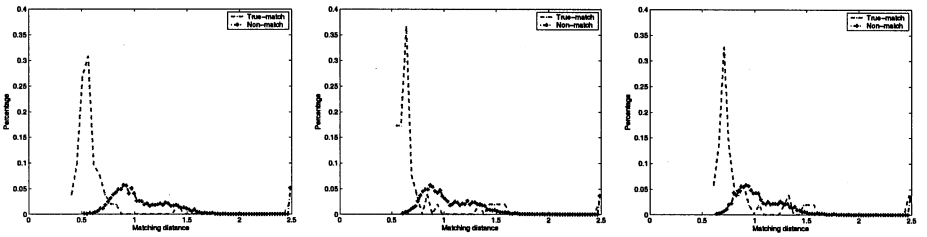
curve which is defined as the plot of genuine acceptance rate against false acceptance rate. Figure 5 shows the ROC curves for three probe sets. From Table 3 and Figure 5, we can see that the performance of the proposed system degrades as the scene noise increases. It's reasonable since Gaussian noise corrupts the surface normals and shape index resulting in the corruption of the local surface patch representation.

3.3 Prediction Results

Every 3D scan in three probe sets is matched to every 3D ear in the gallery and the RMS registration error is calculated using the procedure described in Section 2.4. The RMS registration error is used as the matching distance. Therefore, we obtain 52 true-match distances and 2652 non-match distances for every probe set. The matching distance distribution for true-match and non-match for three probe sets are shown in Figure 6. Based on the distributions, we can predict CMC curve $P(N, r)$ where $r = 1, 2, 3$ and N is 52. We also calculate the CMC curve based on the experimental results for three probe sets. The results of the directly calculated CMC curve and the predicted CMC curve are shown in Table 4. Table 4 shows that the predicted CMC values are close to the calculated CMC values, which demonstrates the effectiveness of our prediction model. We'd like to predict CMC values for larger galleries from the original range image database of 52 subjects. Table 5 shows the predicted CMC values for three probe

Table 4. Predicted and calculated CMC values for three probe sets on 52 subjects

Probe set	Rank-1		Rank-2		Rank-3	
	Predicted	Calculated	Predicted	Calculated	Predicted	Calculated
No added noise	92.5%	90.4%	94.6%	96.2%	95.7%	96.2%
$N(0, \sigma = 0.6mm)$	80.4%	76.9%	83.9%	86.5%	85.8%	86.5%
$N(0, \sigma = 1.0mm)$	51.5%	44.2%	60.2%	61.5%	66.1%	67.3%



(a) (b) (c)

Fig. 6. Matching distance distribution for match and non-match pairs for three probe sets: (a) without added noise, (b) with Gaussian noise $N(0, \sigma = 0.6mm)$, (c) with Gaussian noise $N(0, \sigma = 1.0mm)$

Table 5. Predicted CMC values for three probe sets for larger galleries (Recognition rate shown as percentage)

Probe set	N=100			N=200			N=300			N=400		
	R-1	R-2	R-3	R-1	R-2	R-3	R-1	R-2	R-3	R-1	R-2	R-3
No added noise	91.2	92.8	94.4	90.5	90.9	91.8	90.4	90.5	90.8	90.4	90.4	90.5
$N(0, \sigma = 0.6mm)$	78.3	80.9	83.6	77.1	77.9	79.3	76.9	77.1	77.6	76.9	76.9	77.1
$N(0, \sigma = 1.0mm)$	46.8	52.1	57.9	44.6	45.9	48.6	44.3	44.6	45.4	44.2	44.3	44.5

sets for different gallery size (N=100, 200, 300, 400). Confidence in prediction is of interest and we plan to work on it.

4 Conclusions

In this paper, we first propose an integrated local descriptor for representation to recognize human ears in 3D. We evaluate the proposed ear recognition performance by means of CMC and ROC curves on three different probe sets using a real range image database of 52 subjects. One probe set has no added Gaussian noise; the second probe set has Gaussian noise $N(0, \sigma = 0.6mm)$; the third probe set has Gaussian noise $N(0, \sigma = 1.0mm)$. We obtain rank-one recognition rate of 90.4% for test scans without added noise and the system's performance degrades as the scene noise increases. We also predict the ear recognition performance on larger galleries by modeling cumulative match characteristic curve as a binomial distribution. The predicted rank-one recognition rate is 90.4% on test scans without added noise for a database of 400 subjects. Table 5 demonstrates that we can predict the recognition performace for lager galleries.

References

1. Iannarelli, A.: Ear Identification. Forensic Identification Series. Paramont Publishing Company (1989)
2. Burge, M., Burger, W.: Ear biometrics in computer vision. Proc. Int. Conf. on Pattern Recognition **2** (2000) 822–826

3. Hurley, D., Nixon, M., Carter, J.: Automatic ear recognition by force field transformations. IEE Colloquium on Visual Biometrics (2000) 7/1 –7/5
4. Chang, K.C., Bowyer, K.W., Sarkar, S., Victor, B.: Comparison and combination of ear and face images in appearance-based biometrics. IEEE Trans. Pattern Analysis and Machine Intelligence **25** (2003) 1160–1165
5. Bhanu, B., Chen, H.: Human ear recognition in 3D. Workshop on Multimodal User Authentication (2003) 91–98
6. Bronstein, A., Bronstein, M., Kimmel, R.: Expression-invariant 3D face recognition. Audio and Video based Biometric Person Authentication (2003) 62–70
7. Chang, K.C., Bowyer, K.W., Flynn, P.J.: Multi-modal 2D and 3D biometrics for face recognition. IEEE Int. Workshop on Analysis and Modeling of Faces and Gestures (2003) 187–194
8. Chua, C.S., Han, F., Ho, Y.: 3D human face recognition using point signatures. Int. Conf. on Automatic Face and Gesture Recognition (2000) 233–238
9. Lee, J.C., Milios, E.: Matching range images of human faces. Proc. Int. Conf. on Computer Vision (1990) 722–726
10. Lu, X., Colbry, D., Jain, A.K.: Three-dimensional model based face recognition. Proc. Int. Conf. on Pattern Recognition **1** (2004) 362–366
11. Bhanu, B., Wang, R., Tan, X.: Predicting fingerprint recognition performance from a small gallery. ICPR Workshop on Biometrics: Challenges arising from Theory to Practice (2004) 47–50
12. Tan, X., Bhanu, B.: On the fundamental performance for fingerprint matching. Proc. IEEE Conf. Computer Vision and Pattern Recognition **2** (2003) 499–504
13. Johnson, A.Y., Sun, J., Boick, A.F.: Predicting large population data cumulative match characteristic performance from small population data. Audio and Video based Biometric Person Authentication (2003) 821–829
14. Wayman, J.L.: Error-rate equations for the general biometric system. IEEE Robotics & Automation Magazine **6** (1999) 35–48
15. Daugman, J.: The importance of being random: statistical principles of iris recognition. Pattern Recognition **36** (2003) 279–291
16. Johnson, A.Y., Sun, J., Boick, A.F.: Using similarity scores from a small gallery to estimate recognition performance for large galleries. IEEE Int. Workshop on Analysis and Modeling of Faces and Gestures (2003) 100–103
17. Grother, P., Phillips, P.J.: Models of large population recognition performance. Proc. IEEE Conf. Computer Vision and Pattern Recognition **2** (2004) 68–75
18. Besl, P., Mckay, N.D.: A method of registration of 3-D shapes. IEEE Trans. Pattern Analysis and Machine Intelligence **14** (1992) 239–256
19. Flynn, P., Jain, A.: On reliable curvature estimation. Proc. IEEE Conf. Computer Vision and Pattern Recognition (1989) 110–116
20. Dorai, C., Jain, A.: COSMOS-A representation scheme for free-form surfaces. Proc. Int. Conf. on Computer Vision (1995) 1024–1029
21. Johnson, A., Hebert, M.: Using spin images for efficient object recognition in cluttered 3D scenes. IEEE Trans. Pattern Analysis and Machine Intelligence **21** (1999) 433–449
22. Koenderink, J.J., Doorn, A.V.: Surface shape and curvature scales. Image Vision Computing **10** (1992) 557–565
23. Schiele, B., Crowley, J.: Recognition without correspondence using multidimensional receptive field histograms. International Journal of Computer Vision **36** (2000) 31–50
24. Horn, B.: Close-form solution of absolute orientation using unit quaternions. Journal of the Optical Society of America **4** (1987) 629–642