Superquadric-Based 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 re-fitting 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 modeling, range images

1 Introduction and Motivation

In computer vision, many different models have been used for describing different aspects of objects and scenes. Part-level models are one way of representing 3D objects, when particular entities that they describe, correspond to perceptual equivalents of parts. Several part-level shape descriptions are required to represent an articulated object. One of the more popular types of volumetric part-level descriptions are superquadrics [1,11,13,3,12,5]. They are solid models that represent standard geometrical solids as well as shapes in between.

Pentland [11] was the first who used superquadrics in the context of computer vision. However, Solina and Bajcsy's method for recovery of superquadrics from pre-segmented range images became more widespread [5]. Also, several methods for segmentation with superquadrics have been developed. A tight integration of segmentation and model recovery was achieved [9] by combining the "recover-and-select" paradigm [8] with the superquadric recovery method [13]. The paradigm works by recovering independently superquadric part models everywhere on the image, and selecting a subset which gives a compact description of the underlying data. *Segmentor* is an object-based implementation of the "recover-and-select" segmentation paradigm using superquadrics and other parametric models [5]. Superquadrics, their mathematical properties, recovery from images and its applications are presented in detail in [5].

Segmentor lacks the reliability of segmentation of rough, natural shapes, which are not very close to ideal superquadric shapes. The superquadric models can not expand easily on rough surfaces and complex shapes as easily as on smooth regular objects which results generally in over-segmentation. Nevertheless, our starting hypothesis was that the part-level description obtained by the *Segmentor* system (see Fig. 2a and 2d) is good and stable enough for recognition of part-level models. The configuration of parts and their rough shape should provide enough constraints for successful matching with the models of possible objects. The object hypotheses can be subsequently verified by fitting the object model directly to the range data.

2   Object Recognition Scheme

The results of processing a range image with the *Segmentor* is a set of superquadric models, each with known position, orientation, size and shape. Reconstructed superquadrics represent parts that constitute the scene. An object recognition system that searches for matches between parts of the scene and parts of the modeled object can be used, called "model based matching" [3,10, 2,4]. For identifying the object in the scene we adopted the interpretation tree method [4,14]. Our object recognition system uses 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 hypothesis that do not make sense.

2.1   Object Model

We decided to use human figures as generic articulated test objects for our recognition task. We were not interested in the specific problem of modeling human form, although systems using superquadrics for modeling humans do exist [6,7]. Since the workspace of our range scanner is rather small, toy figures representing "Commander Data" from the Star Trek series (Fig. 1 a) were used. Their arms and legs are flexible and the figurines can thus be configured into many different poses.

We built the model of the figurine manually. The model consists of superquadrics (Fig. 1 b). Each superquadric represents one of the major body parts: head, torso, a pair of upper and lower arms, and a pair of upper and lower legs. Due to the limited scale of parts which can be recovered on the selected range image resolution by the *Segmentor*, the model does not include distinct models of hands and feet. Each body part is described by a superquadric of a particular size and shape. The torso is given a central role in the model. The head and upper arms and legs are attached to it via joints (Fig. 1 c)). For each
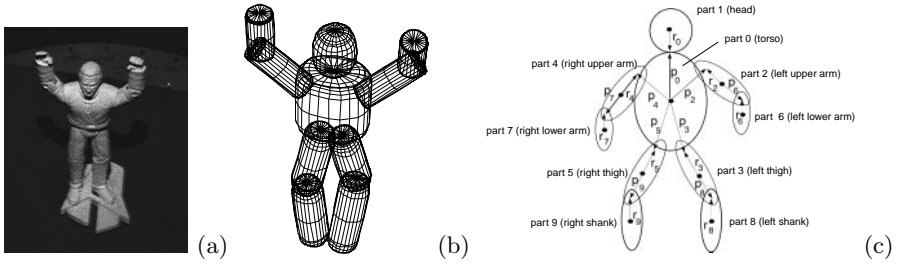
Fig. 1. (a) The object. The object model consists of two levels: (a) superquadric part models define the size and shape of individual parts, (b) the structural level defines how parts are connected to each other.

of those parts a joint position relatively to the center of the part itself (r_i) and to the center of the torso (p_i) is defined. Similar is true for lower extremities. The parameter values for all parts were obtained by measuring the figurine.

In this paper, p_i and r_i stand for the relative positions of the center of the part i in the model, as described above, c_i is the center of superquadric that matches, or should match part i, $\mathbf{R}$ is a ZYZ rotation matrix, and ϕ_i, θ_i and ψ_i are rotation angles for part i.

The test figurine is interesting in several ways. It is fairly realistic and therefore cannot be perfectly modeled by 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.

2.2   Model Matching

The input to model matching is the set of superquadric models resulting from the *Segmentor* and the stored object model. The process of recognizing an object can be viewed as matching scene parts with part models of the stored body model which are arranged in an interpretation tree. The search for correct interpretation begins at the root. The root expands to all possible matches for the first model part. When the search reaches a leaf one gets a consistent interpretation. But because the constraints involved in checking consistency of a match are local in nature, the interpretation does not have to make sense globally. In general, there is no guarantee that a found interpretation makes global sense. These interpretations must therefore be taken only as hypotheses and should be further verified.

Checking the Local Consistency of a Match. Reconstructions of superquadrics on the object in different poses and different viewpoints differ greatly, except in the case of the head. Since superquadric parameters cannot be directly used for the comparison of two superquadrics [5], we decided to base the consistency check merely on the volume of superquadrics. If a volume of a scene part

is within the interval given by the model part, the two parts represent a possible match.

Matching the Head. The analysis of superquadric reconstructions of the human body showed that the head was the most consistently reconstructed body part. At the same time, the head is also the only part that does not change significantly its relative position in relation to the torso. Therefore, we found it reasonable to define constraints for the size and shape parameters of superquadrics, in order to reject as many unsuitable parts as possible. The values for the size and shape parameters were defined on the basis of thirty reconstructions.

Matching the Torso. Superquadrics reconstructed on the torso region differ the most from the torso's model superquadric. On this region several possibly overlapping superquadrics can be recovered, which can partially cover even regions belonging to extremities. Different volume interval was used for the cases of a single superquadric and the case of overlapping superquadrics

If there is also a head in the interpretation, one can more precisely compute the position and (although not complete) rotation parameters for the whole model. This information can be used to further constrain the search for interpretations.

The analysis of the reconstruction showed, that the centers of superquadrics from the torso region were fairly close to the real vertical axis of the torso, whereas the distance to the real center of the torso was not. This can be corrected by taking into account the distance between the centers of the head and the torso. Orientation parameters can be resolved by computing the rotation, that transforms the z-axis into the vector $\boldsymbol{c_0} - \boldsymbol{c_1}$. The parameter ψ_1 defines the rotation around the z-axis of the object itself.

Let $\boldsymbol{c_0}$ be the center of the head's superquadric, and $\boldsymbol{c_1}'$ approximate center of the torso, as mentioned earlier. Then a better approximate center of the torso $\boldsymbol{c_1}$ is

$$\boldsymbol{c_1} = \boldsymbol{c_0} - |\boldsymbol{s_0}|\boldsymbol{s} \tag{1}$$

$$\phi_1 = \arctan \frac{-s_x}{-s_y} \tag{2}$$

$$\theta_1 = -\arctan \frac{\sqrt{s_x^2 + s_y^2}}{s_z}, \tag{3}$$

where

$$\boldsymbol{s_0} = \boldsymbol{r_h} - \boldsymbol{p_h}$$
$$\boldsymbol{s} = \frac{\boldsymbol{c_0} - \boldsymbol{c_1}'}{|\boldsymbol{c_0} - \boldsymbol{c_1}'|}. \tag{4}$$

$\boldsymbol{c_1}'$, $\boldsymbol{c_1}$ and $\boldsymbol{c_0}$ are 3D vectors of the form $[x, y, z]^T$, $\boldsymbol{s}$ is a 3D vector $[s_x, s_y, s_z]^T$, whereas ϕ_1, θ_1 and ψ_1 are the angles that define the torso's orientation and thus the object's orientation.

Matching the Extremities. **For** consistency check of matches for upper arms and legs, two cases were distinct. In the first, there is a real match for the head as well as for the torso included in the interpretation. Therefore the parameters for position and (partially) orientation for the whole model is known and can be used for checking the consistency of a given match. When a superquadric from a scene is in the same volume range as the upper arm or leg part model, parameter ψ_1 can be computed. Parameter ψ_1 rotates the object part model so that the joint position on the object model approximately overlaps with the possible upper arm or leg

$$\psi_1 = \arctan \frac{s_y}{s_x} \tag{5}$$
$$s = \mathbf{R}^{-1}(c_x - c_1),$$

where c_x is the superquadric center, whose match with the upper arm or leg part model is being checked, and $\mathbf{R}$ is a ZYZ rotation matrix with parameters $(\phi_1, \theta_1, 0)$.

If the part in question is really an extremity, its center has to be approximately as distant from the joint as is the case with the object model. If the conditions are met, the match is consistent and rotation parameters for the part can be computed by

$$\phi_x = \arctan \frac{-s_x}{-s_y}, \tag{6}$$

$$\theta_x = -\arctan \frac{\sqrt{s_x^2 + s_y^2}}{s_z}, \tag{7}$$

$$\psi_x = 0. \tag{8}$$

In the second case, there is a wildcard match for head or torso, so there is no information about the position and orientation of the model (or at least about orientation, if there is a real match for the torso in the interpretation).

The lower extremities were not included in the interpretation tree search because the interpretation can be verified quite well without them. Inclusion of lower extremities in the tree search would only increase the search space.

Interpretation Verification. The checking of the global consistency of an interpretation means that the system should answer the question: *"Does the given set of parts really represent object X?"*

We decided that the system should reject all interpretations that include less than four real matches. 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 human body.

To verify a hypothesis obtained by the interpretation tree we fitted individual superquadrics of the stored model to corresponding regions of the range image,

which were defined by the superquadrics included into the tested interpretation. To fit individual superquadric models to such part regions the standard fitting method was used [13]. 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.

As in consistency check, there are two cases of interpretation verification that need to be considered. In the first case, there are real matches for both the head and the torso, and in the second case, at least one of those matches is a wild card. In the first case the position and orientation parameters for the model can easily be computed (Eqs. 1, 5). It turns out that those parameters fairly accurately describe the position of the model. In the second case, the position and orientation of the model could be computed based on the position of joints, that connect the extremities to the torso. But, since the position of the joints cannot be accurately defined based on the reconstructions, the position of the joints cannot be accurately computed as well. In fact, the computation did not deliver any reasonable results and thus we decided that the interpretation is only good if it includes real matches for both, the torso and the head.

Let us return to the interpretation verification by fitting the model parts to the corresponding regions of the range image. The model part parameters were used as initial parameters. The fitting is only performed for superquadrics that model the extremities, since they are elongated, and the fitted model defines the position of the joints well. When the superquadric is fitted, the distance between the joint position on the fitted superquadric and the initial one can be computed as

$$d = |(\mathbf{R_i} r_i + c_i) - (\mathbf{R_f} r_i + c_f)|, \tag{9}$$

where $\mathbf{R_i}$ is a rotation matrix, that rotates part i into initial position, r_i is the relative position of the joint, c_i is the center of initial superquadric, $\mathbf{R_f}$ is the rotation matrix, that rotates part i into the fitted position, and c_f is the center of the fitted superquadric. The computed distance d is compared with a threshold p, which was set to value $p = 11.0$, based on the analysis of reconstructions. If the distance d is greater than the threshold with at least one extremity, the interpretation is rejected.

Interpretation tree search does not include matches for lower extremities, that is lower arms and shanks. The presence of those parts is searched for only if the part attached to it (upper arm, thigh) is present, otherwise the results cannot be verified well. This parts are matched based on volume and joint distance.

3   Experimental Results

Range images which were obtained by a structured light range scanner and processed with the *Segmentor* system [5]. The resulting set of superquadric models were processed as described in the previous sections.

First, we tried to systematically test the system's performace for isolated figurines. A single figure was configured in seven different poses and eight range

images were taken from different viewpoints. Figure 2(a-c) shows one of the results.

(a)   (b)   (c)

(d)   (e)   (f)
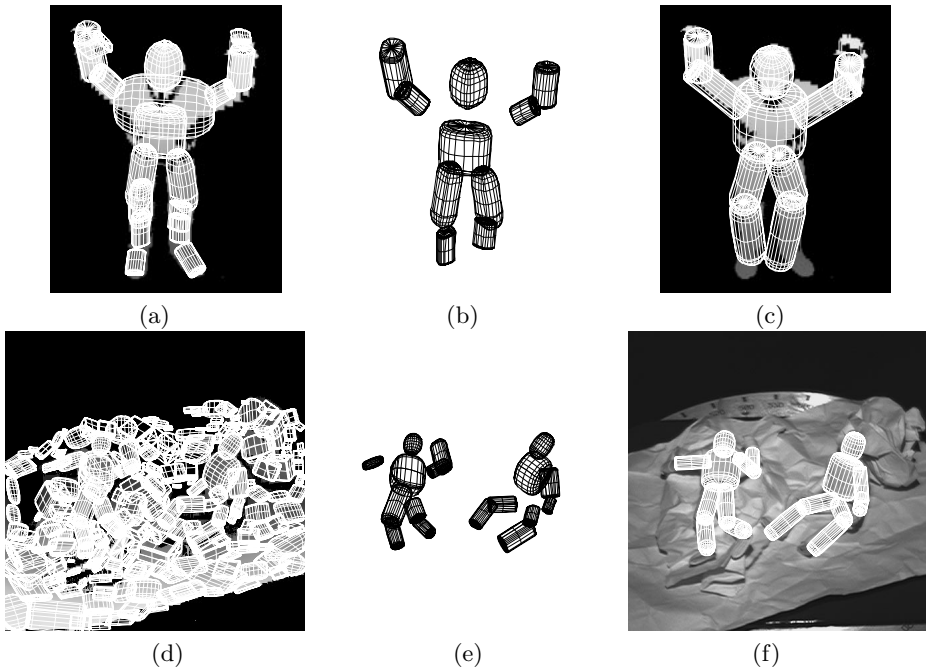
Fig. 2. Interpretation of a single-figurine scene (a,b,c) and a complex scene (d,e,f): (a,d) input range image with superimposed reconstructed superquadrics, (b,e) superquadrics from two hypothesis interpretations, (c,f) verified interpretations.

The object was detected in 39 out of 56 cases. In 24 of those 39 cases, the model computed from the best interpretation fitted the object very well. Interpretation included on the average 7.2 real matches. The object was not detected in 17 cases. In 9 of those cases, the reason for that was an occluded head or torso, so that the reconstructed superquadrics on those regions were not even close to the ones from the model. In 8 other cases the best interpretation found included less than four real matches, and was therefore rejected.

The system's performance was also tested on 20 different complex scenes. Complex scenes included several appearances of the toy figure, as well as many unknown objects (Fig. 2d-f). Nevertheless, there were no false positive recognitions of the human form, although there were many at least partially misleading configurations. It is much harder to test a complex scene in a systematic fashion because of so many possible variables. One can observe that the reconstructions of the supporting surfaces in complex scenes were not appropriate, because such surfaces cannot be modeled well by superquadrics.

4   Conclusions

In this paper, we have investigated if superquadric based shape decomposition can be used for object recognition. 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 modeled by superquadrics which is demonstrated by the recognition of the human figure.

References

1. Barr, A.H.: Superquadrics and Angle-Preserving Transformations. IEEE Computer Graphics and Applications, **1** (January 1981) 11–23
2. Bolles R.C., Horaud. P.: 3DPO: A Three-Dimensional Part Orientation System. The International Journal of Robotic Research, **5** (1986) 3–26
3. Dickinson, S.J., Pentland, A.P., Rosenfeld, A.: From volumes to views: An approach to 3-D object recognition. CVGIP: Image Understanding, **55** (1992) 130–154
4. Grimson W.E.L.: Object Recognition by Computer. MIT Press, Cambridge, MA (1990)
5. Jaklič A., Leonardis A., Solina F.: Segmentation and Recovery of Superquadrics. Kluwer Academic Publishers, Dordrecht (2000)
6. Jojić N., Huang T.S.: Computer vision and graphics techniques for modeling dressed humans. In: Leonardis A., Solina F., Bajcsy R. (eds.): The confluence of computer vision and computer graphics. Kluwer, Dordrecht (2000) 179–200
7. Chella A., Frixione M., Gaglio S.: Understanding dynamic scenes. Artificial Intelligence, **123** (2000) 89–132
8. Leonardis A., Gupta A., Bajcsy R.: Segmentation of range images as the search for geometric parametric models. International Journal of Computer Vision, **14** (1995) 253–277
9. Leonardis A., Jaklič A., Solina F.: Superquadrics for segmentation and modeling range data. IEEE Transactions on Pattern Recognition and Machine Intelligence, **19** (1997) 1289–1295
10. Nevatia R., Binford T.: Description and Recognition of Curved Objects. Artificial Intelligence, **38** (1977) 77–98
11. Pentland A.P.: Perceptual organization and the representation of natural form. Artificial Intelligence, **28** (1986) 293–331
12. Raja N.S., Jain A.K.: Recognizing geons from superquadrics fitted to range data. Image and Vision Computing, **10** (1992) 179–190
13. Solina F., Bajcsy R.: Recovery of parametric models from range images: The case for superquadrics with global deformations. IEEE Trans. Pattern Anal. Machine Intell., **12** (1990) 131–147
14. Trucco E., Verri A.: Introductory techniques for 3-D computer vision. Prentice Hall, NJ (1998)