Using Partial Models for Localising Objects in Range Images

Alan M. McIvor

Imaging and Sensing Team, Industrial Research Limited
P. O. Box 2225, Auckland, New Zealand
a.mcivor@irl.cri.nz

Abstract: This paper describes a method for recognising and localising objects from a known class. Objects are modelled as a collection of simple geometric features and their spatial arrangement. Probability distributions on the parameters of the geometric features and their positions are used to handle the variability between members of the object class. Range images are processed to extract features (planes, spheres, cylinders) and these are matched against the model features using an interpretation tree based method to identify and locate the objects in a scene.

Keywords: Range images, object recognition, partial models.

1 Introduction

This paper describes how a model of certain features of an object and their spatial relationships can be used to recognise and determine the position of objects from range data. Such partial models are useful for objects that have free-form surfaces and those which exhibit significant variation between members of the same class. The features used in this paper are surface patches that have an underlying smooth form (planes, spheres, cylinders) with some higher spatial frequency perturbations imposed. The perturbations are modelled as random fluctuations. The features are detected in range data by segmentation operators tuned to extract the underlying smooth form and ignore the random fluctuations. Once extracted, the surface patches are matched with those in the partial model using an interpretation tree based approach. This aligns the partial model with the observed data and can be used to guide further processing of the object.

The system described in this paper can be divided into three main components:

1. Data acquisition and pre-processing.
2. Feature extraction and description.
3. Model-based matching.

The data being acquired is dense 3D positional data from the visible surfaces, as generated by range sensors such as temporally encoded structured light systems [17], laser profiling systems [15], etc.

Features are extracted using a range data segmentation algorithm [12]. This method has been tuned to extract the types of features used in the matching phase. Many other alternatives exist. In [1], a similar scheme is presented that describes the surfaces in terms of polynomial patches up to order 4. However, for other than planar surfaces, the polynomial parameters and order are dependent on the sensor viewpoint. This reduces the utility of this approach for model based matching. The method in [1] uses the principal quadric estimates only during the seed point selection phase of the region growing, and points are added to a regions only if their position is close to the current surface estimate. As well as using the principal quadric estimates for this purpose, the method described herein also uses them during the region growing phase to estimate the surface parameters (e.g., radius) and compares them to the current estimates when decided whether to add points to the current region.

Another approach is given in [14] where it is assumed that the 3D data has already been segmented into "homogeneous" regions of one surface type, but the particular type hasn't been identified. Planes are fitted to the data using total least squares. It is assumed that if the surface is a sphere or cylinder, its radius is from an apriori given finite set of possible values. Spheres and cylinders are fitted using a Hough technique for each radius value. The surface with the least sum of residual errors is selected. In this context, the Hough technique is used as a method for outlier rejection as well as for identifying the pose parameters (centre for a sphere, axis for a cylinder). Restricting the radius to be from a known set means that it is not necessary to estimate the principal quadric. In [5], another method is presented for fitting surfaces to 3D data which has already been segmented into smooth regions. Both of these two approaches suffer from the separation of region identification from region segmentation.

The final component of the system is a interpretation-tree based scheme for matching the features found in the range data with those in the object model, and to estimate the values of the variable parameters in the model. This is done using an approach very similar to that in [8], except that the constraints used aren't as powerful because they need to allow for the variability. A similar approach is used in [16], although this uses edges as the features for matching. Other matching strategies include relational matching [4], pose clustering [9], geometric hashing [10], and extended Gaussian images [3].

The rest of this paper is organised as follows. Section 2 describes the processing of the 3D data. Section 3 defines a partial model, and describes how they are matched with the 3D data. Section 4 gives an example of the system in a meat processing application. Finally, the paper is concluded in Section 5.

2 Range Data Processing

Range data is gathered from the object of interest using an appropriate range sensing device. The first stage of processing is to isolate the data from the object. It is assumed that the object is resting on a surface of known structure, e.g., a planar work-station. The surface of this is scanned when no objects are on it and this data is used to build a model. All data points in subsequent object scans that are within a small distance of the background surface are assumed to be from the background and removed from further processing. The distance threshold is chosen based on the statistics of the background fitting procedure. For example, if a planar model is fitted by total least squares, the threshold could be selected as 2.54 times the standard deviation of the residual error. This isolates surface data from the part of the object's surface that is visible. In this system, a complete object surface is formed by adding the silhouette projected onto the background surface. A more accurate complete surface model could be generated by taking range images from multiple viewpoints and merging the data together.

The next stage is segmentation of the range image data to identify planar, spherical, and cylindrical patches [13]. This segmentation scheme is similar to the variable order surface fitting scheme in [2], with the most significant difference being that object-based surface patch parameterisations are used instead of sensor-based ones. The former are clearly needed for object recognition.

The principal quadric at each data point is estimated by fitting cubic facets to the components of 3D point locations in a neighbourhood of the data point. The coefficients of the fitted facets are then used to estimate the first and second order derivatives of location with respect to the parameters of the range map. And from these the principal quadric can be calculated [11]. The aspects of the principal quadric that are important in the subsequent processing are the surface normal ($\mathbf{n}$), the directions of principal curvature, and the Gaussian (K) and mean (H) curvature. As well as estimating the principal quadric, the above procedure also filters the data to estimate the corresponding point on the underlying smooth surface, and generates an estimate of the noise standard deviation in each data point's neighbourhood. These estimates are averaged over the whole image to generate a global noise standard deviation estimate.

Planar points, at which the Gaussian and mean curvatures are both zero, are identified by thresholding the magnitude of the Gaussian and mean curvatures estimates. The selected thresholds are a function of the global noise standard deviation estimate. Because the thresholds will be non-zero, a connected set of planar points will not necessarily correspond to a planar region. For example, it could be a cylindrical region of large radius, or two adjacent planar regions with a smooth gently curved transition between them. One approach to identifying such situations is to look at the distribution of surface normals over the set of planar points [6]. The approach taken here is to extract planar regions from within the sets of planar points by region growing. A seed point is chosen from points at which $2H^2 - K$ (the sum of the squared principal curvatures) is minimised. Connected planar points are added if their position is consistent with the plane position, and their surface normal is consistent with the normal to the plane. Consistent means that the parameter being tested is within a validation gate centred on the current estimate, where the size of the gate varies with the parameter estimate's standard deviation. As points are added to the plane, the parameter estimates, and the standard deviation thereof, are updated. At this stage of the processing, the 3D data is divided into some subsets corresponding to planar regions, plus all the other points.

From amongst the points not assigned to planar regions, umbilic points are identified. An umbilic point is a point at which the two principal curvatures have the same value, or equivalently at which $H^2 - K$ is zero. Points on a spherical surface are umbilic points. Spherical patches are extracted from the set of umbilic points by region growing in a similar way to the extraction of planar patches. Finally, from amongst the points that are not assigned to a planar or spherical patch, parabolic points are identified, and cylindrical surface patches extracted by region growing from amongst them. Patches whose area is less than a given number of pixels are rejected as noise.

3 Partial Models and Model Matching

A partial model of an object describes parts of its surface in terms of surface patches of the types extractable from range data, namely planes, spheres, and cylinders. Also part of the model are probability distributions on the surface parameters (e.g., cylinder radii) and their position in the object coordinate frame (or alternatively, the relative position between surfaces). Probability distributions are used so that the model describes a class of objects instead of an individual one so that the technique can be applied to naturally varying objects.

An important requirement of a partial model is that given the component parts, the coordinate frame can be identified. That is, the coordinate origin and the direction of the axes can be defined in terms of the intersections of the surface patches, etc.

An interpretation tree [7, 16] based approach is used to match observed surface patches to those in the partial model and thus identify the pose of the object. Two sets have to be matched: the set $\mathcal{O}$ of observed patches plus a node representing “model feature not observed; and the set $\mathcal{M}$ which consists of the model features augmented with a node that represents “not part of model”. The root node of the interpretation tree is the state where no possible matches have been assigned. The first level has nodes representing one observed feature matched against each member of $\mathcal{M}$. The other observed features are similarly matched at subsequent levels. The leaf nodes of this tree thus represent potential model identifications. The key to efficient evaluation of the leaf nodes is pruning the tree at high levels. Branches are eliminated if a node represents a match between surface patches of different type. Further branches are eliminated if the adjacent nodes represent matches which are not simultaneously visible from any viewpoint. Branches are also pruned if other pairwise properties like the distances and angles between the matched surfaces are incompatible with the probability distribution given for these in the model.

The interpretation tree pruning results in one or more potential matches. The one that maximises the likelihood of the probability distribution of model parameters is selected as the interpretation of the data.

4 Example

This section describes a simple example which illustrates some of the components of the system. In this example, 3D data from the shoulder of lamb shown in Figure 1(a) has been obtained with a temporally encoded structured light system [18]. The geometric components of a partial model for this are drawn in Figure 1(b). It consists of a cylinder which represents the shank, a plane which represents part of the thigh, and a plane that matches the artificial “back” of the object flush on the workstation floor. Note that the model contains surfaces of infinite extent - the boundaries shown in Figure 1(b) are illustrative only. The origin of the model coordinate frame is defined as being where the axis of the cylinder intersects with the top plane. The 'Y' axis of the model coordinate frame is aligned with the projection of the cylinder axis onto the floor plane, with the 'Z' axis perpendicular to the floor plane. The free parameters of the model are the radius of the cylinder, the angle between the cylinder axis and the 'Y' axis, and the orientation of the top plane. That is, 4 degrees of freedom. An acceptable range of values for these parameters was built up via experimentation, and represented as independent probability distributions.

The surface patches extracted from the 3D data are shown in Figure 1(c). Since the number of surface patches extracted from the 3D data is so small in this case, the matching problem is simple. Once the patches have been identified, the transformation from the model coordinate frame to the world coordinate frame can be calculated, and the values of the free parameters in the model estimated.

Once the object's location has been determined, the surface data can be used to control subsequent processing of the object. This is illustrated in Figure 1(d). The object's cross-sectional area has been calculated as a function of position on the 'Y' axis, and thus the volume distribution along the 'Y' axis can be estimated. This is used to determine where cutting planes should be placed to divide the object into equal volume pieces (4 in the example). In a more realistic situation, the surface model would be augmented with additional information such as the internal bone structure, and this would be taken into account when planning the cutting operation.

5 Conclusion

This paper has described a method for using partial models in localising objects from range image data. This is done by extracting surface patches of simple types (planes, spheres, cylinders) and using these as features in an interpretation tree based model matching scheme. The object models used are partial models only describing patches of the object's surface that are composed of the surface types extracted from the range data.

The partial model of an object describes features of the surface which can be reliably extracted from range data, their locations in an object coordinate frame, and a probability distribution which defines the range acceptable feature parameters (e.g., cylinder radii) and the relative location of the features. The 3D data from the object is first isolated from that from the background, and then the principal quadric is estimated at each object point. The segmentation algorithm uses the principal quadric information to find points which are planar, and within these, to find the planar patches. This is done using a region growing algorithm. Similarly, spherical patches are identified within the umbilic points, and cylindrical patches are identified within the parabolic points.

The observed surface patches are matched to the model surface patches using an interpretation tree. Once the matching has been done, the free parameters in the model are estimated. Further work needs to be done on this latter phase. This will be based on using a Bayesian network approach, closely coupled with the model matching phase.

Acknowledgement: This work was funded by the New Zealand Foundation for Research, Science, and Technology under Contract TO8806 Objective 1 "Machine Vision Techniques".

References

[1] Paul J. Besl. Active, optical range imaging sensors. *Machine Vision and Applications*, 1:127–152, 1988.

[2] Paul J. Besl and Ramesh C. Jain. Segmentation through variable-order surface segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(2):167–192, March 1988.

[3] Chitra Dorai and Anil K. Jain. COSMOS — a representation scheme for free-form surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(10):1115–1130, October 1997.

[4] Ting-Jun Fan, Gerard Medioni, and Ramakant Nevatia. Recognizing 3-D objects using surface descriptions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:1140–1157, 1989.

[5] P. J. Flynn and A. K. Jain. Surface classification: Hypothesis testing and parameter estimation. In *Computer Vision and Pattern Recognition*, pages 261–267, Ann Arbor, Michigan, June 1988. IEEE Computer Society Press.

[6] Miguel Garcia and Luis Basanez. Fast extraction of surface primitives from range images. In *13th IAPR International Conference on Pattern Recognition*, volume III, Applications and Robotic Systems, pages 568–572, Vienna, Austria, August 1996.

[7] W. E. L. Grimson. On the verification of hypothesized matches in model-based recognition. In *European Conference on Computer Vision*, pages 489–498, 1990.

[8] W. Eric L. Grimson. Object recognition by constrained search. In Hebert Freeman, editor, *Machine Vision for Three-Dimensional Scenes*, pages 73–107. Academic Press, 1990.

[9] Gongzhu Hu and George Stockman. Representation and segmentation of a cluttered scene using fused edge and surface data. In *IEEE CVPR89*, pages 313–318, 1989.

[10] Y. Lamdan and H. J. Wolfson. Geometric hashing: A general and efficient model-based recognition scheme. In *IEEE International Conference on Computer Vision*, pages 238–249, Tarpon Springs, Florida, December 1988.

[11] Alan M. McIvor. A cubic facet based method for estimating the principal quadric. In *Image and Vision Computing New Zealand*. Auckland University, December 1998.

[12] Alan M. McIvor, David W. Penman, and Peter T. Waltenberg. Segmentation of range data using principal quadric information. IRL Report 8710100-1-98, Industrial Research Limited, 1998.

[13] Alan M. McIvor and Peter T. Waltenberg. Recognition of simple curved surfaces from 3d surface data. In *Asian Conference on Computer Vision*, volume 1, pages 434–441. Hong Kong, Springer, January 1998. URL ftp://ftp.vision.irl.cri.nz/pub/doc/1998/accv98-ammptw.ps.

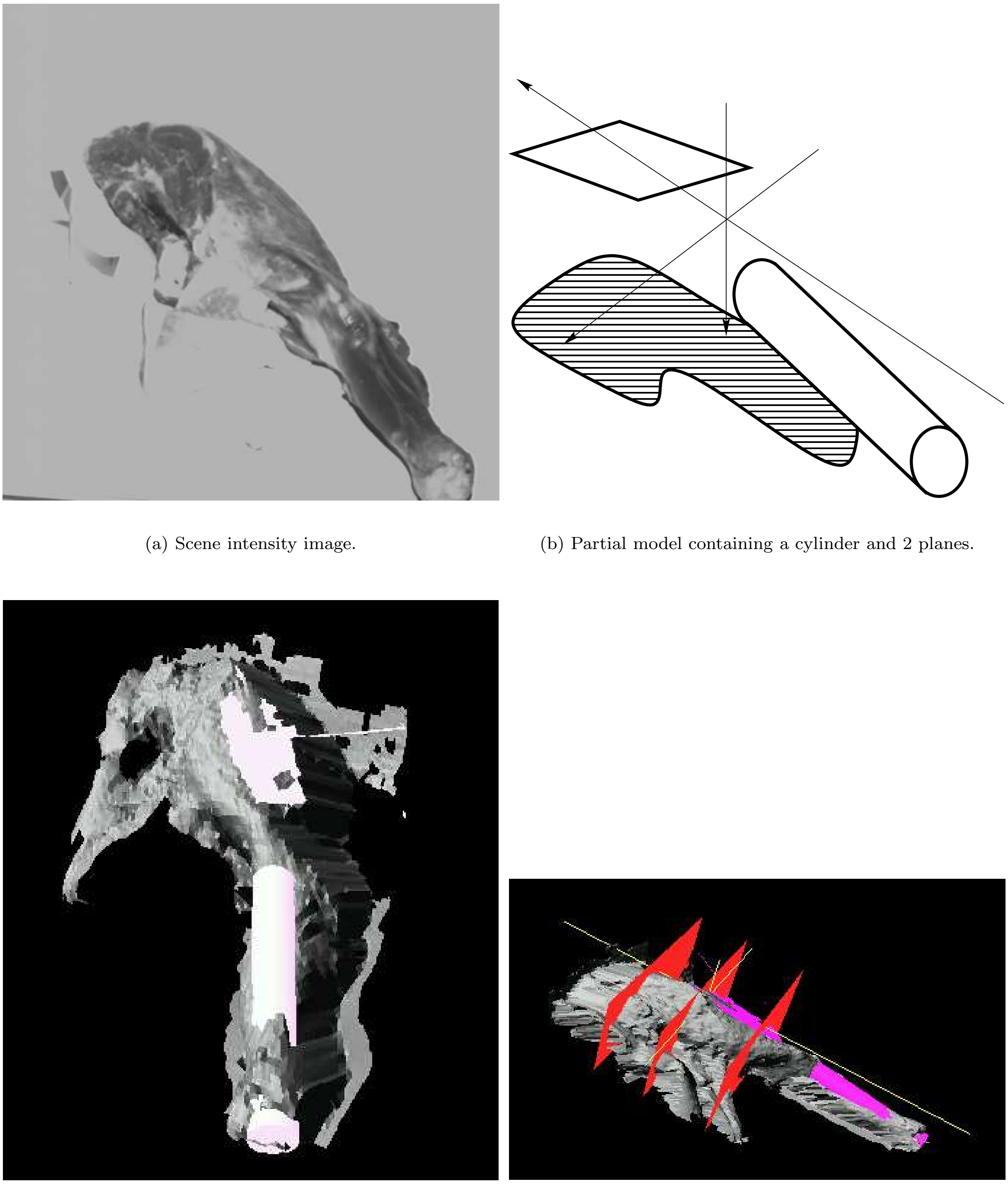
[14] Timothy S. Newman, Patrick J. Flynn, and Anil K. Jain. Model-based classification of quadric surfaces. *CVGIP: Image Understanding*, 58(2):235–249, September 1993.

[15] Tomas Pajdla. Laser plane range finder: The implementation at the CVL. Research Report K335-1995-98, Computer Vision Laboratory, Czech Technical University, CZ-121 35 Praha 2, Karlovo nam. 13, October 1995.

[16] Ian D. Reid and J. Mike Brady. Recognition of object classes from range data. *Artifical Intelligence*, 1995. Special Issue on Computer Vision.

[17] Kosuke Sato, Hiroyuki Yamamoto, and Seiji Inokuchi. Tuned range finder for high precision 3D data. In *International Conference on Pattern Recognition*, pages 1168–1171, Paris, France, October 1986. IAPR, IEEE Computer Society.

[18] R. J. Valkenburg and A. M. McIvor. Accurate 3D measurement using a structured light system. *Image and Vision Computing*, 16(2):99–110, February 1998.



(a) Scene intensity image.

(b) Partial model containing a cylinder and 2 planes.

(c) Surfaces extracted from range data. One planar patch and one cylindrical patch were detected.

(d) Partial model aligned with data and showing cutting planes.

Figure 1: Using the partial model approach to align a model with an object so that cutting planes can be identified for dividing the object into 4 equal volume parts.