# Fast Segmentation and Modeling of Range Data via Steerable Pyramid and Superquadrics

**V. Bruni**[1], **D. Vitulano** [1]

[1] I. A. C. - C. N. R.
Viale del Policlinico 137
00161 Rome, Italy
{bruni,maniscalco,vitulano}@iac.rm.cnr.it

**U. Maniscalco** [1,2]

[2] I. C. A. R. - C. N. R.
Viale delle Scienze
90128 Palermo, Italy
maniscalco@pa.icar.cnr.it

## ABSTRACT

This paper focuses on a fast and effective model for range images segmentation and modeling. The first phase is based on the well-known Simoncelli's steerable pyramid, useful to distinguish image information from noise. Gradient modulus and phase information is then exploited for achieving edges characterizing objects. Modeling is faced through superquadrics recovery. In this case a fast and simple procedure to estimate their free parameters is proposed. Achieved results on simple objects show that our model is simple, fast and robust to noise.

### Keywords

Range image segmentation, Geometric modeling, Steerable pyramid, Superquadrics.

## 1 INTRODUCTION

Recovery of geometric shapes from a real world scene represents a challenging as well as complicated topic in Computer Vision. Its realization comes from the desire of "understanding", or at least, trying to catch and decode the knowledge of a real scene. Nonetheless, it is at the same time very complicated since it is equivalent to link low and high level experience.

A classical way to solve this problem consists of firstly segmenting and then modeling range images. Approaches avoiding segmentation phase proposed in literature usually result quite computationally expensive — see for instance [Leo97a].



***Journal of WSCG, Vol.12, No.1-3, ISSN 1213-6972***
*WSCG'2004, February 2-6, 2004, Plzen, Czech Republic.*


On the other hand, segmentation based approaches [Hoo96a] strongly depend on a correct edge detection. Therefore some preprocessing are often required for achieving satisfying results.

This paper deals with a classical two phases approach and develops the idea proposed in [Che03a]. Simoncelli's wavelet pyramid [Sim92a] is performed for image segmentation while superquadrics are employed for shape recovering based on primitives [Bar81a].

We outline that both segmentation and modeling phases have to mainly account for both fastness and accuracy.

Multiscale edge detection is not a novelty for range images and a lot of papers appear in literature — see [Bur98a] for a useful taxonomy. As a matter of fact, wavelets seem to be an effective tool for this task. They are computationally attractive and robust to noise, often affecting this kind of images. In particular, steerable wavelets provide a suitable object representation filtering the input image along different orientations. This feature is strongly exploited for obtaining cleaner gradient modulus and phase infor-

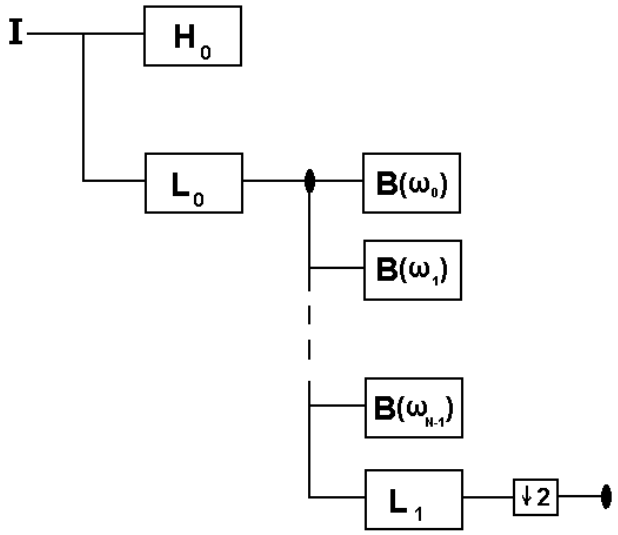


**Figure 1: Block scheme of a steerable decomposition pyramid.**

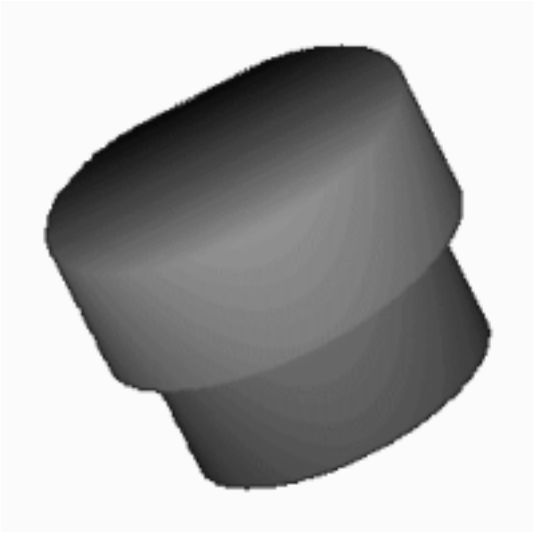

**Figure 2: First test range image: Cap.**

mation [Cal94a]. In fact, range images usually suffer from annoying sampling noise due to sensor accuracy, surface properties and viewing angles [Pag02a, Bou02a]. The gradient modulus is computed from soft-thresholded wavelet bands at first scale level [Mal98a]. The variance of the noise is estimated from residual high pass band of the same pyramid. This way allows us to detect *step edges*: discontinuities on the surface. On the contrary, *crease edges* are discontinuities of the surface normal. They can be achieved by exploiting phase information at third scale level, since quite regularized. It is worth to outline that crease edges are usually the most difficult to detect since information is usually embedded in noise, as it will be more clear later.

Shape recovering phase is based on superquadrics, that have already been adopted by many researchers in robotics [Bar81a, Pen86a, Scl91a, Sol90a]. The proposed approach relies on a fast estimation of the most parameters describing superquadrics. It is performed by exploiting the tensor of inertia information of range points distribution.

First results on simple images sketching mechanical objects are encouraging in terms of both accuracy and low computing time.

The outline of the paper is the following. Next Section focuses on a brief description of steerable wavelets. A detailed presentation of segmentation phase is contained in Section 3 along with some examples. Shape modeling using superquadrics constitutes the topic of Section 4. Finally conclusions and discussions are drawn in Section 5.

## 2 STEERABLE WAVELETS

The Steerable Pyramid [Sim92a] is a linear multiscale, multi-orientation image decomposition. It is like a two dimensional overcomplete wavelet transform (tight frame) where the basis functions are directional derivative operators with different sizes and orientations. In particular, Simoncelli's shiftable multiscale transform makes use of basis functions that are translations, dilations and rotations of a single kernel.

The steerable pyramid performs a polar-separable decomposition in the frequency domain. Then it gives an independent representation of scale and orientation which is *translation invariant* (i.e. the subbands are aliasing-free) and *rotation invariant* (i.e. the subbands are steerable). All these nice properties are of great importance in applications that involve representation of position and/or orientation of image structure. In order to be clearer, in Fig. 1 a block scheme of the pyramidal decomposition is depicted. The image is firstly separated into low ($L_0$) and highpass ($H_0$) subbands. Then the lowpass subband is divided into a set of $N$ oriented bandpass ($B(\omega_k)$, where $k = 0, \ldots, N-1$) subbands and a low-pass ($L_1$) subband. This latter is also subsampled by a factor of 2 in the $x$ and $y$ directions. Finally the recursive pyramid is achieved by performing the same scheme starting from $L_1$ subband. All the properties of a steerable pyramid are guaranteed if the filters, for a fixed frequency band, are constructed under the following constraints:

$$|L_0(\omega)|^2 = |B(\omega)|^2 + |L_1(\omega)|^2 |L_0(2\omega)|^2$$

and

$$H(\omega) = 1 - |L_0(\omega)|^2.$$

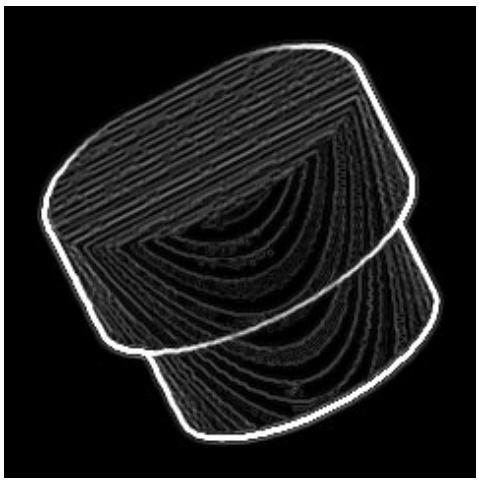

**Figure 3: Modulus of Cap computed using steerable bands at first scale level.**

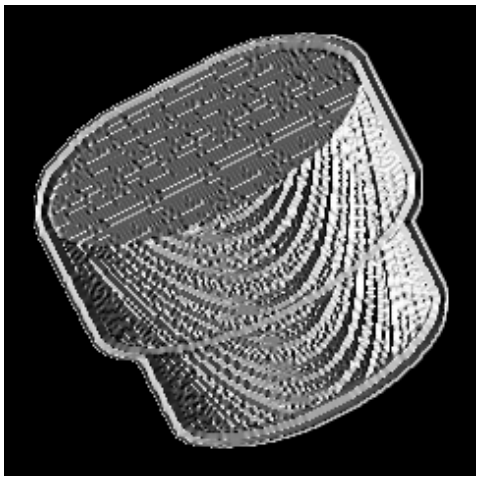

**Figure 4: Phase of Cap computed using steerable bands at first scale level.**

## 2.1 SEGMENTATION

Range images are segmented using modulus and phase information of their oriented decomposition. More precisely, modulus is defined as:

$$M_j = ((\sum_{k=0}^{5} B_j(\omega_k)\cos(\omega_k))^2 + \\ +(\sum_{k=0}^{5} B_j(\omega_k)\sin(\omega_k))^2)^{1/2}, \quad (1)$$

where $\omega_k = \{k\pi/6\}$ and $j$ is a fixed scale level. The corresponding phase is:

$$\Phi j = \arctan\left(\frac{\sum_{k=0}^{5}(B_j(\omega_k)\sin(\omega_k))}{\sum_{k=0}^{5}(B_j(\omega_k)\cos(\omega_k))}\right). \quad (2)$$

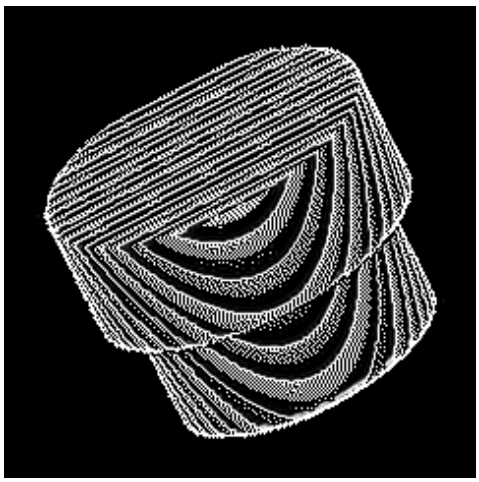

**Figure 5: High pass residual band of the steerable decomposition.**

Modulus and phase of Fig. 2 are shown respectively in Figs. 3 and 4.

Image external contour lines can be easily detected from the binarized modulus, exploiting the dark background of the analysed images.

On the contrary, the detection of internal edges is more complicated since often corrupted by noise; see, for instance, the high frequency residual band of the steerable pyramid in Fig. 5. Therefore, a simple edge detector would also catch a lot of spurious edges.

In order to cope with this problem, eq. (1) is computed at first scale level ($j = 1$), after soft thresholding wavelet bands.

The threshold is tuned computing the Gaussian noise variance from the residual highpass band, using a robust median estimator [Mal98a].

This simple and fast operation allows us to achieve results shown in Fig. 6. Edges coming from this first phase are due to quite evident discontinuities in the surface and, then, rather noticeable in the modulus. This is not the case for remaining edges. In fact, these latter are due to discontinuities of the surface normal, i.e. phase discontinuities. Unfortunately, this latter can result somewhat noisy, since sensitive to irregularities often characterizing this kind of images. An example is depicted in Fig. 7, where column no. 120 of Fig. 4 shows that information is completely embedded in noise. In such conditions, edges can be effectively detected considering them as boundary between flat regions — see for instance [Liu00a] for a very effective but expensive approach. Nevertheless, in our case, steerable pyramid at the coarsest scales gives us a quite regularized phase in a few operations. In Fig. 8, $\Phi_3$ shows that the third scale level reaches good noise reduction also on a quite noisy image

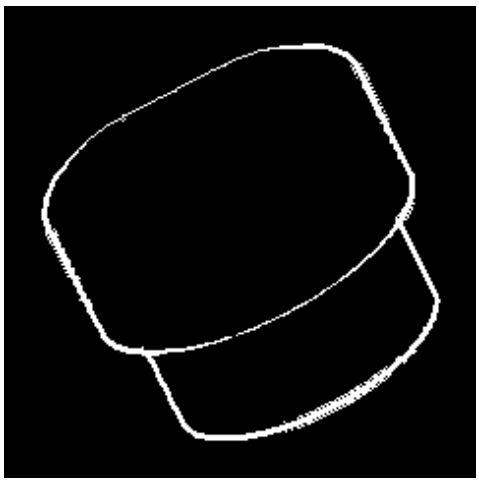

**Figure 6: Cap segmentation composed of external edges (achieved by edge detection on the binarized Fig. 3) and internal edges detected from the modulus of thresholded wavelet bands at first scale level.**

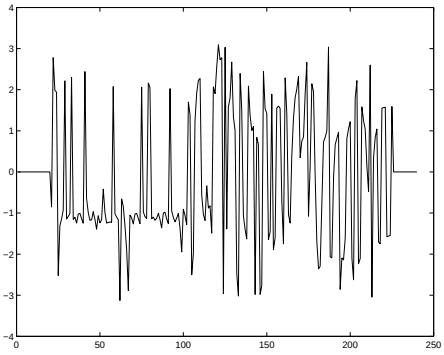


**Figure 7: Column no. 120 of gradient phase of Cap.**

**Figure 8: Phase of Cap at third scale level ($\Phi_3$) of a steerable pyramid.**

as the one under study. It is also arises that $\Phi_3$ contains the already detected edges with annoying ripples. These latter are due to the pyramidal multi-scale algorithm. So, these artefacts are filtered out through a suitable mask, accounting for the wavelet support width — see top of Fig. 9 (Top). Hence, the desired edges are composed of gradient maxima of the remaining information — see bottom of Fig. 9.

Final segmentation is shown in Fig. 10. It is evident that edges coming from phase processing are not closed. This is intrinsically due to the coarser level of investigation. As a result, T-junctions are unconnected. In order to close them, we have applied a simple heuristic technique. In practice, closure points are determined starting from one of the two extremities of a not closed edge and accounting for its curvature. The other extremity point is processed in the same way. Although heuristic, the adopted strategy allows us to manage the intrinsic difficulty of dealing with phase information.

## 2.2 GEOMETRIC MODEL RECOVERING

Geometric modeling is based on superquadrics, in particular super-ellipsoids. They consist of a family of parametric shapes introduced by Gardiener [Gar65a] and successively proposed by Barr [Bar81a] as geometric model for shape representation. Superquadrics ability in modeling natural forms was pointed out by Pentland in [Pen86a] who also showed how their generalized implicit functions can be used in fitting 3D point data [Scl91a]. Each superquadric is represented by eleven parameters: three axes, three centre coordinates, three rotation angles and two form factors. Hence, for modeling segmented range data in a world frame of reference, these parameters are often estimated minimizing an error metric, which is based on implicit function of superquadric surface. Nevertheless, the minimization represents a very hard computational task since involving all eleven superquadric parameters [Sol90a, Wha91a].

On the contrary, a quick and effective estimation of most of them can be straightforwardly computed via the tensor of inertia of range points distribution. We outline that sometimes this estimation could require a further refinement to achieve satisfying results.

In order to achieve a good recovery, previously segmented image is split into its connected components. For each of them, the centre of mass is computed using the discrete version of the Carte-

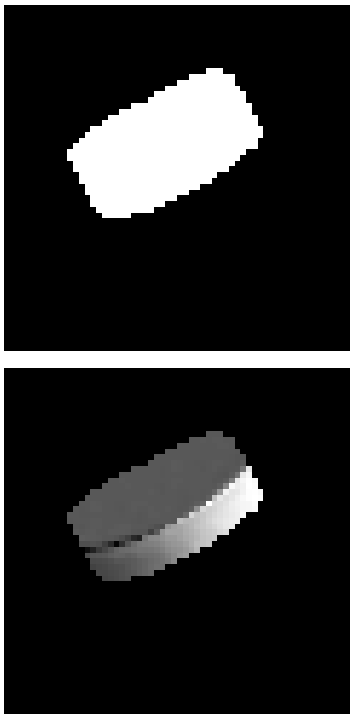

**Figure 9: Top: binary mask achieved leaving out regions closed to external edges of Fig. 6. Bottom: Fig. 8 after multiplication pixel by pixel with the above mask.**

sian moment:

$$C_x = \frac{m_{100}}{m_{000}},\ C_y = \frac{m_{010}}{m_{000}},\ C_z = \frac{m_{001}}{m_{000}}, \quad (3)$$

where $m_{100}, m_{010}, m_{001}$ are first order moments with respect to $x, y$ and $z$ axes while $m_{000}$ is the total mass.
Hence, the pseudo tensor of inertia of range points distribution is

$$I = \begin{bmatrix} I_{xx} & I_{xy} & I_{xz} \\ I_{xy} & I_{yy} & I_{yz} \\ I_{xz} & I_{yz} & I_{zz} \end{bmatrix} \quad (4)$$

where

$$\begin{aligned} I_{xx} &= \sum m((y-C_y)^2 + (z-C_z)^2), \\ I_{yy} &= \sum m((x-C_x)^2 + (z-C_z)^2), \\ I_{zz} &= \sum m((x-C_x)^2 + (y-C_y)^2), \\ I_{xy} &= -\sum m((x-C_x)(y-C_y)), \\ I_{xz} &= -\sum m((x-C_x)(z-C_z)), \\ I_{yz} &= -\sum m((y-C_y)(z-C_z)), \end{aligned}$$

and $m$ equals 1 if there exists a point of coordinates $(x, y, z)$, 0 vice-versa.
Previous tensor is related to a coordinate system translated in the barycenter. Since $I$ is hermitian,

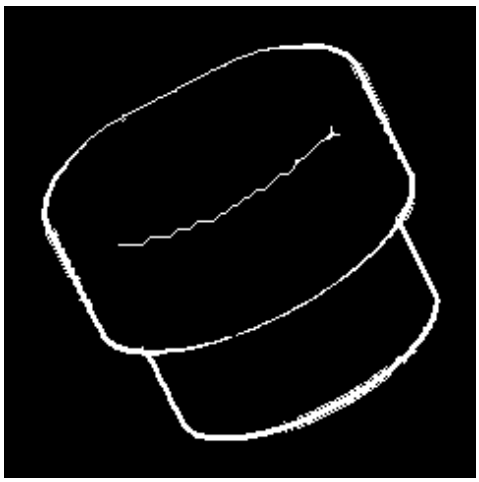

**Figure 10: Cap segmentation before connecting T-junctions.**

it can be diagonalized by a rotation matrix to obtain:

$$I_p = R'IR = \begin{bmatrix} I_x & 0 & 0 \\ 0 & I_y & 0 \\ 0 & 0 & I_z \end{bmatrix}, \quad (5)$$

where $I_x$, $I_y$, $I_z$ are three eigenvalues of $I$ representing principal moments of inertia.
Columns of matrix $R$ are $I$ eigenvectors. They represent the components of the three principal axes in a frame of reference whose origin lies on the centre of mass and whose orientation is parallel to the one of world frame of reference.
Principal axes of inertia can be mapped to the the world frame of reference via the composition of the translation component and the rotation matrix:

$$T = R_o \begin{bmatrix} 1 & 0 & 0 & -C_x \\ 0 & 1 & 0 & -C_y \\ 0 & 0 & 1 & -C_z \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (6)$$

where $R_o$ is the homogeneous version of $R'$.
Thus, six of eleven unknown superquadric parameters have been found: centre coordinates and three rotation angles.
Superellipsoids modeling relies on assumption of object symmetry with respect to image plane — just a half of the object is visible. Therefore, previously computed barycentre needs to be shifted up along z axis for compensating the non visible information. Hence, the length of superellipsoid axes can be easily computed intersecting range points distribution with world frame of reference axes.
The implicit function of superellipsoid surface is

$$\left( (\frac{x}{a_1})^{\frac{2}{\epsilon_2}} + (\frac{y}{a_2})^{\frac{2}{\epsilon_2}} \right)^{\frac{\epsilon_2}{\epsilon_1}} + (\frac{z}{a_3})^{\frac{2}{\epsilon_1}} = 1, \quad (7)$$

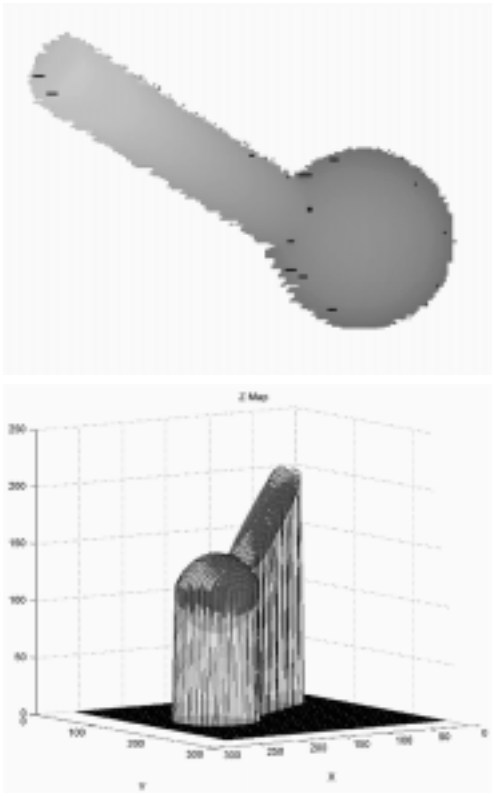

**Figure 11: Top) Shaft: test range image. Bottom) Shaft recovery using the proposed approach: two superquadrics are able to model the object**

where $\epsilon_1$ and $\epsilon_2$ are surface form factors while $a_1, a_2, a_3$ define the superellipsoid size. Hence, the remaining two form factors can be estimated minimizing the error metric [Wha91a]

$$D_3 = (a_1 a_2 a_3)^{1/2}(F^{\epsilon_1}(x, y, z, \epsilon_1, \epsilon_2) - 1), \quad (8)$$

overall range points, where $F$ is the surface implicit function written in eq. (7).
Figs. 11 and 12 depict two examples of object modeling using the above described approach.

## 3 CONCLUSIONS AND DISCUSSIONS

In this paper a fast and effective model for range images segmentation and modeling has been presented. The use of Simoncelli's steerable pyramid allows us to well distinguish image information from noise. Therefore gradient modulus and phase information of the analysed image are suitably exploited for its segmentation. Superquadrics are then used for modeling segmented data. In particular, it has been shown how most of superquadric parameters can be estimated exploiting the tensor of inertia of range points distribution.
Experimental results show that the proposed model is fast, simple and robust to noise.
This paper mainly focused on model effectiveness in terms of computational time. In fact, it is based on a few operations, even though some heuristic steps are employed. This is due to the intrinsic difficulty in analysing this kind of data, since they are acquired from an only one view. Hence, occlusions unavoidably yield lack of information. However, the direct estimation of most of free parameters of superquadrics can be a good starting point for the minimization of the $D_3$ error metric function over all eleven parameters.
Further research will be oriented to reduce required operations for object modeling and to make the proposed framework more robust.

## REFERENCES


[Bar81a] A. H. Barr, Superquadrics and angle-preserving trasformation, IEEE Computer Graphics Application, vol. 1, pp. 11-23, 1981.

[Bou02a] P. Boulanger, O. Jokinen, J.A. Beraldin, Intrinsic filtering of range images using a physically based noise model, Proceedings of International Conference of Vision Interface 2002, Calgary, Canada, May 2002.

[Bur98a] S.G. Burgiss, Range image segmentation through pattern analysis of the multi-scale wavelet transform, Master of Science Degree Thesis, University of Tennessee, Knoxville, August, 1998.

[Cal94a] F. Callari, U. Maniscalco, "New robust approach to image and 3D shape reconstruction", Proceedings of International Conference on Computer Vision and Pattern Recognition, Jerusalem, Israel, pp. 103-107, 1994.

[Che03a] A. Chella, U. Maniscalco, R. Pirrone, A Neural architecture for 3D Segmentation, Proceedings of XIV Italian Workshop on Neural Nets WIRN Vietri03, Vietri Sul Mare (SA), Jun. 5-7, 2003.

[Gar65a] M.Gardiener, The superellipse: A curve that lies between the ellipse and the rectangle, Sci. Amer, vol. 213, pp,222-234, 1965.

[Hoo96a] A. Hoover, G. Jean-Baptiste, X. Jiang, P.J. Flynn, H. Bunke, D. Goldgof, K. Bowyer, D.W. Eggert, A. Fitzgibbon, R.B. Fisher, "An experimental comparison of range image segmentation algorithms", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 18(7), pp. 673–689, 1996.

[Leo97a] A. Leonardis, A. Jaklic, F. Solina, Superquadrics for segmenting and modelling range data, IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 19, no. 11, pp 1289-1295, 1997.

[Liu00a] X.Liu, D.L. Wang, R. Ramirez, Boundary detection by contextual non-linear smoothing, Pattern Recognition, vol. 33, pp. 263-280, February, 2000.

[Mal98a] S. Mallat, "A Wavelet Tour of Signal Processing", Academic Press, 1998.

[Pag02a] D.L. Page, Y. Sun, A.F. Koschan, J.K. Paik, M.A. Adibi, Simultaneous mesh simplification and noise smoothing of range images, Proceedings of IEEE International Conference on Image Processing, Rochester, New York, 2002.

[Pen86a] A. P. Pentland, Perceptual organization and the representation of natural form, Artificial intelligence, vol. 28, no. 3, pp. 293-331, 1986.

[Scl91a] S. Sclaroff, A. P. Pentland, Generalized implicit function for computer graphics, Computer Graphics, vol. 25, no. 4, pp, 247-250, 1991.

[Sim92a] E.P. Simoncelli, W.T. Freeman, E.H. Adelson, D.J. Heeger, "Shiftable Multi-scale Transforms", IEEE Trans. Information Theory, vol. 38, no. 2, pp. 587-607, March 1992.

[Sol90a] F. Solina, R. Bajcsy, Recovery of parametric model from range image: the case for superquadrics with global deformation, IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 12, no. 2, pp 131-147, 1990.

[Wha91a] P. Whaite, F. P. Ferrie, From uncertainty to visual exploration, IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 13, no. 10, pp 1038-1049, 1991.

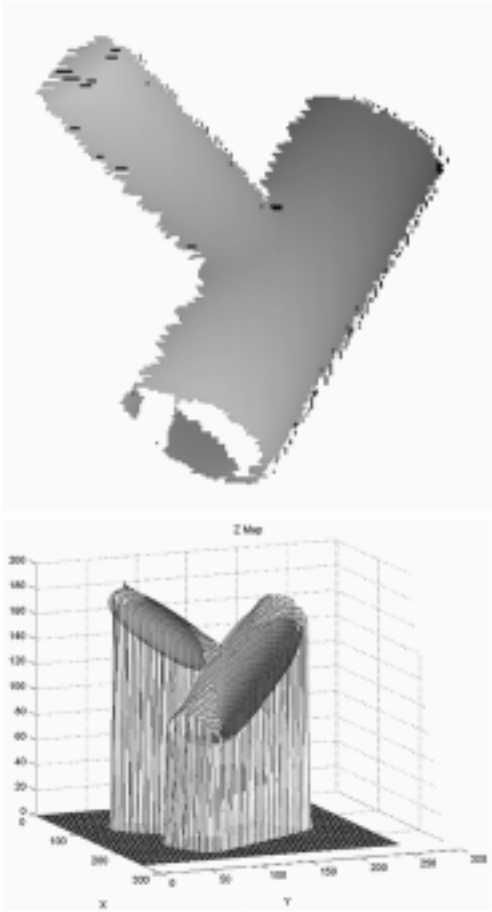

**Figure 12: Top) Tpipe: test range image. Bottom) Tpipe recovery through the proposed approach.**