Recovery of Generic Solid Parts by Parametrically Deformable Aspects*

Maurizio Pilu **Robert B. Fisher**

Department of Artificial Intelligence
The University of Edinburgh
5 Forrest Hill, Edinburgh EH1 2QL
SCOTLAND
email: {maurizp,rbf}@aifh.ed.ac.uk

April 29, 1996

Abstract

This paper presents a novel approach to the recovery of generic solid parts of objects from real 2D images. The part vocabulary chosen is the one of geons, which are qualitative volumetric part primitives that are defined by simple but perceptually relevant properties which are viewpoint quasi-invariant. Most previous works on detection and recognition of geons from 2D images relied on quasi-perfect line drawings. The use of aspects has also been proposed for matching fixed templates of synthetic images. Here we use parametrically deformable aspects as 2D models to be matched to real images of geons in the framework of Model-Based Optimisation. The use of parametric models allows us to efficiently represent geons, whereas the use of topologically different aspects yields more robustness in the optimisation process we use, which is Adaptive Simulated Annealing. A simple control strategy is developed that generates initial aspect hypotheses followed by a maximum *a posteriori* choice of the best one. Experiments are shown that demonstrate the validity of the approach. The proposed method is general, in the sense that it could be easily applicable to other parametrically defined part vocabularies.

*This research paper is extracted for Chapter 6 of [41]. A shorter version of this paper which does not include the aspect-based fitting strategy has been presented at the Fourth European Conference on Computer Vision, Cambridge, April 1996 [42].

1 Introduction

Geons [3] are generic solid primitives defined by qualitative properties of axis and cross-section of generalised cylinders [4] that are invariant under change of view-point.

In this paper a new method is presented for fitting qualitative 3D volumetric parts models to real 2D images that treats geons[1] as *single entities* to be extracted from images. This is done by matching parametrically deformable contour models (PDCM) of geons to edge images in the framework of Model-Based Optimisation (MBO), in which an objective function expressing the global likelihood (goodness) of fit is maximised. The cost function accounts for both matched and unmatched contour portions and is formulated in Bayesian terms.

The potential advantages of such a global approach lie in imposing overall consistency on the image which lead to robustness to cluttering and opens possibilities of direct figure-ground segmentation in the spirit of [30] or the method presented in [41].

Similar approaches to generic part recognition that used deformable superquadrics as generic shape models have been investigated for the 3D case (range data input) in popular work such as [49] and also in [54], [30] and [5]; only in [35] the method was extended to the 2D case as a front-end of the OPTICA system [12].

To date, however, one of the main problems faced by global fitting approaches is their sensitivity to the initial state of the models, which often compromises the quality of the solution. In previous work [42], we used a loosely-constrained optimisation approach which worked well only when the initial model was topologically equivalent to the geon instance being fitted. Here, this deficiency is greatly reduced by using an aspect-based hypothesis generation-and-testing strategy inspired by [14]. The multidimensional parameter space defining the geon PDCM is partitioned into eight topology-equivalent classes that we call Parametrically Deformable Aspects (PDA); the set of eight PDA can be seen as a single deformable model endowed with global topology information. By doing so, the optimisation can independently focus in regions of the parameter space that correspond to models with the same topology, thereby reducing the chances of getting stuck in local minima caused by different interpretations of image features. A simple experimental control strategy suggested by [14] is employed that, by starting from coarse 2D part hypotheses produced as (but not only) in [41], does:

(1) initialises all eight PDA at a representative position for each PDA;

(2) performs the fitting independently for each PDA thus initialised;

(3) chooses the one that achieves the best score.

We will see taht the happy marriage between parametric deformable contour models and the concept of topologically different aspects efficiently represents geons and yields more robustness in the optimisation process we use, which is Simulated Annealing.

The results we achieved from 2D images are very much comparable with the one obtained by using 3D range data (e.g. by [49]), although depth and orientation cannot be obviously recovered from 2D images.

[1] The parts will be still called geons, although they are are a subset of the ones defined in [3]

2 Review of Previous Related Work

In this section, some previous research in model-based optimisation and the use of aspects in recognition are reviewed.

2.1 Model-Based Optimisation

In the context of Computer Vision, Model-Based Optimisation (MBO) aims at finding the best fit of a model by minimising an objective function (or maximising a likelihood function) that can incorporate both high and low level knowledge about the image, object model and goodness of fit. Within this framework, the use of whole boundary models – such as the one used here – is the most natural and effective because [50]: i) the whole structure is imposed to the problem and the task is simplified; ii) gaps are naturally filled and iii) overall consistency is more likely to result.

MBO can be performed in parameter space or in image space and with arbitrary models, fixed templates or deformable models.

Optimisation in image space is done through fitting each composing element (point, lines, etc) of the model more or less separately to the image. Typical models that have been used within this paradigm are fixed templates [14], arbitrary models like snakes [24], bead chains [10], Markov boundaries [17] and parametric shapes like Point Distribution Models (PDM) [8, 41]. As we shall see later, this method allows the model to better track object irregularities but, besides problems of stability, it is often difficult to incorporate high-level knowledge about the overall object shape to guide the fitting process. In most works using this type of models, the high-level knowledge is inspired by physical analogies (such as the smoothness constraint [24]) but very promising results have been achieved by using PDM [8] or finite element models [39], where global information is encoded in the modes of variation.

On the other hand, MBO in parameter space is performed by adaptively changing the parameters of the model and checking the goodness of fit in the image; it implies the use of parametric (deformable) models whose shape variability can be expressed in a compact form by few significant parameters; within this paradigm fit works by Lowe [34], Yuille [55], Staib and Duncan [50] and a wealth of others. Fixed templates have also been used but that is a sort of degenerate case in which the only controlling parameters are the one defining the pose of the object. Although the use of parametric models offers the advantage of compactness of representation and easy classification, often the optimisation in parameter space turns out to be a hard problem (see, e.g., [34]), also because the parameter space is often not as "tight" as for arbitrary models.

As far as the optimisation algorithm goes, that is, the "tactic" for finding the best fit in terms of configuration or parameter values, several methods have been proposed and experimented with but none have provided a reliable and sufficiently general method. Cost functions are often strongly non-linear and present many possibly narrow and shallow local minima that make fast convergence even to a sub-optimal minimum hard. The initial condition, that is, the values of the model parameters before the optimisation starts, often play a crucial role and, where this is not done manually, several heuristic have been timidly proposed (as in [34] and [17]). Commonly used methods include Levenberg-Marquandt (used for instance in [34] and [5]), Simulated Annealing (i.e. in [54]) and hill-climbing in combination with continuation or multi-scale techniques [50].

2.2 Use of Aspects

The concept of *aspects* was first formulated in [27] and a new object representation, called the *aspect graph*, was proposed. An aspect graph is essentially a *"...complete enumeration of topologically distinct views of an object, along with a definition of the region (*cell*) of viewpoint space from which such a view is seen"* [14].

A number of algorithms have been proposed to compute the aspect graphs of polyhedral [51], algebraic surfaces [43] suggestive models [16] or solid of revolution [13], often by approximating the exact solution by tessellating the Gaussian view-sphere. However, the practical use of aspect graphs for recognition has been hindered by the lack of practical implementations and therefore they have been mainly used for feature prediction, that in for checking how a feature combines with others. Relevant works that used such an aspect graph-based recognition strategy are, for instance, [12], [6] and [22].

A major conceptual extension of the use of aspect graphs has been proposed in [14] where the distinct-topology property of aspects is used to constrain an iterative fitting method within a single view-cell, thereby dramatically improving convergence quality and speed.

In most aspect-based works, including [14], CAD models were used because of the difficulty of constructing aspect graphs for general smoothed objects. One of the major contribution of this paper is to show that the use of an aspect based strategy is very beneficial also for the fitting of generic deformable models, such as superquadrics, in which a "topology-blind" strategy often yields poor results.

3 Parametrically Deformable Contour Models of Geons

Geons are volumetric shapes that are defined by qualitative features and are hence subject to high intra-class variability. Within our framework of Model-Based Optimisation, the recognition of geons from 2-D images needs to have a model that can describe in a compact way their appearance in 2D images and, being geon models computed inside the innermost loop of the optimisation process, this must be done as speedily as possible.

Recent works that dealt with the recognition of geons from 3-D range data (e.g. [49], [5], [54], [44]) have associated geons to globally deformable superquadric model [1]. There are mainly two advantages in using superquadric models. Firstly, the distinguishing features characterising geons can be expressed by single parameters such as bending, roundness, swelling and tapering and, secondly, they can represent very compactly a variety of shapes [38].

Here, the use of superquadrics in extended as done in [35] to the 2D case by approximating the contour of the image projection of geons (as opposed to their spatial occupancy) by the apparent contour (outline plus interior edges) of globally deformable superquadrics once they have been properly deformed, roto-translated and projected onto the image [49].

Unfortunately, computing the apparent contour of deformed superquadrics and in general for smooth surfaces is not a trivial job. As classic works in aspect computation show [13, 40], if an exact closed-form solution is sought, huge systems of equations need to be solved and time-expensive search in high-dimensional hyper-spaces has to be carried out. For these reasons, we did not endeavour along this avenue, which has been followed by the (however excellent) work by [35], where the superquadric contour was found by numerical methods at a high computational expense.

A few words must be said about this use of deformable superquadrics. Although they are a good model for representing 3-D shapes, they are extremely clumsy mathematical toys. Their deformations [49] have more of an engineering hack and is their error of fit function has no closed form. If used to compute contours as done in [35], superquadrics' clumsiness is certainly too much a burden for the compactness of representation they can give in exchange.

For our purposes, there is no need to have a precise knowledge of the projected deformable superquadrics contour for the following reasons:

- Contour details such as cusps cannot be reliably detected in real images and if recovered, they would be useful only for structural analysis of the contour such as in [2], which have been proved inapplicable in real cases;

- Very few, if any, actual geons can be properly described by globally deformable superquadrics: they are an arbitrary approximation in the first place, and a *different* approximation does no harm!

- If precise implies expensive, the above reason is even stronger

Therefore, a new, more straightforward approach has been followed, that is, to build a Parametrically Deformable Contour Model (henceforth PDCM) that simulates the deformable superquadric contour – thence without going through the explicit construction the superquadric – in a much more efficient, though approximate way; this constitutes a significant efficiency improvement to the model building method used by [35].

The geon PDCM has been designed following the pragmatic spirit of [55], [7], or [15], where models are designed with recognition in mind, rather than being *inherited* from computer graphics or the mathematics literature, as in the case of superquadrics. For instance, in [15] a parametric 3D wire-frame model of a car was purposely built that was able to represent the essential shape of several vehicles classes through its parameters; the 2D projection was trivially obtained from the 3D model and the fitting was performed using the technique presented in [9] and also used in this work. The approach is rather pragmatic but, if "theoretical" support is sought, it fits in the philosophy of [53], which advocated that *vision has to be driven by structure.*

The model, that is going to be described in the following, is suitable for qualitative geon PDCM and simulates the contour of projected deformable superquadrics in a very efficient way: starting from a cylinder centred on the z axis with superelliptical cross-section (Fig. 1-left), we apply deformations and rotations and find the contour by trivial geometric considerations. In the following the construction of the model is detailed.

The initial superelliptical cylinder $\mathbf{S}$ of height $2 \cdot a_z$ and semi-axes a_x and a_y can be expressed as

$$\mathbf{S} = \begin{bmatrix} \mathbf{x}(\eta) \\ \mathbf{y}(\eta) \\ \mathbf{z} \end{bmatrix} = \begin{bmatrix} a_x cos(\eta)^\epsilon \\ a_y sin(\eta)^\epsilon \\ \mathbf{z} \end{bmatrix} \qquad \begin{matrix} -\pi \leq \eta \leq \pi \\ -a_z \leq \mathbf{z} \leq a_z \end{matrix} \tag{1}$$

where $0 \leq \epsilon \leq 1$ controls the degree of squareness of the cross-section from a rectangle for $\epsilon \to 0$ to an ellipse for $\epsilon \to 1$.

Any curve lying on this cylinder can be variously deformed but for our purpose of representing geons we are particularly interested in three kinds of deformations: *tapering*, *bending* and *swelling* along the principal axis. Below the mathematical definition of these deformations is given. The tapering and bending deformations have been derived from [49] but the latter has been slightly modified by normalising the bending control parameter to a_z and allowing bending on both sides which has also improved the stability of its estimation. The swelling deformation, however, has been introduced here to represent the "expanding and contracting" sweeping rule of geons [3].

Let us indicate by $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ and $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$ the vector of shape points before and after the deformations, respectively.
A *linear* tapering deformation along the z axis is given by

$$Taper(\mathbf{S},\ K_x,\ K_y) = \begin{cases} \mathbf{X} = (\frac{K_x}{a_z}\mathbf{z} + 1)\mathbf{x} \\ \mathbf{Y} = (\frac{K_y}{a_z}\mathbf{z} + 1)\mathbf{y} \\ \mathbf{Z} = \mathbf{z} \end{cases}$$

where $-1 \leq K_x \leq 1$ and $-1 \leq K_y \leq 1$ express the amount of tapering in the x-z and x-z plane, respectively; henceforth we shall assume $K_y = K_x$.
A *circular* bending deformation in the y-z plane is obtained by (see [49] for details):

$$Bend(\mathbf{S}, c) = \begin{cases} \mathbf{X} = \mathbf{x} + sign(c)(R' - r) \\ \mathbf{Y} = \mathbf{y} \\ \mathbf{Z} = \sin(\gamma)(\kappa^{-1} - R') \end{cases} \qquad \text{with} \qquad \begin{cases} r = sign(c)\cos(\beta)\sqrt{\mathbf{x}^2 + \mathbf{y}^2} \\ \beta = \arctan\frac{\mathbf{y}}{\mathbf{x}} \\ R' = \kappa^{-1} - \cos(\gamma)(\kappa^{-1} - r) \\ \gamma = \mathbf{z}/\kappa^{-1} \\ \kappa^{-1} = \frac{a_z}{|c|}; \end{cases}$$

where $-1 \leq c \leq 1$ is the bending control parameter, which, when zero, yields no bending (for $c = 0$ the deformation is not applied).

Finally a *circular* swelling deformation along the z axis is given by:

$$Swell(\mathbf{S}, s) = \begin{cases} \mathbf{X} = \mathbf{x} + sign(\mathbf{x})(R'' \cos\alpha - (R'' - \sigma)) \\ \mathbf{Y} = \mathbf{y} + sign(\mathbf{y})(R'' \cos\alpha - (R'' - \sigma)) \\ \mathbf{Z} = R'' \sin\alpha \end{cases} \quad \text{with} \quad \begin{cases} \sigma = a_x s \\ R'' = (a_z^2 - \sigma^2)/(2\sigma) \\ \alpha = \arctan \frac{\mathbf{z}}{(\mathbf{R}' - \sigma)} \end{cases}$$

where s is the swelling control parameter (zero for no swelling).

Following the suggestion made by [49], the above deformations are applied in the following order: first tapering, then swelling and finally bending.

Once deformed, the shape is roto-translated to simulate the change in viewpoint by applying in sequence pan (about z) and tilt (about x) rotations, orthographic projection ($Proj$) and finally rotation about the optical axis y and translation in the image plane (by P_x and P_z). The whole chain of transformations of the initial 3D shape $\mathbf{S}$ to its full projection onto the image plane z-x $\mathbf{S}'$ is:

$$\mathbf{S}' = \begin{bmatrix} \mathbf{x}' \\ \mathbf{z}' \end{bmatrix} = Trasl(P_x, P_z, Rot_y(\theta_{opt}, Proj(Rot_x(\theta_{tilt}, Rot_z(\theta_{pan}, \\ Bend(c, Swell(s, Taper(K_x, K_x, \mathbf{S}))))))))) \tag{2}$$

Now we are ready to describe the construction of PDCM of geons. The knottiest problem is to determine the occluding contour. For doing this, the following approximation has been employed.

The transformation chain in Eqn. (2) is applied to the two bases of the superelliptical cylinder and take the four outermost points $P1'_a$, $P1'_b$ and $P2'_a$, $P2'_b$ (small circles in Fig. 1-right-B) and find the two corresponding points in the original undeformed superellipses (small circles in Fig. 1-right-A). These two pair of points are linked by two 3D straight lines L_1 and L_2, as shown in Fig. 1-right-A and are then deformed according to Eqn. (2) and the resulting $L1'$ and $L2'$ (Fig. 1-right-B) will then be used as the two sides of the occluding contour.

By checking the projection on the image plane to the normals $\mathbf{n}_a$ and $\mathbf{n}_b$ to the superelliptical ends, it is possible to determine whether each of the two ends are visible or not: if visible, the whole superellipse contour will be added to the geon PDCM; otherwise only its outermost part between $P1'_a$ ($P1'_b$) and $P2'_a$ ($P2'_b$) will be included in the final PDCM.

In the case the geon has square cross-section (small ϵ, say less than 0.5 in the superelliptical cross-section model) the central edge is determined by joining the two corners $P3_a$ and $P3_b$ (Fig. 1-right-C) from the undeformed superelliptical bases occurring at $\eta = \pi/4$ in Eqn. (1) by a 3D straight line and then deforming it by Eqn. (2); the resulting 2D curve is shown in Fig. 1-right-D.

The PDCM described above is controlled by 12 parameters, namely a_x, a_y, a_z, ϵ, K_x, s, c, θ_{pan}, θ_{tilt}, θ_{opt}, P_x, and P_z. All these controlling parameters immediately relate to those of a globally deformable superquadric, therefore they have a 3D meaning as we will see in the experiments, in particular in Sec. 7.2 and 7.3, where deformable superquadrics will be shown.

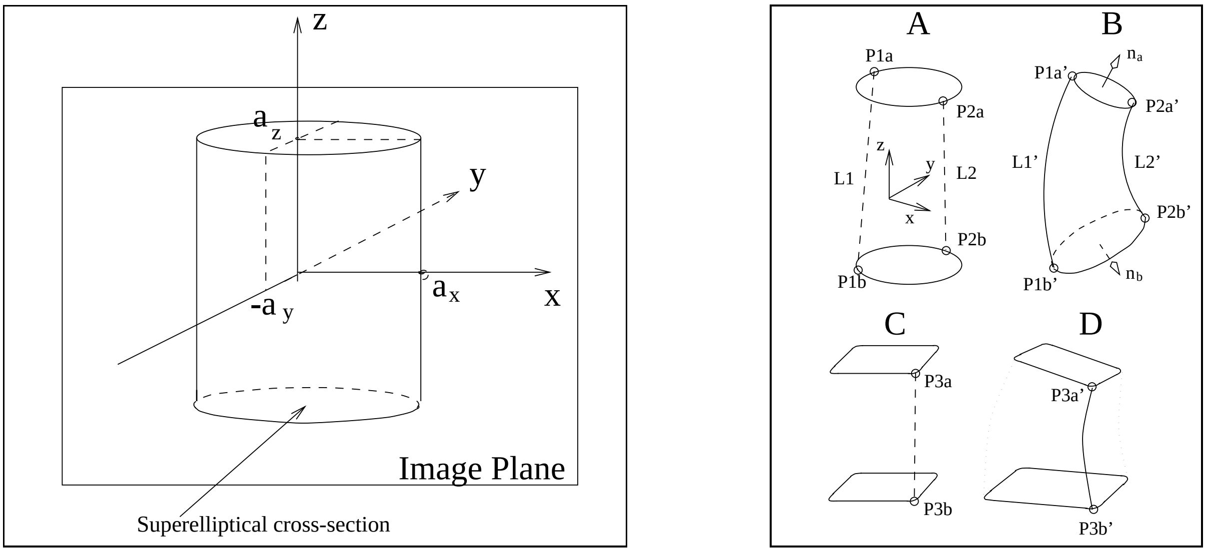


Figure 1: Construction of the parametrically deformable contour model of geons: Initial superelliptical cylinder (left) and determination of occluding contour and central rim (right). See text for details.

By these simple approximated models of geon contour inspired by deformable superquadric modelling, we can represent 12 geon classes with a good level of accuracy. The proposed model could actually represent all 36 geon classes once a certain amount of deformation is introduced that would asymmetrically deform the superelliptical cross section; however this is unnecessary, because it has been shown that such deformations are unrecoverable from 2D images [35].

Some examples of geon PDCMs produced by this method can be seen in Fig. 3 and in the experimental section. The time for creating an instance of such a model is less than 1ms on a SPARC 10 machine, which is over 2 order of magnitudes faster that any other method that would use a direct exact computation of the outline using raster scan techniques or computation of surface normals as in [35].

It is necessary to point out that, although effective, this model becomes rather imprecise with high amount of bending under viewing directions where the tilt is greater than about $\pi/4$; in these situation, however, the geon would be virtually unrecoverable from its contour, unless a precise model of it is known.

One last important remark is due. Geons are, by their very nature, qualitative primitives and one might argue how they can be modelled by simple shapes – such as the one proposed here – or by globally deformable superquadrics. Although this criticism is certainly correct, for the task of recognition and detection these models constitute a valid *low order* approximation of geon shapes and surely good enough to recover their distinguishing features. It is up to the fitting algorithm to be able to cope with this low order-ness and make sure that high-order components do not affect the robustness of the process.

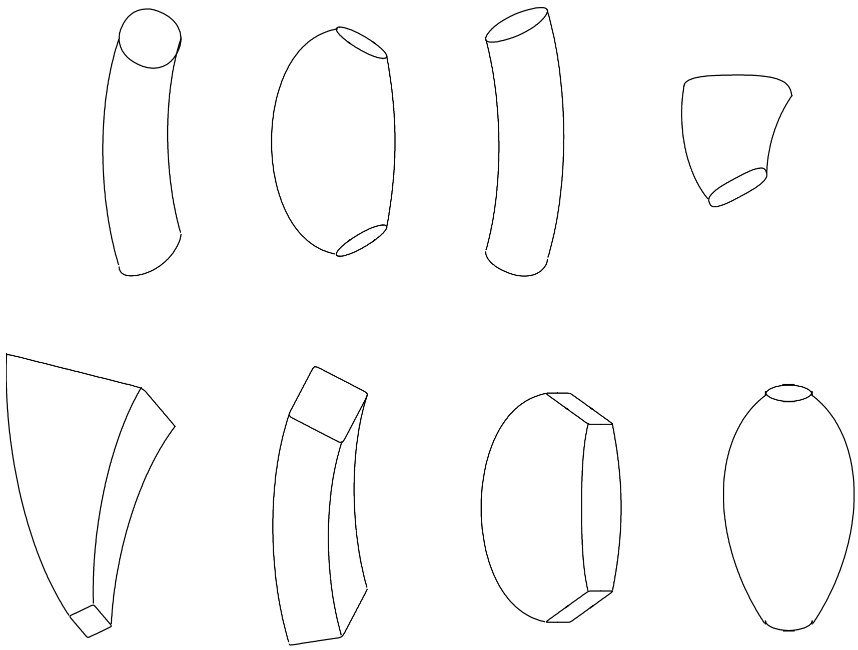


Figure 2: Example of geon contour models generated by the proposed method. The parameters controlling the PDCM shape are the same as the ones that would produce a similar contour projection from a globally deformable superquadric.

4 Aspect Partitioning of PDCM

In the previous subsection, a PDCM has been presented which represents the variable contour of geons through its parameters. This section describes how the PDCM parameter space is partitioned in "cells" that correspond to topologically distinct PDCM aspects.

First, the definition of topological equivalence for geon PDCMs is given, and that will be used to generate distinct aspects. Let us take the model described in the previous section and give it an orientation corresponding to the direction of the positive z axis of the original undeformed superelliptical cylinder.

Now, let us impose a labelling scheme on some features of the geon PDCM. Let $U = \{curved, squared\}$ be two properties of the cross section, and $V_{top} = \{visible/nonvisible\}$ and $V_{bottom} = \{visible/non\_visible\}$ two properties of the two geon ends which indicate whether they are visible or not, the ends being the top and bottom superellipses in Fig. 1-left.

The Cartesian product $U \times V_{top} \times V_{bottom}$ induces 8 PDCM classes. Of the twelve PDCM parameters, only four change the PDCM class, namely ϵ, which affects the cross-section roundness, and c (bending), θ_{tilt} and θ_{pan}, which affect the visibility of the two ends. Cross-section dimensions, length, tapering and swelling do not change the topology as it has been defined . For the topology theory connoisseurs, these equivalence classes

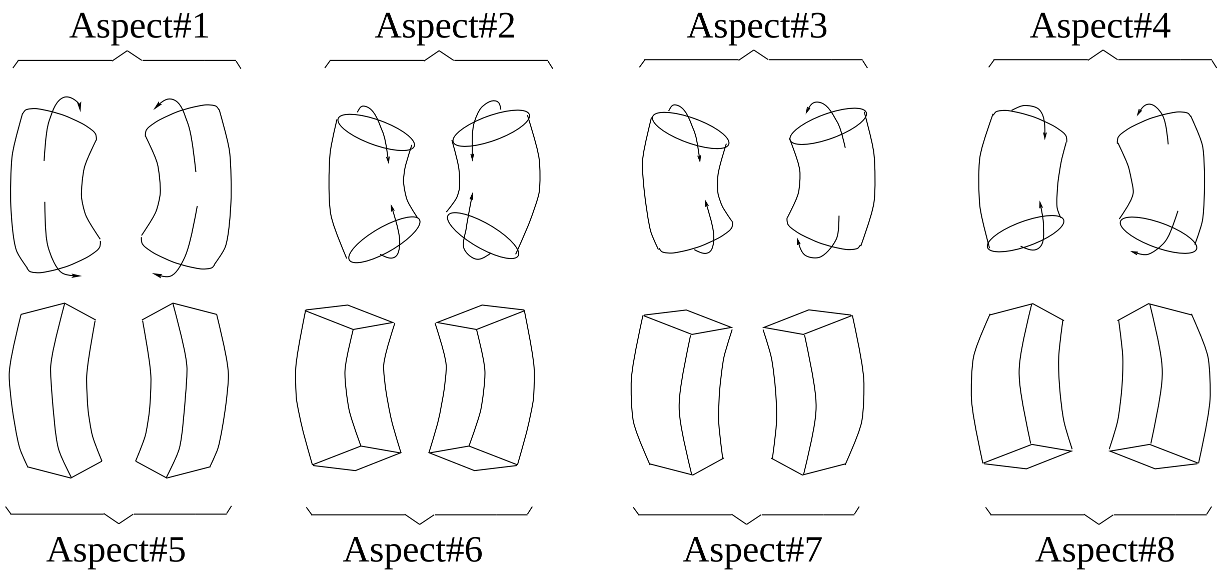


Figure 3: Distinct PDCM topologies and their enumeration. The features defining the topology are the visibility of top and bottom ends and the central rim.

partition the 4D parameter space $S = \{\epsilon, c, \theta_{tilt}, \theta_{pan}\}$ into eight dense simply-connected open subspaces of S, thus creating eight different topologies in the parameters space; each of these topologies correspond to a stable view of the PDCM that preserve the labelling we have imposed; these topologies are known as *aspects* [27] of the PDCM, of which some examples are shown at the top-left of Fig. 3 along with the enumeration that will be used henceforth.

As said in the previous subsection, the property $U = \{curved, squared\}$ is determined by simply setting a threshold $\overline{\epsilon} \in 0.3 .. 0.6$ for ϵ, hence dividing S in two symmetric 3D sub-spaces S' and S''.

The separation from one topology to another in S' (S'') are singularities that are called *visual events surfaces* [27]. By analysing the expressions of the two normals to the ends as functions of c, θ_{tilt} and θ_{pan}, a closed-form for those surfaces has been determined as the zero set of the functions A and B defined as follows:

$$\begin{cases} A = \cos(\theta_{tilt})\sin(\theta_{pan})\sin(\alpha) - \sin(\theta_{tilt})\cos(\alpha) \\ B = \cos(\theta_{tilt})\sin(\theta_{pan})\sin(\alpha) + \sin(\theta_{tilt})\cos(\alpha) \\ \alpha = \arctan(c) \end{cases}$$

The plot in Fig. 4 shows these surfaces. The region within which each aspect is defined is given by the inequalities in the table of Fig. 4.

In principle it should be possible to consider also aspects without one or both ends to model parts that are joined to other parts at their ends. All the discussion so far and what follows can be trivially extended to include these other aspects.

Aspect#1:	$\epsilon > \bar{\epsilon}$	$A > 0$	$B > 0$
Aspect#2:	$\epsilon > \bar{\epsilon}$	$A < 0$	$B < 0$
Aspect#3:	$\epsilon > \bar{\epsilon}$	$A > 0$	$B < 0$
Aspect#4:	$\epsilon > \bar{\epsilon}$	$A < 0$	$B > 0$
Aspect#5:	$\epsilon \leq \bar{\epsilon}$	$A > 0$	$B > 0$
Aspect#6:	$\epsilon \leq \bar{\epsilon}$	$A < 0$	$B < 0$
Aspect#7:	$\epsilon \leq \bar{\epsilon}$	$A > 0$	$B < 0$
Aspect#8:	$\epsilon \leq \bar{\epsilon}$	$A < 0$	$B > 0$

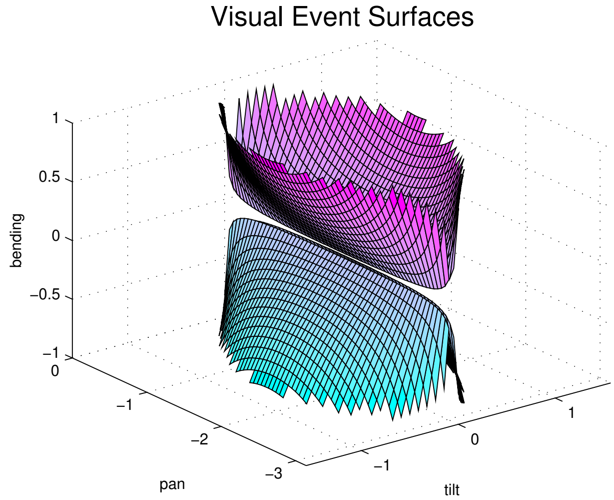


Figure 4: Aspect definition (left table, see text for the definitions) and plot of the visual event surfaces in the bending/pan/tilt parameter subspace (bottom-hull: Aspect#1/5; top-hull: Aspect#2/6; right-part: Aspect#3/7; left-part: Aspect#4/8). The gap between the hulls is a rendering flaw.

5 Matching a Single Aspect

Our Model-Based Optimisation approach to geon detection involves the minimisation of an objective (or cost) function that expresses the quality of the image-model match and other constraints that will be discussed later.

There are many conceptually different ways of designing an objective function suitable for a certain application but they mainly fall in these three categories: Energy Minimisation (EM), Maximum A Posteriori (MAP) or Minimum Description Length (MDL). It has been shown that given a certain problem and a certain fitting quality assessment criterion, they are conceptually equivalent (i.e. in [56, 28]). Practically, however, the nature of the criterion makes the use of a particular method easier. In this work, a MAP philosophy has been followed, but the ideas behind it could be restated in MDL terms.

Let $\mathcal{M}_i = M(\mathbf{x}_i)$ be a geon PDCM instance built as in Sec. 3 expressed in terms of pixels by a set of (i, j) image pixel coordinates and of which we would like the determine the likelihood of fit, and let

$$\mathbf{x_i} = [a_x\ a_y\ a_z\ \epsilon\ K_x\ s\ c\ \theta_{pan}\ \theta_{tilt}\ \theta_{opt}\ P_x\ P_z]^T \tag{3}$$

be the vector of the PDCM parameters. Furthermore, let $\mathcal{I}$ be the original image and $\mathcal{E}$ the binary edge image, which can be produced by a standard Canny edge detector; $\mathcal{E}$ has the same shape as $\mathcal{I}$ and $(i, j) \in \mathcal{E}$ is 1 if an edge has been detected at $(i, j) \in \mathcal{I}$ and 0 otherwise.

The *a posteriori* likelihood of a PDCM matching the image can be expressed in term of *a priori* probabilities by Bayes rule:

$$P(\mathcal{M}_i \mid \mathcal{E}) = \frac{P(\mathcal{E} \mid \mathcal{M}_i)\ P(\mathcal{M}_i)}{\sum_{j=1}^{N_h} P(\mathcal{E} \mid \mathcal{M}_j)\ P(\mathcal{M}_j)} \tag{4}$$

where N_h is the total number of hypotheses produced by the optimisation procedure. The model that best fits the image is the one for which $P(\mathcal{M}_j \mid \mathcal{E})$ is maximum, that is:

$$\mathcal{M}_{best} = M(\mathbf{x}_{best}) = \max_i\{P(\mathcal{M}_i \mid \mathcal{E})\}$$

or, by inverting the sign and expressing probability in term of logarithms:

$$\mathcal{M}_{best} = M(\mathbf{x}_{best}) = \min_i\{-\log(P(\mathcal{M}_i \mid \mathcal{E}))\} \tag{5}$$

Since the denominator of Eqn. (4) is constant over all hypotheses, the minimisation need only be concerned with the numerator. In the two following sections we describe how we defined the model-conditional image and prior probabilities.

5.1 Model-Conditional Image Probability

In Eqn. (4) $P(\mathcal{E} \mid \mathcal{M}_i)$ expresses the conditional probability of having particular image evidence in the presence of the model. Although many ways of defining this probability are possible, we express this probability in terms of how many image edgels "match" the PDCM contour.
Let

$$\mathcal{E}_m(\mathcal{M}_i) = \{(k,l)\ :\ \mid (i,j) - (k,l) \mid \leq d\ ,\ \ (i,j) \in \mathcal{M}_i\}$$

be the d-neighbourhood of the model contour $\mathcal{M}_i$ and $\mathcal{E}_b(\mathcal{M}_i) = \mathcal{E} - \mathcal{E}_m(\mathcal{M}_i)$ the rest of the edge image which is not covered by it; henceforth we drop the $\mathcal{M}_i$ arguments wherever there cannot be ambiguities.

By assuming that the presence/absence of an edge in $\mathcal{E}_b$ and $\mathcal{E}_m$ can be considered independent (this is valid in general) and with different distributions, $P(\mathcal{E} \mid \mathcal{M}_i)$ can be expressed as:

$$P(\mathcal{E} \mid \mathcal{M}_i) = P(\mathcal{E}_b \mid \mathcal{M}_i) \cdot P(\mathcal{E}_m \mid \mathcal{M}_i). \tag{6}$$

$\mathcal{E}_b$ and $\mathcal{E}_m$ can be considered, to a first approximation, as realizations of binary ergodic processes, for which the probability of single local outcomes are all the same, namely p_{b1} and p_{m1}, respectively. The value of p_{b1} is given by the ratio between edge locations and the number of pixels in the image (typical values: 0.02-0.06) and p_{m1} ranges from 0.6 to 0.9, depending on the neighbourhood dimension d and how good the edge detection is expected to be. This ergodicity assumption is simplistic; a Markovian model that would take into account relationships between neighbouring pixels would perhaps be a more accurate model but this is left for future work.

Let N_{b0}, N_{b1}, N_{m0} and N_{m1} be the number of locations (i,j) that are "1" (edge) or "0" (non-edge) in $\mathcal{E}_b$ and $\mathcal{E}_m$, respectively; the probability that a certain number of elements in $\mathcal{E}_b$ and $\mathcal{E}_m$ is "1" or "0" follows a binomial distribution but, since we are interested in a particular realization of the process that is the image itself, the two probabilities in Eqn. (6) can be expressed as:

$$P(\mathcal{E}_b \mid \mathcal{M}_i) = {p_{b1}}^{N_{b1}}(1-p_{b1})^{N_{b0}}$$
$$P(\mathcal{E}_m \mid \mathcal{M}_i) = {p_{m1}}^{N_{m1}}(1-p_{m1})^{N_{m0}}$$

By taking the logarithm of both sides, we obtain:

$$\log(P(\mathcal{E}_b \mid \mathcal{M}_i)) = N_{b1}\log(p_{b1}) + N_{b0}\log(1-p_{b1}))$$
$$\log(P(\mathcal{E}_m \mid \mathcal{M}_i)) = N_{m1}\log(p_{m1}) + N_{m0}\log(1-p_{m1}))$$

which in turn, by letting $N_1 \triangleq (N_{b1}+N_{m1})$ be the overall number of pixels in the image that are edge, are expanded to:

$$\begin{aligned}\log(P(\mathcal{E}_b \mid \mathcal{M}_i)) &= [N_1\log(p_{b1}) + N_1\log(1-p_{b1})] - \\ &\quad (N_{m1}\log(p_{b1}) + N_{m0}\log(1-p_{b1})) \\ \log(P(\mathcal{E}_m \mid \mathcal{M}_i)) &= N_{m1}\log(p_{m1}) + N_{m0}\log(1-p_{m1}).\end{aligned} \tag{7}$$

Then by taking the logarithm of both sides of Eqn. (6) and expanding we obtain:

$$\begin{gathered}\log(P(\mathcal{E} \mid \mathcal{M}_i)) = \log(P(\mathcal{E}_b \mid \mathcal{M}_i)) + \log(P(\mathcal{E}_m \mid \mathcal{M}_i)) = \\ K + [N_{m1}\log(p_{m1}) + N_{m0}\log(1-p_{m1})] - [N_{m1}\log(p_{b1}) + N_{m0}\log(1-p_{b1})] \quad ,\end{gathered} \tag{8}$$

where K is the constant term in square brackets in Eqn. (7) and therefore it will be dropped in the MAP estimation.

In an information theoretical framework this equation has a precise meaning. The term $-\log(P(\mathcal{E} \mid \mathcal{M}_i))$ is the overall number of bits necessary to express the whole edge image $\mathcal{E}$ and $-\log(P(\mathcal{E}_m \mid \mathcal{M}_i))$ and $-\log(P(\mathcal{E}_b \mid \mathcal{M}_i))$ are the number of bits needed to represent the information in the model neighbourhood ($\mathcal{E}_m$) and in the background ($\mathcal{E}_b$) under the ergodicity assumption. The minimisation in Eqn. 5 can then be re-interpreted as the search for the most economical description in term of the edge evidence *and* the model, bringing all into a MDL framework [37][28]. A more formal proof of the MDL/MAP equivalence can be found in [47] and in the context of computer vision in [19, 28]. This information theoretical avenue was followed in [18] but with the fundamental difference that there the *a priori* p_m was computed by looking at the number of pixels matching the *current* instance of the model, therefore making the mistake of using the same data set for both training and estimation; some experiments that we carried out by using their objective function gave unusual high likelihood for bad fits as well, which was somehow expected from what has been just said.

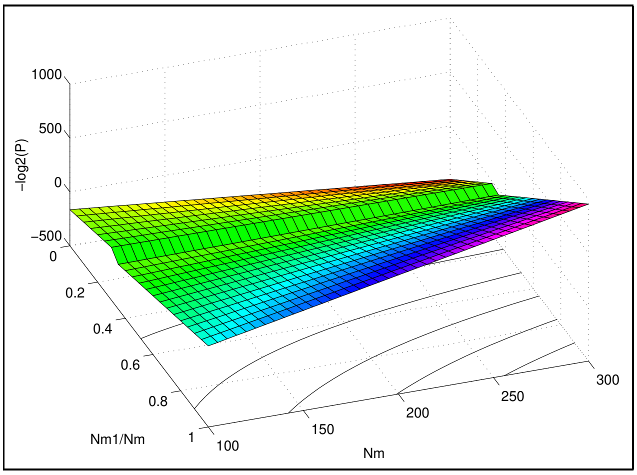


Figure 5: Example of model-conditional image probability $-\log(P(\mathcal{E} \mid \mathcal{M}_i))$ for $p_{m1} = 0.7$, $p_{e1} = 0.06$. See text for details.

Fig. 5 shows an example behaviour of $-\log(P(\mathcal{E} \mid \mathcal{M}_i))$ (not considering K) for $p_{m1} = 0.7$, $p_{e1} = 0.06$ and the total number of model points $N_m = N_{m1} + N_{m0}$ ranging from 100 to 300; the small step has been added in order to show the points at which the absence or presence of the model $\mathcal{M}_i$ is equally likely ($P(\mathcal{E} \mid \mathcal{M}_i) = 0.5$): beyond this line the probability increases with the overall model dimension N_m, that is a preference is given to bigger models.

5.2 Model Prior Probability: A Heuristic

Within a Bayesian framework it is necessary to express the occurrence probability of each instance of the model, called the *model a priori probability*. In most research this probability is neglected (i.e. is considered uniform) but, through experimentation, it has been found that by introducing a heuristic on the prior probabilities, the overall quality of the fitting can be improved.

The reasons for introducing a model prior probability are essentially three: *i*) some parameter configuration are unlikely to occur (such as a bent and swollen object); *ii*) certain configurations of parameters arise from a weird viewpoint that would make detection impossible; and *iii*) it biases the fitting to more perceptually likely shapes. These considerations are both practical and also correspond to sensible assumptions to reduce the quantitative shape ambiguities caused by the projection.

A sensible heuristic has been defined to express these loose constraints. The probability of each aspect is expressed by overlapping (multiplying) marginal densities of parameter values or combinations of them, tacitly assuming independence amongst them. The parameters we took into considerations are the dimension parameters a_x, a_y and a_z, swelling, bending and the pan rotation; the others are given a uniform probability. Below we show

A: $P(c, \theta_{pan}) = K' - N(\mid c \mid -1, \sigma_{c,\theta_{tilt}}) \cdot N(\theta_{pan} + \pi/2, \sigma_{c,\theta_{tilt}})$

B: $P(a_x, a_y) = N(\mid a_x - a_y \mid, \ \sigma_{a_x,a_y})$

C: $P(a_z \mid a_x, a_y) = \begin{cases} N(\tau \frac{\mid a_x + a_y \mid}{2} - a_z, \sigma_{a_z}) & \text{if } a_z \leq \tau \frac{\mid a_x + a_y \mid}{2} \\ \frac{1}{\sqrt{2\pi} \cdot \sigma_{a_z}} & \text{otherwise} \end{cases}$

D: $P(c, s) = N(\mid c \mid \cdot s, \ \sigma_{c,s})$

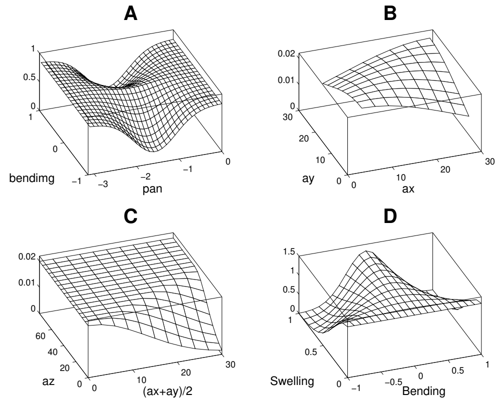


Figure 6: Heuristic model prior probabilities: definitions and plot for each contributing term. The definitions and details are given in the text. These probabilities constitute an heuristic that bias the fitting to perceptually more plausible volumetric shapes corresponding to similar 2D contour projections.

how we defined the probability density functions.

- c and θ_{pan} In the case of Aspects #3, #4, #7 and #8 when θ_{pan} is close to $-\pi/2$, (that is we have frontal view of the only visible ends) bending cannot be detected from the occluding contour and therefore we need to strongly assume straightness of axis, i.e. the only thing we can perceive in these situations. Without this constraint the model could bend forward an arbitrary amount and yielding essentially the same occluding contour. To model this constraint we set up an unnormalised p.d.f. like the one Fig. 6-A, where K' is an additive constant (controlling the desired minimum value of the p.d.f.) and $\sigma_{c,\theta_{tilt}} = 0.5$. In fitting Aspects #1, #2, #5 and #6, the bending is essential for the visibility or invisibility of both ends and this constraint is not used.

- a_x and a_y The projection onto the image plane of a 3D object changes its shape, but our perceptual system is slightly biased to assume more compact cross-sections rather than weird rotation angles [33]. We therefore model the joint p.d.f. as given in Fig. 6-B, which is a constant-height ridge running along the $a_x = a_y$ line. The value of σ_{a_x,a_y} is fairly large because this constraint need not be severe ($\sigma_{a_x,a_y} = 20$ in Fig. 6-B). This constraint assumes that the objects in the scene are not too flat and should be dropped if that is the case.

- a_z The PDA length could take any value but, since it defines the length of allegedly elongated parts like geons, it should be biased to be bigger than the cross-section dimensions by a constant factor τ. A non-normalised p.d.f. as the one given in Fig. 6-C has been set to model this constraint; the figure shown it for $\tau = 1.5$ and $\sigma_{a_z} = 20$.

- c and s High swelling and bending are incompatible. In statistical terms we can express this constraint by a (non-normalised) p.d.f. like the one shown in Fig. 6-D and arising from a Gaussian distribution over the product $c \cdot s$. The plot in Fig. 6-D is given for $\sigma_{c,s} = 0.3$.

Now that we have all the non-normalised probabilities and given the assumption of prior independence between parameters, we just multiply them together to obtain the (non-normalised) *a priori* p.d.f. of the model:

$$\begin{aligned} \log(P(\mathcal{M}_i)) = \quad & H + \log(P(a_z \mid a_x, a_y)) + \log(P(c, s)) + \\ & \log(P(a_x, a_y)) + \log(P(c, \theta_{pan})). \end{aligned} \tag{9}$$

The normalisation constant H is unnecessary because it does not affect the MAP estimate.

This heuristic has improved the perceptual goodness of the recovered shapes but there would be other possible ways of defining the model prior probability, which could also incorporate more detailed specific domain-dependent knowledge about the scene structure.

5.3 MAP Estimation Procedure

The MAP estimation obtained by the minimisation of

$$-\log(P(\mathcal{M}_i \mid \mathcal{E})) = -\log(P(\mathcal{E} \mid \mathcal{M}_i)) - \log(P(\mathcal{M}_i)), \tag{10}$$

where the two terms are given by Equations (9) and (8), is rather difficult to achieve, since it is extremely irregular and presents many shallow and/or narrow minima.

As an example, Figure 7 shows some graphs of the objective function value taken at three orthogonal planar regions of the parameter space (in particular about the initial estimate of the handset upper-piece example of Figure 10): although the three surfaces are rather rugged, three pronounced valleys stand out that correspond to good values of the objective function. In the middle figure, however, two valleys beyond the ripples might jeopardise the fitting procedure.

By trying to minimise Eqn. (10) alone, it was also found that sometimes the optimisation got stuck in local minima because of the step-like nature of the model-conditional probability of Eqn. (8) (remember we used a binary "belonging to the model" criteria). For overcoming this problem, a small smoothing term has been added to the right side of Eqn.10; this term represents the average minimal distance between contour model and image edge points (by using a minimal distance transform computed off-line) and it does not affect the MAP estimate but just helps convergence in cases where image and model are much displaced and the numerical computation of the gradient become meaningless due to the low number of edge points falling inside the model neighbourhood. This term can then be seen as "telling the optimisation where to go" in absence of other information.

In early stages of the work, a Levenberg-Marquandt method with added random perturbations was used, following [5] and other works, but this method led to difficult convergence. The choice fell then to Simulated Annealing (see Appendix A) for a summary

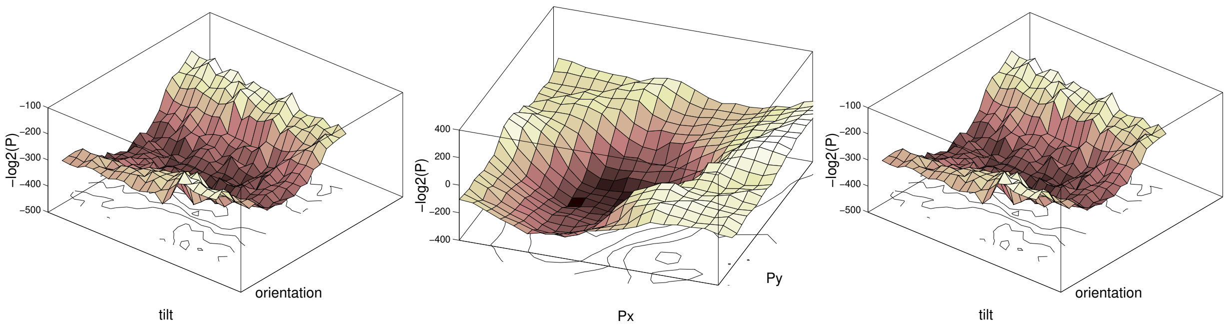


Figure 7: Three graphs of the objective function value taken at three orthogonal planar regions of the parameter space about the initial estimate of the handset upper-piece example of Figure 10: although the three surfaces are rather rugged, three pronounced valleys stand out that correspond to good values of the objective function.

of the method), which is a powerful optimisation tool that efficiently combines gradient descent and controlled random perturbations to perform the minimisation of non-convex functions. The actual implentation is a publicly available version of Simulated Annealing, called Adaptive Simulated Annealing (ASA) [23]. The set-up of the ASA algorithm will be extensively discussed in the next section.

6 Experimental System

In this section outlines the simple experimental system, schematically depicted in Fig. 8, that has been used to carry out the experiments.

Starting (for instance) from the set of hypotheses produced by the method described in [41], for each hypotheses each of the eight PDA are initialised at a representative position and independently fitted to the image. The PDA that obtains the best scores is considered the best fit to the image.

The approach relies on two fundamental assumptions [14]:

1. The MAP estimate that started with the "correct" hypothesis will converge to the correct interpretation of the image;

2. The quality of the fit (score) of this correct interpretation must be higher than any other.

No theoretical proof of convergence and uniqueness of the method is possible since the problem is strongly non-linear and too complex to be analysed as stated also by [14], where rigid models were used. The experiments of the next section will, however, empirically show that the proposed method reasonably complies with these two goals.

In the following three subsections, the PDA initialisation and the optimisation set-up of the experimental system is described.

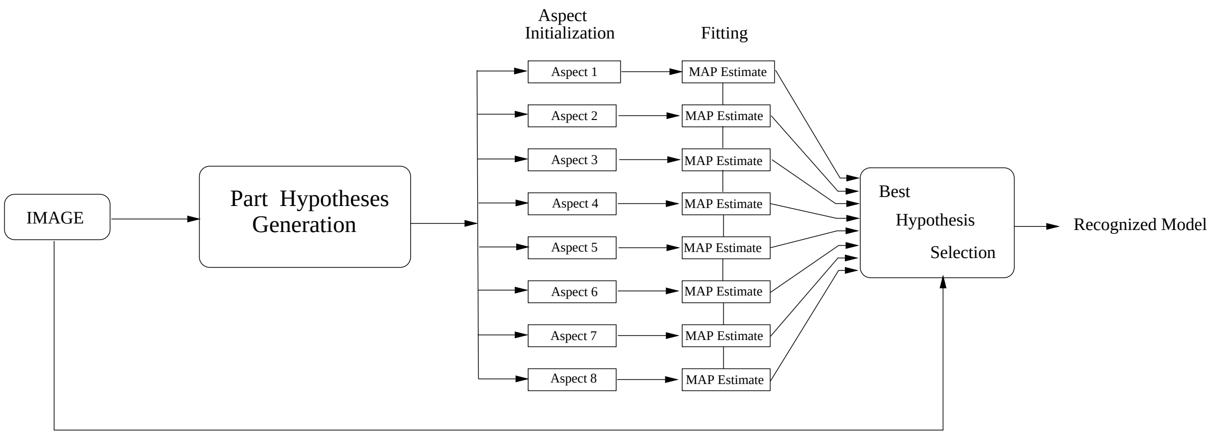


Figure 8: The simple aspect-based control strategy. For each part hypotheses, the eight PDA are independently initialised and fitted to the image. The one that obtain the best fitting score gives the best interpretation of the image.

6.1 Initialisation

The initialisation stage is concerned with estimating coarse part initial hypotheses (sometime called *the frame* [52]) that comprise position, orientation of the major axis and dimensions. These initialisations need not be precise and the degree of allowed inaccuracy depends upon the power of the optimisation procedure. The initial estimates could be produced by the part-grouping and filtering method proposed in [41]. However, the two modules are currently not integrated and in some experiments the intialization have set hy hand to be qualitatively similar to those output by the MDL hypothesis filtering method presented in Chapter 5 of [41].

It is worth highlighting again that the generation of generic part hypotheses from 2D images is a novelty in vision, functionally matched only by the system proposed by [35]. Part hypotheses are physically PDM models fitted onto the image [41] and their position, orientation and respective variation modes are used to initialise the PDCM by assigning them to $\overline{P}_x$, $\overline{P}_z$, $\overline{a}_z$, $\overline{a}_x = \overline{a}_y$ and $\overline{\theta}_{opt}$, respectively. The values of $\overline{a}_x$ and $\overline{a}_y$ are set to be equal because we do not have prior information on the aspect-ratio of the part cross-section. Information about bending, tapering or roundness, which are projective quasi-invariant properties, are not currently used to better initialise PDCMs, but this could be done with relatively little effort.

6.2 Aspect Hypotheses Generation

From each of the initial frames, all eight distinct aspects are instantiated by properly setting the PDCM parameters that control the aspect topology. Referring to Fig. 4, we chose points in the topology-controlling parameter space (Sec. sec:partitioning) that are

	squareness (ϵ)			bending (c)			θ_{tilt}			θ_{pan}		
	init	min	max	init	min	max	init	min	max	init	min	max
Aspect#1:	0.75	0.51	0.99	-0.5	-1	0	0	$-\pi/4$	$\pi/4$	$-\pi/2$	$-\pi$	0
Aspect#2:	0.75	0.51	0.99	0.5	0	1	0	$-\pi/4$	$\pi/4$	$-\pi/2$	$-\pi$	0
Aspect#3:	0.75	0.51	0.99	0	-1	1	$-\pi/8$	$-\pi/4$	0	$-\pi/2$	$-\pi$	0
Aspect#4:	0.75	0.51	0.99	0	-1	1	$\pi/8$	0	$\pi/4$	$-\pi/2$	$-\pi$	0
Aspect#5:	0.25	0.05	0.49	-0.5	-1	0	0	$-\pi/4$	$\pi/4$	$-\pi/2$	$-\pi$	0
Aspect#6:	0.25	0.05	0.49	0.5	0	1	0	$-\pi/4$	$\pi/4$	$-\pi/2$	$-\pi$	0
Aspect#7:	0.25	0.05	0.49	0	-1	1	$-\pi/8$	$-\pi/4$	0	$-\pi/2$	$-\pi$	0
Aspect#8:	0.25	0.05	0.49	0	-1	1	$\pi/8$	0	$\pi/4$	$-\pi/2$	$-\pi$	0

Table 1: Initialisation and bounds for the aspect topology-controlling parameters. See text for details.

more or less equidistant from the visual event surfaces and therefore are placed in a fairly central position within each aspect cell. This choice is a sensible heuristic that reduces the distance between the initial point and any possible true final estimate [2]. Table 1 (along the "init" columns) shows these values for each aspect topology. The other parameters, bending, swelling and tapering, were all set to zero.

6.3 Optimisation Set-Up

The optimisation of strongly non-linear functions is *"typically a non-typical problem"* [45] and therefore no canned optimiser can be used. As pointed out in [23], the set-up of the ASA algorithm is a bit tricky, since no theoretical guide exists, but once the right configuration has been found, the method becomes reasonably robust. Having said that, here we describe the essential set-up of the ASA optimiser.

One of the key decisions when using a constrained optimisation algorithm is the choice of the parameter bounds; the ASA algorithm requires hyper-rectangular bounds defined by a minimum and a maximum for each parameter.

Within our aspect-based control strategy, we basically have two sets of parameters, those controlling the PDCM aspect topology (ϵ, c, θ_{tilt} and θ_{pan}) and those that do not change it (a_x, a_y, a_z, K_x, s, P_x P_z and θ_{opt}).

Section 3 gave a closed-form expression of the visual event surfaces bounding different aspect topologies. In order to make the ASA optimiser to "stay within" a certain aspect topology, we do two things: *(1)* give it a 4D search bounds (given in Table 1) that enclose the true aspect cell; and *(2)* invalidate states (through a specific ASA option) that fall outside the chosen aspect cell by checking the constraints given in the table in Fig. 4. In most of the experiments we carried out, the ratio between invalid and valid generated

[2]The choice of these values can be regarded also as giving maximal disambiguation distance between visual events [25].

states was always less the 5%.

Besides the parameters constraining the aspect topology, the others need bounds too. Bounds for the tapering and swelling deformations are set to their full range (Sec. 3); in the case of position, sizes and orientations, bounds are set as tolerances with respect to the initial values $\overline{P}_x$, $\overline{P}_z$, $\overline{a}_z$, $\overline{a}_x = \overline{a}_y$ and $\overline{\theta}_{opt}$. The following table summarises these bounds (N is the resolution of the image):

	a_x	a_y	a_z	K_x	s	θ_{opt}	P_x	P_z
Min	$\overline{a}_x - 40\%$	$\overline{a}_y - 40\%$	$\overline{a}_z - 40\%$	-1.0	0.0	$\overline{\theta}_{opt} - \pi/8$	$\overline{P}_x - \frac{N}{10}\%$	$\overline{P}_z - \frac{N}{10}\%$
Max	$\overline{a}_x + 40\%$	$\overline{a}_y + 40\%$	$\overline{a}_z + 40\%$	1.0	1.0	$\overline{\theta}_{opt} + \pi/8$	$\overline{P}_x + \frac{N}{10}\%$	$\overline{P}_z + \frac{N}{10}\%$

In order to improve convergence, we need also to specify the deltas for computing the partial pseudo-derivatives of the cost function, which are chosen such that for each parameter, a perturbation equal to its respective delta should produce detectable changes in the image at a given resolution. For 128x128 images, the values of Δ_{a_x},Δ_{a_y},Δ_{a_z}, Δ_{ϵ}, Δ_{K_x}, Δ_s, Δ_c, $\Delta_{\theta_{pan}}$, $\Delta_{\theta_{tilt}}$, $\Delta_{\theta_{opt}}$, Δ_{P_x} and Δ_{P_z} are set to 1.0, 1.0, 1.0, 0.05, 0.2, 0.05, 0.05, 0.01, 0.01, 0.01, 1.0 and 1.0, respectively.

The annealing schedule plays an important role. We have experimentally found that sub-optimal schedules are also related to the aspect topology we are trying to fit, probably because of the different kind and number of features. For good convergence the Temperature_Ratio_Scale parameter [23] has been set to 10^{-12} for Aspect#1 ... Aspect#4 and to 10^{-10} for Aspect#5 ... Aspect#8. Finally the number of iterations has been set to 2000, which we found to be a good trade-off between speed (about 5s for each optimization run on a SPARC 10) and good convergence; moreover, for the experiments carried out with 128x128 images, we set p_{b1}, p_{m1} and d (see Sec. 5.1) to 0.07, 0.85 and 1, respectively.

7 Experimental Results

In this section, three set of experiments are discussed.

In the first set, several fitting experiments of geon PDCM are shown for both synthetic and real images with the purpose of verifying the validity of the cost function and the optimization. The second set aims at assessing the validity of the two premises to the use of aspects given in Section 6. In the final set of experiments, three fitting experiments to the familiar handset test image are given along with interpretation of the results; in particular, an example of what can happen when the aspect-based strategy in not used is also supplied.

7.1 Testing the MAP fitting

In this subsection a number of single fitting experiments are shown that help assess the validity of the cost funciton and the optimization method for fitting the PDCM proposed in this paper. Here the aspect-based strategy is not used but, but in some experiments some of the topology-definining parameters have been constrained, as we shall see later.

The experiments presented here can in turn be divided in two sets, which are described in the following. In both experiments, the initialization is performed manually and is intentionally set to be poor to test for worst cases.

FIRST SET

The set of 18 fitting experiments shown Fig. 9 was designed to assess convergence and viability of the cost function and the optimization procedure. Six geon-like objects were created with some plasticine and an image of them was then taken with a resolution of 512x512 pixels. A Canny edge detector was applied and the resulting cluttered edge image is shown in Image C (left) of Fig. 9. This image has been intentionally used without any post-processing – like cleaning and linking – because we wanted to test the convergence in hard conditions.

Afterwards, two synthetic images mimicking the original one created, one with roundish primitives (Image A of Fig. 9) and the other one with squared cross-sections (Image B of Fig. 9). The initial PDAs are shown in the left column Fig. 9 overlapped to the respective edge images; the initializations for these two synthetic images are rather crude but the right topologic aspects have been imposed to each example.

The manual initializations are the same across the three images except for the roundness parameter, which has been set to "squared cross-section" in image B. The corresponding results of the fitting can be seen in the right column. The neighbourhood dimension was set to 7 (that is $d = 3$) and the other parameters are the same as given in Sec. 6.3; each estimate was produced in about 25 seconds on a networked SPARC 10 machine.

- **Image A (Top of Fig. 9)** The results here are essentially good but in the case of Object 6 the sign of the bending is wrong. All the geon distinquishing features have been correctly detected, as can be visually seen.

- **Image B (Centre of Fig. 9)** In this case the results are better than the one in Image A because the presence of the additional interior edge gives "more information" to the fitting.

- **Image C (Bottom of Fig. 9)** As expected, the results here are not particularly exciting but they can be considered positive, given the intentionally poor edge image quality we have used. Here, the roundness parameter ϵ was set free to check whether a change in the aspect topology would occour. The results for object 2,3 and 5 are very good. The fit of Object 1 is essentially correct (apart from slight tapering), but the spurious edge due to a high shading gradient caused the object to be interpreted as a bent prism. Object 4 too has been fit rather poorly (because the high noise) but the essential orientation, bending and tapering have been recovered. In the case of Object 6 the presence of shadows and poor image contrast has been fatal and the fitting is a complete failure, with a final result that, although obtaining a higher score, looks poorer that the initial estimate.

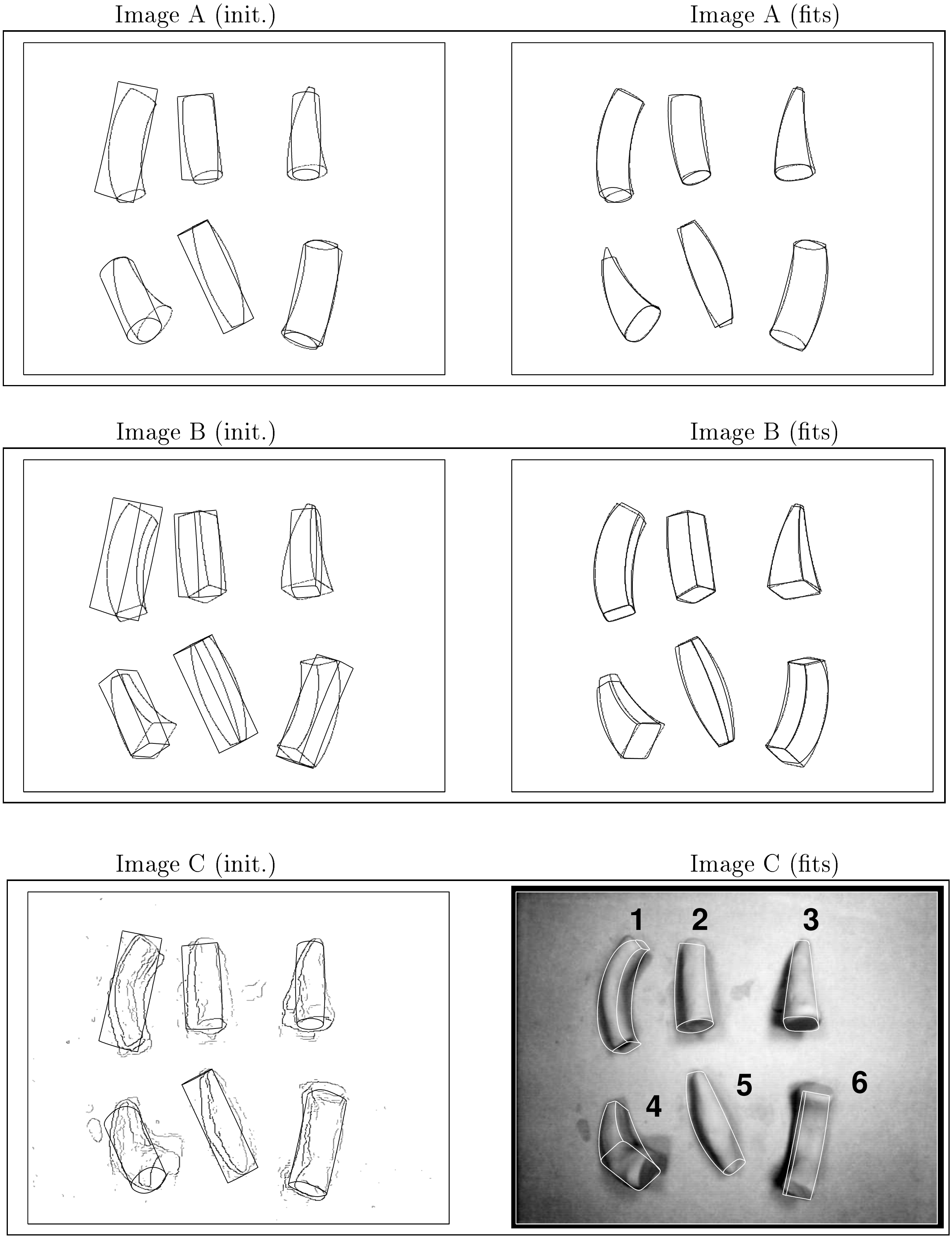


Figure 9: First set of experiments. The purpose is to assess validity of the objective functon and the optimization; the aspect-based strategy is not used here. A description of the eighteen fitting experiments is given in the text. Although only one initialization for each is shown here, many others have been tried that, however, kept the same initial topology as the ones shown.

	Handset Image			Banana&Mug	
	Top Piece	Handle	Bottom Piece	Banana	Cup
Tapering (K_x)	0.09	0.08	0.21	-0.02	-
Swell (s)	0.08	0.28	0.42	0.47	-
Bending (c)	-0.12	0.25	0.15	0.35	-
Squareness (ϵ)	0.84	0.26	0.69	0.45	-

Table 2: Final Parameter estimation. The recovered parameters allow a coarse description of the shape: Top Piece: cylinder; Handle: slightly bent prism; Bottom Piece: Swollen and slighlty tapered cylinder; Banana: bent and swollen prism. See text for details.

SECOND SET

This set of experiments has been carried out with real images of isolated objects, a handset, a mug and a banana. The examples in Fig. 10 are 128x128 gray-level images; the neighbourhood dimension was set to 3 (that is $d = 1$) and the optimization set-up was the same as for the first set of experiments;

This time, the initialization was performed by manually selecting out rectangular regions of the image (top of Fig. 10), thresholding to extract the silhouette and finally by computing the principal moments that gave coarse estimates of position, axes lengths and orientation; the result are the initializations shown at the top of Fig. 10.

- **Handset** The top-left of Fig. 10 shows the original handset image with the initial models instances and their major axes overlapped on it. The two end parts (ear and mouth piece) have a rather poor initial estimate because of their low eccentricity and the shadows cast on the background. On the other hand, the central part is well defined and hence a good initial estimate is achieved; at this point there is no knowledge about the squareness of this part. The centre-left figure shows the edge image. It can be noticed that there is some cluttering, like that caused by circular ridges at the mouth piece. The bottom-left figure shows the results obtained after applying the optimisation to each one of the initial estimates. As it can be seen, the results are rather good. Table 2 shows that the main geons' distinguishing features are captured, with the exception of the top part (ear piece) not being swollen as it should; in this case, however, even for a human it would be difficult to tell the exact shape of such a short part just from that poor edge image. Another remark worth making is that the length of the central part was correctly found despite the rightmost edge that runs along the whole handset. Note that some research has been recently carried out [44, 5] in the classification of geons from parameters such as those that define our PDCM.

- **Banana** The top-right of Fig. 10 shows the initial estimate of the banana shape. The combined effect of a shadow in the right-hand side of the banana and poor resolution has lead to the poor edge image shown in the centre-right image. Here,

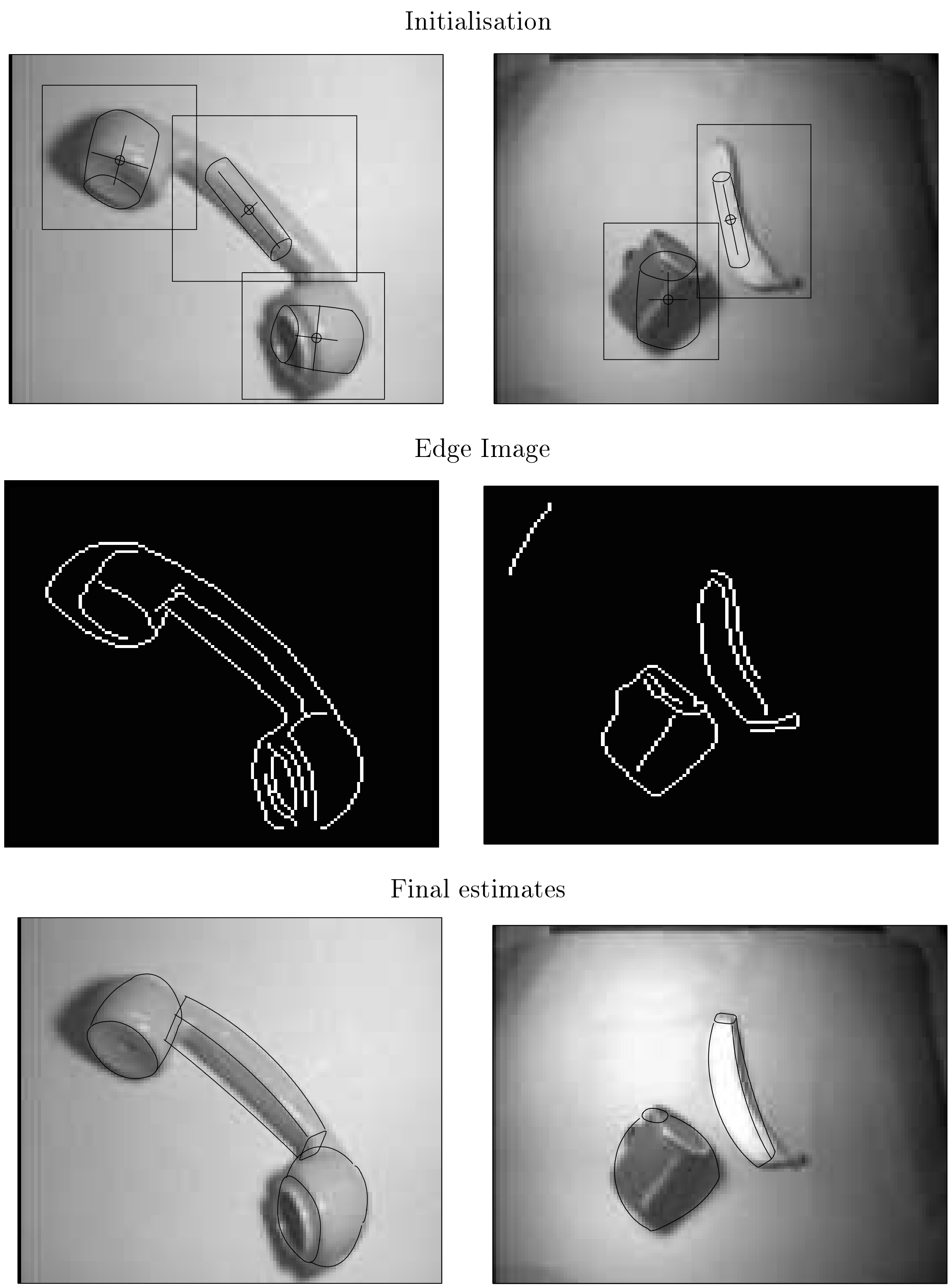


Figure 10: Second set of experiments with semi-automatic initialization again without using the aspect-based strategy (see text for details). The fitting to the handset geons and the banana are reasonably good whereas mug one is a sheer disaster.

the little incomplete square that somehow appears at the top and the double edge running along the right-hand side were interpreted as part of the shape, as shown in the final estimate in the bottom-right image. Table 2 shows that again all the essential features (apart from roundness, as just said) are grasped, such as curvature, swallowness and no tapering.

- **Mug** This experiment is a complete failure. The big shadow, the highlight at the top and poor resolution led to an edge image that is virtually uninterpretable by the human eye. The initial estimate shown at the top-right of Fig. 10 is mis-oriented in the image plane by roughly $\pi/4$ and the estimation procedure produced a very poor result. Only by giving a very good initial estimate, a better result was achieved.

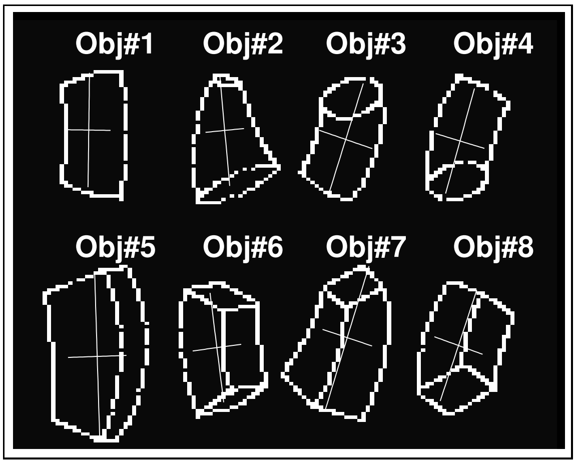
The experiments described above show that the proposed method works reasonably well. The tests with clean images indicate that the optimization converges well. Results with real images show the the method performs well if a coarse initial estimate is given and there is not too much noise or spurious edges. However, as in the mug example, more care must be given at determining the initial estimate for one that is poor can yield dramatically wrong results, especially for low-eccentricity objects and with high noise level.

7.2 Testing the Aspect-based strategy: Synthetic Image

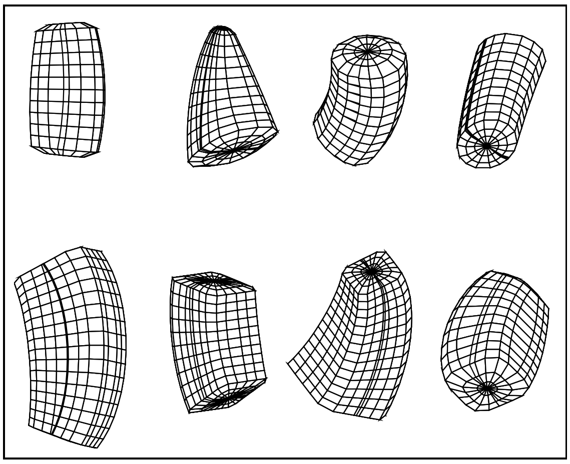
This subsection presents one of the experiments set up for testing the aspect-based strategy, in particular the two premises given at the beginning of Section 6: when starting from the correct PDA, the fitting must both converge and give a better score than the ones obtained from initialization with any of the wrong-topology aspects.

Eight synthetic contours of geons (Obj#1 ...Obj#8), each representing a different apect topology (Aspect#1 ...Aspect#8), have been placed in the same 128x128 image (Fig. 11-A) and a coarse initialization was given using estimates of just orientation, position, length and cross-section dimension; the initializations are represented by the crosses. Then, all eight distinct PDA were initialized by the method given in Section 6.1 and fitted on each of the eight objects, with the same optimization set-up as the one given in Section 6.3; the resulting scores were put in a confusion table (Fig. 11) whose lines represent the scores of fitting an object with all the aspects. These results validate the two main assumptions of the aspect-based control strategy outlined at the beginning of Sec. 6: the boxed scores on the diagonal are the best ones for each geon, that is the correct aspect obtained the best score in all cases. Fig. 11-B show the superquadric representation using the very same parameters results from the fitting of the best aspect. It is worth pointing out that the superquadrics are built using the very same parameters produced by the fitting and used for contructing the PDCM; these volumetric representations are then the ones that once projected onto the image plane would yield the fitted object contours.

An interesting behavior also crops up from the analysis of the scores in the confusion matrix. Let us take the case of Obj#8. The second best score corresponds to the one obtained with Aspect#6, which has a visible bottom end, whereas the third best score



A



B

	Aspect#1	Aspect#2	Aspect#3	Aspect#4	Aspect#5	Aspect#6	Aspect#7	Aspect#8
Obj#1	$\boxed{-264.77}$	-210.16	-170.32	-238.72	-211.14	-121.30	-133.60	-123.24
Obj#2	-242.49	$\boxed{-314.16}$	-252.11	-252.83	-167.71	-170.09	-106.96	-179.19
Obj#3	-177.34	-142.54	$\boxed{-255.22}$	-193.46	-183.00	-116.05	-13.18	-161.31
Obj#4	-236.55	-220.50	-187.26	$\boxed{-261.39}$	-177.06	-151.19	-166.85	-130.36
Obj#5	-252.05	-291.04	-245.37	-284.76	$\boxed{-478.75}$	-171.62	-62.20	-161.26
Obj#6	-290.78	-384.35	-346.86	-295.48	-392.26	$\boxed{-458.83}$	-362.63	-261.09
Obj#7	-284.68	-192.94	-300.22	-135.86	-166.14	-245.18	$\boxed{-437.68}$	-230.28
Obj#8	-249.61	-275.68	-211.26	-322.67	-270.43	-326.21	-241.54	$\boxed{-374.05}$

Figure 11: Experiment with synthetic images of 8 different aspect of geons and the confusion matrix representing the results of the fittings. The boxed results are the highest scoring PDA for each fitting experiment and all correspond to the PDA with the same topology as the respective test contours in fig. A. The superquadric corresponding to these best PDAs are displayed in figure B: the 3D shapes are in well in agreement with the 3D structure that pops up from the contour images when we see them.

correspond to Aspect#4, which is the one that presents a visible bottom face but non-squared cross-section; evidently these features matched well the image and contributed to improve the overall score. Similar considerations can be made for other objects. This behavior suggests a side-effect of this strategy, that is the ranking of aspect hypotheses according to “how well” they fit the instance of objects, at least insofar as synthetic images go. With real images this phenomenon is much smoothed but is still present, as we shall see for the other examples.

7.3 Real Image: a Handset

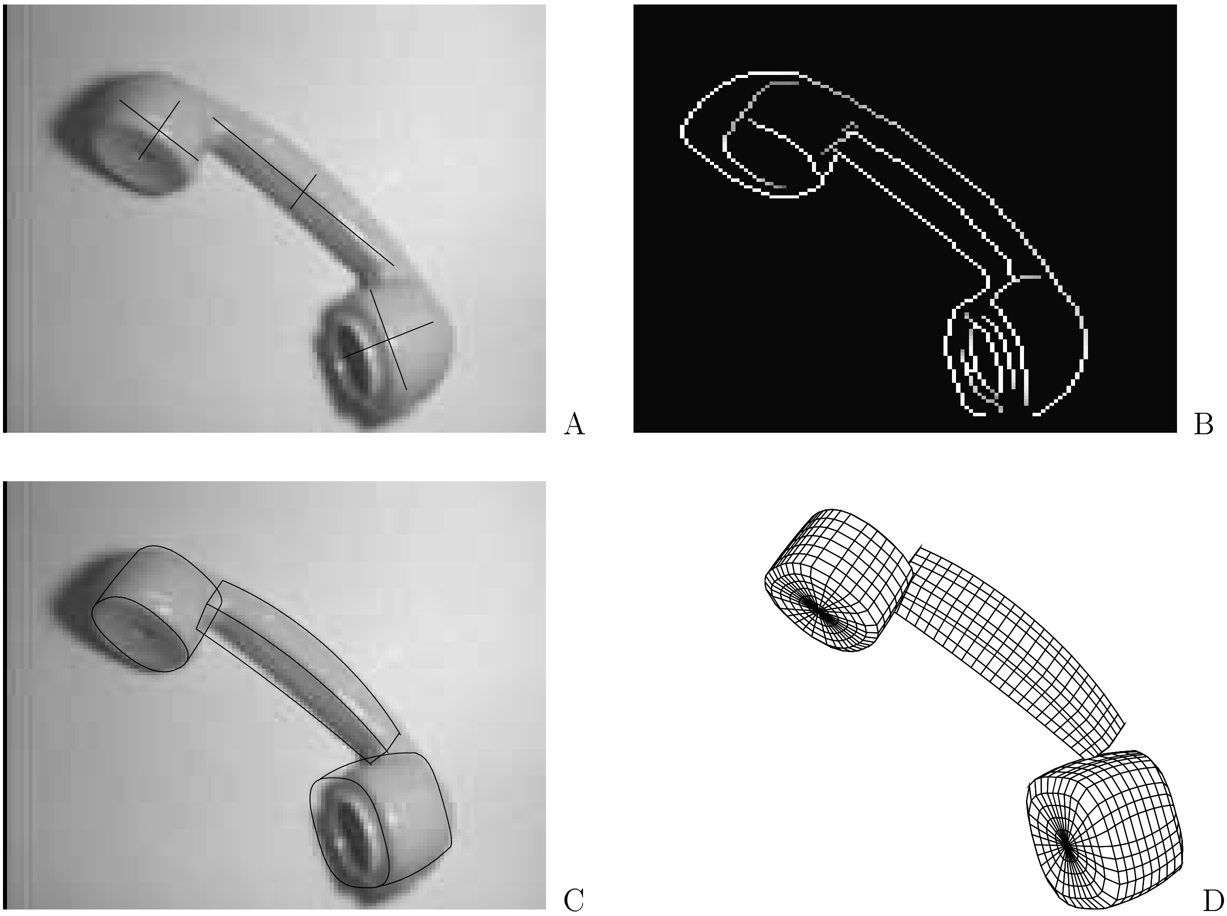
In this experiment, the now familiar 128x128 grey-level image of the handset is used (Fig. 12-A). The corresponding edge image is reported here for convenience in Figure 12-B. The initializations are performed as outlined in Section 6.1 and come from selected hypotheses produced by the part-based grouping and filtering method presented in [41].

Both end-pieces of the handset have almost no eccentricity and therefore it was not possible to determine their natural major axis, which is an essential requirement of geon representation. Which of the two axes was the major one was imposed by hand, but a straightforward automated strategy would just assume that for low-eccentricity blobs both major-axis hypotheses should be tried out and the best be selected. This problems are familiar when the data has rotational symmetries but the models employed are oriented [31].

As in the experiments with the synthetic image, Aspect#1 through Aspect#8 were fitted to the image for each of the initialization hypotheses, again with the same optimization set-up as is Sec. 6; the scores for each fit are given in the table in Fig. 12. The best fits, which correspond to the boxed scores, are displayed both as contours overlapped onto the real image in 12-C and Fig. 12-D.

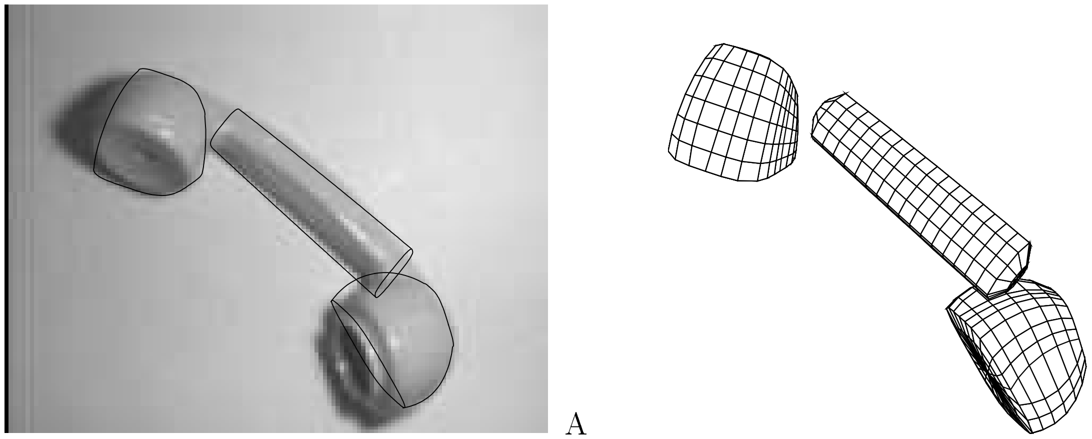
The two correct aspects for the mouth and ear pieces got the highest score as auspicated, since their ends are well visible. The interpretation of the mid-part has turned out to be a bit ambiguous, with the two scores for Aspect#5 and Aspect#7 very close; this is due to the invisibility of its ends and their overall low weight for such an elongated part. The correct aspect scored the highest here but either would have acceptable, given that in this case they are almost indistinguishable. It is worth remembering that we are looking for qualitative features of parts and what really matters is that the model with the right features is selected over other possible alternative ones, that is, the fitting quality need not be absolute but relative.

The aspect-based strategy avoids situations as the real case presented in Fig. 13, where the fitting results are shown that are obtained from the same position/axes initialization as above but when all the parameters governing the aspect topology were left unconstrained (of course always within meaningful ranges). In the experiments of Section 7.1, the fitting was performed by giving a good initialization and good results were obtained; here, pan, tilt and squareness values are set to 0.0 and 0.5, respectively.
Although the number of iterations was increased to compensate for the bigger search-space, the results obtained are rather poor. The top piece is completely mis-interpreted, as well



	Asp. #1	Asp. #2	Asp. #3	Asp. #4	Asp. #5	Asp. #6	Asp. #7	Asp. #8
Upper	-216.78	-193.43	-150.51	$\boxed{-246.30}$	-171.89	-88.31	-158.49	-112.34
Mid	-226.07	-223.15	-223.18	-228.82	$\boxed{-386.32}$	-306.45	-382.91	-301.60
Lower	-213.79	-238.68	-74.19	$\boxed{-286.10}$	-171.95	-156.41	-160.54	-246.30

Figure 12: Real-image experiment with the aspect-based control strategy. Here, the PDA have been initialized automatically from some of the hypotheses produced by the part-based grouping and filtering method presented in [41]. The figure shown initialization (A), edge image (B), contour fits (C) and their volumetric representation (D). The scores of the PDA fittings are shown in the table. See text for more details.



A B

Figure 13: Handset fitting results *without* using the aspect-based strategy and from an initialization where pan, tilt and squareness values are set to 0.0 and 0.5, respectively, and size/position/orientation as the ones in Figure 12. The fitting in all three cases got stuck in deep local minima.

as the mid-part, which was recognized as a cylinder. It can be noticed that in these two cases the fitted models (Fig. 13-left) match very well a considerable amount of the edges, a clear indication that very deep minima of the objective function were found there which the optimizaton algorithm could not escape. The use of topologically distinct aspects has not only the property of reducing the dimension of the search space but also of dramatically boiling down the presence and effect of undesirable minima within it.

8 Discussion

In this paper a novel approach to 3D qualitative part recovery from real 2D images has been presented. A new efficient deformable model is fit to raw edge images in the framework of Model-Based Optimisation, with an objective function expressed in Bayesian terms and the use of topologically distinct aspects has led to more reliability. The results we have presented here show that this method is valid and open to further developments.

In this section, the major contributions of the material presented in this paper are highlighted, followed by some criticisms and the proposition of future work.

8.1 Contribution

There are several contributions to vision research in this paper. All of them were recognized by anonymous reviewers of a paper based on the paper and its early version that appears in [42].

- A new approximated but efficient parametric model of deformable superquadric contour is presented in Section 3. Previously, when performing deformable superquadrics fitting to 2D images as in [35], a very clumsy method was used whereby the whole superquadric was built, deformed and its contour computed by finding zero-crossings of the surface normal component along the optical axis. Here, a leaner, simple geometrical model has been pragmatically designed that approximates the contour of the deformable superquadric in a tiny fraction of the cost needed by the other method. The parameters keep a clear three-dimensional interpretation, as well as for deformable superquadrics.

- The fitting of the aforementioned parametrically deformable contour model is performed through model-based optimization where an objective function is minimized that, in information theoretical terms, essentially expresses the economy obtained by representing groups of edgels in the image by the model contour. Although similar cost functions had been proposed in the past, the one presented here formally accounts for both matched and unmatched contour portions and the background in formal Bayesian terms, whereas previous method (such as [19, 18]) did not do so. Through experimental results, it has been shown that this method copes with a significant amount of cluttering.

- Although its contribution has not yet well quantified, the embedding into the model prior probability of a bias towards more perceptually plausible 3D shapes – described in Sec. 5.2 – is a rather clever idea, as remarked by F. Ferrie in a personal communication.

- The concept of using an aspects-based strategy to deformable contour model fitting has been introduced here for the first time. Previous work had used aspects only for fitting CAD-based models, such as in [14]. The benefits of such a strategy are straightforward: the optimization can independently focus on regions of the parameter space that correspond to models with the same topology, thereby reducing the chances of getting stuck in local minima caused by different interpretations of image features. Due to the simplicity of the geon model defined in this paper, a closed-form solution for the aspect cell subdivision has been found.

- The idea of recognizing generic primitives like geons from 2D images by fitting contour of superquadrics is not a new idea, but the only implementation known to the author is by [35]. However, there the fitting was performed to segmented data and optimization was done in image space (see Sec. 2.1) in a multistage fashion with two *ad hoc* different search strategies for cylindroids and prismoids – probably due to severe fitting problems, also highlighted by the apparent syntheticity of the examples shown in their paper. Here, this topological information has been brought to the fore by employing right different models, which has allowed us to safely utilize a more general optimization algorithm such as Simulated Annealing.

8.2 Criticisms to the Method

There are several issues that need to be addressed in future work to improve the proposed method. Here, we will first discuss criticisms made by anonymous reviews of a submission based on this paper; the overall opinion on the work was rather positive, but some acute criticisms were made which are summarised and commented in the following. Some other criticisms will be added later.

The first criticism was that the method does not constitute a significant advancement with respect to the current state-of-the-art work by [35]. This criticism was mainly attributed to the manual initialisation phase that was then used – now taken over by the automatic part-based grouping presented in [41]. In my opinion the criticism in unjustified. The method proposed in [35] assumes that faces and edges belonging to a single part are pre-segmented by the OPTICA [12] system, which also supplies information about the class of object to be fitted; this allowed them to implement an *ad hoc* strategy for dealing with the fitting of different classes of models. The problem of fitting to unsegmented data was not even taken into consideration, whereas here the fitting is done to *unsegmented* data and, in principle, the initialisation could come from methods other that the one proposed in [41]. Another important remark is that in [35] a clumsy method for determining superquadric contours was used, whereas here a purposely designed model (Sec. 3) has been built that allows much greater efficiency.

Differently from the OPTICA system equipped with the superquadric fitting machinery of [35], the scope of this thesis was not to build a generic-part segmentation recognition system – quite beyond the state of current vision technology – but to explore the possibility of using a global model-driven method to segment out generic parts from ordinary edge image. The imposition of structure on the solution by the parametrically deformable aspects of the geon fitting method presented in this paper is nothing but a natural extension to the *fil rouge* of the line of thought of [41].

Another criticism coming from another anonymous reviewer was that the fitting results were *not* impressive; this is a rather unfair statement that probably was inspired by improbable comparisons between the results given here and parallel works on part segmentation from (often pre-segmented) **range data**, such as [49, 54]. The absence of precise models, image cluttering and, again, the use of unsegmented 2D edge data, would never allow a precise fitting, unless other information is used.

The use of a neighbourhood “in/out” criterion in the design of the cost function of Section 5 has allowed a formal expression in Bayesian terms of the goodness of fit but some troubles can be encountered when the geon being fitted cannot be properly represented by the PDCM given in Section 3. This representation problem is common to all global deformable models fitting methods but it manifests itself more when censored error norms are employed, like [19] [11] [29] or the one presented in this paper.
The possibility of using a different, smoother error norm that would avoid these problems is under investigation. Preliminary experiments showed that the results are much worse than the one presented in this paper but it too early to draw conclusions.

Finally, some doubts could arise regarding the model prior probability given in Sec.

5.2. The definition might look arbitrary but it should be remembered that it was meant to have an heuristic character. In early stages of the work, this probability was uniform over the parameter space, as it is often done in these cases. When such a heuristic was added, the fitting results improved rather significantly especially in regard to the recovery of 3D shape and not only the matching of the contour; further systematic experiments are needed to evaluate how these probabilities affect the final results but, given the stochastic nature of the optimiser (simulated annealing) a large amount of experiments are need to objectively evaluate the effectiveness of such a heuristic. However, this activity was not deemed relevant at this stage and is left to future work.

8.3 Future Work

The technique presented in this paper has opened some problems and interesting prospectives alike. Some issues that would need to be addressed in the near future are the following.

First above all, in [41] *codons* [46] were presented as indivisible pieces of information. Here we are working on raw data, because after the hypotheses generation phase we have only an rough idea about which codons make up the actual part outline, let alone interior edges. In Section 5.1 we saw that edgels falling within a certain neighbourhood were considered as matching the model but some effort could perhaps be spent in trying to use whole codons as data to be matched. This modification would probably prevent spurious chunks of data locally matching the model contour to fool the goodness of fit evaluation, and would also yield a smoother objective functions, thereby easing optimisation. Some preliminary results in this direction look promising.

As said back in Section 2.1, the fitting is performed in parameter space. However, a very exciting prospect would be to use the point-to-point correspondence method by single value decomposition by [48] for fitting each aspect in image space, analogously to PDM fitting of the part-based grouping phase of [41]. For doing so, each PDA would need to be redesigned as a point distribution model, as done in Chapter 3 of [41] for building the generic-part PDM from superellipses. This technique might allow greater robustness, speed of convergence and tolerance to bad initialisation, due to the power of the SVD correspondence method that would *globally* find the best matches between PDA landmarks of the aspects and the data, however cluttered it might be.

Another exciting step to try is to account for interactions between parts. In Chapter 4 of [41] we saw that by taking account of many competing interpretations of local evidence, it is possible to produce a minimal, hopefully correct, interpretation of the image. The same considerations could be done here. In the case of the handset test image, for instance, the fitting could be performed concurrently for the three parts and penalty terms could be introduced for overlapping as in the support competition method prposed in [41]. However, differently from that, in this case the fitting and hypotheses competition would be performed at the same time and the workload would be huge.

A natural extension, which would however present several theoretical problems, would be to integrate other non-edge information in the fitting, specifically in the cost function, such as coarse depth and surface orientation information as it could be produced by a shape-from-shading method [21].

9 Acknowledgements

The first author is partially supported by SGS-THOMSON Microelctronics. Andrew Fitzgibbon and David Eggert have contributed interesting suggestions.

A Simulated Annealing

Simulated Annealing (SA) [26] is a powerful optimisation tool that effectively combines gradient descent and controlled random perturbation to perform the minimisation of non-convex functions and it was developed from the Metropolis algorithm [36], which was originally contrived to simulate the equilibrium state in statistical mechanics; the Metropolis' algorithm can be summarised as follows. Given a solid composed by interacting atoms, small random perturbations are added to the current state; a differential of energy ΔE is computed and if $\Delta E < 0$ the new state is accepted as a valid one. Conversely is $\Delta E \geq 0$ the state is not rejected but it is given a probability $e^{-\Delta E/kT_a}$ (*Metropolis criterion*) , where k is the Boltzmann constant and T_a is the absolute temperature. By keep repeating this procedure for a large number of times the system eventually converges to a thermal equilibrium.

More recently, in [26] an important modification was proposed to the Metropolis' algorithm that consisted of running it with decreasing temperatures (called Boltzmann Annealing) until a low enough temperature is reached. This improvement was inspired by the physical annealing process of a solid and therefore the method was called *simulated annealing* and the way the temperature is lowered called *annealing schedule*.

Optimisation by SA was first introduced to the vision community in the seminal paper by Geman&Geman [20] and more recently used also in [32, 54] and other works.

In this thesis, a recent publicly available implementation of SA has been used, called Adaptive Simulated Annealing (ASA), developed by Ingber at Caltech [23]. As described by Ingber, *"the major difference between ASA and standard Boltzmann SA is that the ergodic sampling takes place in a $n+1$ dimensional space, in term on n state variables and the cost function"*.

References

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

[2] R. Bergevin and M. Levine. Generic object recognition: Building and matching coarse description from line drawings. *IEEE PAMI*, 15(1):19–36, January 1993.

[3] I. Biederman. Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94:115–147, 1987.

[4] T.O. Binford. Visual perception by computer. In *Proceedings of the IEEE System Science and Cybernetic Conference*, Miami, December 1971.

[5] D.L Borges. *Recognizing Three-Dimensional Objects using Parametrized Volumentric Models*. Phd thesis, Department of Artificial Intelligence, University of Edinburgh, 1996.

[6] C. Chen and A. Kak. A robot vision system for recognizing 3D objects in low-order polynomial time. *IEEE Transaction Syst. Man and Cybern.*, 19:1535–1563, 1989.

[7] T.F. Cootes, D. Cooper, C.J. Taylor, and J. Graham. A trainable method of parametric shape description. In *British Machine Vision Conference*, pages 54–61. Springer-Verlag, 1991.

[8] T.F. Cootes, A. Hill, C.J. Taylor, and J. Haslam. Use of active shape models for locating structires in medical images. *Image and Vision Computing*, 12(6):355–365, 1994.

[9] T.F. Cootes and C.J. Taylor. Active shape models - 'smart snakes'. In *British Machine Vision Conference*, page xxxxx, 1992.

[10] C.W.K. Critton and E.A. Parish. Boundary location from an initial plan.: the bead chain alogrithm. *PAMI*, 5(1):8–14, January 1983.

[11] T. Darrell and A.P. Pentland. Cooperative robust estimation using layers of support. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 17(5):474–487, May 1995.

[12] S.J. Dickinson, A.P. Pentland, and A. Rosenfeld. 3-D Shape Recovery Using Distributed Aspect Matching. *IEEE PAMI*, 14(2):130–154, 1992.

[13] D. Eggert and K. Bowyer. Computing the perspective projection aspect graph of solids of revolution. *PAMI*, 15(2):109–128, 1993.

[14] D. Eggert, L. Stark, and K. Bowyer. Aspect graphs and their use in object recognition. *Annals of Mathematics and Artificial Intelligence*, 13:347–375, 1995.

[15] J.M. Ferryman, A.D. Worrall, G.D. Sullivan, and K.D. Backer. A generic deformable model for vehicle recognition. In *BMVC*, volume 1, pages 127–136, Birmingham, September 1995.

[16] A. W. Fitzgibbon and R. B. Fisher. Practical aspect graph derivation incorporating feature segmentation performan ce. In *British Machine Vision Conference*, September 1992.

[17] N.H. Friedland and A. Rosenfeld. Compact object recognition using energy-function-based optimization. *PAMI*, 14(7):770–777, 1992.

[18] P. Fua. Objective functions for feature discrimination: Theory. In *DARPA Image Undertanding Workshop*, 1889.

[19] P. Fua and A.J. Hanson. Objective functions for feature discrimination: Applications to semiautomated and automated feature extraction. In *DARPA Image Understanding Workshop*, pages 676–694, 1989.

[20] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions and the bayesian restoration in images. *IEEE PAMI*, 6:721–741, 1984.

[21] B.K.P. Horn. Shape from shading: a method for obtaining the shape of a smooth opaque object from one view. In M.J. Brooks and B.P.K. Horn, editors, *Shape from Shading*. MIT Press, 1989.

[22] K. Ikeuchi. Generating an interpretation tree from a cad model for 3D object recognition. *International Journal of Computer Vision*, pages 145–165, 1987.

[23] L. Ingber. *Adaptive Simulated Annealing*. Lester Ingber Research, Mc Lean, VA, 1993. [ftp.alumni.caltech.edu./pub/ingber/ASA.tar.zip].

[24] M. Kass, A. Witkin, and D. Terzopulos. Snakes: active contour models. *International Journal of Computer Vision*, 1:321–331, 1988.

[25] J.R. Kender and D.G. Freudenstain. What is a "degenerate" view? In *DARPA Image Understanding Workshops*, pages 589–598, 1987.

[26] S. Kirkpatrick, C.D. Gelatt, and M.P. Vecchi. Optimization by simulated annealing. *Science*, 220:671–680, 1983.

[27] J.J. Koenderink and A.J. van Doorn. The internal representation of solid shape with respect to vision. *Biological Cybernetic*, 32:211–216, 1979.

[28] Y.G. Leclerc. Constructing simple stable description for image partitioning. *International Journal of Computer Vision*, 3:73–102, 1989.

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

[30] A. Leonardis, F. Solina, and A. Macerl. A direct recovery of superquadric models in range images using recover-and-select paradigm. In *ECCV*, pages 309–318. Springer-Verlag, 1994.

[31] J. J. Leou and W.H. Tsai. Automatic rotational symmetry determination for shape analysis. *Pattern Recognition*, 20:571–582, 1987.

[32] K.G. Lim. *Area Based Stereo*. PhD thesis, University of Cambridge, April 1994.

[33] D. Lowe. *Perceptual Organization and Visual Recognition*. Kluwer Academic Publishers, Boston, MA, 1985.

[34] D. Lowe. Fitting parametrized 3D models to images. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 13(5):441–450, May 1991.

[35] D. Metaxas, S.J. Dickinson, R.C., Munck-Fairwood, and L. Du. Integration of quantitative and qualitative techniques for deformable model fitting from orthographic, perspective and stereo projection. In *Fourth International Conference on Computer Vision*, pages 364–371, 1993.

[36] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E.Teller. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092, 1953.

[37] E.P.D. Pednault. Some experiments in applying inductive inference principles to surface reconstruction. In *International Joint Conference on Artificial Intelligence*, pages 1603–1609, Detroit, MI, August 1989.

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

[39] A.P. Pentland and S. Scarloff. Closed-form solutions for physically based shape modelling and recognition. *IEEE PAMI*, 13(7), July 1991.

[40] S. Petitjean, J. Ponce, and D.J. Kriegman. Computing exact aspect graphs of curved objects: Algebraic surfaces". *International Journal of Computer Vision*, 9(3):231–255, 1992.

[41] M. Pilu. *Part-based Grouping and Recogntion: A Model-guided approach*. Phd thesis, Department of Artificial Intelligence, University of Edinburgh, SCOTLAND, 1996. Forthcoming.

[42] M. Pilu and R.B. Fisher. Recognition of geons by parametrically deformable contour models. In R. CIpolla and B. Buxton, editors, *Fourth European Conference on Computer Vision*, volume I of *Lecture Notes in COmputer Science*, Berlin, April 1996. Springer-Verlag.

[43] J. Ponce, S. Petitjean, and K. Kriegman. Computing exact aspect graphs of curved objects: Algebraic surfaces. *European Conference on Computer Vision*, pages 599–614, 1992.

[44] N.S. Raja and A.K. Jain. Recognizing geons from superquadrics fitted to range data. *Image and Vision Computing*, 12(3):179–189, April 1992.

[45] S. S. Rao. *Optimization: Theory and Appolications.* Wiley Estern, 1984. 2nd edition.

[46] W. Richards, J.J. Koenderink, and D.D. Hoffman. Inferring 3D shapes from 2D codons. Technical Report A.I. Memo No. 840, MIT, April 1985.

[47] J. Rissnanen. A universal prior for integers and estimation by minimum description length. *The Annals of Statistics*, 2:416–431, 1983.

[48] G.L. Scott and H.C. Longuet-Higgins. An algorithm for associating the features of two patterns. In *Proc. Royal Society London*, volume B244, pages 21–26, 1991.

[49] F. Solina and R. Bajcsy. Recovery of parametric models from range images: The case of superquadrics with global deformations. *IEEE PAMI*, 12(2):131–147, February 1990.

[50] L.H. Staib and J.S. Duncan. Boundary finding with parametrically deformable models. *PAMI*, 14(11), November 1992.

[51] J. Stewman and K.W. Bowyer. Direct construction of perpective projection aspect graphs for planar-face convex objects. *Computer Vision, Graphics and Image Processing*, 51:20–37, 1990.

[52] J.B. Subirana-Vilanova. Mid-level vision and recognition of non-rigid objects. AI-TR 1442 1442, MIT, January 1993.

[53] A.P. Witkin and J.M. Tenenbaum. On perceptual organization. In A. Pentland, editor, *From Pixels to Predicates.* Ablex, Norwood, NJ, 1985. Original work published in 1981.

[54] K. Wu and M.D. Levine. Recovering of parametric geons from multiview range data. In *IEEE Conference on Computer Vision and Pattern Recognition*, Seattle, WA, 1994.

[55] A. Yuille, P.W. Hallinan, and D.S. Cohen. Feature extraction from faces using deformable templates. *International Journal of Computer Vision*, 8:99–111, 1992.

[56] S. Zhu and A. Yiulle. Region competition: Unifying snakes, region growing, energy/bayes/mdl for multi-band image segmentation. In *International Conference on Computer Vision*, 1995.