Model-Driven Grouping and Recognition of Generic Object Parts from Single Images

Maurizio Pilu* **and Robert B. Fisher****

Department of Artificial Intelligence
University of Edinburgh
5 Forrest Hill, Edinburgh EH1 2QL
SCOTLAND (UK)

Abstract. Grouping is often intended as a general-purpose early vision stage which gathers together image features of perceptual salience, usually having a well-definable structure. This work addresses the problem of generic part-based grouping and recognition from single two-dimensional edge images following a strategy that employs generic part models at all stages: the key underlying idea is to perform a purposive grouping of simple parts and these parts can be conveniently represented by generic part models. This paper outlines the proposed computational method, which is extensively treated in [18].

1 Introduction

Since the early days of computer vision research, part segmentation and recognition has been acknowledged an important role towards the realization of a generic object recognition system. A reliable segmentation of generic objects into their constituent parts would, for instance, tremendously ease object grasping and manipulation, fast indexing to large object databases and so forth. For these reasons, research in part segmentation has been vigorous indeed.

However the large majority of approaches dealt with segmentation from silhouette images, which are normally difficult to extract. In the past few years good works have appeared that use ordinary edge images as input; notably, the method in [6] was region-based and therefore could cope only with clean images. Other excellent approaches have been proposed that are based on Gestaltic perceptual grouping such as [15, 27]; these works are heavily based on the detection of symmetries between part sides and cannot properly cope with very cluttered images.

In this paper, the new paradigm of *part-based grouping* of features is presented that bridges the classical grouping and model-based approaches with the purpose of directly recovering parts from real images, and part-like models are used that both yield low theoretical complexity and reliably recover part-plausible groups of features.

Figure 1 depicts the structure of the proposed computational approach to part-based grouping and recognition by models and at the same time shows how the different topics discussed in this paper relate to each other. From the raw input edge image, *codons* are extracted and then used to form small *seed groups* that allow generic part models (the generic part Point Distribution Model [23]) to be initialised (by ellipse fitting [22]) and then fitted to additional codon evidence. The many hypotheses that are produced by this grouping stage (discussed in Section 2) are subsequently reduced by the Minimum Description Length (MDL) filtering stage (Section 3). Once part segmentation is available, qualitative 3D structure can be recovered by the final parametrically deformable aspect fitting stage (Section 4).

The approach is extensively dealt in [18] and in other publications; this paper, for reason of space, will just describe the underlying philosophy and outline the computational approach.

* Email: `maurizp@aifh.ed.ac.uk`
** Email: `rbf@aifh.ed.ac.uk`

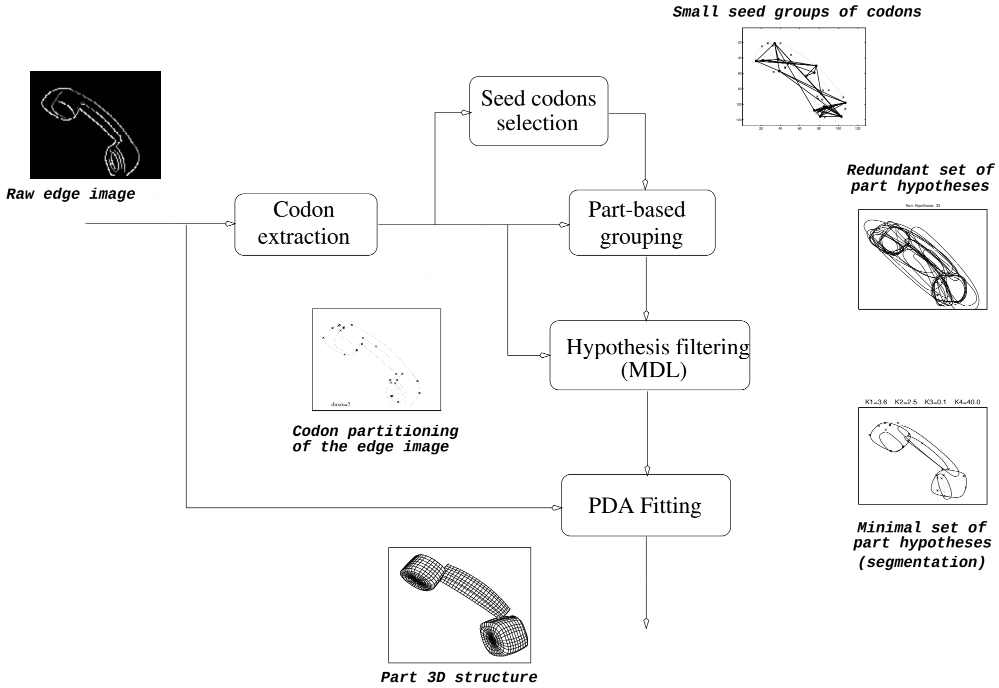

Fig. 1.: The proposed computational approach to part-based grouping and recognition by models. See text for details.

2 Part hypotheses generation by model-driven grouping

Let us think of the noisy contour image of a tree; to the eyes of the human, the tree would be grossly described by two parts: the trunk and the foliage. To achieve this abstract part-level description, a computer system should not only employ some means for "smoothing" the shape but also have a notion of the essential "thing-like" nature of parts. This is also valid for the three-dimensional case as, according to Huffman and Richard's theory of parts [8], solid parts can be inferred from their 2D projection by looking for non-accidental invariant properties in edge images. The "thing-like" nature of parts, also called *objectness*, had often been neglected as a guideline to the computational study of the part segmentation problem until the work of Pentland [17], who argued that objectness can also be expressed by *a set of generically applicable part models* and this line of thought is the hinge of the approach.

Objectness is represented here by the closed contour of the simple generic part Point Distribution Models (PDM) [4] whose training set was built by random deformable superellipses [23]. However, as Pentland put it, there is no known computational model to *"begin immediately with recognition of part models"* [17].

The infeasibility of a method that directly looks for parts in an image suggests that perhaps it is necessary to step one level back from whole-part models in an hypothetical representational hierarchy of objects.

The computational approach that is proposed in this paper to perform model-driven part-based grouping consists of four distinct stages. A synthetic description of the method in terms of *pseudo-code*, is given in Figure 2.

In the first stage *codons*, contour portions of similar curvature [25], are extracted from the raw edge image. They are considered as indivisible image features because they have the desirable property of belonging either to single parts or joints. Codons are represented by second order polynomials, and are recovered from ordinary edge images via a variation of the simple iterative end point fit and split algorithm [24]. An example of codons is given in Figure 3-A.

Once codons are available, a method should be devised for "grouping" codons belonging to single parts. Most works – notably symmetry-based – assume that each codon covers most of the part sides. Unfortunately, this is not the case in real images: often codons are over-segmented, whole boundary segments missing, and

```
Partition image contour into codons
Find small part-plausible seed groups of codons
for each seed group do
       Initialise the part model to the seed group
       Pre-shape the part model to the seed group
       Find supporting codons to the pre-shaped model
       Fit the part model to the additional support
end for
A set of part hypothesis is now available
```

Fig. 2.: Pseudo-code of the model-driven part-based grouping method proposed in this paper. See text for details.

marking, shadows and shading edges are always present. Codons can be considered as *seeds of perception* [3] from which more and more complicated descriptions of the images are constructed. In the same frame of mind and to overcome the above limitations, in the second stage small *seed groups* (currently pairs) of codons are found that give enough structural information for part hypotheses to be created.

The third stage consists in initialising and pre-shaping the models to all the seed groups. First, coarse positions and orientations of the part-like models are determined by fitting ellipses [22] to the pixel belonging to each seed groups of codons. Successively, the PDMs are *pre-shaped* to the seed groups of codons; in this phase, coarse bending and/or tapering estimates are recovered along with positions and dimensions. Note that the concept of pre-shaping to few significant features is a relatively new concept for deformable models that has helped to dramatically increase the robustness of the fitting stage; pre-shaping can also be seen as a way of reducing complexity and facilitating convergence, as much as done in, e.g., hand pre-shaping for robot grasping [28].

Finally, in the fourth stage, a full fitting of the generic part PDMs is performed to a large neighbourhood of each pre-shaped models. Many hypotheses are thus created but the great majority of them will represent the contour data poorly due to the lack of image evidence and can be discarded straight away. However, a number of good or plausible hypotheses end up contending for describing the image evidence, such as those shown in Figure 3-B; the filtering of these hypotheses to produce part segmentation is the subject of the next section.

The outcome of this procedure is also to effectively produce a part-based grouping of edges. It is necessary to stress that this model-driven grouping method is complementary to other grouping techniques, such as symmetry recovery [27] and convex grouping [9], in the sense that it cannot alone solve the grouping problem. These matters are discussed more extensively in [18].

3 Filtering hypotheses by Minimum Description Length

This section presents a novel method for filtering the redundant set of part hypotheses $\mathcal{H}$ produced by the previous grouping stage that retains only those that are likely to correspond to actual parts. The method is inspired by recent work [12, 5] in segmentation using the Minimum Description Length (MDL) criterion [16, 11]. The method has previously been used for segmenting surfaces into patches but, for the first time, here the philosophy is applied to a two-dimensional context. In the proposed approach, supporting evidence for hypotheses is put into competition under the MDL framework to select part hypotheses that most economically represent supporting edges in the "language" of generic parts. The filtering is performed by the maximisation of a quadratic boolean cost function by a genetic algorithm.

The theoretical underpinning of the method is extensively discussed in another paper [19] and here we briefly discuss the implementation.

The method is based on finding the models that most economically *encode* (in terms of bits) the edge image by the contour of the part hypothesis.

Let us indicate by $\mathcal{M}_i$ and $\mathcal{B}_i$ the *supported* and *unsupported* contour portions of each part hypotheses $\mathcal{H}_i \in \mathcal{H}$ (see Figure 3-C), by $\chi^2(\mathcal{M}_i, \mathcal{R}_i)$ the sum of squared orthogonal distances between the hypothesis'

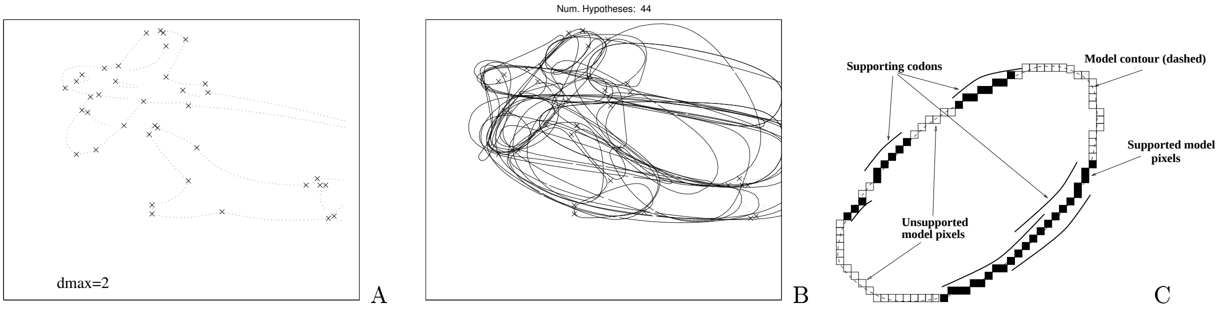

Fig. 3.: A: Segmentation of and edge image into codons. B: Initial redundant set of part hypotheses. C: Illustration of supporting codons and supported contour pixels of a generic part model.

supported contour portions and its supporting codons $\mathcal{R}_i$, and by $\mathcal{M}_{i,j}$ the set of pixels of the hypothesis $\mathcal{H}_i$ (or equivalently $\mathcal{H}_j$) that are supported by the supporting codons $\mathcal{R}_i \cap \mathcal{R}_j$.

Let us now suppose we can determine four constant K_1, K_2, K_3 and K_4 such that K_1 is the average number of bits necessary to represent each supported pixel of a model contour, K_2 is the average number of bits necessary to represent each unsupported pixel of a model contour, K_3 is a constant such that when multiplied by $\chi^2(\cdot,\cdot)$ gives the average encoding length for representing the residuals and, finally, K_4 is the average number of bits needed to specify the parameters of a hypothesis (the shape parameters of the generic part PDM).

If we also presume that in the *final* solution the only kind of model overlapping taken into account is pairwise [17], the best interpretation of the edge image in terms of the hypotheses is obtained by:

$$\hat{\mathbf{m}} = \arg\max_{\mathbf{m}} \left\{ \mathbf{m}^T \mathbf{Q} \mathbf{m} \right\} \tag{1}$$

where $\mathbf{Q}$ is the hypothesis correlation matrix, which will be defined next, and $\mathbf{m} = [m_1 \; m_2 \; \cdots \; m_M]^T$ is the hypothesis presence vector in which each element m_i is "1" or "0" if the model $\mathcal{H}_i$ is present or absent, respectively, in the final image description; any given $\mathbf{m}$ selects a subset $\mathcal{X}$ of the whole set of hypotheses $\mathcal{H}$.

Each diagonal element $q_{i,i}$ expresses the length of encoding the supporting region $\mathcal{R}_i$ of a hypothesis $\mathcal{H}_i$ by $\mathcal{H}_i$ itself:

$$q_{i,i} = K_1|\mathcal{M}_i| - K_2|\mathcal{B}_i| - K_3\chi^2(\mathcal{M}_i, \mathcal{R}_i) - K_4;$$

The off-diagonal elements $q_{i,j}$ deal with interaction between two competing (possibly partially overlapping) hypotheses $\mathcal{H}_i$ and $\mathcal{H}_j$ and ensure that saving and residual overhead due to shared supports are accounted for only once:

$$q_{j,i} = q_{i,j} = \frac{1}{2}\left\{ -K_1 \cdot |\mathcal{M}_{i,j}| \; + \; K_3 \cdot \chi^2(\mathcal{M}_{i,j}, \mathcal{R}_i \cap \mathcal{R}_j) \right\}$$

Intuitively speaking, with this definition $\mathbf{m}^T\mathbf{Q}\mathbf{m}$ is large when the smallest number of models best describe the image and do not have too many unsupported contour portions.

Equation (1) is, technically speaking, a *quadratic boolean optimisation problem*, as the solution space can be represented as the corner of an M-dimensional hypercube. In [12] and [17] this optimisation problem was tackled by using different greedy strategies, which we have found unsuitable to our minimisation because we do not have, in general, good hypotheses. Since our intention was to investigate the real properties and limitations of the proposed segmentation method in the optimal case, a simple genetic algorithm was implemented to perform the boolean optimisation (see [18]).

The MDL principle states that the choice of the constants K_1, K_2, K_3 and K_4 should be theoretically driven by prior probability distributions of edges, gaps, residual and model parameters.

In [20] it is shown that if p_{m1} is the probability that a pixel on a model contour is supported (matching a feature) and if p_{b1} is the probability of detecting an edge at a certain image pixel, and σ^2 is the variance of the model/codon displacements, reasonable values of K_1, K_2, K_3 are given by:

$$
\begin{aligned}
K_1 &\approx \log_2(p_{m1}) - \log_2(p_{b1}) \\
K_2 &\approx -(\ \log_2(1-p_{m1}) + log_2(1-p_{b1})\ ) \\
K_3 &\approx \frac{\log_2 \sigma + \frac{1}{2}\log_2 2\pi e}{\sigma^2}
\end{aligned}
$$

For instance, for the sensible values of $p_{m1} = 0.8$ and $p_{b1} = 0.05$ we obtain $K_1 = 4$ and $K_2 = 2.3$, which are amazingly close to what in the experiments indicated as an optimal combination. In the case of K_3 the experiments show that the above equation slightly overestimates the value found to be optimal in the experiments (with $\sigma = 1 \ldots 3$), probably because the residual distribution is not Gaussian.

The value of K_4 represent the number of bits necessary to represent the model parameters. A good range of K_4 has been experimentally found to be from 40 to 80.

4 Recovery of qualitative 3D structure

In the previous sections, it has been shown that qualitative 3D primitives like geons can be segmented out from real images by looking for their outline but the essence of their 3D structure (the *geon class*, according to [1]) is lost in the process. For instance, in the part segmentation of the handset of Figure 5-A, both pieces and handle hypotheses have a neat 3D structure which could not be recovered by the simple 2D models used in the part-based grouping.

This section outlines the method we used for fitting qualitative 3D volumetric parts models to real 2D images that treats geons[3] 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 sound Bayesian terms [20]. A few examples of geon PDCMs and fitting results can be seen in Figure 5-B. 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 [13] or the MDL method presented in the previous section. 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 works such as [26, 29, 13, 2]; only in [14] the method was extended to the 2D case as a front-end of the OPTICA system [6]. 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 and early work [20], we used a loosely-constrained optimisation approach which worked well only when the initial model was topologically equivalent to the geon instance being fitted. Later [18] this deficiency has been reduced by using an aspect-based hypothesis generation-and-testing strategy inspired by [7]. The multidimensional parameter space defining the geon PDCM is partitioned into eight topology-equivalent classes which have been called *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 [7] is employed that, by starting from coarse 2D part hypotheses produced as in the previous sections, 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.

The 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 [10].

More details about this section can be found in [18, 21].

[3] The parts are called geons here despite they are a subset of the ones defined in [1].

5 Experimental Results

This section presents some experiments that show the principled validity of the proposed approach. More detailed experiments, which include robustness analysis can be found in [18].

Figure 4 shows four experiments in which the original edge image, the initial set of part hypotheses and the final filtered set are given on the left, centre and right figures, respectively.

It can be seen that the initial set includes many poor hypotheses and multiple ambiguous interpretations of the edge data. In all the examples, the part-based grouping managed to produce a redundant set of part hypotheses that includes the actual ones and the MDL filtering method to finally produce the correct part segmentation: the surviving part hypotheses are the minimal set of models that most economically represent the edge image in the "language" of generic parts, right in the spirit of the MDL principle.

Notice that in the four experiments the same set of parameters K_1, K_2, K_3 and K_4 were used. In [18] many more experiments (not included here for reasons of space) are given that show that the method is fairly stable to variations in K_1, K_2, K_3 and K_4 but some problems, mainly due to the well-known figure-ground ambiguity, are reported.

Figure 5 shows an example of how the 3D structure of parts can be recovered by means of parametrically deformable aspects fitting as outlined in Section 4. Figure 5-A shows the initialisations of the PDA in terms of position, size and orientation; Figure 5-B shows the final fitting results and in Figure 5-C these results are rendered by deformable superquadrics; notice that the superquadrics are produced by using the same numerical values of the parameters as those that define the PDAs.

The results we achieved from 2D images are very much comparable with the one obtained by using 3D range data (e.g. by [26]), although depth and orientation cannot be obviously recovered from 2D images.

6 Acknowledgements

We whish to thank A.W. Fitzgibbon and David Eggert for useful discussions. Maurizio Pilu was partially sponsored by SGS-THOMSON Microelectronics.

References

1. I. Biederman. Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94:115–147, 1987.
2. D.L Borges. *Recognizing Three-Dimensional Objects using Parametrized Volumentric Models*. Unpublished PhD Thesis, Department of Artificial Intelligence, University of Edinburgh, May 1996.
3. Michael Brady. Seeds of perception. In *Proceeding of the ALVEY Conference*, pages 259–265, 1987.
4. T.F. Cootes and C.J. Taylor. Active shape models - 'smart snakes'. In *Proceedings of the British Machine Vision Conference*, pages 266–275, 1992.
5. 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.
6. S.J. Dickinson, A.P. Pentland, and A. Rosenfeld. 3-D Shape Recovery Using Distributed Aspect Matching. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 14(2):130–154, 1992.
7. 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.
8. D. Hoffman and W. Richards. Parts of recognition. In A. Pentland, editor, *From Pixels to Predicates*. Ablex, Norwood, NJ, 1985.
9. D. W. Jacobs. Robust and efficient detection of convex groups. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 18(1):23–37, January 1996.
10. S. Kirkpatrick, C.D. Gelatt, and M.P. Vecchi. Optimization by simulated annealing. *Science*, 220:671–680, 1983.
11. Y.G. Leclerc. Constructing simple stable description for image partitioning. *International Journal of Computer Vision*, 3:73–102, 1989.
12. 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.
13. A. Leonardis, F. Solina, and A. Macerl. A direct recovery of superquadric models in range images using recover-and-select paradigm. In *Proceedings of the European Conference on Computer Vision*, pages 309–318. Springer-Verlag, 1994.

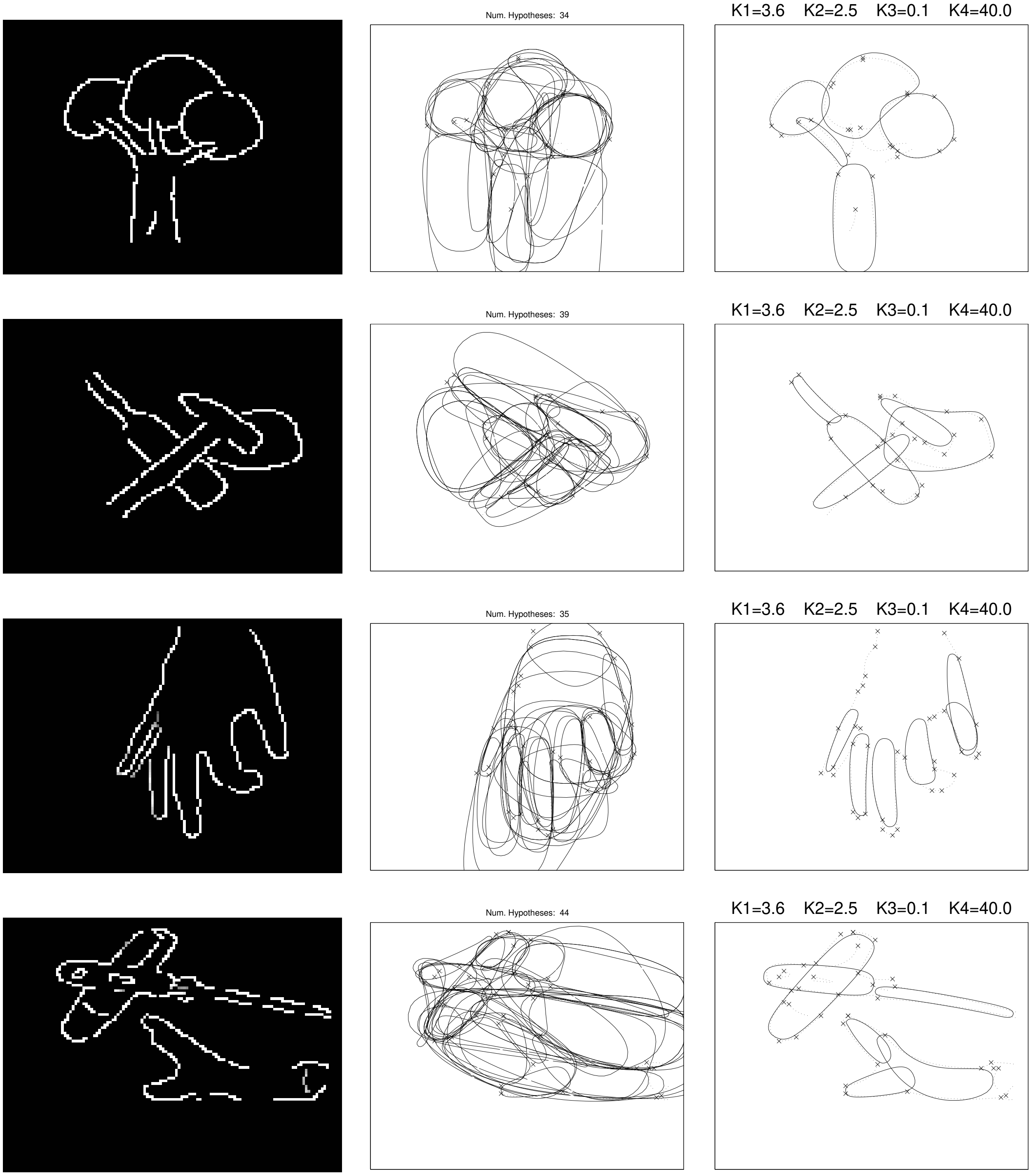

Fig. 4.: Four hypothesis filtering examples. Left: Original edge image; Centre: Redundant set of hypotheses; Right: Hypotheses selected by the MDL filtering method.

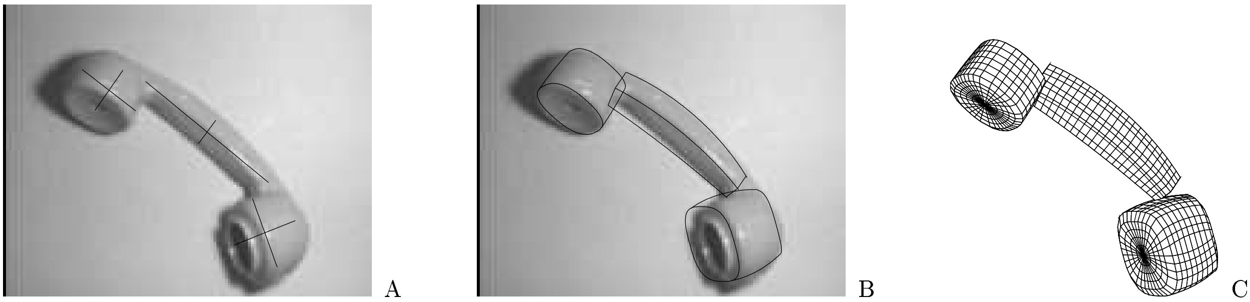

Fig. 5.: Real-image experiment with the aspect-based control strategy. Here, the PDA have been initialised automatically from some of the hypotheses produced by the part-based grouping and MDL filtering method. The figure shows initialisation (A), contour fits (B) and their volumetric representation (C).

14. 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.
15. R. Mohan and R. Nevatia. Perceptual organization for scene segmentation and description. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 14(6):616–635, June 1992.
16. E.P.D. Pednault. Some experiments in applying inductive inference principles to surface reconstruction. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 1603–1609, Detroit, MI, August 1989.
17. A.P. Pentland. Automatic extraction of deformable part models. *International Journal of Computer Vision*, (4):107–126, 1990.
18. M. Pilu. *Part-based Grouping and Recogntion: A Model-Guided Approach.* PhD Thesis, Department of Artificial Intelligence, University of Edinburgh, Scotland, 1996. Forthcoming.
19. M. Pilu and R.B. Fisher. Part segmentation from 2D edge images by the MDL criterion. In *Proceedings of the British Machine Vision Conference*, Edinburgh, September 1996. Submitted.
20. 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*, pages 71–82, Berlin, April 1996. Springer-Verlag.
21. M. Pilu and R.B. Fisher. Recovery of generic parts by parametric deformable aspects. Technical Report 801, Department of Artificial Intelligence, University of Edinburgh, May 1996.
22. M. Pilu, A.W. Fitzgibbon, and R.B. Fisher. Ellipse-specific least-square fitting. In *IEEE International Conference on Image Processing*, Lausanne, Switzerland, September 1996. To Appear.
23. M. Pilu, A.W. Fitzgibbon, and R.B. Fisher. Training PDM on models: The case of deformable superellipses. In *Proceedings of the British Machine Vision Conference*, Edinburgh, September 1996. Submitted.
24. U. Ramer. An iterative procedure for the polygonal approximation of plane curves. *Computer Graphics and Image Processing*, 1:244–256, 1972.
25. 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.
26. F. Solina and R. Bajcsy. Recovery of parametric models from range images: The case of superquadrics with global deformations. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 12(2):131–147, February 1990.
27. F. Stein and G. Medioni. Recognition of 3D objects from 2D groupings. In *Proceeding of the DARPA Image Understanding Workshop*, pages 26–29, 1992.
28. D.O. Wren. *Planning Dextrous Grasps from Range Data using Preshaping and Digit Trajectories.* Unpublished PhD Thesis, Department of Artificial Intelligence, University of Edinburgh, May 1996.
29. K. Wu and M.D. Levine. Recovering of parametric geons from multiview range data. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 159–166, Seattle, WA, 1994.

This article was processed using the TEX macro package with SIRS96 style