Axel Pinz (Ed.)

Pattern Recognition 1996

Proceedings of the 20th Workshop of the Austrian Association for Pattern Recognition (ÖAGM/AAPR)
Schloß Seggau, Leibnitz, 9-10 May 1996

R.Oldenbourg Wien München 1996

Simultaneous Recovery of Surface and Superquadric Models*

Aleš Leonardis[†], *Aleš Jaklič, Bojan Kverh, Franc Solina*

Computer Vision Laboratory, Faculty of Computer and Information Science
University of Ljubljana, Tržaška 25, 1001 Ljubljana, Slovenia
E-mail: {alesl, alesj, bojank, franc}@fri.uni-lj.si

Abstract:

We propose a new approach to simultaneous recovery of surface and volumetric models from range images. To achieve a compact and *accurate shape representation at the same time, only one particular shape model type is usually not enough. A combination of different types of models is thus needed where each model determines its own domain of applicability. Our approach is based on the* recover-and-select *paradigm which consists of two intertwined stages: model-recovery and model-selection. At the model-recovery stage a redundant set of surface patches and superquadrics is initiated in the image and allowed to grow, which involves an iterative procedure combining data classification and parameter estimation. All the recovered models of both types are passed to the model-selection procedure where only the models resulting in the simplest overall description are selected. We present experimental results on real range images.*

Key words: range image interpretation, recovery of surface patches, recovery of superquadrics, recover-and-select paradigm.

1 Introduction and Motivation

It is becoming clear in the vision community that a single complete and general geometrical model that would be applicable for all kinds and levels of shape representation is not feasible. So far, many different models have been used for modeling different aspects of objects and

* This work was supported by the Ministry of Science and Technology of Republic of Slovenia (Project J2-6187), European Union Copernicus Program (Grant 1068 RECCAD), and by U.S.–Slovene Joint Board (Project #95-158). A. L. also acknowledges the support from the Austrian National Fonds zur Förderung der wissenschaftlichen Forschung (No. S7002MAT).

† Also with the Department for Pattern Recognition and Image Processing, Technical University Vienna, Austria.

scenes, e.g., surface models [1] and volumetric models [9]. However, only a single model type has usually been applied at a time or a volumetric model has been built as a composition of surface patches [2]. Volumetric models such as superquadrics are good for capturing the global geometry of objects, especially for part-level description, but they lack the malleability to model small scale features. While it is possible to enchance volumetric models with local deformations, the huge number of parameters makes recovery from images unstable and difficult. We believe that instead of enhancing volumetric models with additional layers of local deformations a better solution is to use two or more different types of complementary models. In this way we can keep the advantages of each model type. Volumetric models, for example are good for abstraction (robust shape indexing, grasping, etc.) while surface models give a rich description needed for recognition.

Let us consider a few examples. Mobile robots that have to perform in a complex environment must concentrate only on those features of the environment that are relevant for their goal and cannot waste their resources on perceiving everything that is within the view of their sensors. If a wheeled mobile robot has to move in a specific direction, its vision system must find flat surfaces, free of obstacles, on which it can drive forward. Appropriate, "natural models" for such a goal could be, for example, planar patches for the driving surfaces and volumetric models for the obstacles. The scale of the planar patches, i.e., what is "flat", is determined by the size of the vehicle (radius of wheels) so that smaller, for the task unrelevant features can be ignored. Therefore, one should select a shape representation according to specific assumptions or for specific tasks [10]. Our goal is to develop an approach that would be able to deal with multiple types of models from the very beginning.

The key to successful combination of different types of models is the *recover-and-select* paradigm which gives the simplest overall description with the selected type of models and which achieves segmentation and model recovery at the same time. We have already presented results of range image interpretation under this paradigm using surface models [6] and superquadrics [8, 7] *separately*. Here we propose to combine these two types of models in a single process. The result is a compact but accurate range image description consisting of both types of models, which are selected simultaneously employing the Minimum description length principle as proposed in [10]. Shapes that can be described as superquadric shapes are modeled as such, while the rest of the data (where applicable) is modeled with surface patches.

A similar multi-model scheme is being developed by the McGill group [11]. They propose to model the range data with superquadrics first and then fit to the residual surface patches at several scales. We think that such consecutive recovery might result in suboptimal description. Our method tries to avoid this problem by parallel recovery of both types of models and MDL-like selection of the most compact description.

2 Recover-and-Select Paradigm

We present here a brief general outline of the recover-and-select paradigm and emphasize those features that are important for dealing simultaneously with models of different types. For details the reader is referred to [5].

The paradigm consists of two intertwined stages: model-recovery and model-selection. At the model-recovery stage, a redundant set of models is *independently* (possibly in parallel) initiated in the image and allowed to grow, which involves an iterative procedure combining data classification and parameter estimation. Since all of the models are initiated independently, they can be of completely different types involving both linear and/or nonlinear estimation procedures to determine the parameters of the models. The final outcome of the model recovery procedure for a model m_i consists of three terms:

1. The region R_i, which represents the domain of the model and encompasses $n_i = |R_i|$ image elements that belong to the model,
2. the set of model parameters $\mathbf{a}_i$ (N_i denotes the cardinality of this set), and
3. the error-of-fit measure ξ_i which evaluates the conformity between the data and the model.

While this description is general, i.e., independent of a particular choice of models, specific procedures designed to operate on individual type of models can differ significantly. This is primarily due to the matching and fitting processes which depend on the type of models and on the choice in defining the measure of the distance between the model and the data.

All the recovered models are then passed to the selection-procedure which is defined as the quadratic Boolean problem based on the MDL principle.

How can we apply this principle to our problem? Let vector $\mathbf{m}^T = [m_1, m_2, \ldots, m_M]$ denote a set of models, where m_i is a *presence-variable* having the value 1 for the presence of a model and 0 for its absence in the final description, and M is the number of all models. The length of encoding of an image L_{image} can be given as the sum of two terms

$$L_{\text{image}}(\mathbf{m}) = L_{\text{pointwise}}(\mathbf{m}) + L_{\text{models}}(\mathbf{m}) \ . \tag{1}$$

$L_{\text{pointwise}}(\mathbf{m})$ is the length of encoding of individual data points that are not described by any model, and $L_{\text{models}}(\mathbf{m})$ is the length of encoding of data described by the selected models. The idea is to select a subset of models that would yield the shortest length description. In other words, we should tend to maximize the *efficiency* of the description, defined as

$$E = 1 - \frac{L_{\text{image}}(\mathbf{m})}{L_{\text{pointwise}}(\mathbf{0})} \ , \tag{2}$$

where $L_{\text{pointwise}}(\mathbf{0})$ denotes the length of encoding of the input data in the absence of models.

Alternatively, we can define a quantity S which represents the *savings* in the length of encoding in the presence of models[1]

$$\begin{aligned} S &= \text{length of encoding of the data in the absence of models} - \\ &\quad \text{length of encoding of the data in the presence of models} = \\ &= L_{\text{pointwise}}(\mathbf{0}) - L_{\text{image}}(\mathbf{m}) \ . \end{aligned} \tag{3}$$

The question is how to translate the above equations into our particular case using the outcome of the model recovery procedure. Remember that the output of the model recovery procedure for the i-th model (M_i, R_i) consists of three terms.

Analogous to equation (1) we can write

$$L_{\text{image}}(\mathbf{m}) = \mathrm{K}_1(n_{\text{all}} - n(\mathbf{m})) + \mathrm{K}_2\xi(\mathbf{m}) + \mathrm{K}_3 N(\mathbf{m}) \ , \tag{4}$$

where n_{all} denotes the number of all data points in the input and $n(\mathbf{m})$ the number of data points that are explained by the selected models. $N(\mathbf{m})$ specifies the number of parameters which are needed to describe the selected models and $\xi(\mathbf{m})$ gives the deviation between the models and the data that these models describe. K_1, K_2, K_3 are weights which can be determined on a purely information-theoretical basis (in terms of bits), or they can be adjusted in order to express the preference for a particular type of description.

Now we can state the task as follows: Find $\hat{\mathbf{m}}$ such that

$$L_{\text{image}}(\hat{\mathbf{m}}) = \min_{\mathbf{m}} L_{\text{image}}(\mathbf{m}) \ . \tag{5}$$

Since n_{all} is constant, minimization of equation (4) is equivalent to maximizing the expression

$$F(\hat{\mathbf{m}}) = \max_{\mathbf{m}} F(\mathbf{m}) = \mathrm{K}_1 n(\mathbf{m}) - \mathrm{K}_2\xi(\mathbf{m}) - \mathrm{K}_3 N(\mathbf{m}) \ . \tag{6}$$

This equation supports our intuitive thinking that an encoding is efficient if

- the number of data points described by a model is large,
- the deviations between the model and the data are low,
- while at the same time the number of model parameters is small.

The optimization function has been presented on a general level. For technical details, as well as for a discussion on different methods that can be used to solve the optimization problem, the reader is referred to [5, 7].

The solution is a subset of all recovered models which yields the simplest overall description. To increase the computational efficiency of the method, model-recovery and model-selection are

[1] A similar definition was also proposed by Fua and Hanson [3].

combined in an iterative fashion. The recovery of currently active models is interrupted by the model-selection procedure which selects a set of currently optimal models that are then passed back to the model-recovery procedure. This process is repeated until the remaining models are completely recovered.

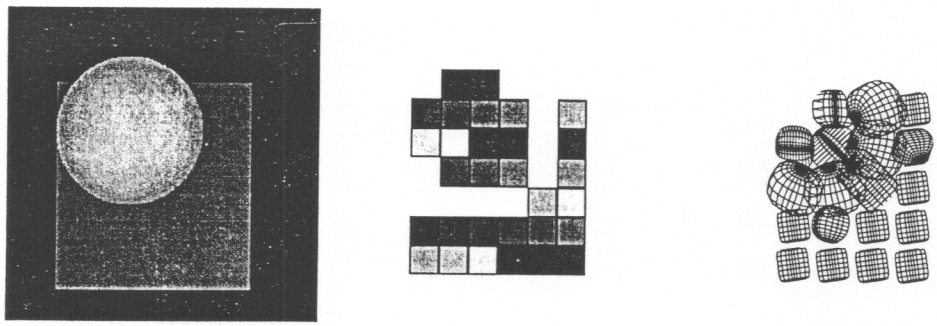
In our specific case, the admissible model types consist of surface patches (planar and second order) and superquadrics. At the model-recovery stage models determine their own domain of applicability by developing (growing) only where there is a close correspondence between the data and the models. Then, at the model-selection stage, the recovered models (many of whom may be partially or even completely overlapped) compete (on the basis of the overall simplicity) to be selected in the final description, meaning that the selection procedure determines which models are better suited to describe different parts of the image.

In the next section we discuss some of the specific issues that pertain to the problem of combining surface and volumetric models in a common framework and show some experimental results.

3 Experimental Results

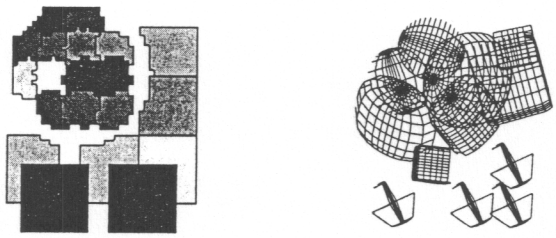
One of the subtle issues when combining different types of models in the recover-and-select paradigm is when to perform the selection. Although the answer is quite well argumented in the case of surface models [5], it is easy to show that the selection process over volumetric and surface models before these descriptions are fully grown can produce spurious results. This can happen because during the growth, a single volumetric model might locally model the image much better than a set of corresponding surface patches and consequently surface patches are rejected. However, if both the volumetric models and the surface patches were fully grown, the surface patches would be selected. This is even more so if a simple greedy algorithm is used for the selection.

The experiments were performed using an object-oriented framework for image segmentation based on the recover-and-select paradigm (Segmentor [4]). To avoid the question when to perform the selection, we postponed this decision until all the surface and volumetric models were fully grown. In short, we first independently recovered the surfaces using the recover-and-select paradigm, then did the same for the volumetric models, and finally selected models from both sets of models. The overall procedure thus exhibits recover-*then*-select paradigm, while recover-*and*-select is used for a particular type of models. We present the results on a synthetic range image (Figs. 1–5) and on a real range image (Figs. 6–10).



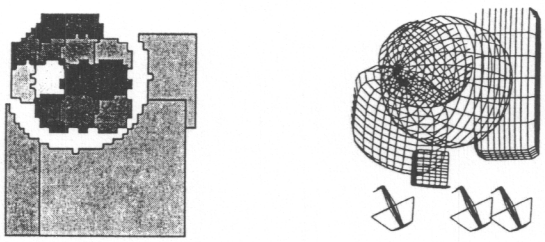
(a) (b) (c)

Figure 1: a) Synthetic range image, b) initial surface models, c) initial volumetric models.



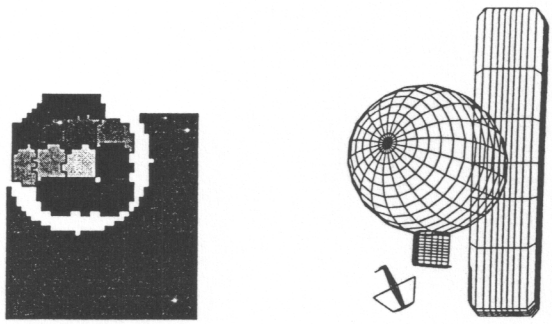
(a) (b)

Figure 2: First recover-and-select iterations for a) surface models, b) volumetric models.



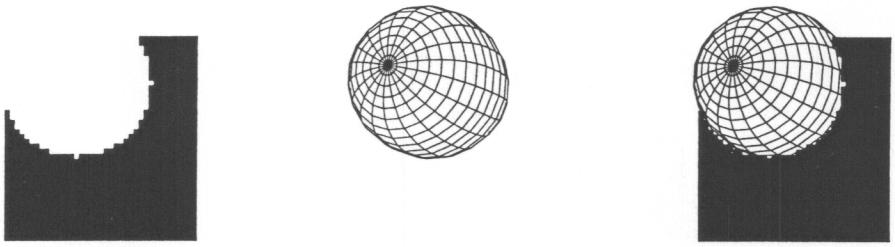
(a) (b)

Figure 3: Second recover-and-select iterations for a) surface models, b) volumetric models.



(a) (b)

Figure 4: Independently recovered a) surface models and b) volumetric models.



(a) (b) (c)

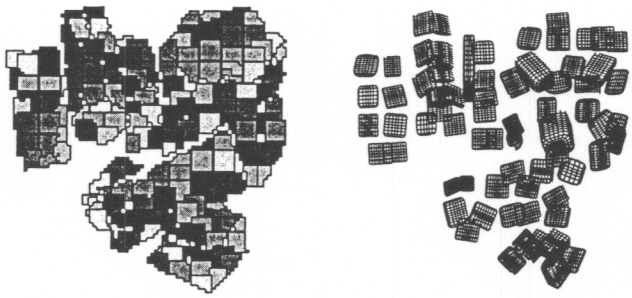
Figure 5: Final result of the selection (recover-then-select): a) surface models after the selection, b) volumetric models after the selection, c) surface and volumetric models.

4 Future work

In our summary of the paper we have proposed an approach to simultaneous recovery of surface and volumetric models from range images. Future work will be directed towards studying the relations between different types of models. Namely, a volumetric model implicitly and explicitly encodes much more information than a surface model and the question is how we can account for this in the objective function based on the MDL principle. Besides, we plan to expand our set of admissible model types since the proposed general framework poses no conceptual barriers to their number.

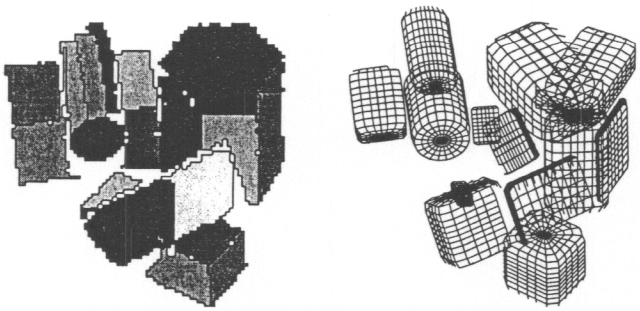
(a) (b) (c)

Figure 6: a) Real range image, b) initial surface models, c) initial volumetric models.



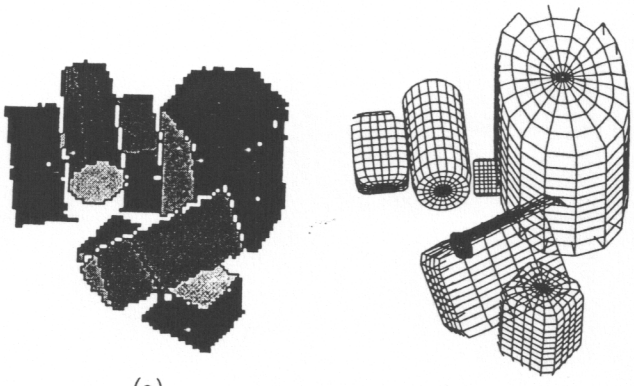
(a) (b)

Figure 7: First recover-and-select iterations for a) surface models, b) volumetric models.



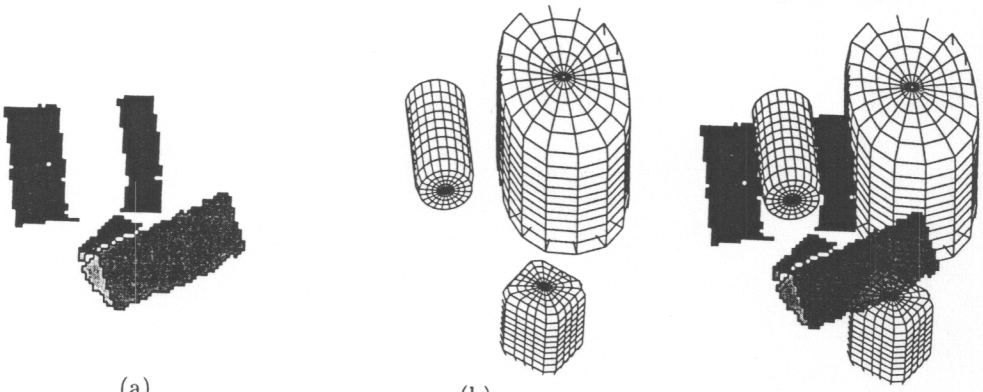
(a) (b)

Figure 8: Second recover-and-select iterations for a) surface models, b) volumetric models.



(a) (b)

Figure 9: Independently recovered a) surface models and b) volumetric models.



(a) (b) (c)

Figure 10: Final result of the selection (recover-then-select): a) surface models after the selection, b) volumetric models after the selection, c) surface and volumetric models.

References

[1] P. J. Besl. *Surfaces in Range Image Understanding.* Springer-Verlag, 1988.

[2] F. P. Ferrie, J. Lagarde, and P. Whaite. Darboux frames, snakes, and super-quadrics: geometry from the bottom up. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(8):771–784, 1993.

[3] P. Fua and A. J. Hanson. Objective functions for feature discrimination. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, pages 1596–1602, Detroit, MI, August 1989. Morgan Kaufman.

[4] A. Jaklič, A. Leonardis, and F. Solina. Segmentor: An object-oriented framework for image segmentation. Technical Report LRV-96-2, Computer Vision Laboratory, University of Ljubljana, Faculty of Computer and Information Science, 1996.

[5] A. Leonardis. *Image analysis using parametric models: model-recovery and model-selection paradigm.* Ph.d. dissertation, University of Ljubljana, Faculty of Electrical Engineering and

Computer Science, 1993.

[6] A. Leonardis, A. Gupta, and R. Bajcsy. Segmentation of range images as the search for geometric parametric models. *International Journal of Computer Vision*, 14:253–277, 1995.

[7] A. Leonardis, A. Jaklič, and F. Solina. Superquadrics for segmenting modeling range data. Technical Report LRV-96-1, Computer Vision Laboratory, University of Ljubljana, Faculty of Computer and Information Science, 1996.

[8] A. Leonardis, F. Solina, and A. Macerl. A direct recovery of superquadric models in range images using recover-and-select paradigm. In J.-O. Eklundh, editor, *Proceedings of Third European Conference on Computer Vision, Vol. I*, Lecture Notes in Computer Science 800, pages 309–318, Stockholm, Sweden, May 1994. Springer.

[9] A. P. Pentland. Automatic extraction of deformable part models. *International Journal of Computer Vision*, 4:107–126, 1990.

[10] F. Solina and A. Leonardis. Selective scene modeling. In *Proceedings of the 11th International Conference on Pattern Recognition*, pages A:87–90, The Hague, The Netherlands, September 1992. IAPR, IEEE Computer Society Press.

[11] N. Ayoung-Chee, G. Dudek, F. P. Ferrie. Enchanced 3D Representation Using Multiple Models In *Proceedings of the International Workshop on Object Representation for Computer Vision*, pages 35–36, Cambridge, England, April 1996.