

Segmentation of 3D Articulated Objects by Dynamic Grouping of Discontinuities

Dibio L. Borges and Robert B. Fisher

Department of Artificial Intelligence, Edinburgh University
5 Forrest Hill, Edinburgh EH1 2QL, Scotland, UK

Abstract

Segmentation of 3D articulated objects into a small number of parts is an important step for recognition of objects in 3D space. Range data provides ready-to-use information to derive a reasonably stable and accurate differential structure of the surfaces, which allow us to track discontinuities between these parts. We present a mixed and comprehensive approach to segmenting and extracting prototypical parts from 3D articulated objects. Discontinuity points detected from the images are combined using a "snake-type" minimization technique, and deformable superquadric models are fitted to the resulting regions afterwards. We compare our approach with others for part-based segmentation.

1 Introduction

One of the key questions in Computer Vision research is how to segment a scene in meaningful chunks which have stable visual interpretation and can be extracted in a reliable way. This segmentation process is important as a first stage of a system which aims to reason about and recognize 3D articulated objects based on the notion of prototypical parts.

To do this we present a mixed and comprehensive approach to identifying separate parts of a natural articulated object by relying on stable estimates of the differential structure of the object's surfaces which prevents from smoothing across discontinuities, followed by a dynamic grouping of the detected discontinuities on the surface and a fitting of deformable superquadric models for the segmented regions.

The final aim of the segmentation is to get a set of 3D deformable superquadric models where each model of the set is related to only one part of the object. Our assumption is that three types of discontinuities (depth, orientation, and loci of negative minima of the Principal Directions) detected upon the local differential structure of the object's surface can be grouped together in a way to become the boundaries of important salient parts of an articulated 3D object. This grouping is formulated as an energy-minimization process of deformable curves which are placed at the discontinuity points are attracted to paths of minimum changes in

the Principal Directions (around a fixed neighbourhood) until they reach other discontinuities. After this stage the set of final curves will indicate the regions on the object's surface to be represented as separate parts. We propose that globally deformable superquadrics can capture well this notion of an object's parts, and in our solution we fit deformable superquadrics to each of the segmented regions to get a final extracted model of the object suited for identification of its overall shape and position.

We work with range data acquired by a laser stripser set up in our laboratory. From the range image we need a stable and accurate way of estimating Gaussian, Mean, and Principal Curvatures as well as Principal Directions for identifying and grouping the discontinuities.

The approach uses a mix of techniques which provides a tailored solution to parts-based segmentation of 3D objects. The overall structure of the approach bears similarities to recent work done by Ferrie *et al* [7] and it differs from their work by the inclusion of discontinuity preserving conditions for smoothing and curvature consistency calculations, dynamic grouping of all the detected discontinuities and by the use of deformable superquadric models for representing the objects' parts. One of the motivations of our work is to establish a direct way to extract a large set of deformable prototypical parts for articulated 3D objects in order to be able to reason about the shape and the location of natural and man-made objects.

Section 2 describes the calculation of stable estimates of the differential structure of the surfaces. Section 3 is on the partitioning of the object by allowing a dynamic grouping of the discontinuities guided by the Principal Directions on the surface. The Superquadric representation and the fitting of globally deformed models are discussed in Section 4. Section 5 shows some images of the segmentation. Section 6 gives some conclusions.

2 Differential Structure of 3D Objects

The use of differential geometry techniques to model and segment range images has become a major topic of investigation in Computer Vision [1]. One approach taken is to derive *qualitative surface descriptions* by using the signs of the Gaussian and Mean Curvatures estimated from the surface patches. There are eight different surface shapes that can be classified by the signs of Gaussian and Mean Curvatures [6]. This curvature-based segmentation has been applied with some success to 3D curved objects [2], [17].

The local structure of a 3D surface is richer though, and instead of using only the Gaussian (K) and the Mean (H) Curvatures one can also obtain from a local fit the Principal Curvatures (κ_1, κ_2) and Principal Directions (\vec{d}_1, \vec{d}_2). The problem with these estimates is that they are very unstable, and not very accurate. In order to overcome these problems we combine a discontinuity preserving smoothing with a two step consistency algorithm to get reliable estimates for the local structure of the surface.

Noise smoothing of the raw range data prior to the computation of the surface curvatures is an essential module due to the effects of quantization. Gaussian smoothing is known to blur discontinuity points on the surface that are important

to preserve in our case. Cai [5] has suggested that diffusion smoothing is an elegant solution to gaussian smoothing, and it can be made to prevent smoothing across the discontinuities. Trucco and Fisher [17] have improved on that by imposing an adaptive boundary condition on the diffusion equation and we use their algorithm to smooth our input data preserving the discontinuities.

The initial estimates of the local structure of the surface, or the Darboux Frames $D(\mathbf{P}) = (\mathbf{P}, \kappa_{1P}, \kappa_{2P}, \mathbf{N}_P, \vec{d}_{1P}, \vec{d}_{2P})$, is similar to other approaches where for each point $\mathbf{P} \in S$ we locally fit a paraboloid of the form $g(\mathbf{u}, \mathbf{v}) = a\mathbf{u}^2 + b\mathbf{u}\mathbf{v} + c\mathbf{v}^2$ to the surface. The local structure $D(\mathbf{P})$ can then be computed from $g(\mathbf{u}, \mathbf{v})$. Two particulars of the fit we adopted are: 1) The origin of the parametric coordinate frame $\langle \mathbf{u}, \mathbf{v}, g \rangle$ is placed at the point \mathbf{P} aligning the g axis with the normal, and 2) an updated map with depth and surface orientation discontinuities is kept and checked at each pass of the estimation in order to provide a more precise local structure of the surface at the end.

The Darboux Frame estimates obtained from the first stage are good for the purpose of HK classification of the surfaces; however, they are not good enough for reliable and consistent Principal Directions. For this we use a method which was first developed for applications in medical imagery by Sander and Zucker [14]. The method is an iterative constraint satisfaction technique where at each point \mathbf{P} on the surface the Darboux Frames $D()$ of its neighbouring points are extrapolated from $D(\mathbf{P})$. From this new set of $D()$ in a neighbourhood of \mathbf{P} , a functional is set up to minimize the variation in $D(\mathbf{P})$ using the error measures:

$$E_1 = \min \sum_{\alpha=1}^n \|\mathbf{N}_P - \mathbf{N}_{()P\alpha}\|^2 + (\kappa_1 - \kappa_{1()P\alpha})^2 + (\kappa_2 - \kappa_{2()P\alpha})^2 + \lambda((\mathbf{N}_P \cdot \mathbf{N}_{()P\alpha}) - 1). \quad (1)$$

$$E_2 = \min_{\theta} \sum_{\alpha=1}^n [1 - (\vec{d}_{1P}(\theta) \cdot \vec{d}_{1P\alpha})^2]. \quad (2)$$

where α indexes over the neighbourhood.

The minimization is subject to these constraints on the Principal Directions $(\vec{d}_{1P}, \vec{d}_{2P})$ and the normal (\mathbf{N}_P) :

$$(\vec{d}_{1P} \cdot \mathbf{N}_P) = 0; \quad (\mathbf{N}_P \cdot \mathbf{N}_{()}) = 1; \quad (\vec{d}_{1P} \cdot \vec{d}_{1P}) = 1 \quad (3)$$

The algorithm to achieve this is rather complex (see [14] for more details).

Either of the Principal Directions can be used in the E_2 measure. The convergence is controlled by a composite measure $\mathcal{U}_S^{(i)}$ of E_1 and E_2 , where a minimum change in the measure throughout the iterations is required.

$$\mathcal{U}_S^{(i)} = \sum_j E_{j1}^{(i)} + E_{j2}^{(i)}; \quad \mathbf{P}_j \in S. \quad (4)$$

A map of the discontinuities is kept and checked at each iteration in order to avoid distortion of the Darboux Frame values during the minimization. We believe the use of this map improves the reliability of the original curvature consistency algorithm (see [7]), resulting in a more precise final local structure of the surface.

3 Parts of Objects

The problem of defining what constitutes a part of an object is very important and yet can be hard to solve. We perceive regularities from our visual world, and these regularities help us to organize chunks of the image into meaningful parts. A set of these parts would constitute a rough model of the object.

Important evidence for defining the extent of an object part is some sort of discontinuity on the surface. Two types of discontinuities useful here are depth discontinuities and orientation discontinuities. As first argued by Hoffman and Richards [8], another important principle for parts decomposition is the *transversality regularity*, which refers to the loci of negative minima of each principal curvature as part boundaries.

The depth discontinuities are found by simply thresholding the range image looking for abrupt ($> \textit{threshold}$) changes between pixel values in a small neighbourhood. The orientation discontinuities are detected as the place where the angle made between the calculated normals to the surface in a small neighbourhood is bigger than a prescribed value. This prescribed angle in our case is set to $\pi/6$. The discontinuities on the Principal Curvatures are the points where [8]:

$$\kappa_1'(x, y) |_{\vec{d}_1} = 0 \quad \text{AND} \quad \kappa_1(x, y) < 0. \quad (5)$$

or,

$$\kappa_2'(x, y) |_{\vec{d}_2} = 0 \quad \text{AND} \quad \kappa_2(x, y) < 0. \quad (6)$$

All of the three kinds of discontinuities, depth, orientation and the negative minima points (referred from now on as *nmp points*) are important in determining the extent of an object's part. The problem we face now is how to combine them to isolate the very salient parts the articulated objects suggest [11].

A suitable framework for deriving these contours is the notion of energy-minimizing curve fitting, or snakes [9]. Here, fixed length snakes are placed initially at the discontinuity points detected above (depth, orientation, and *nmp points*) and they grow to follow the direction of a local tangent field [18] model derived from the differential structure of the 3D surface (Section 2).

The formulation of the problem as energy-minimizing curve fitting gives us a dynamic way of combining the segmentation points of the object. The local structure of the surface enters as an important constraint in the minimization. For a more detailed discussion about the snakes minimization and the equations please see [9]. The process can be seen as a minimization of the following deformable curve

$$\ell(s, t) = (x(s, t); y(s, t)). \quad 0 \leq s \leq 1 \quad (7)$$

with kinetic energy functional $T(\ell)$:

$$T(\ell) = \frac{1}{2} \int_0^1 \mu | \ell_t |^2 ds. \quad (8)$$

where μ is a constant. The potential energy functional $U(\ell)$ is:

$$U(\ell) = \frac{1}{2} \int_0^1 (\omega_1(s) | \ell_s |^2 + \omega_2(s) | \ell_{ss} |^2 + I(\ell) + S(\ell)) ds. \quad (9)$$

where:

- $\omega_1(s)x_s$ is the tension of the curve
- $\omega_2(s)x_{ss}$ is the rigidity of the curve
- $I(\ell)$ is the potential field from the surface local structure
- $S(\ell)$ is the force between curves in close proximity

The solution we look for are the curves $x(s, t)$ and $y(s, t)$ for which

$$\int_{t_1}^{t_2} T(\ell) - U(\ell) dt. \quad (10)$$

is a minimum.

4 Superquadrics Representation

Volumetric primitives capture well the notion of parts in natural [11] and man-made objects [3]. They are also well suited for indexing since they have a relatively small number of parameters compared to other primitives. Basically two types of volumetric primitives have been used in Computer Vision, Generalized Cylinders ([11], [4]) and recently Superquadrics ([12]). Superquadrics are a family of parametric shapes that can be globally deformed, and with this represent a wide range of prototypical parts.

A superquadric surface can be defined by the following vector in 3D space

$$x(\eta, \omega) = \begin{bmatrix} a_1 \cos^{\epsilon_1}(\eta) \cos^{\epsilon_2}(\omega) \\ a_2 \cos^{\epsilon_1}(\eta) \sin^{\epsilon_2}(\omega) \\ a_3 \sin^{\epsilon_1}(\eta) \end{bmatrix} \quad (11)$$

where,

$$-\pi/2 \leq \eta \leq \pi/2. \quad (12)$$

$$-\pi \leq \omega < \pi. \quad (13)$$

The parameters a_1 , a_2 , and a_3 define the size of the superquadrics in the coordinates x , y , and z respectively. ϵ_1 and ϵ_2 are the squareness parameters. The capabilities of the superquadrics for modeling can be enhanced by introducing tapering and bending deformations to them. Cavity [15] and local deformations [16] are also possible, but they require different methods for recovering the parameters from the depth data.

For the purpose of recovering a superquadric surface from the data, equation (11) can be manipulated in order to get the following implicit equation

$$\left(\left(\frac{x}{a_1} \right)^{2/\epsilon_2} + \left(\frac{y}{a_2} \right)^{2/\epsilon_2} \right)^{\epsilon_2/\epsilon_1} + \left(\frac{z}{a_3} \right)^{2/\epsilon_1} = 1. \quad (14)$$

Based on this equation we can define the inside-outside function [15]

$$\mathbf{F}(x, y, z) = \left(\left(\left(\frac{x}{a_1} \right)^{2/\epsilon_2} + \left(\frac{y}{a_2} \right)^{2/\epsilon_2} \right)^{\epsilon_2/\epsilon_1} + \left(\frac{z}{a_3} \right)^{2/\epsilon_1} \right)^{\epsilon_1}. \quad (15)$$

Function (15) determines whether a given point $[x, y, z]$ lies relative to the superquadric surface

- $\mathbf{F}(x, y, z) = 1$, point (x, y, z) is on the surface
- $\mathbf{F}(x, y, z) > 1$, point (x, y, z) is outside the surface
- $\mathbf{F}(x, y, z) < 1$, point (x, y, z) is inside the surface

Equation (15) defines the superquadric surface in an object centered coordinate system (x_s, y_s, z_s) . By using a world coordinate system (with origin at (p_x, p_y, p_z)) and Euler angles (ϕ, θ, ψ) we can express the inside-outside function in general position

$$\mathbf{F}(x_w, y_w, z_w) = \mathbf{F}(x_w, y_w, z_w; a_1, a_2, a_3, \epsilon_1, \epsilon_2, \phi, \theta, \psi, p_x, p_y, p_z). \quad (16)$$

The expanded function \mathbf{F} , equation (16) has now 11 parameters without the inclusion of tapering and bending deformations. Once we include the deformations the number of parameters to be recovered go up to 15

$$\mathbf{F}(x_w, y_w, z_w) = \mathbf{F}(x_w, y_w, z_w; a_1, \dots, a_{11}, b, \beta, T_x, T_y). \quad (17)$$

where,

- (x_w, y_w, z_w) is a point on the surface of the superquadric
- a_1, \dots, a_{11} are the 11 parameters of an undeformed superquadric
- b, β are the bending parameters along axis z
- T_x, T_y are the tapering parameters along axis z

Following Solina [15] we estimate the 15 parameters using the Levenberg-Marquardt method for nonlinear least squares minimization of the expression

$$\sum_{i=1}^N [\mathbf{R}(x_i, y_i, z_i; a_1, \dots, a_{15})]^2. \quad (18)$$

where,

$$\mathbf{R} = \sqrt{a_1 a_2 a_3} (\mathbf{F} - 1). \quad (19)$$

and N is the number of observed points. In our case, we use all of the range data points in each surface patch found by the isolating discontinuity boundaries found in Section 3.

To include the deformations there has to be a priority in the order of recovery imposed on them. According to studies by Leyton [10] the following order preserves better the structure of the object

$$\text{Translation}(\text{Rotation}(\text{Bending}(\text{Tapering}(x))))). \quad (20)$$

5 Examples

This Section includes examples of the segmentation of an articulated 3D object. Figure (1) shows the range image of the object and the discontinuities initially detected. Figure (2) shows respectively the extension of the discontinuities (using the snakes) in order to segment the parts of the object, where regions between the connected boundaries become the segmented surface patches, and the recovered superquadrics for the object. The rendering of the superquadrics was done using the Thingworld Modeling System [13].

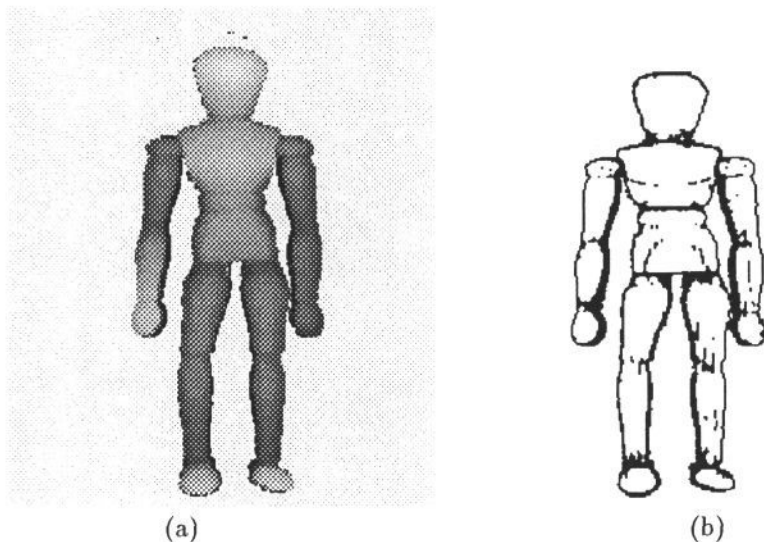


Figure 1: (a) Raw range image of a wooden doll; (b) Image of the discontinuities detected

6 Conclusions

Segmentation of 3D objects in prototypical parts provides a vision system the possibility of reasoning about the location, shape and function of complex natural and man-made objects with articulations. The differential structure of the object's surface gives useful information on where an articulated object might be segmented. A mixed approach which produces a stable estimate of the local structure of the surface without smoothing across discontinuities is presented. This allows a characterization of the local structure more precisely than simple curvature consistency alone. The use of globally deformable superquadrics as a representation for the segmented parts is especially appealing because of their great capability to capture the notion of 3D parts in natural form.

Initial tests based on the segmentation approach described here show promising results. We propose to use the segmentation in a 3D object recognition system for natural and man-made articulated objects. This class represents a wide set of

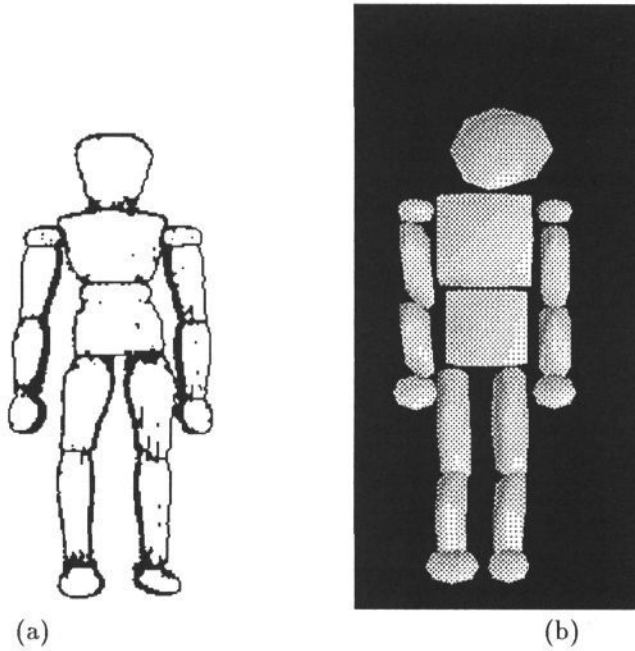


Figure 2: (a) Contours of segmented regions; (b) Globally deformable superquadrics fitted

objects dealt with in many practical situations, and segmenting the objects in a small set of 3D parts is a step in the direction of locating and identifying their shapes by means of a vision understanding system. Investigations on these issues are being carried out by us in our laboratory.

Acknowledgements

Dibio L. Borges is supported by National Council for Research and Scientific Development, Ministry of Science and Technology, Brazil, grant number 200389/88.2. The authors would like to thank S. Sclaroff and A. Pentland from M.I.T. for the licence to use the Thingworld Modeling System. This paper greatly contributed from discussions with S. Zucker, M. Trucco, and A. Fitzgibbon.

References

- [1] P. Besl and R. Jain. Three-dimensional object recognition. *ACM Computing Surveys*, 17(1):75–145, March 1985.

- [2] P. Besl and R. Jain. Invariant surface characteristics for 3D object recognition in range images. *Computer Vision, Graphics, and Image Processing*, 33:33-80, 1986.
- [3] I. Biederman. Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94:115-147, 1987.
- [4] R. Brooks. Model-based 3-D interpretation of 2-D images. In A. Pentland, editor, *From Pixels to Predicates*. Ablex Pub. Co., 1986.
- [5] L. Cai. A diffusion smoothing approach to sculptured surfaces. In D. Handscornb, editor, *The Mathematics of Surfaces III*. Clarendon Press, 1989.
- [6] M. P. do Carmo. *Differential Geometry of Curves and Surfaces*. Prentice-Hall, 1976.
- [7] F. Ferrie, J. Lagarde, and P. Whaite. Recovery of volumetric object descriptions from laser rangefinder images. In *Proc. First European Conference on Computer Vision ECCV-90, Antibes, France*, pages 387-396. Springer-Verlag, 1990.
- [8] D. Hoffman and W. Richards. Parts of recognition. In S. Pinker, editor, *Visual Cognition*, pages 65-96. MIT Press, 1985.
- [9] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *Int. J. Computer Vision*, 1:321-331, 1987.
- [10] M. Leyton. Nested structures of control: An intuitive view. *Computer Vision, Graphics, and Image Processing*, 37:20-53, 1987.
- [11] D. Marr and H. Nishihara. Representation and recognition of the spatial organization of three-dimensional shapes. *Proc. R. Soc. London B*, 200:269-294, 1978.
- [12] A. Pentland. Perceptual organization and the representation of natural form. *Artificial Intelligence*, 28:293-331, 1986.
- [13] A. Pentland and S. Sclaroff. Closed-form solutions for physically based shape modeling and recognition. *IEEE Trans. P.A.M.I.*, 13(7):715-729, 1991.
- [14] P. Sander and S. Zucker. Inferring surface trace and differential structure from 3-D images. *IEEE Trans. P.A.M.I.*, 12(9):833-854, 1990.
- [15] F. Solina and R. Bajcsy. Recovery of parametric models from range images: The case for superquadrics with global deformations. *IEEE Trans. P.A.M.I.*, 12(2):133-147, 1990.
- [16] D. Terzopoulos and D. Metaxas. Dynamic 3D models with local and global deformations: Deformable superquadrics. In *Proc. Third Int. Conf. on Computer Vision ICCV, Osaka, Japan*, pages 606-615. IEEE Press, 1990.
- [17] E. Trucco and R. Fisher. Computing surface-based representations from range images. In *Proc. IEEE Int. Symposium on Intelligent Control ISIC-92, Glasgow*, pages 275-280. IEEE Press, 1992.
- [18] S. Zucker, C. David, A. Dobbins, and L. Iverson. The organization of curve detection: Coarse tangent fields and fine spline coverings. In *Proc. Second Int. Conf. Computer Vision, Florida*, pages 568-577. IEEE Press, 1988.

