PART-LEVEL OBJECT RECOGNITION

Jaka Krivic and Franc Solina

University of Ljubljana
Faculty of Computer and Information Science
Computer Vision Laboratory
Tržaška 25, 1000 Ljubljana, Slovenia
{jakak, franc}@lrv.fri.uni-lj.si
http://lrv.fri.uni-lj.si

Abstract: *This paper proposes a technique for object recognition using superquadric built models. Superquadrics, which are three dimensional models suitable for part-level representation of objects, are reconstructed from range images using the recover-and-select paradigm. Using an interpretation tree, the presence of an object in the scene from the model database can be hypothesized. These hypotheses are verified by projecting and refitting the object model to the range image which at the same time enables a better localization of the object in the scene.*

Keywords: *superquadrics, part-level object modelling, range images*

1 INTRODUCTION

Part-level models are useful for describing 3D objects in computer vision. An articulated object can be represented by several part-level shape descriptions. Superquadrics are volumetric models, that can be used for part-level shape descriptions. They are mathematical solids that can represent standard geometric shapes and shapes in between. Part-level description of a scene can be obtained by partitioning the image into segments, where each segment corresponds to individual part. Also, a part model for each segment has to be recovered. In the case of superquadric part models, several recovery methods were developed [5, 6]. A tight integration of segmentation and model recovery was achieved [4] by combining the *recover-and-select* paradigm [3] with the superquadric recovery method [6]. *Segmentor* is an object-based implementation of the *recover-and-select* segmentation using superquadrics and other parametric models [2]. Segmentation and shape recovery with *segmentor* is fairly stable on smooth and regular man made objects, that can be easily represented with superquadric shapes. On the other hand, the superquadric models can not expand easily on rough surfaces and natural shapes, which are not very close to the superquadric ideal, usually resulting in over-segmentation. Despite those deficits we decided to test the applicability of the *Segmentor* system for object recognition of articulated objects in complex scenes.

Our aim was to investigate possible use of part-level description obtained by the *Segmentor* system for recognition of part-level models. We hypothesized that the configuration of parts and their rough shape should provide enough constraints for successful matching with the models of known objects. Recognition system would search for matches between scene and model parts, a procedure known as *model based matching*. We used

interpretation tree method developed by Grimson [1]. The method works by arranging all possible matches of scene part with model part in a tree structure. The problem of recognizing an object is to find consistent interpretations without exploring all possible way of matching scene and model parts. Consistent interpretations, or object hypotheses, can subsequently be verified by fitting the object model directly to the range data of a scene.

2 OBJECT RECOGNITION SYSTEM

Output of a *Segmentor* system is a set of recovered superquadrics, which represent the parts of the input scene. On this set the search for feasible interpretations of the stored model is employed. When an interpretation is found, it can be verified by projecting object model into the scene.

The system we propose consists of the following three steps:

1. range image segmentation and superquadric recovery using the *Segmentor*,

2. search for feasible interpretations of the stored model in the scene using interpretation trees, and

3. hypothesis verification by projecting object models into the scene.

The second and third step can be interleaved, to early eliminate those hypotheses that do not make sense.

2.1 Object Model

The object is modelled with superquadric parts. A part is connected to another part by a joint. There are two types of joints: flexible and rigid. Flexible joints contain information about the position relative to each of the two parts they connect. In the process of hypothesis verification they are also given the rotational parameters, which rotate the two parts into the right configuration. Such joints connect non rigid objects such as a human figurine on figure 1, where all joints are flexible. Rigid joints, on the contrary, have, besides positional parameters, predefined rotational parameters, and therefore 'glue' the two parts together. The object from figure 4 contains two rigid joints.

Figure 1: Toy figurine.

Since we focused our work on the recognition phase, we built the models manually, by measuring the parts and approximating the superquadric and other parameters for each part and joint (e.g. Fig. 2). Nevertheless, one can easily imagine the supervised building of models. Supervisor would first have to label the parts on several superquadric reconstructions of the object, as well as enter their connectional relations. The system would then compute the valid intervals for the parameters involved.

(a)   (b)
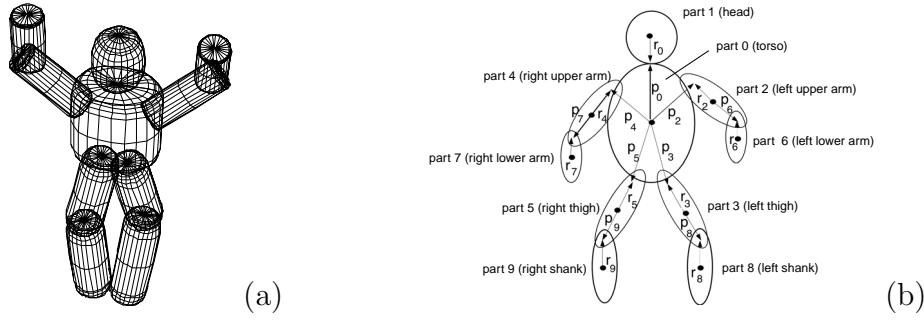
Figure 2: An object from fig. 1 is modelled in two levels: (a) superquadric part models define the size and shape of individual parts, (b) the structural level defines how parts are connected to each other. p_i and r_i stand for positional parameters of joint i.

2.2 Model Matching

One can easily imagine the process of recognizing an object as matching scene parts with part models of the stored body model. All possible matches arranged in a tree structure are called *interpretation tree*. Nodes in an interpretation tree represent a match between a part of the scene and a part of the model. The search for correct interpretation begins at the root of the interpretation tree. The root expands to all possible matches for the first model part. The search continues from a node in depth only if the match represented by that node is consistent, i.e. if the two parts represented by that node are similar. In our case, a match is consistent if the scene part's superquadric parameters are in the model part's intervals. We also applied part volume test for parts with less reliable superquadric recovery. When there are several other parts in the interpretation, connection relations are also verified, such as part distance and approximate joint distance.
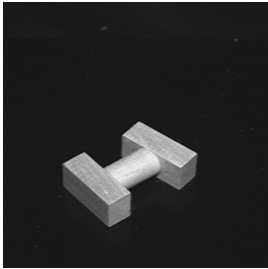

Figure 3: Simple three part object.

In real scenes some parts of an object may be hidden to the viewer and some occluded by other objects or parts. Also, the part detector can miss some parts or introduce some spurious ones. To enable the system to deal with such cases, a fictitious scene part that matches every model part is introduced. It is simply appended to the list of image parts.

When the search reaches a leaf one gets a consistent interpretation. But because the constraints involved in checking match consistency are local in nature, the interpretation does not have to make sense globally and has to be further verified.

2.3 Interpretation Verification.

First, the system can reject interpretations that include too few real part matches. By setting a threshold on interpretation size the system may, therefore, reject some correct interpretations (false negatives), but it will reject many more wrong ones (false positives), since there is a low probability that some parts will "randomly" form a structure similar to the structure of the object.
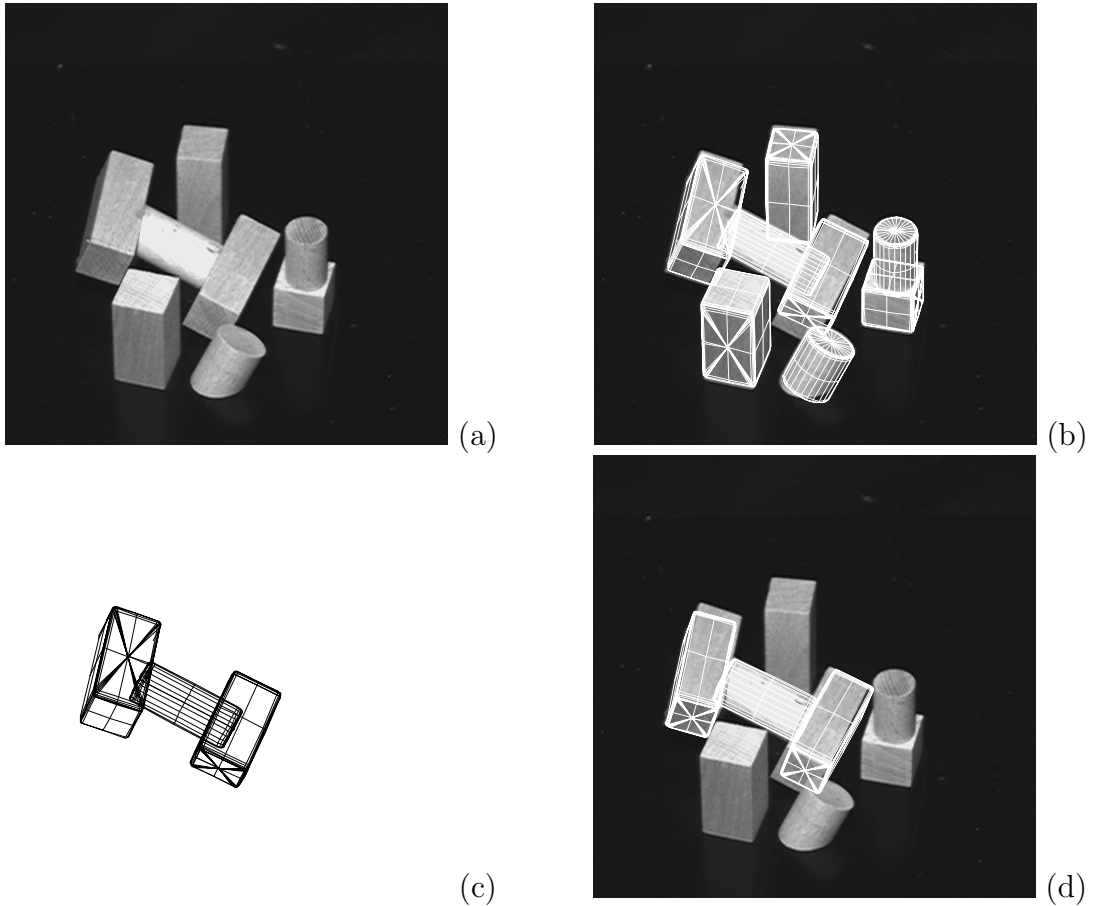
Figure 4: Interpretation of a simple scene: (a) intensity image of a scene, (b) input range image with superimposed reconstructed superquadrics, (c) superquadrics from hypothesis interpretation, (d) verified interpretation.

Second, from an interpretation, the object's position and configuration can be estimated. The individual superquadric parts of the model are then fitted to the corresponding regions of the range image, defined by the superquadrics included in the interpretation being verified. To fit individual superquadric models to such part regions the standard fitting method was used [6]. The fitting function was minimized only for the position and orientation parameters, the size and shape parameters were fixed to the values of the tested model part superquadric. If the initial (ie. estimated) moved or rotated too much in the process of fitting, the interpretation was rejected.

3   EXPERIMENTAL RESULTS

We present the system's results on two examples. The object to be recognized in the first example consists of two boxes and a connecting cylinder in a rigid configuration and can be seen in figure 3. Figures 4a to 4d show the scene, superquadric reconstruction of the scene, the best hypothesized interpretation and verified interpretation, respectively.

The second example shows the system's performance on a scene of a toy figurine from fig. 2. The figurine is fairly realistic and therefore cannot be perfectly modelled by
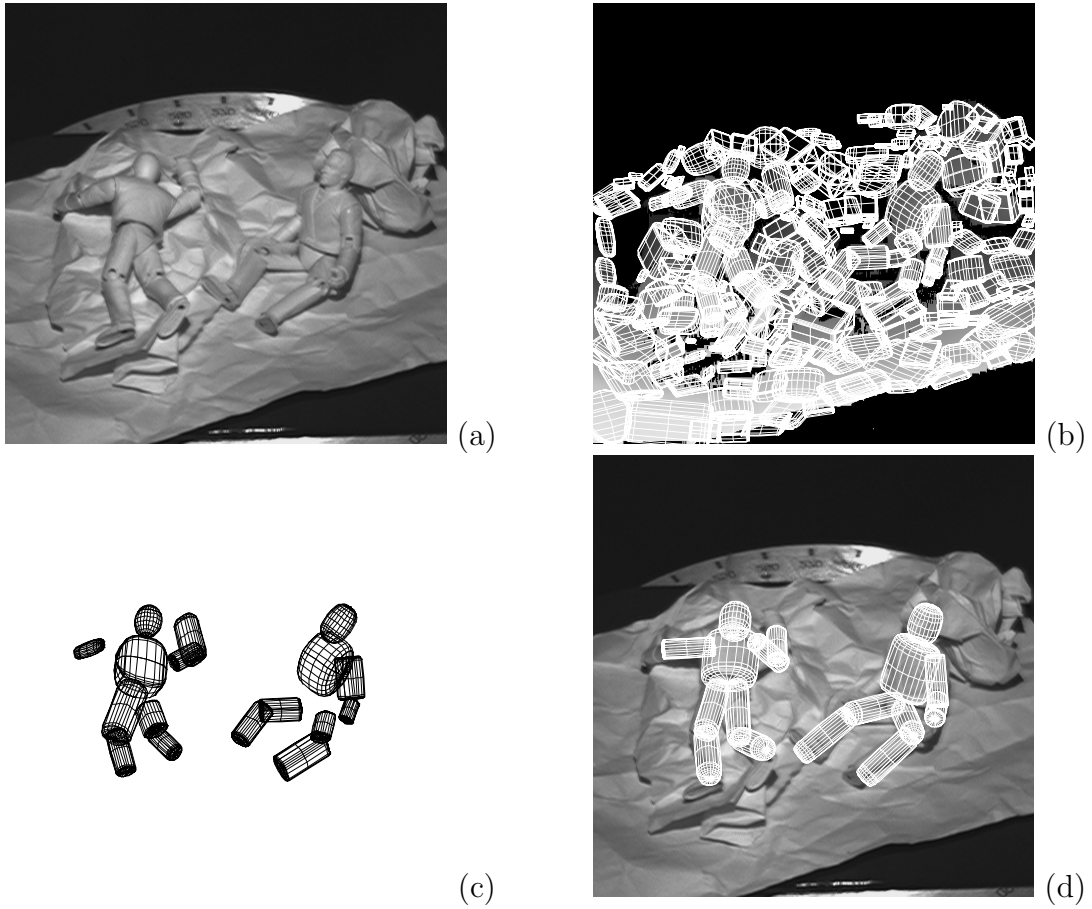
Figure 5: Interpretation of a complex scene: (a) intensity image of a scene, (b) input range image with superimposed reconstructed superquadrics, (c) superquadrics from two hypothesis interpretations, (d) verified interpretations.

superquadrics. Since the surfaces are not smooth, the reconstruction of superquadrics on their range images is less stable. The flexibility of body joints makes the matching problem more complex than if the object part configuration would be rigid. The figurine was modelled with ten superquadric parts (Fig. 2). Since the figurine's head can not be moved (relatively to the torso), the joint between the two is rigid. All other joints are flexible and are assigned rotational parameters during hypothesis verification. System's performance was tested on simple (containing only the figurine) as well as complex scenes (such as the one in fig. 5. With the former we tried to determine the system's ability to recognize the object. The figurine was put in seven different poses and for each pose range images from eight viewpoints were captured (56 range images altogether). The object was detected in 69%. With the latter test we tried to determine the system's false positive recognition rate. A dozen range images containing one or two figurines as well as other objects were captured and processed. In no case were there any false positive recognitions, although there were many at least partially misleading configurations.

Example result of a complex scene can be seen in fig. 5. One can observe, that the reconstructions of the supporting surfaces in complex scenes were not appropriate, because such surfaces cannot be modelled well by superquadrics.

4 CONCLUSIONS

In this paper, object recognition using superquadric based shape decomposition was investigated. The system is based on interpretation trees. We have shown, that despite very rough and somewhat unstable part description, superquadrics can be used in an object recognition scheme. The system can handle flexible articulated objects that cannot be perfectly modelled by superquadrics which is demonstrated by the recognition of the human figure.

REFERENCES

[1] W. E. L. Grimson. *Object Recognition by Computer*, MIT Press, Cambridge (MA), 1990.

[2] Ales Jaklic, Ales Leonardis, Franc Solina. *Segmentation and Recovery of Superquadrics.* Computational imaging and vision 20, Kluwer, Dordrecht, 2000.

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

[4] A. Leonardis, A. Jaklič and F. Solina. Superquadrics for segmentation and modeling range data. *IEEE Transactions on Pattern Recognition and Machine Intelligence* 19(11):1289–1295, 1997.

[5] A.P. Pentland. Perceptual organization and the representation of natural form, *Artificial Intelligence* 28:293–331, 1986.

[6] F. Solina in R. Bajcsy. Recovery of parametric models from range images: The case for superquadrics with global deformations, *IEEE Trans. Pattern Anal. Machine Intell.* 12:131–147, 1990.