Part based Segmentation and Modeling of Range Data by Moving Target

Roberto Pirrone

Dipartimento di Ingegneria Automatica e Informatica
University of Palermo, and CERE-CNR,
Viale delle Scienze, I-90128 Palermo, Italy
e.mail: pirrone@unipa.it

Abstract

A system for part-based segmentation of range data and their interpretation as a composition of deformable superquadrics is described. Segmentation and reconstruction phases are performed using the same algorithm at different scales. First, the data set is partitioned in regions corresponding approximately to simple convex objects, and then single deformable models are fitted to each region. Refinements of the model can be achieved by recursively applying the method. The proposed Moving Target (MT) algorithm is an original variation of the well known approach by Solina, ideated to avoid the classical inconvenient of the minimization procedure where the solution escapes the global minimum and/or gets stuck in a local one. Motivations of the proposed approach along with its theoretical formulation are presented and discussed in detail. The whole system has been validated using a several range images.

Key Words: Object recognition; Recognition by parts; Moving target; 3D Segmentation; Fitting of range data; Superquadrics.

1. INTRODUCTION

The development of the vision system is a crucial part of the whole architectural design for an autonomous robot: vision processes has to be at the same time fast, robust and accurate to guarantee the correct perception of the essential elements that are present in the operating environment. Moreover, the kind of images processed from the visual component of the robot, and the features that can be extracted from them, affect the other sensors equipment, the shape and, to some extent, the mission abilities of the robot itself.

Object recognition is one of the most intriguing visual processes to be modeled in a robot vision system. Really, it is not so clear if we recognize an object using some sort of 3D mental model encoding structural relations between its parts, or if we learn and store in our memory different views of the object itself, that in turn allow recognition using a global matching with our actual perception.

These two ways of thinking produced an interesting debate, during the last twenty years, both in the psychological and in the computer vision communities, giving rise to two main group of theories. On one hand, there is a group of theories that are commonly referred to as “recognition by parts” which assume that human beings perform a sort of volumetric segmentation of the perceived image where each part is related to the others by structural relations such as *above()*, *larger()*, *side()* and so on (Marr 1982, Marr and Nishihiara 1978). The object shape is thus described with few flexible primitives that can be related to each other in several ways. The most famous system in this area is JIM and its evolutions proposed by Biederman and his colleagues (Biederman 1987, Biederman and Hummel 1992, Hummel and Stankiewicz 1996). JIM uses "geons" as geometric primitives, but other systems have been poposed in the computer vision community that use superquadrics (Pentland 1987, Dickinson Pentland and Rosenfeld 1992).

On the other hand several view-based theories of object recognition have been developed following the key idea of a global match between the perceived image and some image-like "views" stored in our long term memory. Several vision systems have been proposed in literature, each with its own definition of the concept of view, depending on the particular theory to be validated. In general, one can say that a view is a vector containing the spatial coordinates of some image features. Such coordinates are expressed relative to a common reference point. The two main approaches to view matching are the global (*holistic*) image matching proposed by Poggio, Edelman and their collegues (Poggio and Edelman 1990, Edelman and Poggio 1991, Edelman and Weishalll 1991, Bülthoff and Edelman 1995, Edelman 1998) and the theory of alignment, proposed by Ullman, where two-dimensional image features are geometrically aligned with a three-dimensional object model (Ullman 1989 and 1996, Ullman and Basri 1989).

Both recognition by parts and view-based recognition have been validated by psychological findings. One of the strongest arguments supporting recognition by parts is that humans can easily recognize a novel object as belonging to a particular class, despite slight variations in the displacement and in the shape of its parts. This finding can be explained with the dynamic binding between parts in the object structural description: few simple primitives can be arranged in different ways at recognition time to obtain the whole model (Hummel 2000). In contrast, view-based approaches suffer from static binding of features to their position in the vector defining the view, and are not able to catch consistent rearrangement of the features in the reference space. The argument in favor of view-based approaches is the recognition invariance to rotations in the image plane and, to some extent, rotations in depth. See the example of the structural description of a coffee mug and a bucket in (Biederman 1987) for an explanation of this topic. The sensitivity of recognition by parts to rotation have been an arguments for several debates between exponents of the two theories (Biederman and Gerhardstein 1993 and 1995, Tarr and Bülthoff 1995, Tarr 1995).

In principle, both the approaches described above can be used in a robot vision system. Two-dimensional feature matching is a classical paradigm in robot vision where the main task is navigation and exploration of the environment (Arkin 1998). Even in the case of manipulators, view alignment can be a good solution for the robot visual servoing problem (Hutchinson Hager Corke 1996) where the environment is totally controlled as in plants or factories. The framework of the system presented in this work is the development of the visual component for a robot manipulator that operates in a partially known environment, so it has to learn objects' structure, moving the actuator in a space not well a priori defined, and classify novel objects. Such requirements can be satisfied only with a rigorous knowledge about the geometry and the spatial relations between objects in a true 3D model of the world.

The presented system has been developed in the framework of a cognitive architecture for robot vision (Ardizzone Chella Frixione Gaglio 1992, Chella Frixione and Gaglio 1997 and 2000, Chella Gaglio

Pirrone 2001). This architecture aims to integrate the perception of static and dynamic scenes with their symbolic representation in order to understand the perceived environment and to guide the actions of the robot in a feedback loop where novel perceptions enrich the symbolic knowledge about the operating domain.

The integration of the presented system in the general architecture depends in a crucial manner from the segmentation step. The model recovered from raw data can be useful to the symbolic component of the robot only if it fulfills some criteria about the goodness of the obtained reconstruction. In particular I claim tat a good reconstruction is *minimal* and *meaningful*. In what follows these two terms are explained in detail.

A reconstruction is said minimal when the number of recovered models matches exactly the number of the parts obtained with segmentation.

Meaningfulness means that each recovered model has to be recognizable (that is matched with a high goodness-of-fit) as a convex primitive at most affected by a global deformation. FIGURE 1 provides a graphical example of minimality and meaningfulness of the reconstruction.

FIGURE 1 NEAR HERE

Meaningfulness ensures that symbolic descriptions of the objects' structures are coherent with each other and that new ones can be added to a global knowledge base. On the other hand, a minimal reconstruction helps the autonomous agent to generate hypotheses about the object nature when it has to disambiguate a particular reconstruction. In other words, meaningfulness acts as a syntactic constraint over the reconstruction, making its symbolic description expressed using words that are taken from a well defined vocabulary. In the same way, a minimal reconstruction is to be regarded as a semantic constraint over the object description because its expressive power is limited to a particular form (i.e. a hammer can be only described as *above(box, cylinder)*).

In this context, the vision system presented in this work has been developed by the author with the previous principles in mind. It is intended for 3D segmentation and modeling of range data as a composition of deformable superquadrics, in order to provide detailed geometric and structural description of the modeled object. The system is the perceptual component of a cognitive architecture for the supervision and control of an autonomous robot performing grasp and manipulation tasks in a partially structured operating environment. The robot needs detailed 3D information in order to perform movements and to plan manipulation strategies to classify unknown objects.

The whole system relies on a novel algorithm proposed by the author to perform data fitting. The Moving Target (MT) algorithm extends and generalizes the approach to fit superquadrics to range data developed by Solina (Solina and Bajcsy 1990). It is applied repeatedly at different scales to the input data, first to obtain part segmentation and then to fit each segment with a 3D model. In this context, a "part" has to be intended as a region of the data set that can be modeled with a simple convex primitive at most affected by a global deformation like bending or tapering. As a consequence the system searches for regions with the maximum concavity in the data set and uses them to subdivide the scene in convex segments.

Another source of inspiration is the work by (Whaite and Ferrie 1991). Whaite and Ferrie point out the general non-uniqueness of data interpretation in the least-squares fitting procedure, regardless to the metric that is being used to measure goodness-of-fit. The main problem with misinterpretation of data is in the initial estimate of the model, above all when the viewpoint is prone to generate ambiguities (one of the major axes of the model results aligned with the line of sight). In the MT approach, points are projected from their true position onto the smallest enclosing sphere positioned in the data center of mass. Then the fitting process is performed in multiple steps, each of them using the previous estimate as the starting point to fit the model to a data set a bit more shrunk towards the original configuration. In this way, the error surface is constrained to progressively deform towards its final shape. Convergence of the solution is induced by the intrinsic convexity of the error function near the global minimum (Solina 1987). The solution is always close to the global minimum and moves towards it avoiding local ones and/or flat zones.

From the above description, it is straightforward to argue that maximum concavity points converge more quickly then others to their true positions, so they can be easily detected as the ones with the highest shrinking velocity. The whole system performs as follows: first the MT algorithm is applied to the entire data set to separate convex regions; then the same algorithm is used to fit a 3D model onto each part. Residual analysis has been used both to refine boundaries between convex regions and to measure the goodness-of-fit for each recovered model. In case of poor fitting, the algorithm can be applied repeatedly both to find new convex segments and to recover a better model for a single part.

The rest of the paper is arranged as follows. In section 2 the mathematical formulation of the problem is reported along with the motivations for the choice of superquadrics as geometric primitives. Section 3 describes in detail the MT algorithm for fitting a convex data set and the global segmentation and fitting procedure. In section 4 experimental results are reported with several typologies of depth images. Finally, section 5 reports conclusion and some discussion about future work.

2. MATHEMATICAL FORMULATION OF THE PROBLEM

The following mathematical formulation has been developed taking into account the literature about model selection techniques following the Minimal Description Length principle (MDL) and its applications to visual data segmentation for object modeling (Darrel Sclaroff and Pentland 1990, Solina and Leonardis 1998).

Given some range data acquired from an object under a generic viewpoint, they can be regarded as a set of three-dimensional points $O = \{\mathbf{x}_i : \mathbf{x}_i = (x_i \quad y_i \quad z_i)\}_{i=1}^{N}$ which defines the object itself at the lowest level of abstraction. The segmentation process is aimed to subdivide O in a number of non-overlapping parts P_i such that:

$$O = \bigcup_i P_i,\ \forall k\ P_k = \{\mathbf{x}_i\}_{i=1}^{M \le N},\ P_k \subset O$$
$$\forall P_k, P_l \subset O \Rightarrow P_k \cap P_l \equiv 0\,.$$

A generic geometric primitive used to model raw data, is defined in the form $\mathcal{M}(\mathbf{x}, \ddot{\mathbf{e}})$, where $\ddot{\mathbf{e}}$ is a suitable vector of tunable parameters, while $\mathbf{x}$ is a generic 3D point. Starting from the last definition a reconstruction $\mathcal{R}(O)$ is defined as:

$$\mathcal{R}(O)=\langle \mathcal{M}(\mathbf{x},\ddot{\mathbf{e}}_i)\rangle_{i=1}^{M}.$$

A reconstruction is therefore a tuple of instances of the model; each instance is characterized by a particular configuration $\ddot{\mathbf{e}}_i$ of the elements in the parameter vector, and it fits one part of the data set:

$$\mathcal{M}(\mathbf{x},\ddot{\mathbf{e}}_i)\equiv \mathcal{R}(P_i)$$

Using the previous statements it is possible to provide a definition for a reconstruction $\mathcal{R}^*$ that is both minimal and meaningful. The former property is expressed claiming that $\mathcal{R}^*$ must have the same number of models as the number of parts obtained from the segmentation step:

$$O=\bigcup_{i=1}^{k}P_i \Rightarrow \ \mathcal{R}^*(O)=\langle \mathcal{M}(\mathbf{x},\ddot{\mathbf{e}}_i)\rangle_{i=1}^{k}. \tag{1}$$

The equation (1) is just a formalization of the MDL principle applied to the domain under investigation, but in this formulation no a priori probability distribution is taken into account for the segmentation and the reconstruction processes. The guiding principle is that the whole object is segmented in simple convex parts, searching the data set for the regions with maximum convexity.

A general, but too restrictive, formulation of meaningfulness implies that each part P_i has one and only one reconstruction corresponding with the instance of the model defined by $\ddot{\mathbf{e}}_i$, assuming that each P_i corresponds to a convex region.

$$\forall P_i \ \exists!\ \mathcal{M}(\mathbf{x},\ddot{\mathbf{e}}_i) : \mathcal{M}(\mathbf{x},\ddot{\mathbf{e}}_i)\equiv \mathcal{R}^*(P_i). \tag{2}$$

The above statements are too general to be useful in a computational model. In practice, minimality is satisfied using the constraint of modeling each part with a single instance of $\mathcal{M}$. Meaningfulness is obtained as a constraint on the reconstruction goodness-of-fit. In what follows the reconstruction error $epr(\mathcal{R})$ relative to a reconstruction will be used as inverse measure of goodness-of-fit. A general expression for *epr* relative to the k-th part $P_k \subset O$ can be provided as:

$$epr(\mathcal{R}(P_k))=\sum_{\mathbf{x}_i\in P_k} D(\mathcal{M}(\mathbf{x}_i,\ddot{\mathbf{e}}_k)). \tag{3}$$

In (3) *D* is a suitable function depending on the analytical form of the metric one wants to use to measure goodness-of-fit, while $\mathcal{M}(\mathbf{x}_i,\ddot{\mathbf{e}}_k)$ is the result of evaluating the model instance defined by $\ddot{\mathbf{e}}_k$ in the point $\mathbf{x}_i$. Meaningfulness can now be expressed by the following:

$$\forall P_j \subset O \ epr^*(\mathcal{R}(P_j))\equiv \min_{\ddot{\mathbf{e}}\in\Lambda} epr(\mathcal{R}(P_j))=\min_{\ddot{\mathbf{e}}\in\Lambda}\sum_{\mathbf{x}_i\in P_j} D(\mathcal{M}(\mathbf{x}_i,\ddot{\mathbf{e}})),\ j=1,\ldots,k. \tag{4}$$

Here Λ is the parameter space. The formulation of (4) is a theoretical limit for two reasons. A minimization procedure provides, in general, a sub-optimal solution in order to save computational time. Moreover, the choice of *D* can influence the result depending on the constraints one has used in the construction of its analytical form (Gross and Boult 1988, Solina and Bajcsy 1990, Whaite and Ferrie 1991).

In the working system, $\mathcal{M}$ has been implemented using deformable superquadrics. Superquadrics, in their original form, have been introduced by Barr (Barr 1981) as a simple extension of the quadrics where two real exponents can tune the shape profile both in latitude and in longitude. A superquadric inside-outside function is expressed as:

$$f(x,y,z)=\left(\left(\frac{x}{a_1}\right)^{2/e_2}+\left(\frac{y}{a_2}\right)^{2/e_2}\right)^{e_2/e_1}+\left(\frac{z}{a_3}\right)^{2/e_1}=1.$$

Superquadrics are a powerful tool for expressing different 3D shapes (see FIGURE 2) and have been widely used in computer vision, since the first eighties, as good primitives for object description by parts (Pentland 1986).

FIGURE 2 NEAR HERE

Some systems for 3D reconstruction from range data make use of other primitives like hyperquadrics (Kumar Han Goldgof and Bowyer 1995, Kumar and Goldgof 1995) or extend the superquadric model in order to cope with local shape deformations (Terzopoulos and Metaxas 1991, Bardinet Cohen and Ayache 1998, Zhou and Kambhamettu 1999). The author preferred to use superquadrics than hyperquadrics due to computational reasons: a hyperquadric requires 30 parameters to be fully defined in 3D space, while a superquadric needs only 15 parameters including the ones describing deformations. Moreover, the hyperquadric fitting procedure requires a rather complex initialization where some parameters must be set using heuristic criteria that are not well specified. Computational reasons also hold for the use of extended superquadric models to deal with local and/or global deformations. These kinds of systems are more suited to computer graphics applications than for robot vision: in fact, model computation becomes too heavy to allow fast recognition, above all in those cases where parametric surfaces are used to describe object's details. Starting from the expression of the inside-outside function, the model is enriched with parameters describing attitude in 3D space and eventual deformations. The parameter vector can be thus split in the following form:

$$\ddot{\mathbf{e}}=\left[\ddot{\mathbf{e}}_{\mathrm{int}}\,|\,\ddot{\mathbf{e}}_{c}\,|\,\ddot{\mathbf{e}}_{or}\,|\,\ddot{\mathbf{e}}_{b}\,|\,\ddot{\mathbf{e}}_{t}\right]$$

Previous expression accounts separately for the parameters in the inside-outside function, the center coordinates of the object, its orientation angles, the bending and tapering factors. The model $\mathcal{M}$ results defined (according to Solina 1990):

$$\begin{gathered}\mathcal{M}(\mathbf{x},\ddot{\mathbf{e}})\equiv\Im(\mathbf{x},\ddot{\mathbf{e}}_{\mathrm{int}},\ddot{\mathbf{e}}_{c},\ddot{\mathbf{e}}_{or},\ddot{\mathbf{e}}_{b},\ddot{\mathbf{e}}_{t})=f^{e_1}(\mathbf{X},\ddot{\mathbf{e}}_{\mathrm{int}}),\\ \mathbf{X}=\mathrm{Trans}(\ddot{\mathbf{e}}_{c},\mathrm{Rot}(\ddot{\mathbf{e}}_{or},\mathrm{Bend}(\ddot{\mathbf{e}}_{b},\mathrm{Taper}(\ddot{\mathbf{e}}_{t},\mathbf{x})))).\end{gathered}$$

The selected metric is the "minimal volume" one adopted by Solina and Bajcsy, so the expression for D is:

$$D(\Im(\mathbf{x},\ddot{\mathbf{e}}))=\sqrt{a_1a_2a_3}\,(1-\Im(\mathbf{x},\ddot{\mathbf{e}})).$$

Here a_1, a_2, and a_3 represent the model size along its three main directions. The reconstruction error *epr* is:

$$epr = \frac{1}{N}\sum_{i=1}^{N}\left[\sqrt{a_1 a_2 a_3}\left(1-\Im\left(\mathbf{x},\ddot{\mathbf{e}}\right)\right)\right]^2 .$$

Here *N* is the total number of data points. Finally the meaningfulness constraint is expressed as:

$$\min_{\ddot{\mathbf{e}}\in\Lambda}\frac{1}{N}\sum_{i=1}^{N}\left[\sqrt{a_1 a_2 a_3}\left(1-\Im\left(\mathbf{x},\ddot{\mathbf{e}}\right)\right)\right]^2 . \qquad (5)$$

This expression is exactly the minimization criterion adopted by Solina and Bajcsy. The author wants to derive an extension of this method that copes with the segmentation problem and the procedure has proved to work well if the object's parts are extracted as convex regions where at most some deformations of the model are needed in order to improve the reconstruction goodness-of-fit. Moreover, Solina proposes this metric as a way to ensure "perceptual acceptability" of the model: this constraint matches with the definition of meaningfulness previously given. Finally, in case of poor reconstructions, an autonomous robot is able to perform disambiguation varying interactively the viewpoint.

Despite previous considerations, different metrics have been used in the experimental phase without any remarkable performance improvement. This topic will be discussed in detail in the next two sections.

3. THE MOVING TARGET ALGORITHM

In this section the description of the MT algorithm will be provided along with a detailed discussion of the implementation choices. First, the use of the MT algorithm to retrieve a single convex segment with a deformable superquadric will be exposed, making a comparison with the original approach by Solina. Then the general procedure to obtain segmentation and reconstruction of the whole object will be discussed. Finally a comparison will be made with other approaches proposed in literature.

3.1 Fitting a Single Segment

The MT algorithm is based essentially on a sequence of steps where a model is fitted to a "moving" data set. The concept of moving data set means that the points configuration is slightly deformed at each step, from a sphere placed in the data center of mass, to the actual shape, as it has been acquired by the sensor.

FIGURE 3 NEAR HERE

With reference to FIGURE 3, the generic point **x** projects itself onto a point $\mathbf{x}_p$ on the smallest enclosing sphere by the data center of mass $\mathbf{x}_0$. Denoting with $d = \left|\mathbf{x}-\mathbf{x}_0\right|$ the Euclidean distance between **x** and $\mathbf{x}_0$, and expressing the radius *R* of the sphere as $R = \max_{\mathbf{x}} d$, it is possible to obtain $\mathbf{x}_p$ as

$\mathbf{x}_p = \mathbf{x}_0 + \mathbf{p}$. In turn, **p** has the following expression $\mathbf{p} = R(\mathbf{x} - \mathbf{x}_0)/d$ meaning that it has a modulus equal to R and the same direction of the difference (**x** - **x**$_0$). Finally, **x**$_p$ has the form:

$$\mathbf{x}_p = \mathbf{x}_0 + \frac{R}{d}(\mathbf{x} - \mathbf{x}_0).$$

Shrinking deformation from **x**$_p$ back to **x** is performed using the parametric equation of the straight line passing by these two points, so the point **x**$_t$ at deformation step *t* is obtained as $\mathbf{x}_t = \mathbf{x}_p + (t/N_t)(\mathbf{x} - \mathbf{x}_p)$. Here N_t is the total number of deformation steps. Combining the expression of **x**$_t$ with the one derived for **x**$_p$, it is possible to obtain an iterative solution to compute intermediate configurations without projecting points onto the smallest enclosing sphere:

$$\mathbf{x}_t = \mathbf{x}_0 + k(\mathbf{x} - \mathbf{x}_0),\ k = \frac{R}{d} + \frac{t(d-R)}{dN_t}. \qquad (6)$$

Intermediate deformations along with an enhanced minimization strategy, with respect to the Levenberg-Marquardt implementation used by Solina, allow the algorithm to find meaningful solutions even in the case of poor viewpoints. The working system has the possibility to switch between the implementation of the Levenberg-Marquardt method provided by Morè (Morè 1977) and a trust region algorithm that is based on the interior-reflective Newton method described in (Coleman and Y. Lin 1994 and 1996). Here each iteration involves the approximate solution of a large linear system using the method of preconditioned conjugate gradients (PCG). The choice for these two algorithms derives directly from the use of the moving target strategy. Both the cited approaches are based on the key idea that the error surface can be approximated with a simpler function in a suitable neighborhood of the solution, while the Levenberg-Marquardt or the PCG algorithms are used to move one step on this surface towards the solution itself. In the presented system this is the case: Solina has demonstrated that the error surface is concave near the global minimum, thus it can be locally approximated with a second order surface, and the MT algorithm allows the solution to be always close to it, due to the progressive deformation imposed by data shrinking (see FIGURE 4).

FIGURE 4 NEAR HERE

As regards the use of two different minimization algorithms, some authors (Bardinet Cohen and Ayache 1998) have proposed the conjugate gradients approach with respect to Levenberg-Marquardt because the former is faster, requiring only to compute the Jacobian of the function to be minimized, while the latter requires also the Hessian. In the implemented system the two approaches exhibit comparable performance as it is showed in FIGURE 5. A non-iterative method has been proposed in literature to estimate single superquadrics from depth maps (Cotronei and Salvato 1996) but, if it saves computational time, it is limited to un-deformed models, and needs exact knowledge of the projected contours onto the coordinate planes, so it fails in case of noisy real data. In what follows the control flow of the algorithm is reported.

FIGURE 5 NEAR HERE

```
ALGORITHM I: Moving Target strategy to fit a single model.
STEP 1. Reduce the noise in the input data by applying a 7x7 median filter in order to remove isolated points.
STEP 2. Transform range image points $z(x, y)$ from image coordinate system to world coordinate system $\mathbf{x}=(x, y, z)$ centered in the lower left corner of the input range image.
STEP 3. Compute the center of mass $\mathbf{x}_0$ of the data set, and project each point $\mathbf{x}$ onto $\mathbf{x}_p$ using the (6) with $t=0$.
STEP 4. Perform initial model estimate $\mathcal{M}_0$
    4a) Compute eigenvectors and eigenvalues of the moment matrix of the data points, put the $z$ axis along the eigenvector with the minimum inertia and estimate the Euler angles to determine the pose.
    4b) Compute the size of each axis as the extremity points of the data set along each eigenvector computed at the step 4a), set the form factors to 1.0 and set all the deformation parameters to 0.0.
STEP 5. Loop until $t=N_t$
    5a) $t=t+1$.
    5b) For each point in the data set compute $\mathbf{x}_{t+1}$.
    5c) Fit the data using whatever minimization method, with $\mathcal{M}_t$ as initial estimate; exit when the number $n$ of iterations is reached or $(epr|_{k-1} – epr|_k)$ goes below a suitable threshold at step $k$, thus obtaining the model $\mathcal{M}_{t+1}$.
```

3.2 The general approach

The MT algorithm described in the previous sub-section can be extended to perform multi-part segmentation of the data set on the basis of some simple considerations that will be stressed in the following.

According to the definitions for a minimal reconstruction and a meaningful one, the segmentation process is aimed to isolate convex parts in the range data, each of them being modeled by one (at most) deformed superquadric. As a consequence, the algorithm has to search the range image for contour points and for points laying in regions with maximum concavity. The last ones will be referred to as “junction points”.

After noise and isolated points are reduced with a 7x7 median filter, a Canny edge detector (Canny 1986) is used to find contours. This is a good choice due to the nature of the images: background (too deep points) is black, surfaces vary smoothly, and boundary and occluding contours are easy to detect. A simple contour following algorithm is used to detect closed and open contours, T-shaped junctions between them, corners and termination points. The same algorithm removes small segments and small closed contours (holes in the data set). Small gaps between contour extreme points are filled with straight line segments. The flow control of the contour following algorithm is reported at the end of this sub-section. Regions inside a closed contour are segmented immediately. In general, the edge detector is not able to close contours in correspondence of junction points because their intensity values in the range image vary too smoothly. In practice, junction points represent a discontinuity contour connecting occluding contours with either boundary ones or other occluding contours (see FIGURE 6).

Junction points will be searched only in those regions lying between the ends of an open contour segment, and the MT algorithm provides a way to initialize this search.

Consider a two parts object as the one depicted in FIGURE 6. Applying the MT algorithm to the data as a whole, gives rise to two consequences. First, one can observe that the junction points have the highest shrinking velocity; moreover junction points exhibit the highest *epr* value, after reconstruction has been performed. It is possible to express the first property in a mathematical form. Recalling equation (6) the shrinking velocity can be expressed as the first derivative of $\mathbf{x}_t$ with respect to the parameter *t*. Because of *t* is a discrete variable, the first derivative of $\mathbf{x}_t$ will be coincident with the finite difference:

$$\frac{\Delta \mathbf{x}_t}{\Delta t} = \frac{R-d}{dN_t}(\mathbf{x}_0 - \mathbf{x}).$$

Remembering that $|\mathbf{x}_0 - \mathbf{x}| \equiv d$, the scalar shrinking velocity v_s is:

$$v_s = \left|\frac{\Delta \mathbf{x}_t}{\Delta t}\right| = \frac{R-d}{N_t}. \tag{7}$$

In the implementation, v_s has been scaled relative to *d* in order to obtain a sharper distribution near the maxima, so the actual value is $v_s^{'} = (R-d)/d$ where the constant N_t has been neglected. According to the previous statements, the criterion used to select junction points is to search for those points in the data set having $v'_s > \tau$, where τ is a suitable percentage of the maximum value for the v'_s distribution. These points will form a junction region. For a multi-part object, several small isolated junction regions are obtained in correspondence to different sets of concavity points. In general, $J_{Pk,i}$ will indicate the i-th junction region of the k-th part P_k of the object. A junction region is accepted only if it contains both the termination points for a given contour segment. Otherwise, a false positive has been detected: the surface has no local concavity but is a rather flat pacth or has small local concavity that can be fitted using a bend deformation. The range image made only by the junction regions is partitioned using again the edge detector to find each region's contour. Junction regions that do not contain the two termination points for any given contour segment are discarded. Each junction region image is used to initialize a snake (Blake and Isard 1998) aimed to interpolate exactely the junction points (see Appendix A).

FIGURE 6 NEAR HERE

The whole algorithm works as follows. When some junction regions are detected, a snake is fitted for each of them, and the data set is split in as much parts as the number of snakes plus 1. The fitting procedure is applied to each part: if *epr* is not so good for some part the splitting procedure is repeated, searching for some other junction region. In the following, flow control of the whole algorithm is reported.

CONTEXTR: Function for contour extraction, labeling and segmentation of closed regions
STEP 1. Apply the Canny edge detector to the range image.
STEP 2. Initialize the set of parts *P*={}

```
Step 3. Detect the first contour point and follow the contour to find each corner, T-shaped junction and termination, thus computing $E=\{\mathbf{e}i\}$ that is the endpoint set.
Step 4. Compute the set of contour segments $C=\{c_i=\langle \mathbf{e}_{1i},\mathbf{e}_{2i},l_i\rangle\}$ where each contour segment is a triple defined by the endpoints and the contour label.
Step 5. Remove the contours that are too small.
Step 6. if there are some $c_i$ such that $\mathbf{e}_{1i}\equiv \mathbf{e}_{2i}$
        6a) Obtain the part $P_k$ as the set of all the points belonging to the region bounded by $c_i$.
        6b) $P=P\cup P_k$.
        6c) $C=C-\{c_i\}$
Step 7. Merge chained contour segments in order to find closed contours or extended open contours.
Step 8. Repeat Step 6.
Step 9. Return $\langle P,C\rangle$.
```

```
Algorithm II: Global segmentation and reconstruction procedure using the MT algorithm.
Step 1. Reduce the noise and remove isolated points.
Step 2. Transform range image points $z(x, y)$ from image coordinate system to world coordinate system $\mathbf{x}=(x, y, z)$ centered in the lower left corner of the input range image.
Step 3. Compute the set of parts $P=\{P_k\}$, and the set of contours $C=\{c_j=\langle \mathbf{t}_{1j},\mathbf{t}_{2j},l_j\rangle\}$ $\langle P,C\rangle \leftarrow$ ContExtr.
Step 4. Set $epr(\mathcal{R}(P_k))$ = MAXVAL for each $k$.
Step 5. For each part $P_k$ such that $epr(\mathcal{R}(P_k)) > th$
        5a) Compute the center of mass $\mathbf{x}_0$ of $P_k$, and project each point $\mathbf{x}$ onto $\mathbf{x}_p$ using the (6) with $t$=0.
        5b) Compute the distribution $v'_s(x, y)$ for each point in $P_k$.
        5c) If there are some points $z(x, y) \in P_k$ such that $v'_s(x, y) > \tau$.
                5c.1) Apply the zero-crossing operator on the junction range image to obtain junction regions $J_{Pk,i}$, $i$=1,…,n.
                5c.2) Reject all the $J_{Pk,i}$ such that $\neg\exists(\mathbf{t}_{1j},\mathbf{t}_{2j})\in c_j : \mathbf{t}_{1j}\in J_{Pk,i} \wedge \mathbf{t}_{2j}\in J_{Pk,i}$ (false positives).
                5c.3) Fit a snake to each $J_{Pk,i}$, close the open contours whose endpoints belong to each junction region and separate points inside each closed region, thus obtaining parts $P_{Pk,j}$, $j$=1,…,n+1.
                5c.4) Update the set of parts $P=(P-P_k)\cup\{P_{Pk,j}\}$.
        5d) For each $P_k \subset P$ apply steps 4 and 5 of Algorithm I and obtain $epr(\mathcal{R}(P_k))$.
```

3.3 Discussion

Several proposals have been made, during the last years, to face the problem of object segmentation and modeling. In what follows, a general discussion about their performance is reported. Pentland devised the first solution (Pentland 1989 and 1990) using a 2D contour-based blob segmentation of range data and modal dynamics to represent deformations, but this approach is heavy from the computational point of view and makes very restrictive assumptions on the true shape of data. Some other systems (Gupta and Bajcsy 1993, Ferrie Lagarde and Whaite 1993, H. Zha T. Hoshide T. Hasegawa 1998) have intersections with the one presented by the author, while the widely accepted solution proposed by Solina and his collegues (Leonardis Jalik and Solina 1997) uses a dual approach.

Solina claims that the segmentation effort in the modeling procedure is not useful because there are some particular cases where surface properties of the range data do not allow to separate the object into different parts. The method starts with some model seeds placed inside the range points envelop, and alternates a fitting and a model selection phase. Models with the highest goodness-of-fit are selected to grow and fit again their data plus some suitable neighborhood of points. Models not satisfying this requirement are discarded. Segmentation is achieved automatically, but with a great computational effort that is only partially counterbalanced by the small number of data points for each model. Moreover, reconstructions are often not minimal, depending on the initial placement of seeds.

The author claims that the MT strategy is good in allowing both efficiency, meaningfulness and a minimal reconstruction. The idea of a deforming 3D blob that eventually breaks into more parts to improve fitting is a very intuitive one, and goes in the same direction as the paradigm proposed by Marr (Marr 1982) regarding volumetric segmentation of data followed by structural description. Moreover, in the context of a general architecture for part-based recognition equipping a mobile robot, ambiguities can be resolved by changing the viewpoint.

Among the other segmentation-based approaches cited above, SUPERSEG (Gupta and Bajcsy 1993) seems to be the more similar one to the MT solution, due to the idea of a global fit of the object followed by residual analysis. SUPERSEG uses also a surface-based segmentation technique, and surface labeling supports decision making in the residual analysis phase to obtain 3D segments.

Snakes are used by (Ferrie Lagarde and Whaite 1993) but these are true 3D curves fitting concavity points that were previously detected using differential geometry considerations. This approach results very heavy in contrast to the MT algorithm where the range data are used as a 2D image, and classical filtering techniques are adopted to extract features. This choice is supported by a fast search for junction points that emerge automatically form the analysis of the points shrinking velocity.

Similar considerations apply to (H. Zha T. Hoshide T. Hasegawa 1998). They use directly residual analysis to select junction points, and this technique leads to false positives. Think, as an example, to the case of two tangent spheres with different radius, globally fitted by a tapered ellipsoid: points in the bottom part of the larger one exhibit large residuals. Finally, fitting plane interpolation implies minimization procedure with a greater computational load than the one required by snakes.

4. EXPERIMENTAL RESULTS

Several experiments have been performed to provide global validation of the proposed method. Data have been mainly obtained from the range images provided by Solina as test platform for his system, but also syntethic range data and shape-from-shading images have been used (see FIGURE 7). The adopted shape-from-shading algorithm was the one proposed in (Zhang Tsai Cryer and Shah 1994).

The whole system has been implemented in a prototypical form, using MATLAB 5.3 on a Pentium III 1000 MHz processor with 256 MB RAM, running under Windows Me. All the auxiliary procedures, like the Canny edge detector, the snake and the minimization algorithms, are taken from official MathWorks or user-contributed toolboxes. The author is currently working to the definition of a MT toolbox along with a set of executable programs for Windows platform.

In all the experiments, the size of the median filter has been set to 7x7, and the threshold to select points in a junction region has been set to the 80% of the v'_s peak.

FIGURE 7 NEAR HERE

Two experiments have been performed to check theoretical assertions about the algorithm. The first experiment is related to the choice of the best expression for *epr* in order to obtain model estimation with the maximum accuracy in case of viewpoint ambiguity. Several expressions exist in literature. The "minimal volume" metric epr_S proposed by Solina (equation (5)) prefers the smallest model fitting to data as the most perceptually acceptable. Gross and Boult (Gross and Boult 1988) propose a metric epr_{GB} defined as the mean distance between each data point $\mathbf{x}(\eta,\omega)$ and the corresponding superquadric point $\mathbf{x}_s(\eta,\omega)$ measured along the segment connecting $\mathbf{x}(\eta,\omega)$ with the superquadric center:

$$epr_{GB} = \frac{1}{N}\sum_{i=1}^{N}\left|\mathbf{x}_i - \mathbf{x}_{s,i}\right|$$

Finally, Whaite and Ferrie (Whaite and Ferrie 1991) propose a variation of the previous metric, using the inside-outside superquadric function:

$$epr_{WF} = \frac{1}{N}\sum_{i=1}^{N}\left|\mathbf{x}_i\right|\left|1 - f^{-e_1/2}\left(\mathbf{x}_i, \ddot{\mathbf{e}}\right)\right|$$

In the experiment (see FIGURE 8) a deformed model under varying viewpoints was fitted using 10 moving target steps and the Levenberg-Marquardt minimization algorithm, trying all the three metric values. Results confirmed the better perceptual acceptability of epr_S, while epr_{GB} and epr_{WF} provide better numerical results. In practice, the minimal volume metric is the better choice to obtain both a minimal and meaningful reconstruction

FIGURE 8 NEAR HERE

The second experiments regards the choice of the correct number of shrinking steps in the moving target algorithm when it is used to fit a single segment. In this experiment, three different data sets have been reconstructed with varying shrinking steps, using both Levenberg-Marquardt and the PCG-based algorithm (see FIGURE 9). The experiment shows that when the number of steps becomes too large, there is a sort of saturation because the difference between two models in two subsequent steps is negligible, while the number of iterations per shrinking step decreases significantly. FIGURE 9 shows that there is no significant difference between final models while increasing the number of shrinking steps beyond 10.

FIGURE 9 NEAR HERE

A suitable numerical index of the computational load has been derived to measure performance with varying shrinking steps. The index is computed as the product of the shrinking steps N_t by the average number of minimization iterations performed at each step:

$$I = N_t \cdot \overline{It},\ \overline{It} = \frac{1}{N_t}\sum_{i=1}^{N_t} It_i$$

This index decreases till a minimum, and then tends to become asymptotic with I=N_t (see FIGURE 10). Even in this case, the plot of I vs. N_t attains the minimum for N_t=10.

FIGURE 10 NEAR HERE

5. CONCLUSIONS

A complete implementation of the recognition by parts theory applied to range data segmentation and modeling with deformable superquadric has been presented. The architecture has been motivated both from the psychological and computational point of view, and comparisons have been made with the most cited approaches.

The experimental results prove that the system performs well in a wide range of situations, but some critical situations cannot be disambiguated without some a priori knowledge about the structure of the perceived object. The requirements for a minimal and meaningful reconstruction are a suitable way to code the structural knowledge about the objects in the operating domain. As stated above, they act respectively as syntactic and semantic constraints on the symbolic description of the model. In turn, a rigorous symbolic description can be used to perform disambiguation.

The author and his collegues are working to the definition of a comprehensive robotic architecture for manipulators acting in partially known environments, where the robot has to discover both novel objects and the actual size of the operating space. Structural knowledge can be used, in this case, to guide the actions of the robot on the basis of the expectations generated at the symbolic level by the current perception. This feedback loop ends when perception is stable.

It is possible to draw a scenario in which the robot tries a model-driven fitting step in order to refine the reconstruction. An interesting proposal comes from (L.H. Chen Y.T. Liu and H.Y. Liao 1997) where a volumetric similarity measure for superquadrics is presented. The measure is expressed as the volume difference between the two superquadrics that is the total volume of those regions that belong either to the interior of one object or to that of the other, but not to those of both. In the cited work, similarity is computed for superquadrics in canonical pose. Suppose to store, along with the symbolic description of the model, the geometric description of all the superquadrics representing its parts expressed in a normalized space, where at least one component is in the canonical pose. This normalized model can be regarded as an implementation of the meaningfulness constraint. The volume metric could be used to perform matching between the actual model and the stored one, while the constraint of minimal reconstruction is used to prune the actual reconstruction from exceeding parts, discarding those with the poorest matching degree. Obviously, translation, rotation and scale of the perceived model is required to perform matching, so the entire procedure can be regarded as a sort of 3D alignment.

Matching with a normalized model of the object and interactive viewpoint change can also used in the case of novel object discovery, when the robot has first to create a new model. In this case, the robot could learn by some different examples, refining the stored model in an adaptive way, till it works for an as large as possible class of similar objects.

Another future extension of the system is the creation of a general framework where a minimal reconstruction is obtained taking into account the a priori probability distribution of different segmentation outcomes. In this scenario, global perceptual features will be also used to segment data instead of searching only for concavities. The obtained probabilities, in turn, will affect the superquadric fitting scheme.

As a final remark, it is to be noted that the presented system is inspired by some psychological considerations, as stated before, but it is essentially grounded on computer vision techniques. The creation of an experimental set-up to perform comparisons between human and artificial abilities in providing the structural description of an object in terms of its composing parts is no doubt an interesting activity. In particular, it will be of interest to verify if the artificial system is able to discover parts in the object that are similar or are the same of those selected by a human being. From the segmentation ability depend the model selection the symbolic description of the object. In the case of similar segmentation outcomes, the human and the robot will be able to share the same symbolic description of the object, thus allowing their communication interface to be more flexible and efficient.

REFERENCES

Ardizzone, E., Chella A., Frizione, M. and Gaglio, S. 1992. Integrating Subsymbolic and Symbolic Processing in Artificial Vision, *Journal of Intelligent Systems*, **1**(4), 273-308.

Arkin, R.C. 1998. Behavior-Based Robotics, Cambridge, MA, MIT Press.

Barr, A.H. 1981. Superquadrics and angle-preserving transformations, *IEEE Computer Graphics Applications*, **1**(1), 11-23.

Bardinet, E., Cohen, L.D. and Ayache, N. 1998. A parametric deformable model to fit unstructured 3D data, *Computer Vision and Image Understanding*, **71**(1), 39-54.

Biederman, I. 1987. Recognition-by-components: A theory of human image understanding. *Psychological Review*, **94**(2), 115-147.

Biederman, I. and Gerhardstein, P.C. 1993. Recognizing depth-rotated objects: Evidence and conditions for 3-dimensional viewpoint invariance, *Journal of Experimental Psychology: Human Perception and Performance*, **19**, 1162-1182.

Biederman, I. and Gerhardstein, P.C. 1995. Viewpoint-dependent mechanisms in visual object recognition: A critical analysis, *Journal of Experimental Psychology: Human Perception and Performance*, **21**, 1506-1514.

Blake, A. and Isard, M. 1998. *Active Contour*s, Berlin, Springer-Verlag.

Bülthoff, H.H., Edelman, S.Y. and Tarr, M.J. 1995. How are three-dimensional objects represented in the brain?, *Cerebral Cortex*, **3**, 247-260.

Canny, J. 1986. A Computational Approach to Edge Detection, *IEEE Trans. Patt. Anal. Mach. Intell.*, **8**(6), 679-698.

Chella, A, Frixione, M. and Gaglio, S. 1997. A cognitive architecture for artificial vision, *Artificial Intelligence,* **89**, 73-111.

Chella, A, Frixione, M. and Gaglio, S. 2000. Visual Knowledge Representation of Moving Scenes, *Journal of Intelligent Systems*, **10**(4), 377-404.

Chella, A., Gaglio, S. and Pirrone, R. 2001. Conceptual representations of actions for autonomous robots, *Robotics and Autonomus Systems*, **34**, 251-263.

L.H. Chen, Y.T. Liu and H.Y. Liao 1997. Similarity measure for superquadrics, *IEE Proc. – Vision Image Signal Processing*, **144**(4), 237-243.

Coleman, T.F. and Y. Li 1994. On the Convergence of Reflective Newton Methods for Large-Scale Nonlinear Minimization Subject to Bounds, *Mathematical Programming*, **67**(2), 189-224.

Coleman, T.F. and Y. Li 1996. An Interior, Trust Region Approach for Nonlinear Minimization Subject to Bounds, *SIAM Journal on Optimizatio*n, **6**, 418-445.

Cotronei, M. and Salvato, G. 1996. A Non-Iterative Method for the Estimation of Superquadric Parameters from Depth Maps, *Journal of Intelligent Systems*, **6**(2), 115-132.

Darrell, T., Sclaroff, S. and Pentland A.P. 1990. Segmentation by minimal description, *Proc. of 3rd International Conference on Computer Vision*, Osaka, Japan, 112-116.

Dickinson, S.J., Pentland, A.P. and Rosenfeld, A. 1992. 3-D Shape Recovery Using Distributed Aspect Matching, *IEEE Trans. Patt. Anal. Mach. Intell.*, **14**(2), 174-198.

Edelman, S. and Poggio, T. 1991. Bringing the grandmother back into the picture: A memory-based view of object recognition, *MIT A.I. Memo No. 1181*.

Edelman, S. and Weinshall, D. 1991. A self-organizing multiple-view representation of 3-D objects, *Biological Cybernetics*, **64**, 209-219.

Edelman, S. 1998. Representation is representation of similarities. *Behavioral & Brain Sciences*, **21**, 449-498.

Ferrie, F.P., Lagarde, J. and Whaite, P. 1993. Darboux Frames, Snakes, and Super-Quadrics: Geometry From the Bottom Up, *IEEE Trans. Patt. Anal. Mach. Intell.*, **15**(8), 771-784.

Gross, A.D. and Boult, T.E. 1988. Error of Fit Measures for Recovering Parametric Solids, *Proc. of Second ICCV,* Tampa, FL, 690-694.

Gupta, A. and Bajcsy, R. 1993. Volumetric Segmentation of Range Images of 3-D Objects Using Superquadric Models, *Computer Vision Graphics and Image Processing – Image Understanding*, **58**(3), 302-326.

Hummel, J.E. and Biederman I. 1992. Dynamic binding in a neural network for shape recognition, *Psychological Review*, **99**, 480-517.

Hummel, J.E. and Stankiewicz, B.J. 1996. An architecture for rapid, hierarchical structural description, In T. Inui and J. McClellland (Eds.) *Attention Performance XVI: Information Integration in Perception and Communication*, Cambridge, MA, MIT Press, 93-121.

Hummel, J.E. 2000. Where view-based theories break down: The role of structure in shape perception and object recognition, In E. Dietrich and A. Markman (Eds.) *Cognitive Dynamics: Conceptual Change in Humans and Machines*, Hillsdale, NJ, Erlbaum, 157-185.

Hutchinson, S., Hager, G.D. and Corke, P.I. 1996, A Tutorial on Visual Servo Control, *IEEE Robotics and Automation*, **12**(5), 651-670.

Kumar, S., S. Han, Goldgof D. and Bowyer K. 1995. On Recovering Hyperquadrics from range Data, *IEEE Trans. Patt. Anal. Mach. Intell.*, **17**(11), 1079-1083.

Kumar, S. and Goldgof D. 1995. Model Based part Segmentation of Range Data – Hyperquadrics and Dividing Planes, *Proc. of IEEE Workshop on Physics-Based Modeling in Computer Vision*, 17-23.

Leonardis, A., Jaklic, A. and Solina, F. 1997. Superquadrics for Segmenting and Modeling Range Data, *IEEE Trans. Patt. Anal. Mach. Intell.*, **19**(11), 1289-1295.

Marr, D. and Nishihara, H.K. 1978. Representation and recognition of three dimensional shapes, *Proc. of the Royal Society of London B*, **200**, 269-294.

Marr, D. 1982. *Vision*, W.H. Freeman & Co.

Moré, J.J. 1977. The Levenberg–Marquardt Algorithm: Implementation and Theory, In G.A.Watson (ed.) *Numerical Analysi*s *- Lecture Notes in Mathematics*, **630**, Berlin, Springer Verlag, 105-116.

Pentland, A.P. 1986. Perceptual organization and the representation of natural forms, Artif. Intell., **28**, 293-331.

Pentland, A.P. 1987. Recognition by Parts, *Proc. of First ICCV*, London, 612-620.

Pentland, A.P. 1989. Part Segmentation for Object Recognition, *Neural Computation*, **1**, 82-91.

Pentland, A.P. 1990. Automatic Extraction of Deformable Part Models, *International Journal of Computer Vision*, **4**, 107-126.

Solina, F. 1987. *Shape Recovery and Segmentation with Deformable Part Models*, PhD Thesis, University of Pennsylvania, Tech. Rep. MS-CIS-87-111.

Solina, F. and Bajcsy, R. 1990. Recovery of parametric models from range images: The case for superquadrics with global deformations, *IEEE Trans. Patt. Anal. Mach. Intell.*, **12**(2), 131-147.

Solina, F. and Leonardis, A. 1998. Proper Scale for modeling visual data, *Image and Vision Computing*, **16**(2), 89-98.

Tarr, M.J. 1995. Rotating objects to recognize them: A case study on the role of the viewpoint dependency in the recognition of three-dimensional objects, *Psychonomic Bullettin & Review*, **2**(1), 55-82.

Tarr, M.J. and Bülthoff, H.H. 1995. Is human object recognition better described by geon structural descriptions or by multiple view? Comment on Biederman and Gerhardstein (1993), *Journal of Experimental Psychology: Human Perception and Performance*, **21**, 1494-1505.

Terzopoulos, D., Metaxas, D. 1991. Dynamic 3D Models With Local and Global Deformations: Deformable Superquadrics, *IEEE Trans. Patt. Anal. Mach. Intell.*, **13**(7), 703-714.

Ullman, S. 1989. Aligning pictorial description: An approach to object recognition, *Cognition*, **32**, 193-254.

Ullman, S. and Basri, R. 1991. Recognition by linear combinations of models. *IEEE Trans. Patt. Anal. Mach. Intell.*, **13**, 992-1006.

Ullman, S. 1996. *High-level Vision: Object Recognition and Visual Cognition*, Cambridge, MA, MIT Press.

Whaite, P. and Ferrie, F.P. 1991. From Uncertainty to Visual Exploration, Pentland, A.P. and Horowitz, B. 1991. Recovery of nonrigid motion and structure, *IEEE Trans. Patt. Anal. Mach. Intell.*, **13**(10), 1038-1049.

H. Zha, T. Hoshide and T. Hasegawa 1998. A Recursive Fitting-and-Splitting Algorithm for 3-D Object Modeling Using Superquadrics, *Proc. of ICPR*, 658-662.

Zhang, R., Tsai, P.-S., Cryer, J.E., Shah, M. 1994. Analysis of shape from shading techniques, *Proc. of CVPR'94,* Seattle, Whashington, USA, 377-384.

L. Zhou and Kambhamettu C. 1999. Extending Superquadrics with Exponent Functions: Modeling and Reconstruction, *Proc. of CVPR*, 73-78.

APPENDIX A: THEORETICAL REMARKS ON SNAKES

A snake is a deformable curve that moves in the image under the influence of forces related to the local distribution of the gray levels. In this case gray levels have the meaning of depth values. When the snake reaches an object contour, it is adapted to its shape. Formally, a snake as an open or closed contour is described in a parametric form $v(s)=(x(s), y(s))$ where $x(s)$ and $y(s)$ are the coordinates along the shape contour and s is the normalized arc length ranging in the [0 ,1] interval.
The snake model defines the snake energy of a contour as:

$$E_{snake}(v(s)) = \int_0^1 \left(E_{int}(v(s)) + E_{image}(v(s))\right) ds \,.$$

The energy integral is a functional since its variable s is a function (the shape contour). The internal energy E_{int} is formed from a Tikhonov stabilizer and is defined by:

$$E_{int}(v(s)) = a(s)\left|\frac{dv(s)}{ds}\right|^2 + b(s)\left|\frac{d^2v(s)}{ds^2}\right|^2 \,.$$

The first order continuity term, weighted by $a(s)$, let the contours behave elastically, whilst the second order curvature term, weighted by $b(s)$, let it be resistant to bending. For example, setting $b(s)=0$ at point s, allows the snake to become second-order discontinuous at point and to generate a corner. The image functional determines the features which will have a low image energy and hence the features that attract the contours. In general, this functional is made up by three terms:

$$E_{image} = w_l E_l + w_e E_e + w_t E_t$$

where ù denote a weighting constant. The three terms account for lines, edges and end points energy respectively. In the current implementation, the snake is initialized as an open curve following the junction region contour, while the only term used in E_{image} is the edge functional that has been implemented using a suitably defined operator searching for the junction region minima. When there are no minima, the operator searches for those points with a significant difference between left and right local gradient. This choice is justified by the shape of a generic junction region that is either a valley or is made by two rather smooth pieces of surface $z_1(x, y)$ and $z_2(x, y)$ intersecting along a line whose points (x_l, y_l) are such that Ñ$z_1(x_l, y_l)$ ≠Ñ$z_2(x_l, y_l)$.

FIGURE CAPTIONS

Fig.1. From left to right. An example of good (minimal and meaningful) reconstruction; a non-minimal reconstruction; a non-meaningful reconstruction.

Fig. 2. Some superquadrics encompassing different shapes. From top to bottom and from left to right, the two form factors take the values: 0.1, 0.5, 1.0, 1.5, 2.0.

Fig. 3. A 2D example of the geometry of the data set projection onto the smallest enclosing sphere.

Fig. 4. An example of deformation of the data set and of the error surface for a synthetic range image of a cylinder: t=0, 5, 10, 20. For simplicity only the two form factors have been let free to change.

Fig. 5. Three examples of reconstruction using both minimization algorithms along with the value of *epr*. From left to right: the original model, reconstruction with Levenberg-Marquardt and the PCG based method.

Fig. 6. The MT approach to segmentation. From left to right: the input range image, junction region, junction points fitted by the snake, model fitting of the two parts of the object.

Fig. 7. Some examples of the whole procedure using different data sources (true range data, and shape-from-shading image). For each row: input data, contours and snakes (in gray) to perform segmentation, final reconstruction..

Fig. 8. Recovery of the model for the same data set under three different viewpoints (τ,σ) varying the metric for *epr*. For each row: the model recovered using the metric by Solina, Gross and Boult, Whaite and Ferrie. The parameters of the original model (depicted in the first row) are $\mathbf{l}_{\mathrm{int}}$=[0.8, 1.5, 3, 4, 5].

Fig. 9. Recovery of the model for two different data sets, varying the number of shrinking steps. For each row the original model and those recovered after 1, 2, 5, 10, 20, 50, 100 shrinking steps are reported.

Fig. 10. Plot of the index I vs. the number of shrinking steps N_t for the first model depicted in figure 9. The minimum is attained approximately for N_t=10.

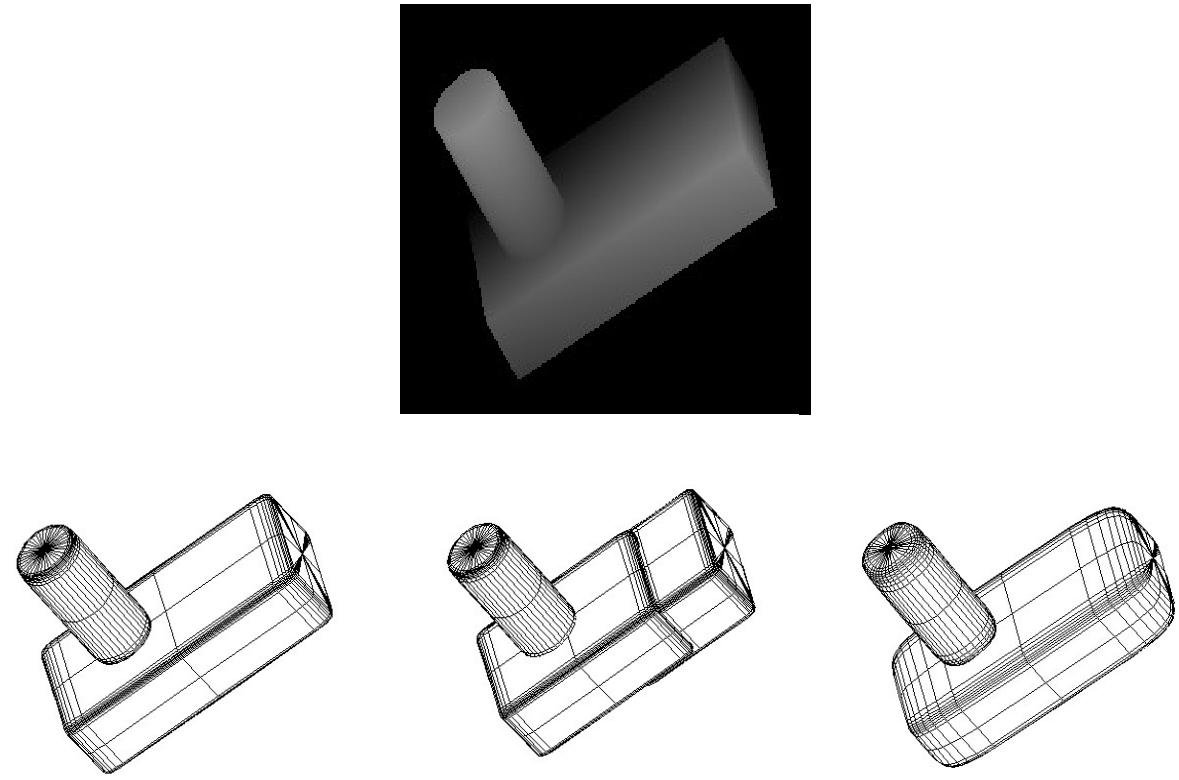

FIGURE 1

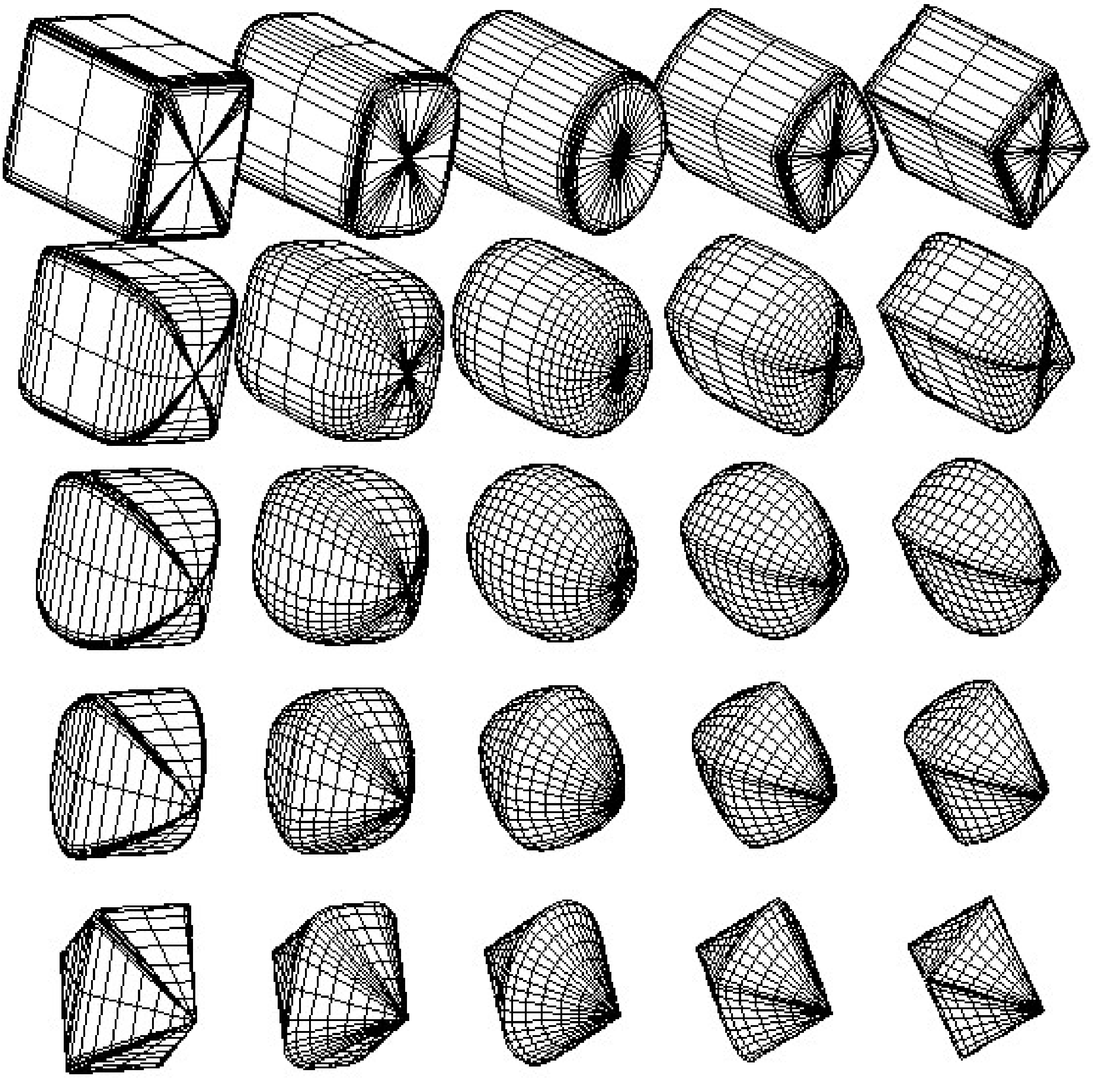

FIGURE 2

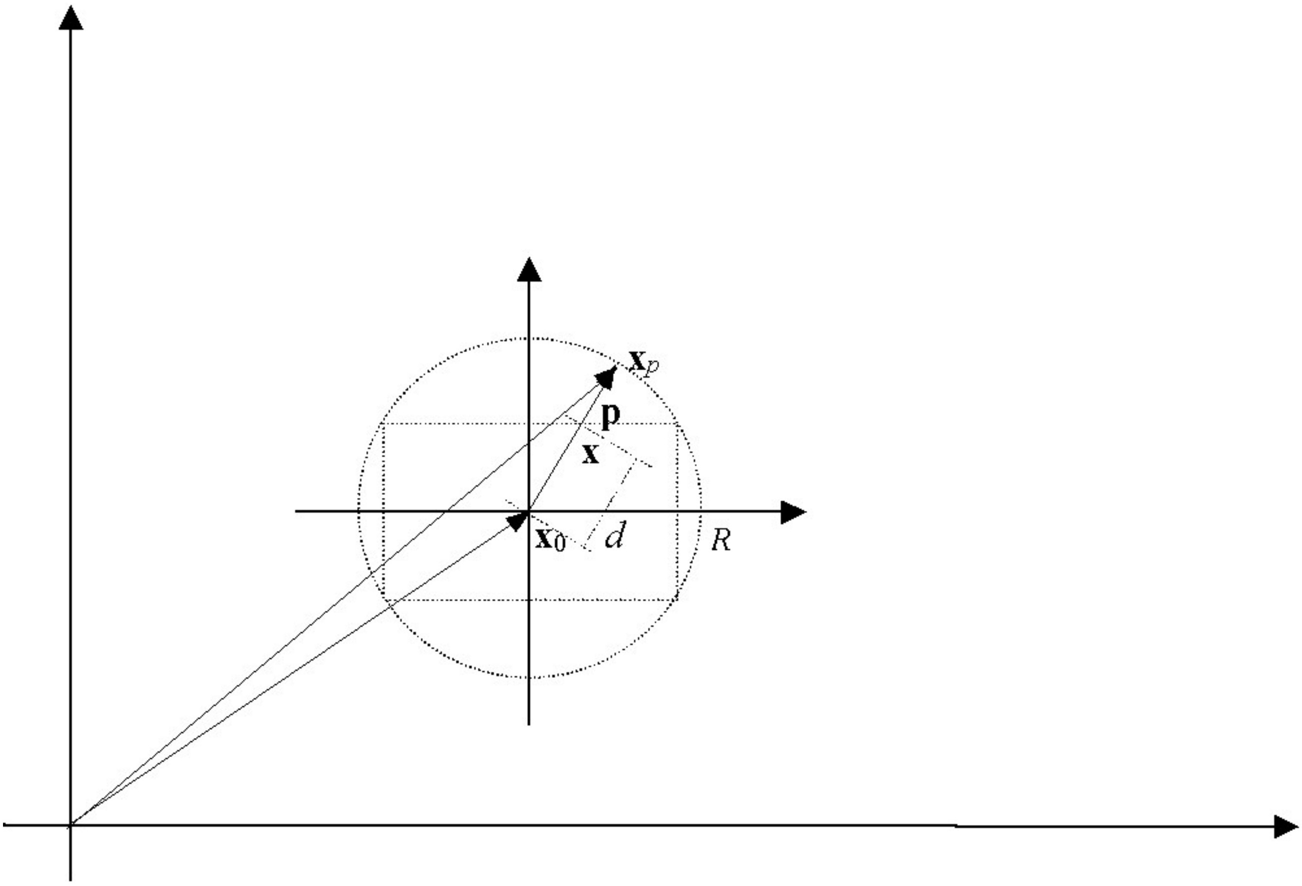

FIGURE 3

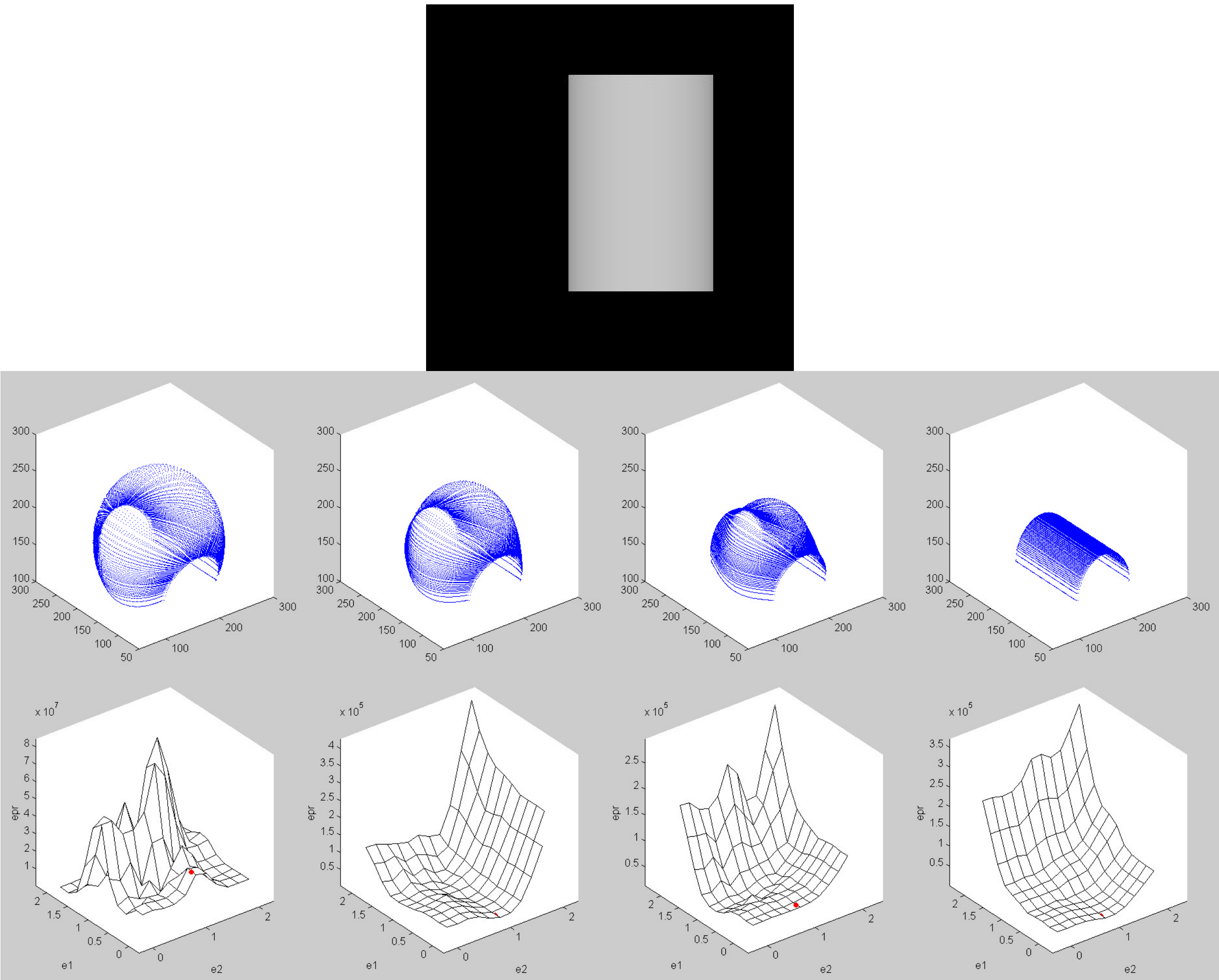
300
250
200
150
100
300
250
200
150
100
50
100
200
300
x 10^7
8
7
6
5
4
3
2
1
epr
2
1.5
1
0.5
0
e1
0
1
2
e2
x 10^5
4
3.5
3
2.5
2
1.5
1
0.5
x 10^5
2.5
2
1.5
1
0.5
x 10^5
3.5
3
2.5
2
1.5
1
0.5

FIGURE 4

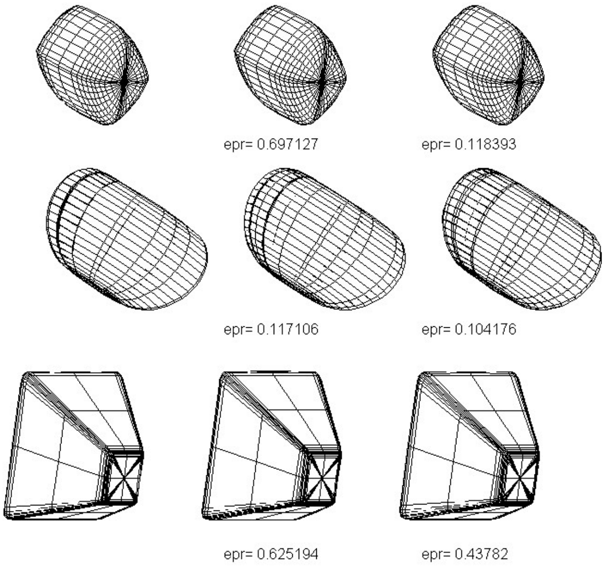
epr= 0.697127
epr= 0.118393
epr= 0.117106
epr= 0.104176
epr= 0.625194
epr= 0.43782

FIGURE 5

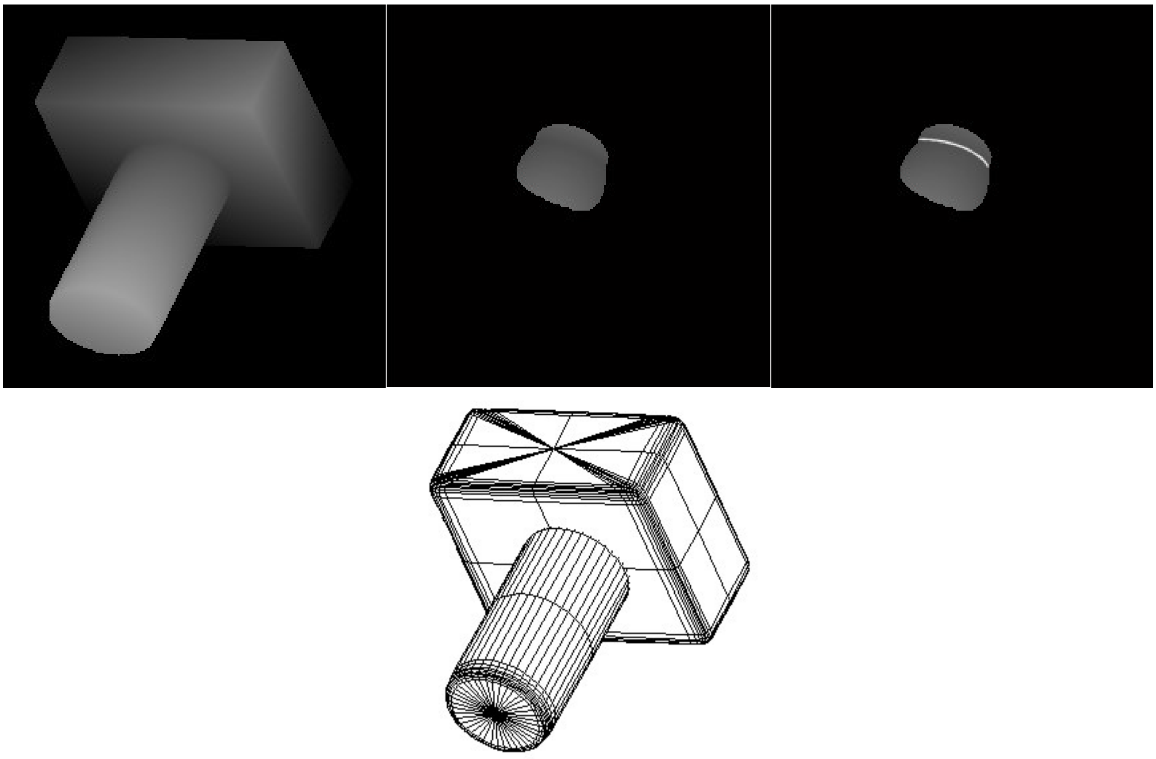

FIGURE 6

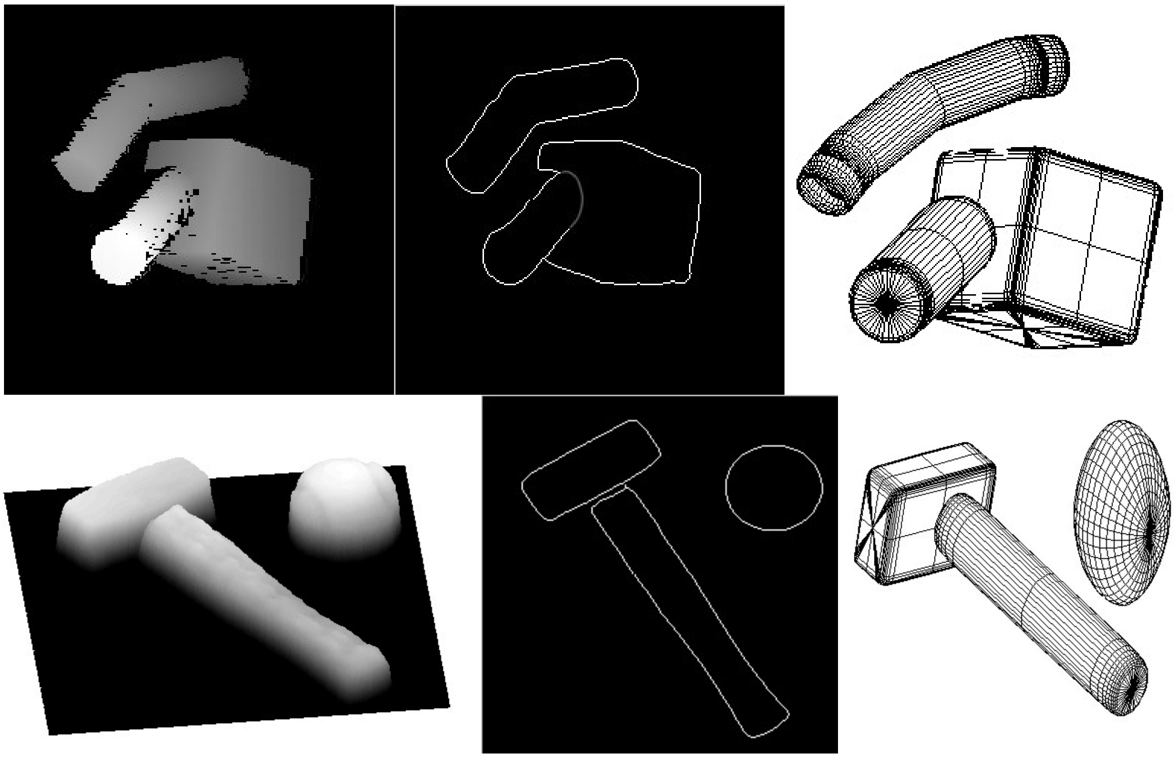

FIGURE 7

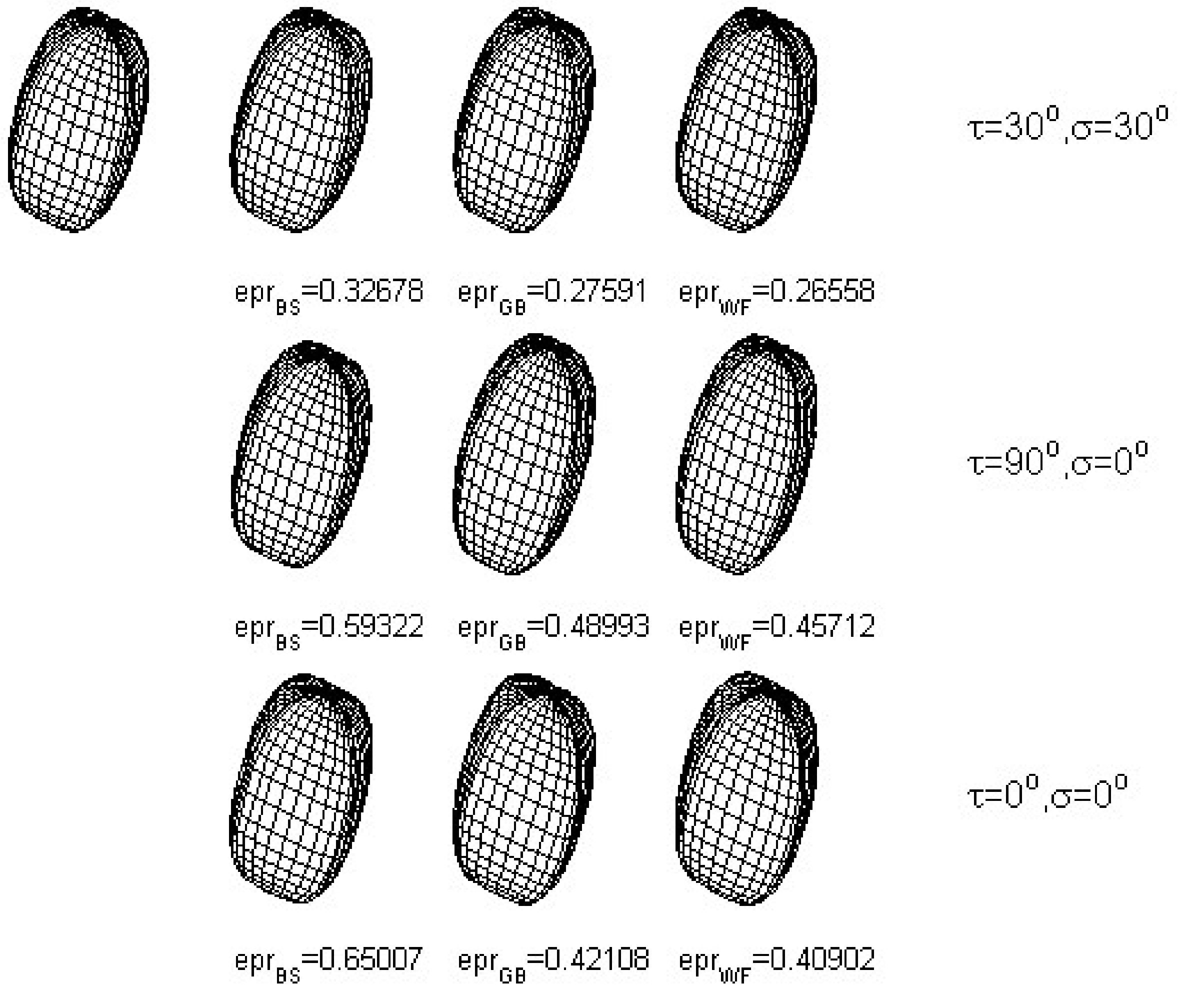
$\tau=30^o,\sigma=30^o$
$\mathrm{epr}_{BS}=0.32678$
$\mathrm{epr}_{GB}=0.27591$
$\mathrm{epr}_{WF}=0.26558$
$\tau=90^o,\sigma=0^o$
$\mathrm{epr}_{BS}=0.59322$
$\mathrm{epr}_{GB}=0.48993$
$\mathrm{epr}_{WF}=0.45712$
$\tau=0^o,\sigma=0^o$
$\mathrm{epr}_{BS}=0.65007$
$\mathrm{epr}_{GB}=0.42108$
$\mathrm{epr}_{WF}=0.40902$

FIGURE 8

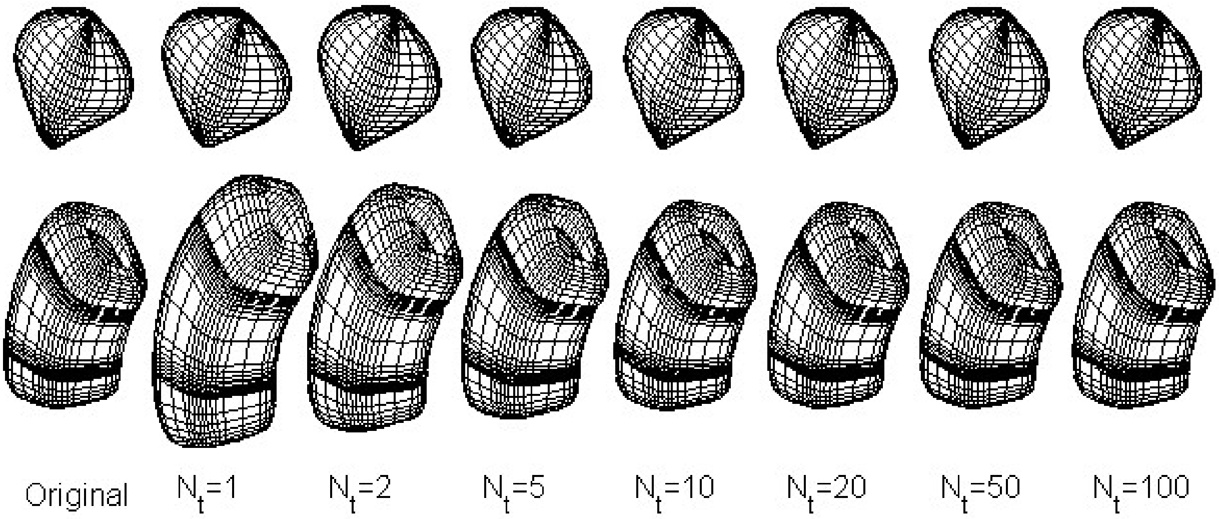

FIGURE 9

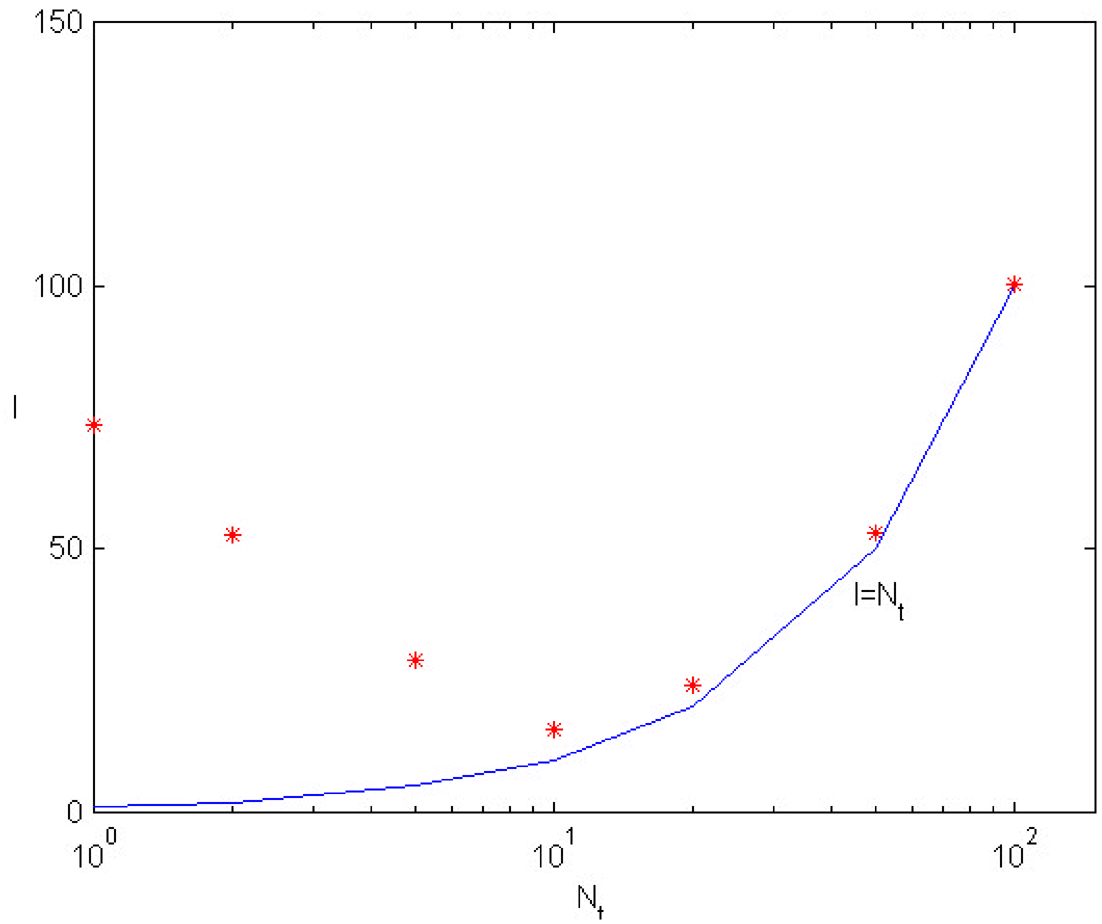

FIGURE 10