5

Recognizing 3D Ears Using Local Surface Patches

In this chapter, we use the local surface patch (LSP) representation for matching 3D ears [1, 12]. The LSP representation, a new local surface descriptor, is characterized by a centroid, a local surface type and a 2D histogram. The 2D histogram shows the frequency of occurrence of shape index values vs. the angles between the normal of reference feature point and that of its neighbors. The proposed human recognition system using the LSP representation is illustrated in Figure 5.1. The local surface descriptors are computed for the feature points which are defined as either the local minimum or the local maximum of shape indexes. By comparing the local surface patches for a gallery and a probe image, the potential corresponding local surface patches are established and then filtered by geometric constraints. Based on the filtered correspondences, the initial rigid transformation is estimated. Once this transformation is obtained, it is then applied to randomly selected control points of the hypothesized gallery ear in the database. A modified iterative closest point (ICP) algorithm is run to improve the transformation which brings a gallery ear and a probe ear into the best alignment, for every gallery-probe pair. The root mean square (RMS) registration error is used as the matching error criterion. The subject in the gallery with the minimum RMS error is declared as the recognized person in the probe image.

5.1 Local Surface Patch Representation (LSP)

In 3D object recognition, the key problems are how to represent free-form surfaces effectively and how to match the surfaces using the

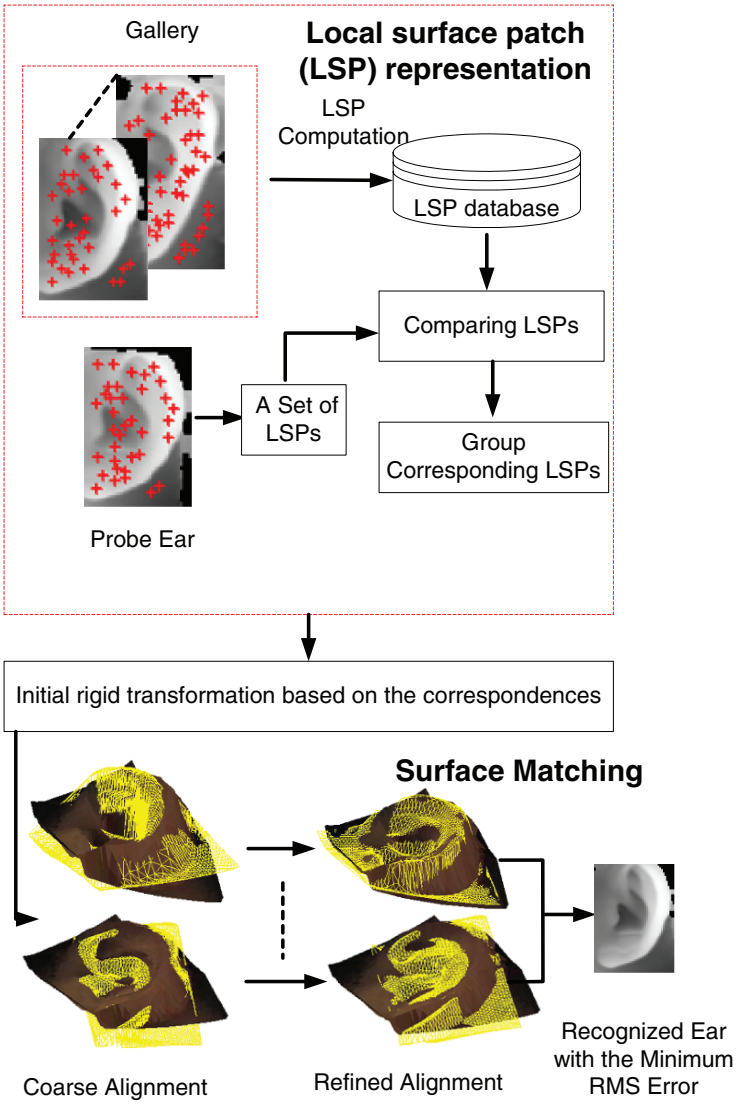


Fig. 5.1. The 3D ear recognition using the local surface patch (LSP) representation.

selected representation. Researchers have proposed various surface signatures for recognizing 3D free-form objects which are reviewed in [12, 51]. In the following, we present a new surface representation, called the local surface patch (LSP), investigate its properties and use it for ear recognition.

5.1.1 Definition of LSP

We define a "local surface patch" (LSP) as the region consisting of a feature point P and its neighbors N. The LSP representation includes feature point P, its surface type, centroid of the patch, and a histogram of shape index values vs. dot product of the surface normal at point P and its neighbors. A local surface patch is shown in Figure 5.2. The neighbors satisfy the following conditions,

$$\begin{aligned} N = \{&\text{pixels} \quad N, ||N - P|| \leq \epsilon_1\} \\ &\text{and acos}(n_p \bullet n_n < A), \end{aligned} \tag{5.1}$$

where $\bullet$ denotes the dot product between the surface normal vectors n_p and n_n at point P and N and acos denotes the inverse cosine function. The two parameters ϵ_1 and A (ϵ_1 = 5.8mm, A = 0.5) are important since they determine the descriptiveness of the local surface patch representation. A local surface patch is not computed at every pixel in a range image, but only at selected feature points.

The feature points are defined as the local minimum and the maximum of shape indexes, which can be calculated from principal curvatures. The curvatures can be estimated as described in Section 3.1.1.

Shape index (S_i), a quantitative measure of the shape of a surface at a point P, is defined by equation (5.2),

$$S_i(P) = \frac{1}{2} - \frac{1}{\pi}\tan^{-1}\frac{k_1(P) + k_2(P)}{k_1(P) - k_2(P)} \tag{5.2}$$

where k_1 and k_2 are maximum and minimum principal curvatures, respectively. With this definition, all shapes are mapped into the interval $[0, 1]$ [77]. The shape categories and corresponding shape index ranges are listed in Table 3.1. Figure 5.3 shows original ear range images and their shape index images for two people. In this figure, the brighter pixels denote large shape index values which correspond to ridge and

Table 5.1. Surface type Tp based on the value of 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

dome surfaces while the darker pixels denote small shape index values that correspond to valley and cup surfaces. Within a $b \times b$ ($b = 5$) window, the center point is marked as a feature point if its shape index is higher or lower than those of its neighbors. The results of feature points extraction are shown in Figure 5.4 where the feature points are marked by red plus sign. In order to see the feature points' location, we enlarge the two images. We can clearly see that some feature points corresponding to the same physical area appear in both images.

For every local surface patch, we compute the shape indexes and normal angles between point P and its neighbors. Then we form a 2D histogram by accumulating points in particular bins along the two axes. 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 noise, we use bilinear interpolation when we calculate the 2D histogram. One example of 2D histogram is shown as a grayscale image in Figure 5.2(c); the brighter areas in the image correspond to bins with more points falling into them. In the implementation, the number of bins for the shape index axis is 17 and the number of bins for the other axis is 34.

We classify surface shape of a local surface patch into three types: concave (Tp = 0), saddle (Tp = 1) and convex (Tp = 2) based on the shape index value of the feature point. The shape index range and its corresponding surface type are listed in Table 5.1. We also compute the centroid of a local surface patch. Note that a feature point and the centroid of a patch may not coincide.

In summary, every local surface patch is described by a 2D histogram, surface type, and the centroid. The 2D histogram and surface type are used for comparison of LSPs and the centroid is used for computing the rigid transformation. The patch encodes the geometric information of a local surface.

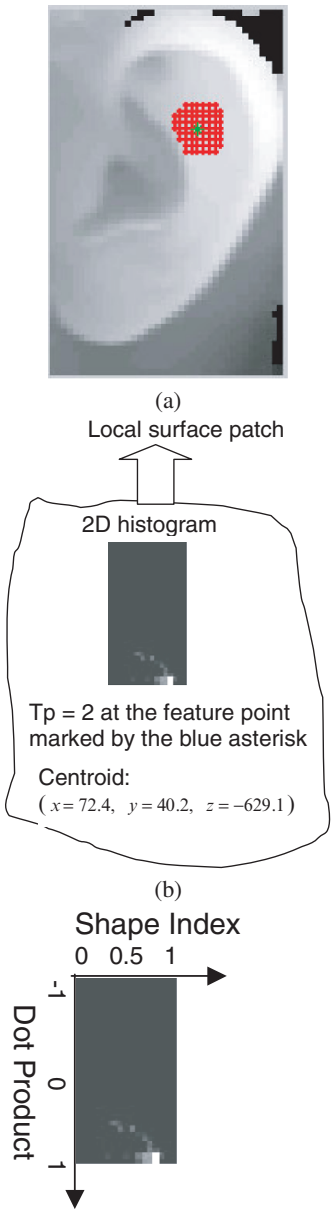


Fig. 5.2. Illustration of a local surface patch (LSP). (a) Feature point P is marked by the asterisk and its neighbors N are marked by the interconnected dots. (b) LSP representation includes a 2D histogram, a surface type and centroid coordinates. (c) The 2D histogram is shown as a gray image in which the brighter areas correspond to bins with the high frequency of occurrence.

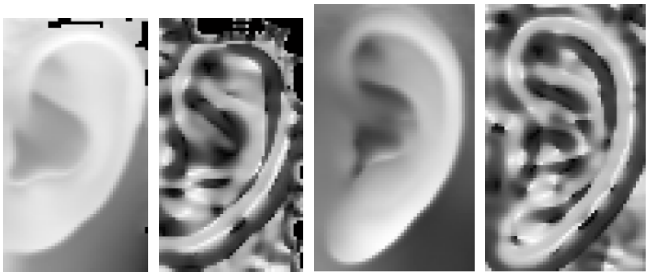


Fig. 5.3. Two examples of ear range images ((a), (c)) and their corresponding shape index images ((b), (d)). In images (a) and (c), the darker pixels are away from the camera and the lighter ones are closer. In images (b) and (d), the darker pixels correspond to concave surfaces and lighter ones correspond to convex surfaces.

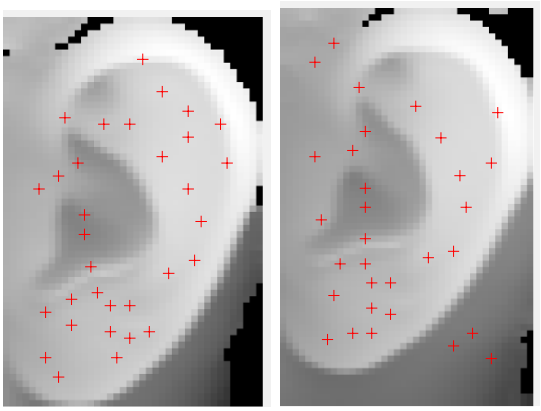


Fig. 5.4. Feature points location (+) in two range images shown as grayscale images of the same ear taken at different viewpoints.

5.1.2 Comparing Local Surface Patches

Given a probe range image, we extract feature points and get local surface patches. Considering the inaccuracy of feature points' location, we also extract local surface patches from the neighbors of feature points. Then we compare them with all of the local surface patches saved in the gallery based on the surface type and histogram dissimilarity. We use a statistical method to assess the dissimilarity between the two probability density functions since a histogram can be thought of as an unnormalized approximation to it. 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, as follows [62]:

$$\chi^2(Q,V) = \sum_i \frac{(q_i - v_i)^2}{q_i + v_i} \tag{5.3}$$

From equation (5.3), we know the dissimilarity is between 0 and 2. If the two histograms are exactly the same, the dissimilarity will be zero. If the two histograms do not overlap with each other, it will achieve the maximum value of 2.

Figure 5.5 shows an experimental validation that the local surface patch has the discriminative power to distinguish shapes. We do experiments under three cases:

- 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 from a different viewpoint; in this case a low dissimilarity exists and both LSPs have the same surface type.
- The LSP1 is compared to LSP3 which lies in a different area of the same ear; the dissimilarity is high and they have different surface type.
- The LSP1 is compared to LSP4 which lies in the similar area as the LSP1 but it is not the same ear; there exists a higher dissimilarity than the first case and they also have the different surface type.

These experimental results suggest that the local surface patch provides distinguishable features and it can be used for differentiation among ears. Table 5.2 shows the comparison results.

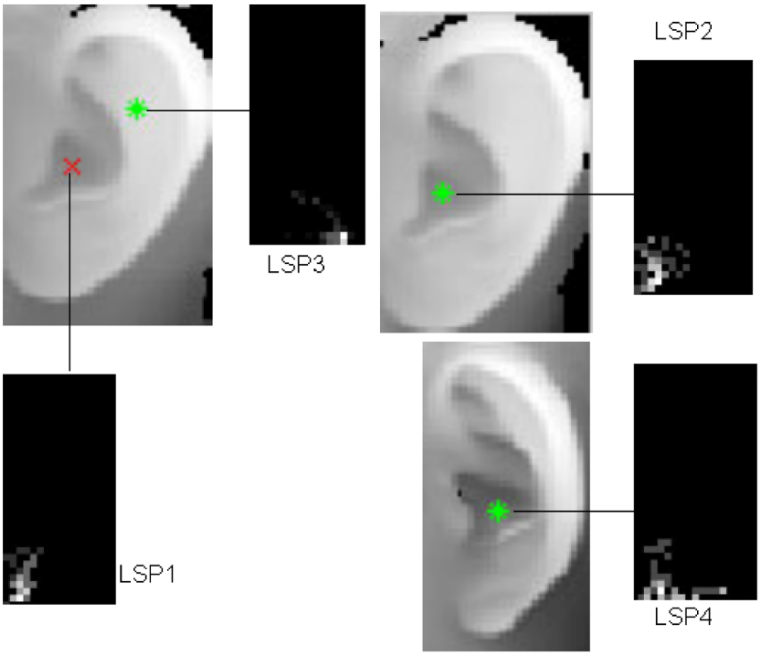


Fig. 5.5. The demonstration of discriminatory power of local surface patches. The ear images in the first row are from the same person but with different viewpoints. The ear image shown in the second row is from a different person. The histograms of four local surface patches (LSP1 to LSP4) are also shown for the comparison.

Table 5.2. Comparison results for four local surface patches shown in Figure 5.5.

Surface Type	LSP1	LSP2	LSP3	LSP4
	Tp=0	Tp=0	Tp=2	Tp=1

χ^2 – divergence	$\chi^2(LSP1, LSP2)$	$\chi^2(LSP1, LSP3)$	$\chi^2(LSP1, LSP4)$
	0.479	1.99	0.984

5.1.3 Generating Parameters of LSP

There are four parameters that control the generation of an LSP: the number of bins for the shape index, the number of bins for the dot product of surface normal vectors, the distance constraint ϵ_1, and the angle constraint A. Figure 5.6 and 5.7 show quantitative analysis of the effect of these four parameters. We compute χ^2 dissimilarity for corresponding and non-corresponding LSPs versus different generating parameters. Since the gallery/probe pair is registered, we can easily determine correspondence in the probe range image for every extracted feature point in the gallery range image. If we only consider the dissimilarity between corresponding LSPs, it is not sufficient to analyze the effect of generating parameters on comparing LSPs. We need to take into account the dissimilarity for both corresponding LSPs and non-corresponding LSPs to analyze the effect of generating parameters. In order to get non-corresponding points, we move their correspondences in the probe image to other locations to make them at least 15mm apart (the average length of ear is about 62mm). For every corresponding LSP, we compute the dissimilarity for corresponding LSPs and get the mean. By repeating the same procedure for different generating parameters, we create the plot of χ^2 dissimilarity for corresponding LSPs versus the varying parameter. We repeat the same procedure for non-corresponding LSPs and obtain the plot of χ^2 dissimilarity versus the varying parameter. We analyze the effect of each parameter while keeping other three fixed.

In Figure 5.6 and 5.7 there are three curves: two of them are plots of the mean of dissimilarity versus the varying parameter; the other one is the plot of the separability versus the varying parameter. Assuming the distributions for corresponding and non-corresponding LSPs are Gaussian, the separability is defined by $\frac{|\mu_1-\mu_2|}{\sqrt{(\sigma_1^2+\sigma_2^2)}}$, where μ_1, σ_1 and μ_2, σ_2 are the mean and standard deviation of dissimilarity for corresponding and non-corresponding LSPs respectively. The separability measures the distance between the dissimilarity distribution for corresponding and non-corresponding LSPs to some degree.

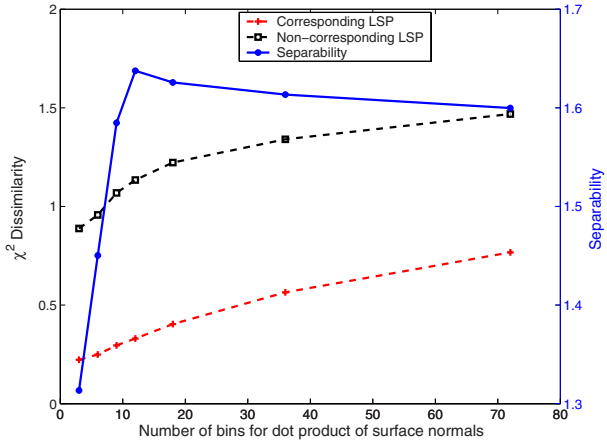
The *number of bins* is an important parameter since it controls the size of 2D histogram and it also affects the descriptiveness of LSPs. If the number of bins is small, the possibility of points falling into the same bin is large, and the χ^2 dissimilarity for corresponding and

non-corresponding LSPs will be small. As the number of bins increases, the χ^2 dissimilarity for corresponding and non-corresponding LSPs increases since the chances of points falling into different bins are higher, which results in the decrease in the overlap of histograms corresponding to two LSPs. Figure 5.6(a), (b) and 5.7(a), (b) verify the analysis. We can also observe that the separability increases with the increasing number of bins and then decreases, which confirms the intuition that too few bins and too many bins are not good choices for generating surface signatures.

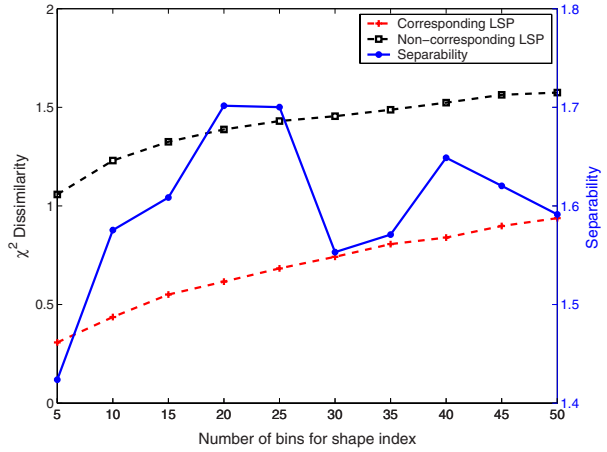
The *distance constraint* controls the number of points contributing to the generation of a LSP. Larger distance constraint allows for more points to be included in a LSP. Intuitively the larger distance constraint results in more discriminating LSPs since LSPs encode more surface information. The intuition is confirmed by the curve of dissimilarity versus the distance constraint in Figure 5.6(c) and 5.7(c), which show that the dissimilarity drops with the increase in distance constraint. It may seem that we should set the distance constraint as large as possible. However, the dissimilarity for non-corresponding LSPs also drops with the distance constraint increasing. From Figure 5.6(c) and 5.7(c), we also observe that the separability increases then drops when the distance constraint increases, which suggests that there is a tradeoff between the descriptiveness of a LSP and the distance constraint.

The *angle constraint* controls the effect of self occlusion. The small value of angle constraint allows for a small number of points contributing to the generation of a LSP, resulting in the large dissimilarity. As the angle constraint is relaxed, the number of points included as a part of a LSP increases. Therefore, in this case an LSP encodes more shape information and the dissimilarity should decrease. As the angle constraint becomes more relaxed the dissimilarity may increase since the shape becomes more and more unique. Figure 5.6(d) and Figure 5.7(d) show that initially dissimilarity decreases and then increases slowly as the angle constraint increases. The separability increases as the angle constraint increases. It seems that we should set the angle constraint as large as possible. However, a small angle constraint is necessary for limiting the effect of occlusion.

The surface type, another component of LSP representation, only depends on the reference point of LSP, and is not affected by the four generating parameters.

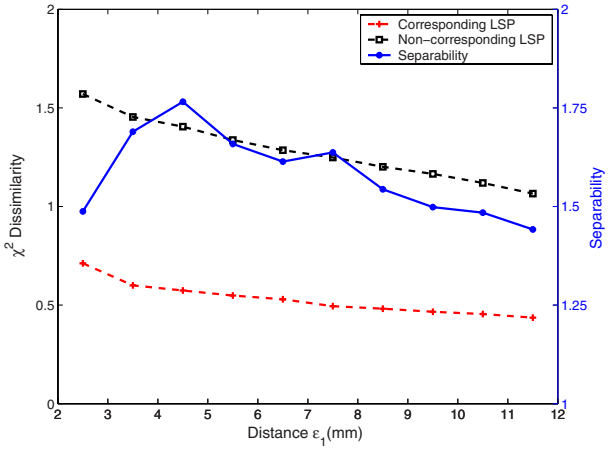


(a) Varying the number of bins for dot product of surface normals

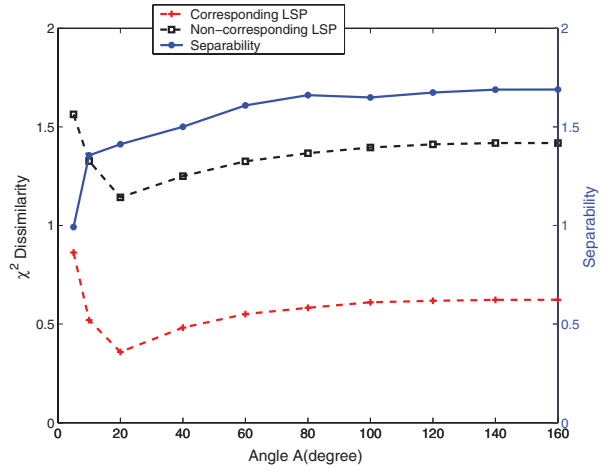


(b) Varying the number of bins for shape index

Fig. 5.6. Effect of generating parameters of LSP on comparing LSPs obtained from one gallery/probe pair ((a)-(d)).

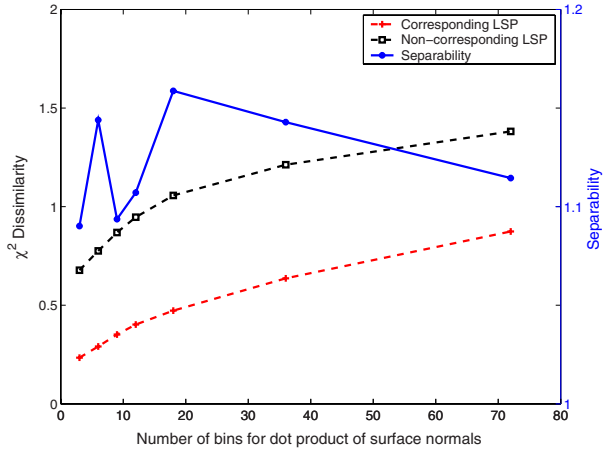


(c) Varying the distance constraint

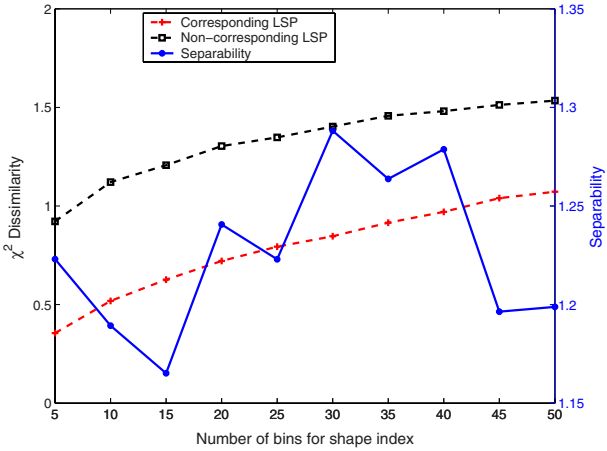


(d) Varying the angle constraint

Fig. 5.6. Figure 5.6 Continued.

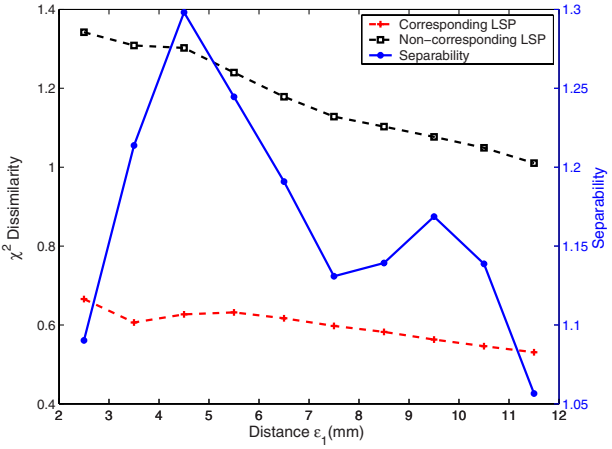


(a) Varying the number of bins for dot product of surface normals

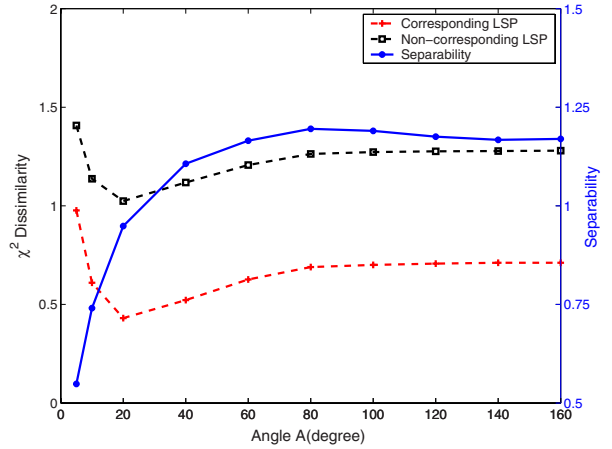


(b) Varying the number of bins for shape index

Fig. 5.7. Effect of generating parameters of LSP on comparing LSPs obtained from another gallery/probe pair ((a)-(d)).



(c) Varying the distance constraint



(d) Varying the angle constraint

Fig. 5.7. Figure 5.7 Continued.

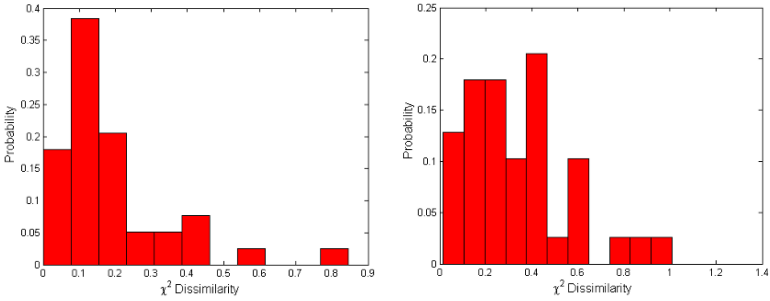
5.1.4 Invariance of Local Surface Patches to Rigid Transformation

The LSP representation consists of histogram of shape index and surface normal angle which are invariant to rigid transformation. To verify this, we compute the χ^2 dissimilarity between reference LSPs and their corresponding LSPs after rigid transformation. We synthetically generate range images at different views by applying 3D rigid transformation. Given a range image $\{v = \{x, y, z\}\}$, we apply the transformation ($tv = R(v - v_0) + v_0$) to generate new views where v_0 is the centroid of 3D vertices in the original range image, R is the rotation matrix and tv are new 3D coordinates after transformation. The rotation matrix R can be written as $R = R_\phi * R_\beta * R_\alpha$ where R_ϕ, R_β, R_α are rotation matrices along x-axis, y-axis, and z-axis, respectively. We calculate shape index and surface normals for the synthetic range image, compute LSPs at the same location as the reference LSPs, and compute the χ^2 dissimilarity between the two corresponding LSPs for the extracted feature points. The surface type for the corresponding LSP is not changed by the rigid transformation. The dissimilarity distributions are shown in Figure 5.8.

Figure 5.8(a) and (b) show the distributions for two different rotations. From this figure, we observe the dissimilarity does not change much and LSP representation is invariant to rigid transformation. Furthermore, as described in Section 5.3, we performed experiments on the UCR dataset which has pose variations ($\pm 35^\circ$) for six different shots of the same subject and on a subset of the UND dataset Collection G which has pose variations (up to 45°) for four shots of the same subject. We achieved good performance. These results show the robustness and view-point invariance of the LSP representation.

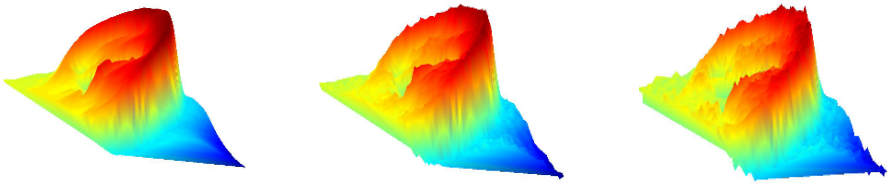
5.1.5 Robustness of Local Surface Patches in the Presence of Noise

In order to use the proposed LSP representation for recognizing 3D objects from real data, we need to address the problem of robustness of LSP in the presence of noise. Since LSP is a local surface descriptor, it will be robust to certain levels of noise. We verify this hypothesis experimentally. While the noise model for different range sensors may be different, we use the Gaussian noise model for the Minolta Vivid camera [78]. Other examples of additive Gaussian noise model with the range data can be found in [79, 80]. Therefore, we inject zero



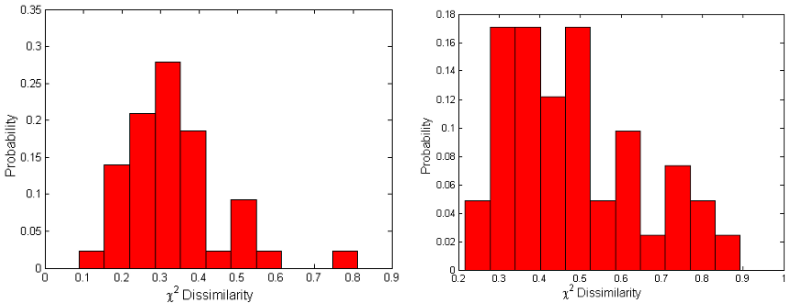
(a) $\phi = 5°,\ \beta = 10°,\ \alpha = 10°$ (b) $\phi = 10°,\ \beta = 30°,\ \alpha = 30°$

Fig. 5.8. χ^2 dissimilarity for corresponding LSPs with respect to the rigid transformation.



(a) (b) (c)

Fig. 5.9. Visualization of a surface corrupted by Gaussian noise $N(\mu, \sigma)$. (a) No noise added. (b) $N(\mu = 0, \sigma = 0.6\text{mm})$. (c) $N(\mu = 0, \sigma = 1.0\text{mm})$.



(a) $N(0, \sigma = 0.6mm)$ (b) $N(0, \sigma = 1.0mm)$

Fig. 5.10. χ^2 dissimilarity for corresponding LSPs in the presence of noise.

mean Gaussian noise $N(0, \sigma^2)$ to range images along the viewing direction (Z-axis). The standard deviation of Gaussian noise that we add depends on the mesh resolution of range scans. However the mesh resolution is not well defined. We use the Johnson's definition [3] according to which "Mesh resolution is defined as the median of all edge lengths in a mesh." Given a 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 range scans is about 1.25mm. Adding noise to range images will corrupt the surface and examples of a range image corrupted with Gaussian noise are shown in Figure 5.9.

Given a range image, we add zero mean Gaussian noise with $\sigma = 0.6$mm and $\sigma = 1.0$mm, then we compute the LSP at the same location as the reference LSP on the corrupted range image, next we calculate the χ^2 dissimilarity between corresponding LSPs. Figure 5.10 shows the distribution of dissimilarity with added Gaussian noise. From Figure 5.10, we can see that the LSP representation is robust to Gaussian noise with $\sigma = 0.6$mm and performs a little worse to Gaussian noise with $\sigma = 1.0$mm. Since the noise corrupts the surface, 14% of corresponding LSPs change the surface type. The dissimilarity for corresponding and non-corresponding LSPs and the separability versus the increased corruption for two different ears are shown in Figure 5.11. We can observe that the dissimilarity for corresponding LSPs increases slowly since the noise corrupts the surface shape and the separability decreases with the increase in corruption, which suggests that LSP matching degrades slowly as the noise level increases.

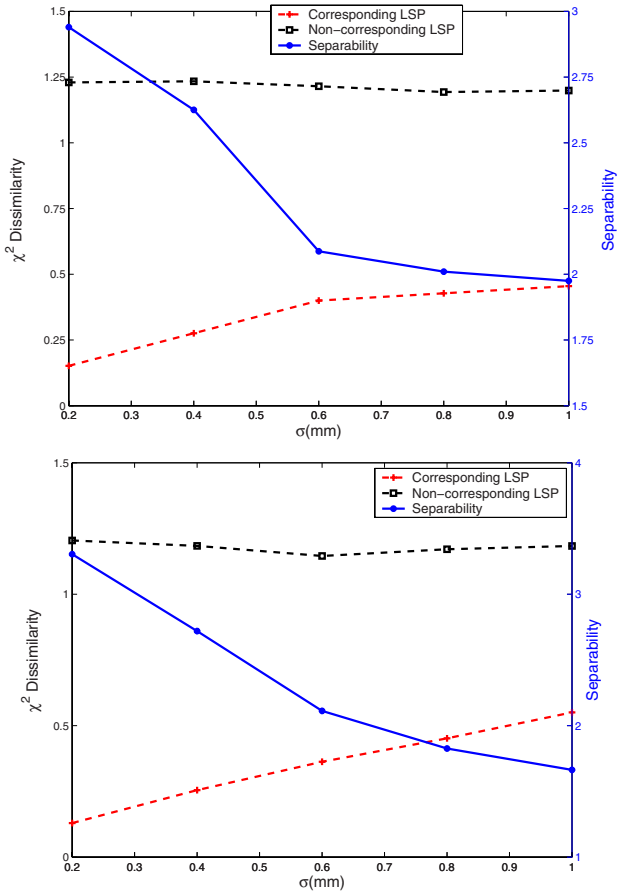


Fig. 5.11. χ^2 dissimilarity for corresponding LSPs with respect to noise of different σ.

5.1.6 Comparison with Spin Image Representation

It is to be noted that our LSP representation is different from the spin image representation [3]. Unlike the LSP representation, spin image is a 2D histogram described by two parameters: distance to the tangent plane of the oriented point from its neighbors and the distance to the normal vector of the oriented point. As described above, we compute LSPs for feature points, while the spin image is computed for every vertex on the surface of an object [3]. In Section 5.3.4, we provide a comparison of the LSP and the spin image representations on a ear dataset.

5.2 Surface Matching

5.2.1 Grouping Corresponding Pairs of LSPs

Given a probe range image, we extract feature points 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 χ^2 dissimilarity mentioned in Section 5.1.2.

For every local surface patch from the probe 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.

$$\begin{aligned} d_{C_1,C_2} = |d_{S_1,S_2} - d_{M_1,M_2}| < \epsilon_2 \\ \max(d_{S_1,S_2}, d_{M_1,M_2}) > \epsilon_3 \end{aligned} \tag{5.4}$$

where d_{S_1,S_2} and d_{M_1,M_2} are Euclidean distances between centroids of two surface patches. The first constraint guarantees that distances d_{S_1,S_2} and d_{M_1,M_2} are consistent; the second constraint removes the correspondences which are too close. For two correspondences $C_1 = \{S_1, M_1\}$ and $C_2 = \{S_2, M_2\}$ where S_i is probe surface patch and M_i is gallery surface patch, they should satisfy (5.4) if they are consistent corresponding pairs. Therefore, we use simple geometric constraints to partition the potential corresponding pairs into different groups. The

larger the group is, the more likely it contains the true corresponding pairs.

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 (5.4).
- Repeat the same procedure for every group.
- Sort the groups in the ascending order based on the size of groups.
- Select the groups on the top of the list.

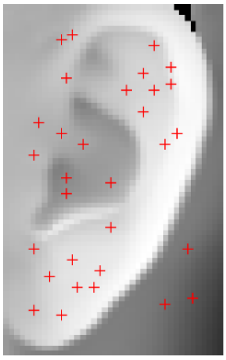
Figures 5.12 and 5.13 show two examples of partitioning corresponding pairs into different groups. Figure 5.12 shows a pair of ears with a small pose variation and Figure 5.13 shows a pair of ears with a large pose variation. Figures 5.12 and 5.13(a) show the feature point extraction results for the probe ears. Comparing the local surface patches with LSPs on the gallery ear, the initial corresponding pairs are shown in Figure 5.12 and 5.13(b), in which every pair is represented by the same number superimposed on the probe and gallery images. We observe that both the true and false corresponding pairs are found. Applying the simple geometric constraints (5.4), examples of filtered groups are shown in Figures 5.12 and 5.13(c) (d), respectively. We can see that the true corresponding pairs are obtained by comparing local surface patches and using the simple geometric constraints for ear pairs with small or large pose variations.

5.2.2 Alignment of Gallery with Probe Ears

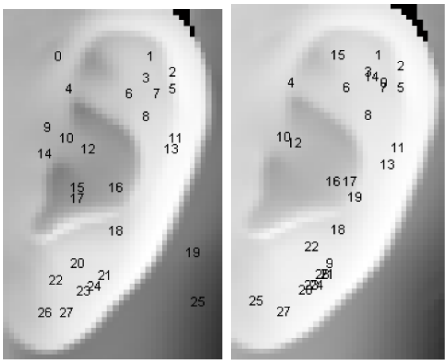
Once the corresponding LSPs between the gallery and probe are established, the initial rigid transformation is estimated and the coarse-to-fine surface matching strategy is followed (see Section 4.2).

5.3 Experimental Results

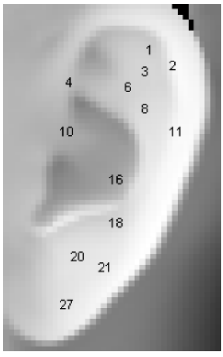
As described in Section 4.3 about the dataset, we perform experiments on the UCR dataset ES_1 and ES_2 and on the UND dataset Collection F and a subset of Collection G (24 subjects with four orientations). The experimental results on the UND dataset Collection F and G are obtained by using the *same* parameters of the ear recognition algorithm as those used on the UCR dataset.



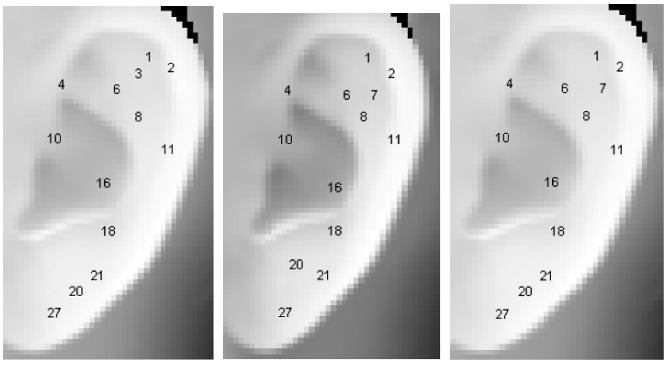
(a) Feature points extraction from a probe ear



(b) Initial corresponding pairs

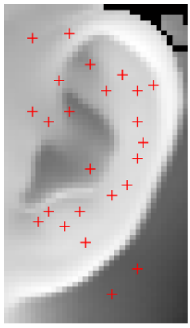


(c) Example of filtered corresponding pairs

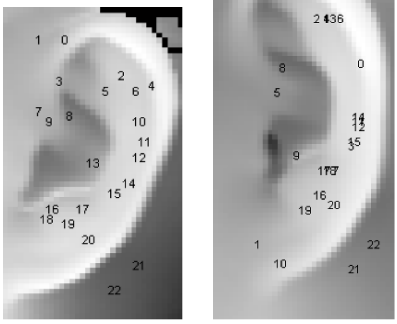


(d) Example of filtered corresponding pairs

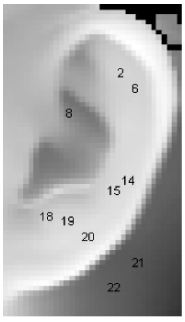
Fig. 5.12. Example of grouping corresponding LSPs for a pair of ears with a small pose variation. The probe ear is shown as the left image in (b), (c) and (d).

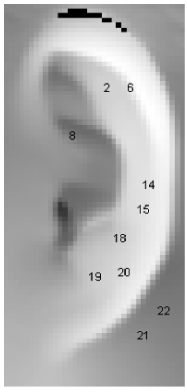


(a) Feature points extraction from a probe ear

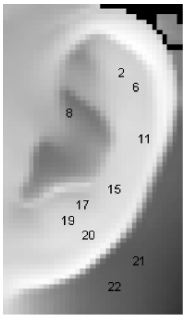


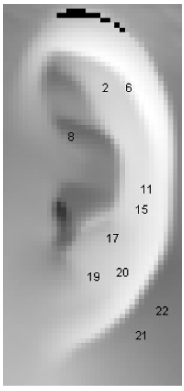
(b) Initial corresponding pairs





(c) Example of filtered corresponding pairs





(d) Example of filtered corresponding pairs

Fig. 5.13. Example of grouping corresponding LSPs for a pair of ears with a large pose variation. The probe ear is shown on the left image in (b), (c) and (d).

5.3.1 Identification Performance

Every probe is matched to every 3D ear in the gallery set and the RMS registration error is calculated using the procedure described in Section 4.2. The matching error matrix $\{ME(i,j), i = 1, 2, \ldots, N_t, j = 1, 2, \ldots, N_m\}$ on the UCR dataset ES_1 using the LSP representation, where N_t and N_m are the number of probe ears and gallery ears, respectively, is displayed as an intensity image shown in Figure 5.14. The average time to match a pair of ears, which includes the coarse and fine alignment, is about 3.7 seconds with C++ implementation on a Linux machine with an *AMD Opteron* 1.8GHz CPU.

The identification performance is evaluated by the cumulative match characteristics (CMC). Table 5.3 shows the rank-r recognition rates using the LSP representation for the UCR dataset and the UND dataset Collection F. In Table 5.3, the numbers of images in the gallery and the probe sets are listed in the parenthesis following the name of the dataset. We achieve 94.84% rank-1 recognition rate (150 out of 155) on the UCR dataset ES_1 and 96.36% rank-1 recognition rate (291 out of 302) on the UND dataset Collection F. As expected, the system performs better on ES_1 with the same pose and the performance degrades slightly on ES_2 with pose variations.

We show four special cases of correctly recognizing gallery-probe ear pairs using the LSP representation in Figures 5.15. In this figure, each probe ear is rendered as a textured 3D surface and each gallery ear is displayed as a mesh. In order to examine the results visually, we display the pre-aligned gallery ear and the probe ear in the same image (Figures 5.15(b)) and also the post-aligned (transformed) gallery and the probe ear in the same image (Figures 5.15(c)). From Figure 5.15, we observe that the ear recognition system can handle partial occlusion. Twelve more examples of correctly recognized gallery-probe ear pairs are shown in Figure 5.16. From Figure 5.16, we observe that each gallery ear is well aligned with the corresponding probe ear.

One error case is illustrated in Figure 5.17. Figure 5.17 (a) and (b) show the color images of two visually similar probe and gallery ears that belong to different subjects; Figure 5.17(c) shows the corresponding gallery ear overlaid on the textured 3D probe ear after registration; Figure 5.17(d) shows the falsely recognized gallery ear overlaid on the textured 3D probe ear after alignment. In Figure 5.17(d), the root mean square error is 1.002mm, which is smaller than 1.079mm in

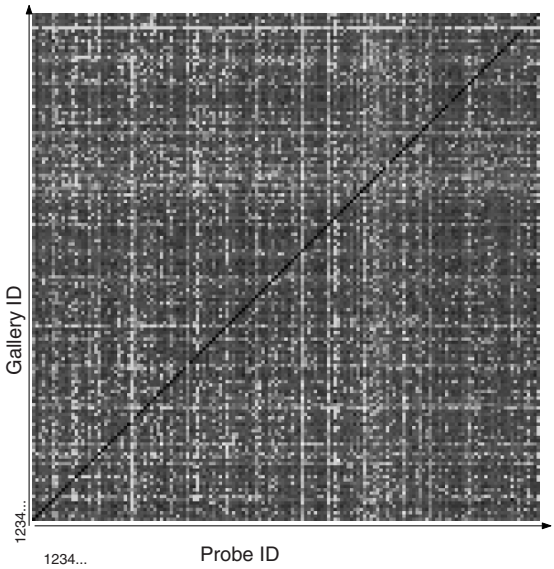


Fig. 5.14. Matching error matrix on the UCR dataset ES_1 using the LSP representation displayed as an intensity image (smaller values correspond to darker pixels). The gallery ID is labeled horizontally and the probe ID is labeled vertically.

Table 5.3. Cumulative matching performance on the UCR dataset and the UND dataset Collection F using the LSP representation.

Dataset	Rank-1	Rank-2	Rank-3	Rank-4	Rank-5
$UCR\ ES_1(155, 155)$	96.77%	96.77%	96.77%	96.77%	96.77%
$UCR\ ES_2(310, 592)$	94.43%	96.96%	97.30%	97.64%	97.80%
$UND(302, 302)$	96.36%	98.01%	98.34%	98.34%	98.34%

Figure 5.17(c). Since the ears of these two subjects are quite similar in 3D, the error occurs and it is an example of false recognition.

Figure 5.17 (a) and (b) show the color images of two visually similar probe and gallery ears that belong to different subjects; Figure 5.17(c) shows the true gallery ear overlaid on the textured 3D probe ear after registration; Figure 5.17(d) shows the falsely recognized gallery ear overlaid on the textured 3D probe ear after alignment. In Figure 5.17(d), the root mean square error for the falsely recognized ear is smaller than the error for the correct ear in Figure 5.17(c). Since we pick up the gallery ear with the minimum RMS error as the recognized ear, we made the errors. In this figure, we obtain good alignment between the gallery and model ears from different persons since these ears are quite similar in 3D.

5.3.2 Verification Performance

The verification performance of the proposed system is evaluated in terms of the two popular methods, the receiver operating characteristic (ROC) curve and the equal error rate (EER). Figure 5.18 shows the ROC curves on the UCR and the UND dataset Collection F using the LSP representation for surface matching. As expected, the system performs better on ES_1 than on ES_2 using the the LSP representations. We obtain the best performance with a 0.023 EER on the UND dataset using the LSP representation. It is clearly seen that without retuning the parameters of the proposed algorithms we achieved good verification performance on the UND dataset.

5.3.3 Evaluation of Verification Performance

We discuss and evaluate the accuracy of the ear verification system by applying the method in [38, 81, 82]. As described in Section 1.3, the UCR dataset has 155 subjects. There are 155 probes providing 155 user claims and 23,870 (155×154) imposter claims for the UCR dataset ES_1. For the UCR dataset ES_2 there are 592 probes providing 592 user claims and 91,168 (592×154) imposter claims. The UND dataset Collection F has 302 subjects. There are 302 pairs of images providing 302 user claims and 90,902 (302×301) imposter claims.

We calculate the number of user claims and imposter claims that gives statistically significant results. Let ζ_μ denote the $\mu-$percentile of

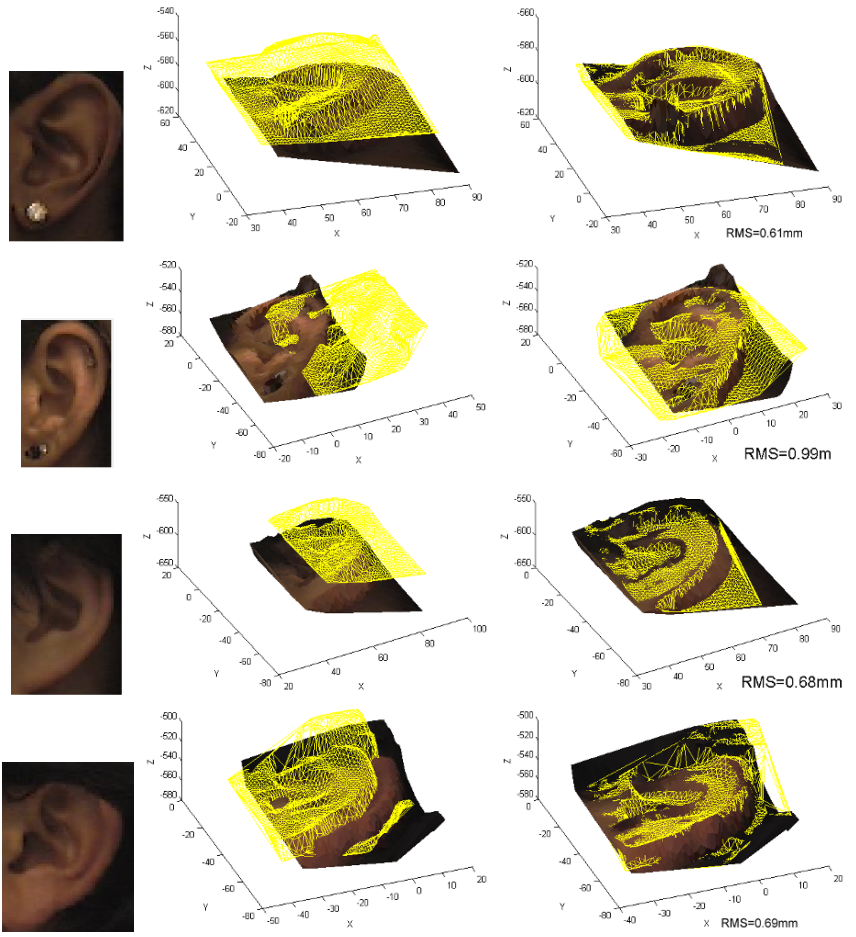


Fig. 5.15. UCR dataset: Four examples of the correctly recognized gallery-probe pairs using the LSP representation. Two ears have earrings and the other two ears are partially occluded by the hair. Images in column (a) show color images of ears. Images in column (b) and (c) show the probe ear with the corresponding gallery ear before the alignment and after the alignment, respectively. The gallery ears represented by the mesh are overlaid on the textured 3D probe ears. The units of x, y and z are in millimeters (mm).

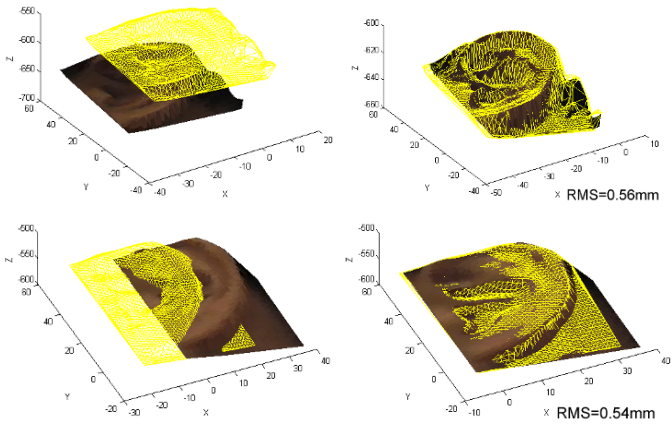


Fig. 5.16. UCR dataset: Twelve cases of the correctly recognized gallery-probe pairs using the LSP representation. (a) Examples of probe ears with the corresponding gallery ears before alignment. (b) Examples of probe ears with the correctly recognized gallery ears after alignment. The gallery ear represented by the mesh is overlaid on the textured 3D probe ear. The units of x, y and z are in millimeters (mm). Two example cases are shown on a page.

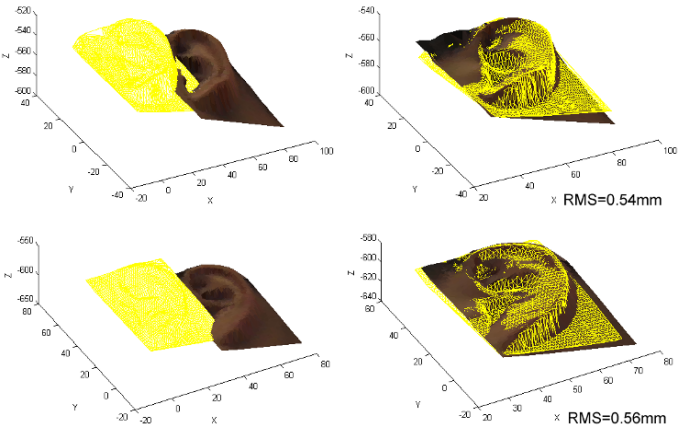


Fig. 5.16. Figure 5.16 Continued.

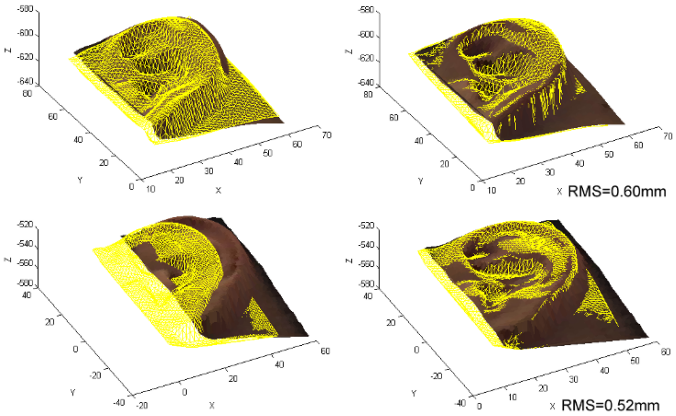
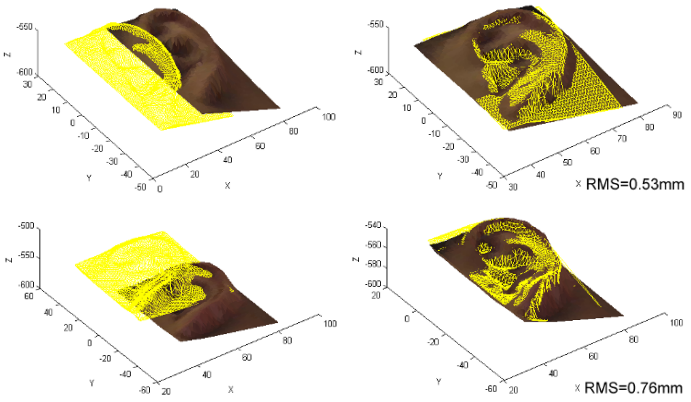


Fig. 5.16. Figure 5.16 Continued.



(a) (b)

Fig. 5.16. Figure 5.16 Continued.

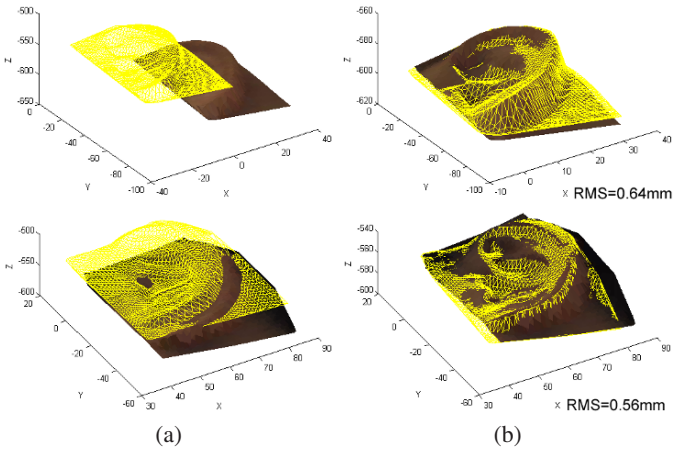


Fig. 5.16. Figure 5.16 Continued.

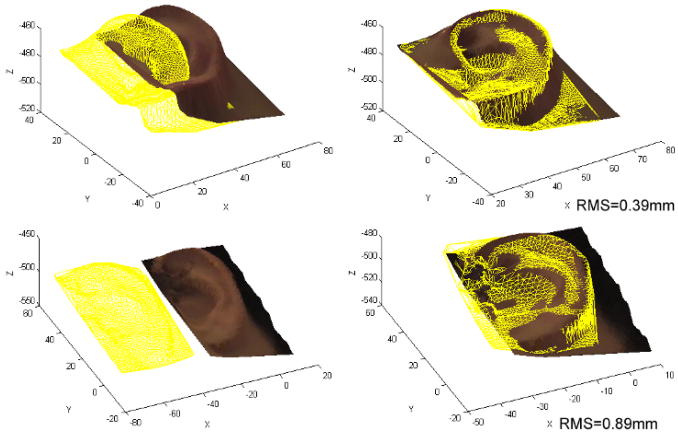


Fig. 5.16. Figure 5.16 Continued.

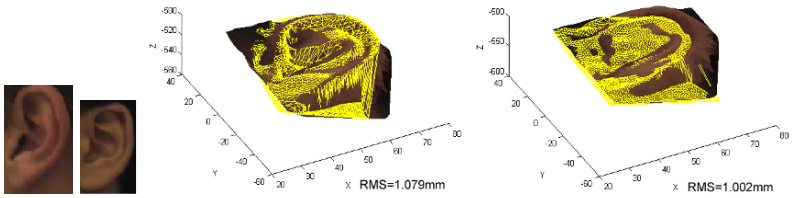


Fig. 5.17. UCR dataset: An incorrectly recognized model-test pair using the LSP representation. The model ear represented by the yellow mesh is overlaid on the textured 3D test ear. The units of x, y and z are in millimeters (mm). (a) Color image of the probe ear. (b) Color image of the falsely recognized gallery ear. (c) The gallery ear after alignment is overlaid on the texture 3D probe ear. (d) The falsely recognized gallery ear after alignment is overlaid on the textured 3D probe ear. Note that for the incorrect match the gallery ear in column (d) achieves a smaller value of RMS error than the gallery ear in column (c).

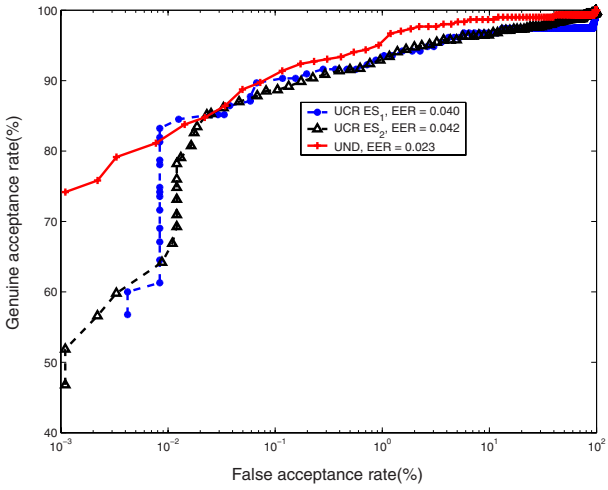


Fig. 5.18. UCR dataset and UND dataset Collection F: Verification performance as a ROC curve. ROC curves on the UCR dataset ES_1, ES_2 and the UND dataset using the LSP representation.

the standard Gaussian distribution with zero mean and unit variance. Since the verification tests can be thought of as Bernoulli trials, we can assert that with confidence $\delta = 1 - \mu$ that the minimum number of user claims which ensures that the expected value of FRR (p_{FR}) and the empirical value ($\widehat{p}_{FR}$) are related by $|p_{FR} - \widehat{p}_{FR}| \leq \xi$, is given by

$$\upsilon_C = \left(\frac{\zeta_{\mu/2}}{\xi}\right)^2 p_{FR}(1 - p_{FR}). \tag{5.5}$$

The number υ_I of the imposter claims [82] which is sufficient to ensure that the expected value of FAR (p_{FA}) and the empirical value ($\widehat{p}_{FA}$) are related by $|p_{FA} - \widehat{p}_{FA}| \leq \xi$ is given by

$$\upsilon_I = \left(\frac{\zeta_{\mu/2}}{\xi'}\right)^2 p_o(1 - p_o), \qquad \xi' = \frac{\xi}{k}, \qquad (1 - p_o)^k = 1 - p_{FA} \tag{5.6}$$

where p_o is the probability that one imposter is falsely accepted as an authorized user and k is the number of imposter claims ($k = 154$ for the UCR dataset and $k = 301$ for the UND dataset Collection F). By setting the desired EER, δ and ξ, we can compute υ_C and υ_I.

For the UCR dataset ES_1, we find $\upsilon_C = 149$ and $\upsilon_I = 24,759$ with $EER = 5\%$, $\delta = 95\%$ and $\xi = 3.5\%$.

For the UCR dataset ES_2, we find $\upsilon_C = 456$ and $\upsilon_I = 75,826$ with $EER = 5\%$, $\delta = 95\%$ and $\xi = 2\%$.

For the UND dataset Collection F, we find $\upsilon_C = 292$ and $\upsilon_I = 94,874$ with $EER = 5\%$, $\delta = 95\%$ and $\xi = 2.5\%$.

Note that the order of magnitude of these numbers are the same as those provided by the test scenarios. The values of ξ on the UCR dataset ES_2 ($\xi = 2\%$) and the UND dataset Collection F ($\xi = 2.5\%$) are smaller than the value on the UCR dataset ES_1 ($\xi = 3.5\%$) since the UCR dataset ES_2 and the UND dataset Collection F are larger in size than the size of the UCR dataset ES_1.

5.3.4 Comparison of LSP and Spin Image Representations

In order to compare the distinctive power of the LSP and the spin image representations, we follow the same procedures as described in this chapter to recognize ears using the spin image representation. In the experiments, the size of the spin image is 15×15. We perform experiments on the UCR dataset ES_1 (155 shots in the gallery and 155

shots in the probe) to compare the performance of these two representations in terms of the CMC and the ROC curves.

Table 5.4 shows CMC values using the LSP and the spin image representations for ear recognition. Figure 5.19 shows the ROC curves using the LSP and the spin image representations for matching ears. From Table 5.4 and Figure 5.19, we observe that the LSP representation achieved a slightly better performance than the spin image representation.

5.3.5 Identification and Verification Performance with Occlusion

In order to show the recognition performance with occlusions, we simulate the occluded probe ear images by selecting a fraction of the extracted LSPs in the images. The area covered by the occluded LSPs is removed. For a probe ear, we randomly remove 10%, 20%, 30%, 40%, and 50% of the extracted LSPs and then match it against every ear in the gallery. The identification performance is given in Table 5.5 and the ROC curves for different occlusion ratios are shown in Figure 5.20. We observe that the recognition performance degrades slowly with the increased occlusion.

5.4 Comparison with Yan and Bowyer's Approach

The most important difference is that we propose the ear helix/anti-helix and the local surface patch (LSP) representations to estimate the initial rotation and translation between a gallery-probe pair while Yan and Bowyer do not estimate the initial transformation. The initial transformation is critical for the success of ICP algorithm, which can be seen from Table 5.6 that shows the rank-1 recognition rates. As described above, in the UND dataset Collection G, there are 24 subjects whose images are taken at four different poses, straight-on, 15° off center, 30° off center and 45° off center. For each angle of an ear image, we match it against all the images at different angles. This is the same experiment performed in [34, Chap. 8.3]. We observe that the results obtained by the helix/anti-helix representation are better than Yan and Bowyer's results [34, Chap. 8.3] for all the cases except the case (45° probe against 30° gallery). The results obtained by the LSP representation outperform the results obtained in [34, Chap. 8.3] for the

Table 5.4. Cumulative matching performance on the UCR dataset ES_1 using the LSP and the spin image representations.

Representation	Rank-1	Rank-2	Rank-3	Rank-4	Rank-5
LSP	94.84%	96.77%	96.77%	96.77%	96.77%
Spin Image	92.90%	95.48%	95.48%	95.48%	95.48%

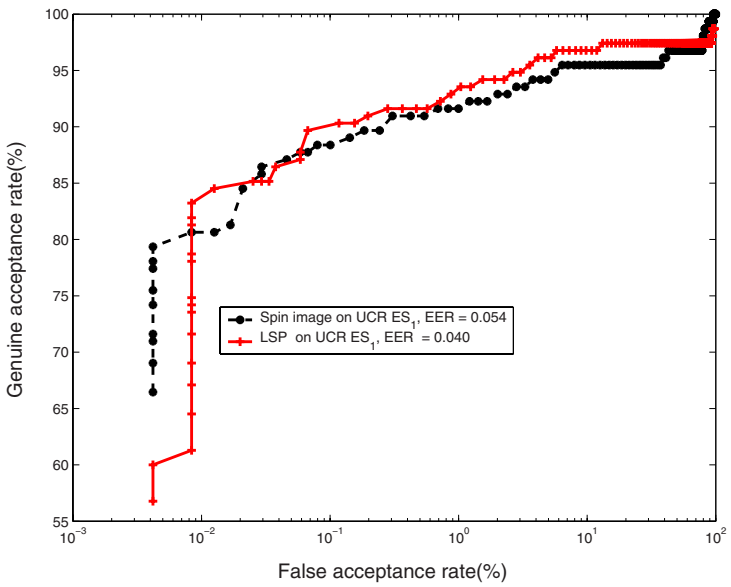


Fig. 5.19. Verification performance as ROC curves using the LSP and the spin image representations on the UCR dataset ES_1.

Table 5.5. Cumulative matching performance on the UCR dataset ES_1 with different occlusion ratios using the LSP representation.

Occlusion ratio	Rank-1	Rank-2	Rank-3	Rank-4	Rank-5
0	94.84%	96.77%	96.77%	96.77%	96.77%
0.1	94.19%	96.13%	96.13%	96.13%	96.13%
0.2	92.26%	94.84%	94.84%	95.48%	95.48%
0.3	90.97%	92.90%	92.90%	94.19%	95.48%
0.4	87.74%	92.90%	92.90%	92.90%	93.55%
0.5	84.52%	88.39%	88.39%	88.39%	88.39%

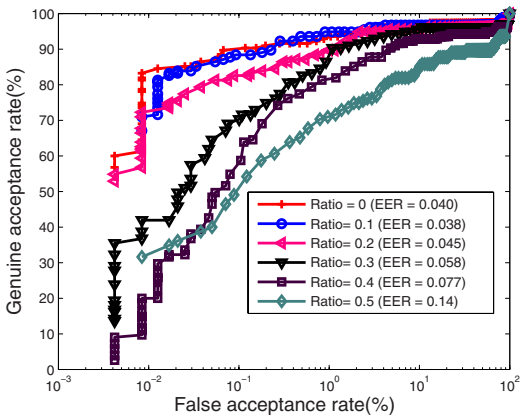


Fig. 5.20. Verification performance as ROC curves with different occlusion ratios on the UCR dataset ES_1 using the LSP representation.

Table 5.6. Comparison between Chen and Bhanu's approach [1] and Yan and Bowyer's approach [2] on a subset of Collection G. Our rank-1 identification results are put inside the brackets. The first number is obtained using the helix/anti-helix representation and the second one is obtained using the LSP representation. The number outside the bracket is obtained from [34, Chap. 8.3].

Probe\Gallery	**Straight-on**	15° **off**	30° **off**	45° **off**	**Average**
Straight-on		[**100%, 100%**], 100%	[**91.7%, 87.5%**], 87.5%	[**87.5%, 83.3%**], 70.8%	[**93.1%, 90.3%**], 86.1%
15° **off**	[**100%, 100%**], 100%		[**100%, 100%**], 100%	[**87.5%, 91.7%**], 87.5%	[**95.8%, 97.2%**], 95.8%
30° **off**	[**91.7%, 91.7%**], 87.5%	[**100%, 100%**], 100%		[**95.8%, 91.7%**], 95.8%	[**95.8%, 94.4%**], 94.4%
45° **off**	[**87.5%, 87.5%**], 79.2%	[**91.7%, 87.5%**], 87.5%	[**95.8%, 87.5%**], 100%		[**91.7%, 87.5%**], 88.9%
Average	[**93.1%, 93.1%**], 88.9%	[**95.8%, 95.8%**], 95.8%	[**97.2%, 91.7%**], 95.8%	[**90.3%, 88.9%**], 84.7%	[**94.1%, 92.4%**], 91.3%

cases with large pose variations (45° and 30° probes against straight-on gallery, straight-on and 15° probes against 45° gallery). From Table 5.6, we see that our representations can reasonably handle the pose variation up to 45°.

5.5 Conclusions

In this chapter, we proposed a new local surface patch (LSP) representation for 3D matching. This representation is invariant to rotation and translation. We used this representation for finding initial correspondences between a gallery-probe pair. Then a modified iterative closest point (ICP) algorithm iteratively refined the transformation which brings the hypothesized gallery and a probe image into the best alignment. The root mean square (RMS) registration error is used as the matching error criterion. The experimental results on two real ear range and color image datasets demonstrated the potential of the proposed algorithms for robust ear recognition in 3D. Extensive experiments are performed on the UCR dataset (155 subjects with 902 images under pose variations), the UND dataset Collection F (302 subjects with 302 time-lapse gallery-probe pairs) and a subset of the UND dataset G for evaluating the performance with respect to pose variations without retuning the parameters of the proposed algorithms. These results showed that the proposed ear recognition system is capable of recognizing ears under pose variations, partial occlusions and time lapse effects. The proposed representation is less sensitive to pose variations.

We also provided a comparison of the LSP representation with the spin image representation for identification and verification. This comparison showed that the LSP representation achieved a slightly better performance than the spin image representation.