Visually Salient 3D Model Acquisition from Range Data

Edvaldo M. Bispo, Andrew W. Fitzgibbon and Robert B. Fisher
Dept. of Artificial Intelligence,
University of Edinburgh,
5 Forrest Hill, Edinburgh EH1 2QL

Abstract

Automatic model building is a crucial requirement of any model-based vision system which must work in unknown environments. Even in the specialised environments where CAD models are available for the small number of parts, these models often lack visual saliency, impacting on the robustness (more than the mean accuracy) of the system using them. Moreover, the time needed to make such models by hand may be prohibitive, particularly with freeform curved objects; hence automatic model acquisition is greatly desirable.

We present a procedure which merges multiple range images of an unmodelled object to create a 3-D body-centred model of the part. By explicitly considering the specific problems of vision systems, we achieve a high-level, robust and accurate description of the unknown scene, where visual applicability of our generated model is guaranteed. The models are characterized by a pleasant 'intuitive' feel which allows easy operator intervention if they must be altered, perhaps to generate a new part from a similar example.

1 Introduction

Acquiring models of previously unknown objects which can later be used to describe, recognise and manipulate these objects is an ability of the human visual system which has been very poorly duplicated in machines. In both the graphics and computer vision fields, much work has been done on defining internal representations (largely on the graphics side), and automatically acquiring models from sensor information (being 'vision' for the most part). In both cases, however, the models record rather than evoke: models used by machines can *describe* the observed objects far better than a human could, but *recognition* is hampered by current limitations of representation and insufficient research into the problems of viewpoint stability and descriptive robustness — considerations to which we shall apply the description *visual saliency*. A system which purports to generate models for use as the visual memory of an autonomous system must carefully balance the dual goals of representational richness and ease of recognition.

The model acquisition problem has been approached many times in the past, *e.g.* the wire-frame models of Mayhew, Porril and Pollard [10]; here we will deal only with range data based systems. Agin [1] described a program for fitting generalized cylinders to real data. Popplestone and Brown [11] created planar and cylindrical surface models from light-stripe range data. Vermuri and Aggarwal [13] proposed the construction of multiple-view three dimensional models using range and intensity data. Potmesil [12] demonstrated a methodology for the automatic generation of computer models of the surfaces of arbitrary three-dimensional objects. Chen and Medioni [4] showed a technique for fusing range image data from different object views in order to build complete descriptions of the object.

The main example of this paper uses a Renault truck part, to demonstrate the advances in model acquisition compared to the landmark work of Faugeras and Hebert [5]. They acquired a planar patch tessellation of complex objects having curved and planar surfaces which was then used for object recognition and part location. The research reported here improves on their work by using curved surface primitives and annotating the model with other information useful for recognition.

Our system for automatically building models from range data, called SMSGEN (Sugestive Modelling System Generator), can be illustrated by the block diagram of Figure 1 that shows our basic strategy towards model acquisition:

1. Acquire range images of different views of the object being modelled.
2. Build models corresponding to each of these views using the results of the segmentation of the range images.
3. Match the different models to calculate the registration between them.
4. Merge the different models in an unique model.

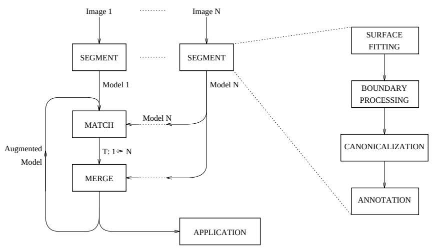

Figure 1: Overall system architecture.

In the next sections, we describe the main modules of the SMSGEN, present some current examples of acquired models and an example of the recognition of a complex object by the Imagine 2 recognition system [6] using an automatic acquired model.

2 Segmentation

Range images of an object are acquired using two-camera laser triangulation. Depth values from the triangulation are measured to 0.1 mm accuracy with the aid of a sub-pixel stripe detection algorithm [8].

The acquired range image $G(x, y)$ is segmented into bivariate polynomial patches in order to extract the major component features of the scene. The segmentation algorithm is based on that described by Besl and Jain [3] with some important modifications. The basic algorithm is region growing with well-chosen seeds:

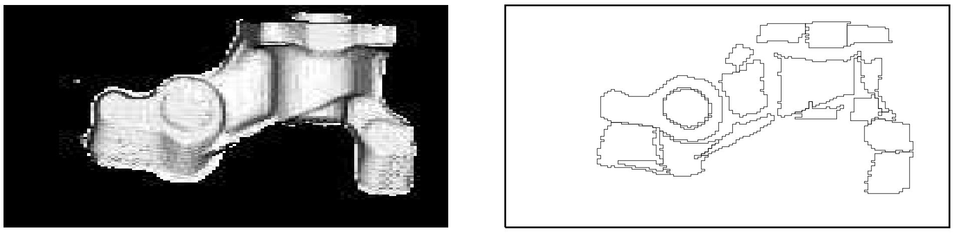

Figure 2: Example cosine shaded range image and the resulting segmentation (Renault truck part).

1. Classify each image pixel according to the signs of locally-estimated mean and Gaussian curvatures. Regions of similarly classified pixels are morphologically eroded for use as seed regions for the region growing process.

2. Using the seed regions from (1) iteratively least-squares fit to the region pixels, expand to include any connected pixels which are 'near' to the surface defined by the least squares solution, and re-fit. When no new pixels are included by the new fit parameters, the process terminates.

While Besl's algorithm was largely concerned with data reduction and reconstruction, object recognition has the additional requirements of repeatability and stability under changes of viewpoint. Comparing the segmentation in Figure 2 to that shown in Besl's paper [3], we see that our modifications do indeed provide a more 'intuitive' segmentation.

The output of segmentation is a list of (canonical *surface*, reference-frame *transformation*) descriptors. The surface is either a plane, cylinder or general biquadratic patch in canonical position, related to the image region by the supplied transformation. In addition, properties such as area, average curvatures and elongation are calculated and added to the descriptions.

3 Object Models

The models considered for analysis are assemblies of non-infinite 2^{nd} order surfaces, described using the Suggestive Modelling System (SMS) [6]. SMS surface models are characterised by the separation of the description into *shape, extent* and *position* [7]. An *assembly* is a set of surfaces and reference-frame transformations:

$$\mathcal{A} \;=\; \left\{(\mathcal{S}_i,\; {}^{\mathcal{A}}T_{\mathcal{S}_i})\right\}_{i=1}^{m}$$

In this notation, the transformation ${}^{\mathcal{A}}T_{\mathcal{S}}$ transforms points in the surface $\mathcal{S}_i$'s reference frame into the assembly $\mathcal{A}$'s reference frame. The surface primitive $\mathcal{S}$ is on one side of the infinite surface defined by its equation $\mathcal{F}$:

$$\mathcal{F}(x,y,z) = a_{xx}x^2 + a_{yy}y^2 + a_{zz}z^2 + a_{xy}xy + a_{yz}yz + a_{zx}zx + a_x x + a_y y + a_z z + a_0 \;=\; 0$$

While this implicit equation allows us to express any second order surface, in practice we are limited to those surfaces which can be reliably extracted from the sensor data. Currently this means restriction to Planes, Cylinders and Ellipsoids and Hyperbolic Paraboloids. Given this restriction, the surfaces chosen will have certain limitations on the values of their parameters.

Besides describing the object being modelled in terms of its surfaces, SMS surface based models also contain some extra information, called *properties*, that is added to the model in order to make the recognition process more efficient. The SMS paradigm allows the inclusion of a very large set of different properties in the model [6]. Among all the possible properties that could be added to the model, we decided to use the ones that are most easily extracted from the range data. Also, because using all the properties that could be obtained from the segmented image would mean an excessively complex model, we restrict the number of the properties added in order to obtain a SMS model that is complete without being excessively detailed. The SMS properties added to the model were: classification of the patches according to their shape (plane, ellipsoid, cylinder, hyperboloid) and curvature, area of patches, maximum and minimum curvatures of patches, adjacency information, relative size of adjacent surfaces, relative orientation of planes.

4 Building single view surface-based models

The construction of a SMS model involves two steps:

1. Enumeration of all the relevant surfaces in the object and definition of their:

   shape (classification of the surface as plane, cylinder, elliptical or hyperbolic paraboloids),

   extent (surface boundary and one point belonging to the surface) and

   position (rigid transformation associating the position of the surface in the object and the canonical position of the surface in the SMS paradigm).

2. Addition of the SMS properties to the model.

In the SMSGEN system the two steps described above are carried out using the output of the segmentation process.

The classification of the surfaces *shape* is directly computed during the segmentation through the analysis of the coefficients of the extracted surface primitives from the image.

The segmentation also produces polylines approximations of the boundary for each surface segmented in the image. These boundaries and the use of one point belonging to each surface define the *extent* of the surfaces in the model.

The *position* of each of the surfaces of the model is also explicitly calculated during the segmentation process by incorporating ${}^{\mathcal{A}}T_{\mathcal{S}}$ into the least-square fitting process.

Finally, the SMS properties are defined by considering the information in the segmentation output and assuming some *ad hoc* tolerances dictated by the experience: 20% of variation around nominal value.

5 View merging

To create a complete model of a physical object, one has to merge different models or descriptions of the object taken from different points of view.

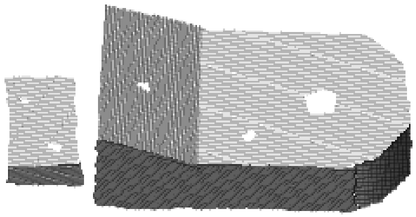

Figure 3: Automatically generated model from first view of widget

The first step in merging two different views of an object is solving the *registration problem*, *i.e.* finding the reference frame transformation that relates the two different views of the object being modelled. In our work the calculation of the geometric transformation between the two object model views uses pairs of *corresponding directions* and *corresponding points* derived from pairs of patches supplied as input to the merging program.

The pairs of *corresponding directions* were used to calculate the rotation between the views and the pairs of *corresponding points* were used to calculate the translation. Planar patch centroids were used as the corresponding points, which given the accuracy of segmentation and in the absence of occlusion are sufficiently accurate. The corresponding directions were obtained by using the planar patch normals or the directions defined by two different planar patch center points in a same view. The registration between the two object views was calculated using a technique suggested in [2] based in the use of SVD decomposition. Improved estimates might use this as an initial guess and then use the method of [4] to improve the accuracy.

After the *registration problem* is solved the next step consists in *integrating* the two different views of the object. Patches that appear just in one view are just added to the merged model without any alteration. When a patch appear in both views it is made a manual choice of which view of the patch to use in the merged model.

6 Results

Examples of the results obtained with the SMSGEN system can be seen in Figures 3, 4, 5 that show SMS models corresponding to a widget. Figures 6 and 7 show the SMS models corresponding to two different views of a Renault truck part (see Figure 2). The model corresponding to the merging of the models of the two different views of the Renault truck part can be seen in Figure 8.

The precision of the registration process was reasonably good: after two views

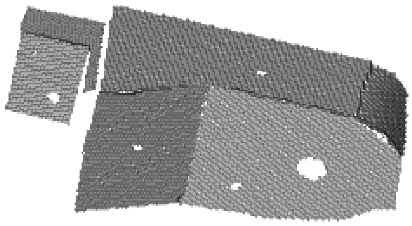

Figure 4: Model generated from second view of widget

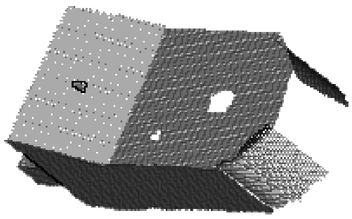

Figure 5: Model created by merging descriptions from the two views shown in Figures 3 and 4.

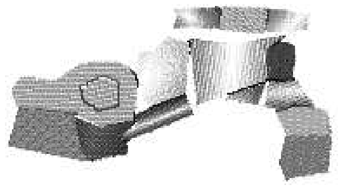

Figure 6: Automatically generated model from first view

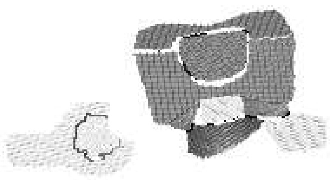

Figure 7: Model generated from second view

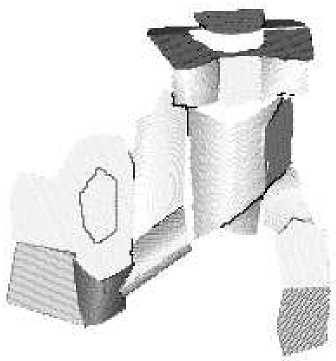

Figure 8: Model created by merging descriptions from the two views shown in Figures 6 and 7.

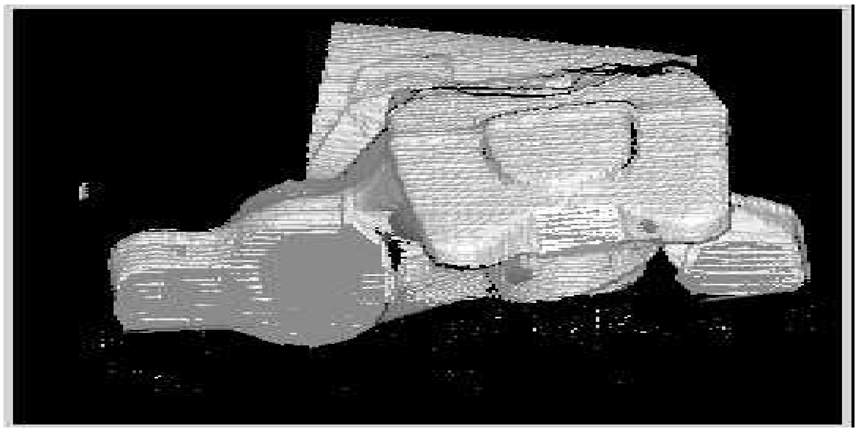

Figure 9: Overlay of model derived from view 1 onto new data, using position estimated by Imagine 2 scene analysis program.

were aligned we observed a maximum error angle of 2.2° between the pairs of corresponding directions and a maximum error distance of 1.7 mm between the corresponding points.

To illustrate the use of the models in the recognition of the test part, we used the model from Figure 6 to recognize the object in the scene from which Figure 7 was extracted in the Imagine 2 recognition system [6]. The Imagine system was able to recognize the object, and also to estimate its pose with a reasonable accuracy. To illustrate the accuracy of the pose estimation, Figure 9 shows the model of Figure 6 projected onto a new image. Note that the position accuracy is such that the range data (light grey) interleaves the projected model (dark grey) suggesting position accuracy within the noise level of the range data (0.15 mm).

7 Conclusions

There are still many improvements to be done to the SMSGEN system:

Improvement in the boundary quality: development of more sophisticated algorithms able to fit lines and second order curves in the polyline approximations of the boundaries.

Registration between views: implementation of algorithms for improving the accuracy of the registration and allowing its completely automatisation.

Integration of different views: development of integration techniques to integrate patches of different views of the same surface into one unique patch. A possible approach to fuse these patches would be an extension of Orr's algorithm [9] for combining two views of the same polygonal surface boundary that considered other kinds of surfaces besides planes.

Operational analysis: study of the variation of the model performance in object recognition with the set of SMS properties added to it. The objective would be to find the ideal set of SMS properties to improve the efficiency of the recognition process.

However, we have successfully developed an automatic model acquisition process that produces 3D geometric models of objects with curved surfaces while ensuring that the models remain useful for object recognition. The models are geometrically complex compared to previous curved surface models, and are richer and more accurate than the complex planar patch approximation of Faugeras and Hebert [5]. Attention to the dictum of visual saliency means that subsequent object recognition using the model is greatly simplified.

8 Acknowledgements

We would like to acknowledge the financial support of the CNPq in Brazil to Edvaldo Marques Bispo.

References

[1] G. J. Agin. *Representation and description of curved objects.* PhD thesis, AIM-173, Stanford AI Lab, 1972.

[2] K. S. Arun, T. S. Huang, and S. D. Blostein. Least-squares fitting of two 3-d point sets. *IEEE Transactions on Pattern Analyses and Machine Intelligence*, 9(5), September 1987.

[3] P. J. Besl and R. C. Jain. Segmentation through variable-order surface fitting. *IEEE T-PAMI*, 10(2):167–192, Mar. 1988.

[4] Y. Chen and G. Medioni. Object modelling by registration of multiple range images. *Imagine and Vision Computing*, 10(3), Apr. 1992.

[5] O. Faugeras and M. Hebert. The representation, recognition, and locating of 3-d objects. *Int. J. Robotics Research*, 5(3):27–52, 1986.

[6] R. B. Fisher. SMS: A suggestive modelling system for computer vision. *IVC*, 5(2):98–104, May 1986.

[7] A. W. Fitzgibbon and R. B. Fisher. Suggestive modeling for machine vision. In *Proceedings, OE/Technology '92*. SPIE, November 1992.

[8] D. K. Naidu and R. B. Fisher. A comparative analysis of algorithms for determining the position of a stripe to subpixel accuracy. In *Proceedings, BMVC*. pp 217-225, 1991.

[9] M. Orr, J. Hallam, and R. Fisher. Fusion through interpretation. In *Proc. 2nd European Conf. on Computer Vision*. St. Margherita Ligure, Italy, pp 801-805, 1992.

[10] A. P. Pollard, S.B., J. Porril, and J. E. W. Mayhew. Recovering partial 3d wire frames descriptions from stereo data. In *Proc. 1990 Britsh Machine Vision Conference*. pp 31-36, 1990.

[11] R. J. Popplestone, C. M. Brown, A. P. Ambler, and G. F. Crawford. Forming models of plane-and-cylinder faceted bodies from light stripes. In *Proceedings, 4th IJCAI*. pp 664-668, 1975.

[12] M. Potmesil. Generating models of solid objects by matching 3d surface segments. In *Proceedings, 8th IJCAI*. Karlshure, Germany, pp 1089-1093, Aug. 1983.

[13] B. C. Vermuri and J. K. Aggarwal. 3-d model construction from multiple views using range and intensity data. In *Proceedings, Conf. on Comput. Vision and Pattern Recognition*. Miami Beach, Florida, pp 435-437, 1986.