Finding Surface Correspondence for Object Recognition and Registration using Pairwise Geometric Histograms

A. P. Ashbrook, R. B. Fisher, C. Robertson and N. Werghi

Department of Artificial Intelligence
The University of Edinburgh
5, Forrest Hill, Edinburgh, EH1 2QL
Telephone: +44 131 650 4504
Fax: +44 131 650 6899
anthonya@dai.ed.ac.uk

Abstract. Pairwise geometric histograms have been demonstrated as an effective descriptor of arbitrary 2-dimensional shape which enable robust and efficient object recognition in complex scenes. In this paper we describe how the approach can be extended to allow the representation and classification of arbitrary $2\frac{1}{2}$- and 3-dimensional surface shape. This novel representation can be used in important vision tasks such as the recognition of objects with complex free-form surfaces and the registration of surfaces for building 3-dimensional models from multiple views. We apply this new representation to both of these tasks and present some promising results.

1 Introduction

Finding a correspondence between two or more surfaces is a frequently encountered problem in many computer vision tasks. When surface based descriptions are used for object recognition, the hypothesis that a particular object is in a scene is confirmed by finding a good correspondence between scene and model surfaces [6]. When constructing geometric models of objects by merging multiple range images taken from different viewpoints, the surfaces described by each range image require registration into a common coordinate frame [3, 1]. This can be done by finding the correspondence between portions of the object's surface which is common to two or more views.

In this paper we present a novel representation for arbitrary $2\frac{1}{2}$- and 3-dimensional surface data which enables correspondences to be found reliably and efficiently. The representation is based on pairwise geometric histograms which have previously been demonstrated as a representation for 2-dimensional shape data for object recognition applications [4].

The approach that we are proposing determines whether two surfaces have a correspondence as follows:

1. Each of the surfaces is approximated by a triangular mesh. The details of this approximation and the algorithms we have employed for this are presented in Section 3.1.
2. Each triangular mesh facet is represented by a pairwise geometric histogram which records the relationship between this facet and the surrounding facets within some specified neighbourhood. This representation is discussed in Section 3.2.
3. Correspondences between individual facets are found by matching their respective geometric histograms. These *local* correspondences provide hypotheses for the correspondence between the two surfaces. The metric employed for matching geometric histograms is described in Section 4.
4. The *global* surface correspondence is found by finding consistent local hypotheses using a probabilistic Hough transform. This is discussed in Section 5.

2 Background

A number of approaches to the problem surface registration have been developed from the "iterated closest point" (ICP) algorithm proposed by Besl and McKay [2]. These algorithms have been popular for registering multiple views of an object for model construction and for refining pose in object recognition tasks. The central idea behind this algorithm is that by forming correspondences between points on one surface and their nearest neighbours on another and then minimising the distances between them, the registration of the two surfaces is improved. If this process is iterated the registration of the surfaces often converges. The approach is computationally expensive because of its use of raw surface point data and because of the iterative nature of the algorithm. A more serious problem is that the algorithm is not guaranteed to converge, sometimes getting caught in local minima, and typically requires good initial alignment of the surfaces to get a reasonable solution. One of the advantages of the ICP approach is that, because it uses all of the surface data available, when it does converge the registration can be very accurate. The algorithm is also suitable for arbitrary classes of surface.

Other researchers have used interest points on the surface instead of all of the surface data and formed correspondences by matching geometric descriptors of those points. Thirion [13] proposes the use of extremal points on 3-dimensional surfaces which can be characterised by a number of properties such as their curvature. Interest points with similar properties are treated as potential correspondents and the transformation that aligns the surfaces is determined from triplets of corresponding pairs. Recently, Johnson and Hebert [9] have proposed a novel interest point descriptor which allows point correspondences to be formed between surfaces. In their approach the interest points are defined by the vertices of a polygonal mesh fitted to the surface. At each vertex the geometric relationship with all of the other mesh vertices are recorded in a 2-dimensional *spin-image* which is invariant to rigid transformations of the surface. Interest point correspondences are found by identifying points with similar spin-images.

Local surface features such as edges and surface patches have also been used to determine the correspondence between two surfaces [5]. Initially all features on the first surface are considered as potential correspondents of features of the same class on the second surface. The number of potential correspondences is then quickly reduced using approaches based on geometric constraints such as the interpretation tree. Each pair of matched features provides a constraint on the transformation that aligns the surfaces and these are used to determine the best global alignment. The motivation for using features is to reduce the amount of data to be processed whilst maintaining valuable information needed to perform matching and constrain the alignment transformation. The disadvantage is that a particular choice of features can limit the scope of the algorithm to particular classes of surfaces.

3 A Novel Surface Shape Representation

3.1 Surface Reconstruction and Approximation

Initially a given surface S, acquired using a range sensor, is described by a set of points samples $P = \{p_1, \ldots, p_N\}$. The points may represent a single view of the surface or a number of different views, for example from different viewpoints around an object. The point set is then used to construct a triangular mesh approximation $\hat{S}$ to the original surface, where $\hat{S} = \{t_1, \ldots, t_M\}$ and t_i is a triangular facet of the mesh.

It is important to clarify at this stage that the only requirement of the mesh is that it is a good approximation of the surface shape. No assumptions are made about the actual distribution of facets over the surface as this is unlikely to be repeatable. To minimise the amount of memory and computation needed to solve the correspondence problem, the mesh should also contain the smallest number of facets needed to give a good approximation of the surface.

A number of algorithms have been proposed for reconstructing a triangular faceted mesh from a set of points. In the work presented here an initial, regular mesh was constructed from the sampled point data using a reconstruction algorithm by Hoppe *et al* [8]. The resulting regular mesh was then refined to minimise the number of facets whilst maintaining most of the surface shape using a surface simplification algorithm by Garland and Heckbert [7].

There are a number of advantages in using a triangular mesh to approximate the surface to be represented instead of more complex features such as quadric patches, the most obvious being efficiency. Constructing a mesh is also significantly more straightforward than segmenting a surface into more complex features. A second important issue is scope. Any surface can be approximated by a triangular mesh but selecting a fixed set of features can impose limitations on the types of surfaces that can be described. Another important issue is that of stability. If surface patches are assigned to different classes based on their shape then borderline cases can result in sudden changes in the representation because of slightly different viewing conditions or noise.

The disadvantage of using a triangular mesh is that it requires many facets to describe surfaces with high curvature to a high degree of accuracy. By statistically modelling the shape error introduced by the triangular shape approximation, it is still possible to obtain a good shape representation when only a relatively small number of facets are used.

3.2 Histogram Construction

A pairwise geometric histogram h_i is constructed for each triangular facet t_i in a given mesh which describes its pairwise relationship with each of the other surrounding facets within a predefined distance. This distance controls degree to which the representation is a local description of shape. The histogram is defined such that it encodes the surrounding shape geometry in a manner which is invariant to rigid transformations of the surface data and which is stable in the presence of surface clutter and missing surface data.

Figure 1(a) shows the measurements used to characterise the relationship between facet t_i and one of its neighbouring facets t_j. These measurements are the relative angle, α, between the facet normals and the range of perpendicular distances, d, from the plane in which facet t_i lies to all points on facet t_j. These measurements are accumulated in a 2-dimensional frequency histogram, weighted by the product of the areas of the two facets as shown in Figure 1(b). The weight of the entry is spread along the perpendicular distance axis in proportion to the area of the facet t_j at each distance. To compensate for the difference between the measurements taken from the mesh and the true measurements for the original surface, the entry is blurred into the histogram. For the work presented here a Gaussian blurring function has been used, but we intend to investigate more appropriate error models in the future. Certainly the scale of the blurring function relates to the coarseness of the mesh. The complete pairwise geometric histogram for facet t_i is constructed by accumulating these entries for each of the neighbouring facets.

For clarity, an example of a pairwise geometric histogram is presented in Figure 2(a). This has been constructed for the highlighted facet on the hemispherical mesh presented in Figure 2(b). Note that the representation only depends upon the surface shape and not on the placement of facets over the surface. This independence on the placement of the facets is important because recovering exactly the same mesh for the same surface under different viewing conditions is very unlikely, particularly if there is some surface occlusion.

4 Generating Correspondence Hypotheses

Given two surface meshes, $\hat{S}^A$ and $\hat{S}^B$, the geometric histogram representation allows correspondences between all facets, t_i^A and t_j^B, from each of the meshes to be determined. A match for facet t_i^A is determined by finding the best match between its respective pairwise geometric histogram and all of the histograms

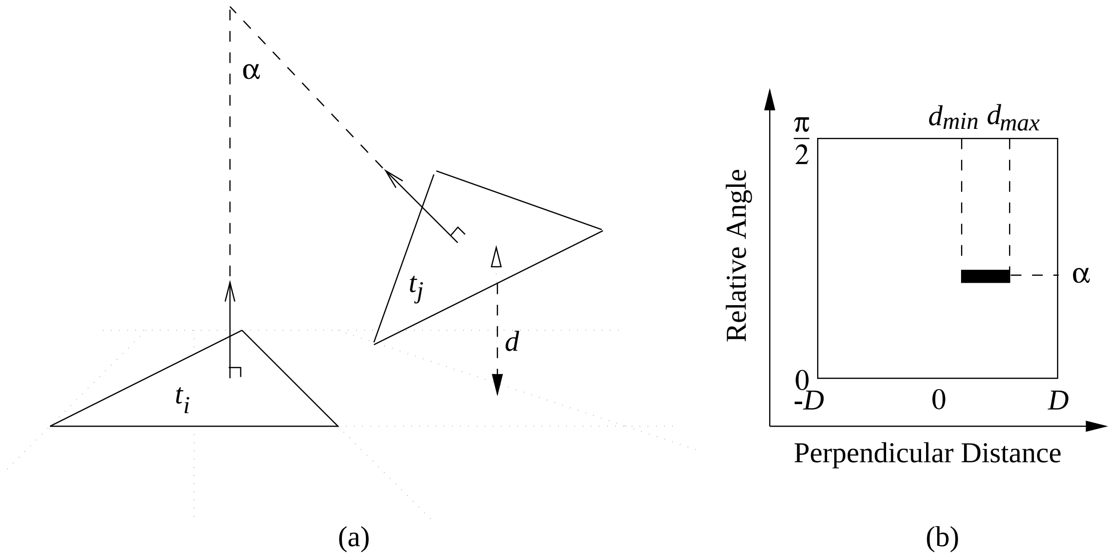

Fig. 1. (a) The geometric measurements used to characterise the relationship between two facets t_i and t_j. (b) The entry made into the pairwise geometric histogram to represent this relationship.

representing the facets in surface $\hat{S}^B$. These *local* correspondences are treated as hypotheses for the correspondence between the two surfaces S^A and S^B.

The similarity, D_{ij}, between two pairwise geometric histograms h_i and h_j is defined using the Bhattacharyya metric. This is given by the expression:

$$D_{ij} = \sum_{\alpha,d} \sqrt{h_i(\alpha,d)}\sqrt{h_j(\alpha,d)} \tag{1}$$

The Bhattacharyya metric is appropriate when the error on the data can be described using a Poisson distribution. This is a reasonable assumption for measured frequency distributions such as a geometric histogram [12]. A derivation of this metric is presented in Appendix A.

5 Hypothesis Verification

Each pair of matched mesh facets provides evidence that the surfaces to which they belong have the same shape, at least locally, and can therefore be registered. The transformation that aligns the paired facets also provides a constraint on the transformation that aligns the complete surfaces. The problem then is to determine whether there is enough evidence to support these hypotheses and, if so, to determine the transformation that aligns the surface data.

We have used an approach taken by other researchers in which N-tuples of matched features, in our case paired mesh facets, are used to estimate the

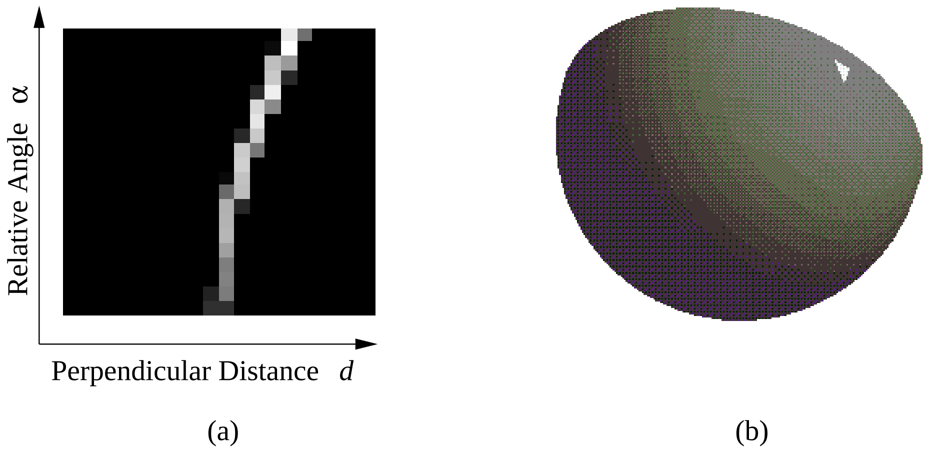

Fig. 2. (a) The geometric histogram that characterises the relationship between highlighted facet and the other facets in the mesh in (b).

alignment transformation. These estimates are then accumulated in a Hough transform resulting in a peak where there is consistency. As an improvement to this scheme we have adopted a probabilistic approach in which the error on the estimated transformation is integrated into the Hough accumulator [11]. This error is determined by statistically modelling the error between the facets and the true surface and propagating this error through the transformation estimator.

Initially 2-tuples of paired facets are used to estimate the rotation component of the alignment transformation and votes are placed in a 3-dimensional Hough transform. The number of 2-tuples can be very large so only a proportion of the largest paired facets are used. If a significant peak is found in this space then 3-tuples of paired facets are used to estimate the translation component of the alignment transformation. Again, only a proportion of the largest facets are used to allow fast operation. If a significant peak is found in the translation space then the hypothesis that the surfaces can be registered is accepted.

6 Experiments

Two applications of the proposed surface representation are presented here. The first application is the registration of two different views of an object with a complex surface. The second application is the identification and localisation of known objects in a scene. All of the data were acquired using a laser stripe range scanner with an accuracy of approximately 0.1mm. The pairwise geometric histogram parameters selected for both of these experiments are presented in Table 1.

Quantisation of Relative Angle Axis	20 bins
Quantisation of Perpendicular Distance Axis	20 bins
Maximum Perpendicular Distance	$\pm$ 100mm
Maximum Relative Angle	$\frac{\pi}{2}$ radians

Table 1. The pairwise geometric histogram parameters used in the experiments presented here.

6.1 Registration of Free-form Surfaces

In this experiment the objective is to find the correspondence between two surfaces constructed from different views of an object. The surface meshes, presented in Figure 3, describe the surface of a farm animal model and consist of 1000 facets each. It should be noted that the model has quite complex, free-form surfaces which are difficult to describe using features such as quadric patches or edges.

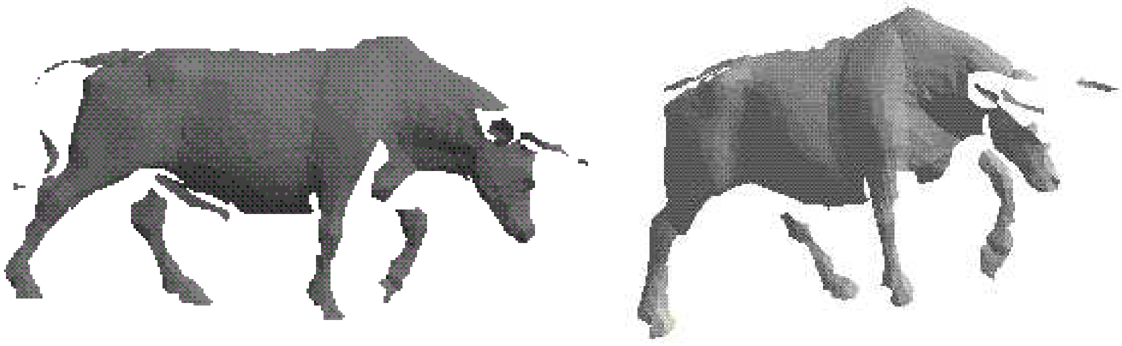

Fig. 3. The triangular meshes for two different views of the surface of a farm animal model.

Figure 4(a) presents the two surfaces in their registered positions. Certainly, from a qualitative point of view, the registration seems to have been successful. This is emphasised by the inter-meshing of the two surfaces on the rear leg of the model shown in close-up in Figure 4(b). The fact that this inter-meshing is not visible over all of the surface suggests that there is some registration error, however.

Only the largest 5% of the facets were matched and used to determine the alignment transformation. The entire registration process took approximately 4 minutes 24 seconds on a 200MHz Sun Ultra. A breakdown of these times is presented in Table 2.

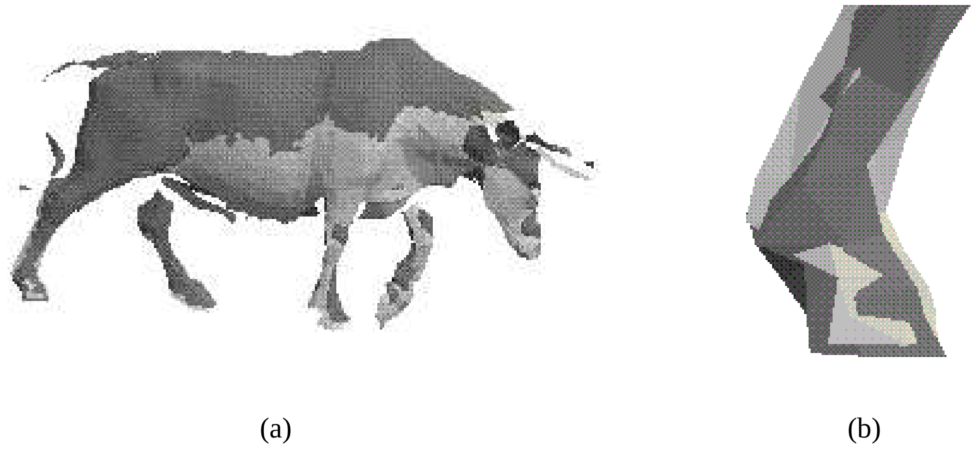

Fig. 4. (a) The two meshed surfaces in their registered positions. (b) A close-up of the rear leg of the model. The light and dark shades of grey represent the two different surfaces.

Triangular Mesh Construction	110 seconds
Geometric Histogram Construction	212 seconds
Geometric Histogram Matching	6 seconds
Resolving Hypotheses	126 seconds

Table 2. A breakdown of the time to complete the registration for each of the main algorithm stages.

6.2 Object Recognition and Pose Estimation

The objective of this experiment is to identify known objects in a scene and estimate the pose of those objects. The object models, presented in Figure 5, have been constructed from multiple views to produce a complete 3-dimensional description of all of the surfaces. Each model is represented by 1000 facets.

Figure 6 presents a scene containing two of the known models. The scene has been captured with a single range image and represented by 1000 facets.

The classification of each of the scene facets is presented in Figure 7. In each of the three images the scene facets which best match a facet from the respective model have been drawn. It can be seen that most of the facets have been classified as belonging to the correct models. Most of the incorrectly classified facets lie very close to surface discontinuities where the recovery of the surface normal is very poor. This is largely due to the mesh construction algorithm which has problems preserving discontinuities in the range data. There are also some problems with the classification of the underside of the cylinder model. This is likely to be because this surface is almost parallel to the viewing direction which makes recovery of the surface normal prone to error.

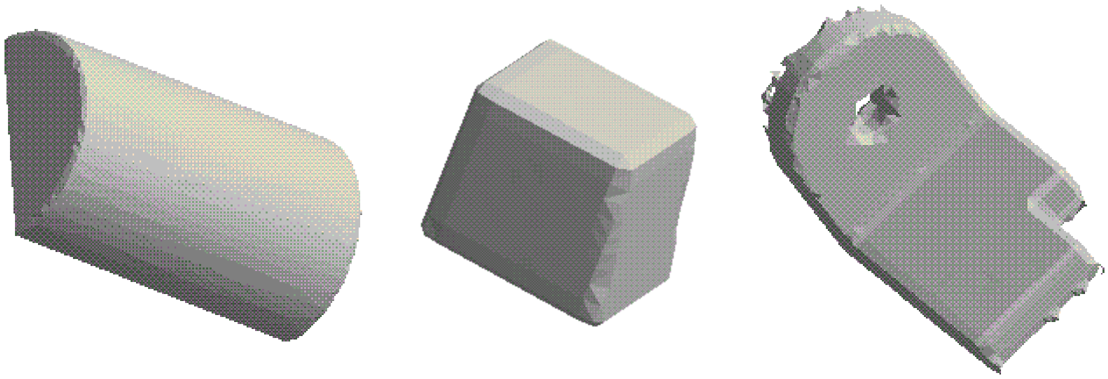

Fig. 5. The three model objects used in the recognition experiment.

Figure 8 presents the results of the recognition of pose estimation process. The original scene data is shown in the darker shade and the recognised models are shown in their estimated positions in the lighter shade. The algorithm has both determined the objects present in the scene and formed a reasonable estimation of their positions.

All of the facets were matched and then the largest 5% from each class were used to determine the model poses. The entire object recognition process took approximately 14 minutes 3 seconds on a 200MHz Sun Ultra. A breakdown of these times is presented in Table 3.

Triangular Mesh Construction	54 seconds
Geometric Histogram Construction	96 seconds
Geometric Histogram Matching	329 seconds
Resolving Hypotheses	364 seconds

Table 3. A breakdown of the time to complete the recognition process for each of the main algorithm stages.

7 Conclusions

The problem of finding a correspondence between two or more surfaces has been investigated by a number of researchers and several solutions have been proposed. The most reliable approaches are based on finding point-feature or

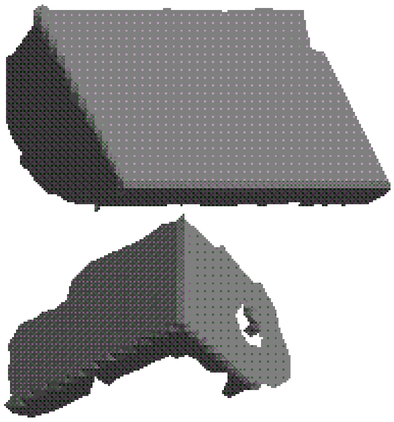

Fig. 6. The scene data used in the recognition experiment.

surface-feature correspondences between the surfaces being registered and using these to estimate the transformation that aligns the complete surfaces.

In this paper we have proposed a novel representation for surface data which enables local surface correspondences to be determined. This representation is invariant to rigid transformations of the surface data and, because of its statistical nature, allows errors in the approximation of the surfaces by triangular meshes to be modelled.

Having established local correspondences we have shown that the transformation that aligns complete surfaces can be determined using a Hough voting scheme. The advantage of using Hough voting is that it is possible to model transformation errors present in the local correspondences by adopting a probabilistic Hough transform.

To demonstrate the effectiveness of the new representation and the algorithm that determines the alignment transformation, we have presented two experiments. In the first experiment two surfaces of a complex curved surfaced object taken from different viewpoints are successfully registered. In the second experiment, known objects are successfully identified and located in a scene.

Acknowledgements

The work presented in this paper was funded by a UK EPSRC grant GR/H86905. The author would also like to thank Dr. Peter Rockett for the derivation of the Bhattacharyya distance presented in Appendix A.

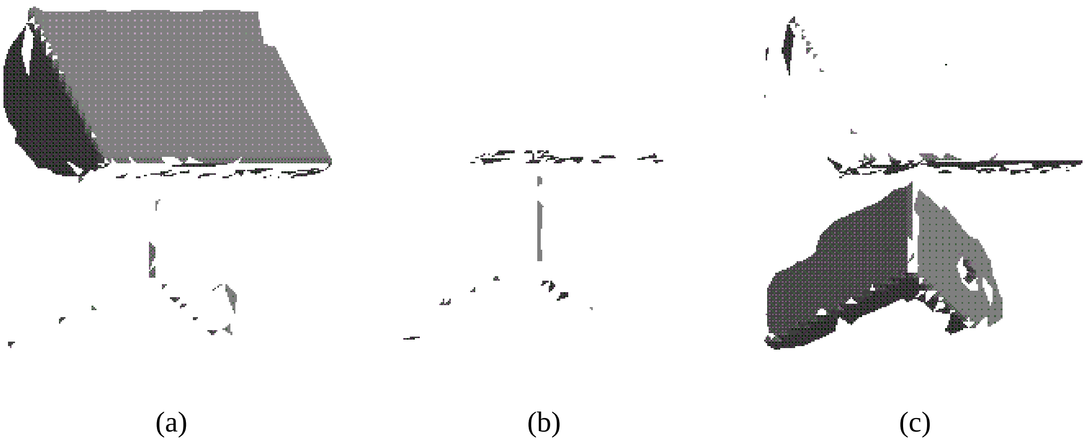

Fig. 7. (a),(b) & (c) present the scene facets which best match facets in Models 1, 2 & 3 respectively.

A Derivation of the Similarity Metric

In this section the derivation of a statistical metric for comparing binned measurements is presented. Given a random variable X, a statistical measure of the distance D between the endpoints $X = x$ and $X = x + \delta x$ of a short line is obtained by normalising by the standard deviation σ.

$$D = \frac{\delta x}{\sigma} \tag{2}$$

In general then, the statistical distance between any two points $X = s$ and $X = m$ can be determined by the definite integral:

$$D = \int_{s}^{m} \frac{dx}{\sigma} \tag{3}$$

For N independent measurements the statistical distance is given by a sum of squared components:

$$D^2 = \sum_i \left(\int_{s_i}^{m_i} \frac{dx_i}{\sigma_i}\right)^2 \tag{4}$$

It is well known that binned data conforms to a Poisson distribution and that the variance of a Poisson variable is equal to its mean. A statistical distance metric for binned data is then obtained by substitution of $\sigma_i = \sqrt{x_i}$.

$$D^2 = \sum_i \left(\int_{s_i}^{m_i} \frac{dx_i}{\sqrt{x_i}}\right)^2 \tag{5}$$

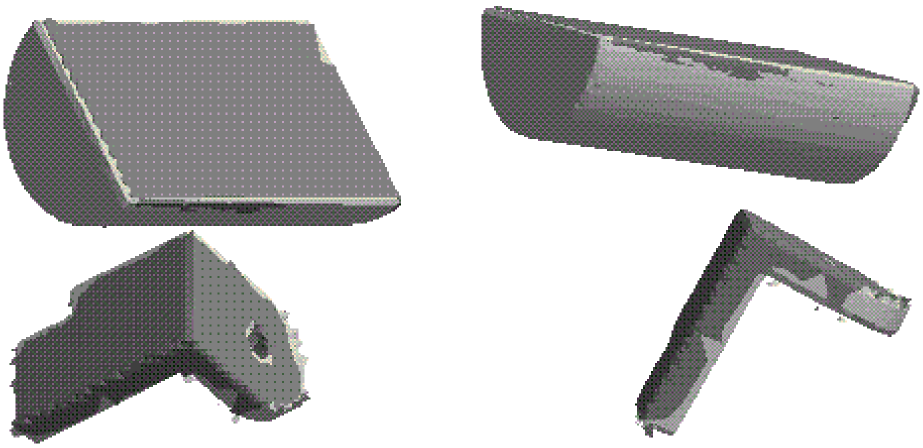

Fig. 8. The identification and localisation of the two objects in the scene. The scene data is presented in the darker shade and the models in the lighter shade. The second image presents the scene from a different view-point.

$$= 4 \sum_i (\sqrt{s_i} - \sqrt{m_i})^2 \tag{6}$$

Removing the constant factor in this expression gives the statistical metric proposed by Matusita [10] which is known as the Matusita distance.

$$D_{matusita} = \sum_i (\sqrt{s_i} - \sqrt{m_i})^2 \tag{7}$$

Expanding this expression gives:

$$D_{matusita} = \sum_i s_i + \sum_i m_i - \sum_i \sqrt{s_i}\sqrt{m_i} \tag{8}$$

If both m and s are normalised, or when using this metric to compare a single *scene* histogram with a set of normalised *model* histograms, this is simply:

$$D_{matusita} = const - \sum_i \sqrt{s_i}\sqrt{m_i} \tag{9}$$

Removing the constant results in the Bhattacharyya distance.

$$D_{bhattacharyya} = \sum_i \sqrt{s_i}\sqrt{m_i} \tag{10}$$

References

1. Bergevin, R., Laurendeau, D. and Poussart, D., “Registering Range Views of Multipart Objects”, CVIU, 61(1), pp1-16, 1995.
2. Besl, P. J. and McKay, N. D., ”A method for registration of 3-D shapes”, IEEE PAMI, 14(2), pp 239-256, 1992.
3. Eggert, D., Fitzgibbon, A. W. and Fisher, R. B., “Simultaneous registration of multiple range views for use in reverse engineering”, Proc. ICPR96, pp243-247, Vienna, 1996.
4. Evans, A. C., Thacker, N. A. and Mayhew, J. E. W., “The Use of Geometric Histograms for Model-Based Object Recognition”, Proc. BMVC93, pp429, 1993.
5. Faugeras, O. D. and Hebert, M., “A 3-D Recognition and Positioning Algorithm using Geometric Matching between Primitive Surfaces”, Proc. 8th IJCAI, pp-996-1002, 1983.
6. Fisher, R. B., “From Surfaces to Objects: Computer Vision and Three Dimensional Scene Analysis”, John Wiley & Sons, 1989.
7. Garland, M. and Heckbert, P. S., “Surface Simplification using Quadric Error Metrics”, SIGGRAPH97, pp209-216, 1997.
8. Hoppe, H., DeRose, T., Duchamp, T., McDonald, J. and Stuetzle, W., “Surface Reconstruction from Unorganised Points”, Computer Graphics, 26(2), pp71-78, 1992.
9. Johnson, A. E. and Hebert, M., “Recognizing Objects by Matching Oriented Points”, Proc. CVPR97, pp684-689, 1997.
10. Matusita, K., “Decision Rules Based on Distance for Problems of Fit, Two Samples and Estimation”, Ann. Mathematical Statistics, Vol. 26, pp. 631-641, 1955.
11. Stephens, R. S., “A Probabilistic Approach to the Hough Transform”, Proc. BMVC90, pp55-59, 1990.
12. Thacker, N. A., Aherne, F. J. and Rockett, P. I., “The Bhattacharyya Metric as an Absolute Similarity Measure for Frequency Coded Data”, STIPR97, 1st International Workshop on Statistical Techniques in Pattern Recognition, Prague, Czech Republic, 1997.
13. Thirion, J., “New Feature Points based on Geometric Invariants for 3D Image Registration”, IJCV, 18(2), pp121-137, 1996.