Part Segmentation from 2D Edge Images by the MDL Criterion

M. Pilu R.B. Fisher

Department of Artificial Intelligence
The University of Edinburgh
5 Forrest Hill, Edinburgh EH1 2QL
SCOTLAND

Abstract

In the context of part segmentation from 2D edge images, this paper presents some interesting results with a novel method that addresses the problem of filtering a redundant set of part hypotheses that retains only those that are likely to correspond to actual parts. In the proposed method, supporting evidence for hypotheses are put in competition in a Minimum Description-Length (MDL) framework to select part hypotheses that most economically represent supporting edges in the "language" of generic parts.

1 Introduction

The recovery of generic solid parts is a hard fundamental step towards the realization of general-purpose vision systems.

A wealth of previous works has been proposed in the past to perform part segmentations that, with a couple of exceptions, always used silhouette-like data, such as symmetry axes or skeleton techniques, contour-based techniques, and many more.

However, in a primitive-based framework suggested by Pentland [6] the procedure of producing the final interpretation of an image in terms of generic part models can be recast into a robust estimation framework where we "fit" each of those models to the right data regardless of extraneous disturbances originating from all sorts of noise, cluttering and right distinct elements in the image. For instance, in the noisy tree edge image of Figure 1-A, we would like to find the two models (dashed in the figure) representing the trunk and the foliage.

The key to many robust estimation methods is the notion of *support region*, which defines the data deemed to be originated by a single process, i.e. an object part in our case. Although many works make an implicit use of support regions, three roughly concomitant seminal works proposed the explicit use of support regions in computer vision, namely by Pednault [5], Leclerc [3] and Pentland [6]. The introduction of the concept of support regions in computer vision allows multiple processes to be naturally dealt with; support regions can also be disconnected and hence, in principle, occlusions could be handled in a rather natural way.

The unifying idea behind all these works is that a number of concurrent hypotheses are weighted against each other, and accepted or rejected in order to produce an "economic" representation of the image based on the Occam's razor (simplicity) criterion. For this purpose, they all make use of information theoretical arguments under the umbrella of the Minimum Description-Length (MDL) framework [9, 3].

Thus far, MDL-based methods have been very promisingly used in the context of surface segmentation, as in [1], and to achieve part segmentation from range data [4]; In both these two works, the MDL principle was used for selecting the best representation out of a large set of competing hypotheses produced by earlier fitting stages.

Following the same strategy, in this paper we present some interesting results with a novel method that addresses the problem of filtering a redundant set of part hypotheses from *single unsegmented edge images* that retains only the hypotheses that are likely to correspond to actual parts. In this context, the method is pushed to the limits in that it has to cope with incomplete data, coarse models and multiple objects.

Although good results have been obtained, some principled limitations have been discovered that have not been mentioned in previous MDL works and an account of them is given in Section 6.

2 Overview

The method presented in this paper is part of a larger work [7] addressing the grouping, segmentation and recognition of generic solid parts from two-dimensional edge images.

The computational approach for generating part hypotheses that is proposed in [7] is called *part-based grouping* and consists of four distinct stages, which are very succinctly summarised in the following.

In the first stage *codons*, contour portions of similar curvature, are extracted from the raw edge image; codons are considered as indivisible image features because they have the desirable property of belonging either to single parts or joints. In the second stage, small *seed groups* (currently pairs) of codons are selected that normally give enough structural information for a part hypotheses to be created. The third stage consists in initialising and *pre-shaping* the generic part models – purposely trained PDM as presented in [8] (in this proceedings) – to all the seed groups and then performing a full fitting to a large neighbourhood of the pre-shaped model. At this point we have a set of part models – such as the one in Fig. 1-B – which hopefully contains (possibly duplicated) actual parts hypotheses along with meaningless hypotheses.

This paper deals with the fourth, significant final stage of global filtering all the generated hypotheses in order to produce a correct part segmentation of the edge image. For doing so, supporting evidence of hypotheses are put in competition in a Minimum Description-Length framework to select part hypotheses that most economically represent supporting edges in the "language" of generic parts. The filtering is actually performed by the maximisation of a boolean quadratic objective

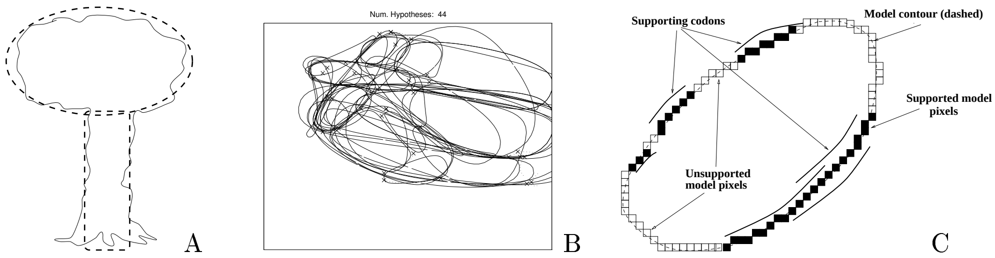

Figure 1: A: Example of segmentation by part primitives. B: Typical initial set of part hypotheses. C: Illustration of supporting codons and supported contour pixels of a generic part model.

function by a genetic algorithm; in the next section we describe the rationale, the implementation and the optimisation of this MDL-based cost function.

3 Filtering hypotheses by MDL of supports

The method presented in this section is inspired by a recently developed segmentation technique based on the Minimum Description-Length (MDL) criterion which is used in [4] and [1] to segment range data into 3D patches (albeit other applications are proposed); here, its basic principles are for the first time applied to the segmentation of geometric primitives from real unsegmented 2D edge images.

Let us first introduce the notation that is going to be used to describe the MDL based cost function whose maximisation yields an MDL description of the image edges in term of the generic part models.

$\mathcal{E}$: the edge image; $\mathcal{E}$ has the same form of the original image $\mathcal{I}$ and $(i,j) \in \mathcal{E}$ is 1 if an edge has been detected at $(i,j) \in \mathcal{I}$ and 0 otherwise;

$\mathcal{C}$: the set of N codons $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_N\}$, which are the indivisible entities by which the original edge image $\mathcal{E}$ is expressed at this stage; each C_i is a connected chain of edge points (i,j).

$\mathcal{B}$: the set of background (non-edge) pixels; it is $\mathcal{B} \subseteq \mathcal{E}$ and $\mathcal{E} = \mathcal{B} + \mathcal{C}$;

$\mathcal{H}$: the initial set of M generic primitive model hypotheses $\mathcal{H} = \{\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_M\}$ produced as in [7]; Fig. 1-B gives an example of a such a set;

$\mathcal{X}$: a set of h model hypotheses $\mathcal{X} \subseteq \mathcal{H}$

$\mathcal{R}_i$: the set of supporting codons $\mathcal{R}_i = \{\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_k\}$ of a model hypothesis $\mathcal{H}_i \in \mathcal{H}$. Supporting codons are found by thresholding a proper distance norm to the model contour (see Fig. 1-C);

$\mathcal{R}_\mathcal{X}$: the set of support regions $\mathcal{R} = \{\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_h\}$ of a set $\mathcal{X}$ of h model hypotheses $\mathcal{X} \subseteq \mathcal{H}$;

$\mathcal{B}_i$**:** the set of unsupported pixels covered by the contour of a model $\mathcal{H}_i$ (dashed portions of model contour as illustrated in Fig. 1-C;

$\mathcal{B}_\mathcal{X}$**:** the set of *unsupported* pixels that are covered by the contours of the set of hypotheses $\mathcal{X}$. It is $\mathcal{B}_\mathcal{X} = \bigcup_{\mathcal{H}_i \in \mathcal{X}} \mathcal{B}_i$;

$\mathcal{M}_i$**:** the set of supported pixels of the hypothesis $\mathcal{H}_i$ symbolically denoted by $\mathcal{M}_i = \mathcal{R}_i \dashv \mathcal{H}_i$[1] and illustrated in Fig. 1-C;

$\mathcal{M}_\mathcal{X}$**:** the set of *supported* pixels of the set of hypotheses $\mathcal{X} \subset \mathcal{H}$, that is $\mathcal{M}_\mathcal{X} = \mathcal{R} \dashv \mathcal{X}$ or, equivalently, $\mathcal{M}_\mathcal{X} = \bigcup_{\mathcal{H}_i \in \mathcal{X}} \mathcal{M}_i$;

$\mathcal{M}_{i,j}$**:** the set of pixels of the hypothesis $\mathcal{H}_i$ (or equivalently $\mathcal{H}_j$) that are supported by the shared codons, that is $\mathcal{M}_{i,j} = (\mathcal{R}_i \cap \mathcal{R}_j) \dashv \mathcal{H}_i$;

$\chi^2(\mathcal{M}_i, \mathcal{C}_j)$**:** the error of fit function which expresses the displacement between the supported pixels $\mathcal{M}_i$ of a hypothesis $\mathcal{H}_i$ and one of its supporting codons $\mathcal{C}_j \in \mathcal{R}_i$. By indicating with $d(h_k, \mathcal{C}_i)$ the geometric distance of a model pixel $h_k \in \mathcal{M}_i$, to a codon $\mathcal{C}_j$, the error of fit function is defined as $\chi^2(\mathcal{M}_i, \mathcal{C}_j) = \sum_{h_k \in \mathcal{M}_i} d(h_k, \mathcal{C}_j)^2$;

$\chi^2(\mathcal{M}_\mathcal{X}, \mathcal{R}_\mathcal{X})$**:** the error of fit function which expresses the displacement between the supported pixels $\mathcal{M}_\mathcal{X}$ of set of models $\mathcal{X} \subset \mathcal{H}$ and its supporting codons $\mathcal{R}_\mathcal{X}$. $\chi^2(\mathcal{M}_\mathcal{X}, \mathcal{R}_\mathcal{X}) = \sum_{\mathcal{M}_i \in \mathcal{M}_\mathcal{X}} \sum_{\mathcal{C}_j \in \mathcal{R}_i} \chi^2(\mathcal{M}_i, \mathcal{C}_j)$.

Let us now explicitly formulate the problem in MDL terms.

Let us indicate by $L(\cdot)$ a generic function that gives the number of bits needed to represent a certain entity. Since the edge image $\mathcal{E}$ can be decomposed into two distinct elements, namely the background and the codons, the number of bits needed to represent it can be written as:

$$L(\mathcal{E}) = L(\mathcal{C}) + L(\mathcal{B})$$

When we interpret part of the edge image $\mathcal{E}$ by a set of models $\mathcal{X}$, the encoding length changes to that indicated by $L(\mathcal{E}|\mathcal{X})$:

$$L(\mathcal{E}|\mathcal{X}) = L(\mathcal{E}) - L(\mathcal{M}_\mathcal{X}) + L(\mathcal{B}_\mathcal{X}) + L(\chi^2(\mathcal{M}_\mathcal{X}, \mathcal{R}_\mathcal{X})) + L(\mathcal{X}) \tag{1}$$

where $-L(\mathcal{M}_\mathcal{X})$ is a negative term representing the saving due to support regions being now described by supported portions $\mathcal{M}_\mathcal{X}$ of the set of models $\mathcal{X}$, $L(\mathcal{B}_\mathcal{X})$ represents the additional cost of having to encode unsupported background portions of the contour model, $L(\chi^2(\mathcal{M}_\mathcal{X}, \mathcal{R}_\mathcal{X}))$ expresses the additional number of bits needed to encode the displacement between support regions and supported model contours and, finally, $L(\mathcal{X})$ (called model overhead) expresses the additional cost of having to encode the parameters of the models.

In the MDL framework, the minimal subset of models $\hat{\mathcal{H}}$ of $\mathcal{H}$ that most economically represents the image is given by:

[1] The *symbolic* notation $y \dashv x$ introduced here indicates the elements of y that are related or represented by x. In model matching, for instance, x could be a model and y image features and $y \dashv x$ indicates the set of image features y that match model x. Mnemonically, it can be interpreted as "projection".

$$\hat{\mathcal{H}} = \arg\min_{\mathcal{X}\subseteq\mathcal{H}} \{L(\mathcal{E}|\mathcal{X})\}$$

By using the definition given in Eqn. (1) and by noticing that the term $L(\mathcal{E})$ is constant, the above minimisation becomes:

$$\hat{\mathcal{H}} = \arg\max_{\mathcal{X}\subset\mathcal{H}} \left\{L(\mathcal{M}_{\mathcal{X}}) - L(\mathcal{B}_{\mathcal{X}}) - L(\chi^2(\mathcal{X},\mathcal{R}_{\mathcal{X}})) - L(\mathcal{X})\right\} \tag{2}$$

The maximiser expression in braces, which we call $S(\mathcal{E}|\mathcal{X})$, is normally termed as "bit saving", because in fact it represents the decrease of encoding length due to the use of models.

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 encode each *supported* pixel of a model contour, K_2 is the average number of bits necessary to encode each *unsupported* pixel of a model contour, K_3 is a constant such that when multiplied by $\chi^2(\mathcal{X},\mathcal{C})$ gives the average encoding length for representing the residuals and, finally, K_4 is the average number of bits for specifying the parameters of a model.

Then, following the philosophy of [4], we can rewrite the bit saving $S(\mathcal{E}|\mathcal{X})$ as follows:

$$S(\mathcal{E}|\mathcal{X}) = K_1\cdot|\mathcal{M}_{\mathcal{X}}| - K_2\cdot|\mathcal{B}_{\mathcal{X}}| - K_3\cdot\chi^2(\mathcal{M}_{\mathcal{X}},\mathcal{R}_{\mathcal{X}}) - \sum_{\mathcal{H}_i\in\mathcal{X}} K_4 \tag{3}$$

where $|\cdot|$ indicates the number of image pixels represented by $\mathcal{M}_{\mathcal{X}}$ and $\mathcal{B}_{\mathcal{X}}$, respectively. It is fundamental that savings due to supports and overheads caused by the residuals are not accounted for more than once when portions of contour are shared by the same models in the final description [6].

The inclusion of the term $K_2\cdot|\mathcal{B}_{\mathcal{X}}|$ gives favour to models which have higher contour covering and constitute a fundamental variation with respect to the MDL cost functions used in [1] and [4]. Without this term, models could be selected regardless of the amount of unsupported contour portions.

A practical method for the minimisation of a similar cost function was proposed by Pentland [6] and here we follow his footsteps.

If we presume that in the *final* solution the only kind of model overlapping taken into account is pairwise[2] [6], the maximisation in Eqn. (2) can be operatively achieved by transforming Eqn. (3) into a more compact matrix form, which is derived from [4]. This pairwise overlapping assumption is a fairly sensible choice that helps keep the computational cost down, eases optimisation and is also justified by the fact that three or more parts are very seldom jointly together in the same region.

Under this assumption, the maximisation can be rewritten as:

$$\hat{\mathbf{m}} = \arg\max_{\mathbf{m}} \left\{\mathbf{m}^T\mathbf{Q}\mathbf{m}\right\} \tag{4}$$

where $\mathbf{Q}$ is the hypotheses correlation matrix, which will be defined next, and $\mathbf{m} = [m_1\ m_2\ \cdots\ m_M]^T$ is the hypotheses presence vector in which each element

[2] Differently from [6], overlapping here refers to sharing codons.

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:

$$q_{i,i} = K_1|\mathcal{M}_i| - K_2|\mathcal{B}_i| - K_3\chi^2(\mathcal{M}_i, \mathcal{R}_i) - K_4; \tag{5}$$

An off-diagonal element $q_{i,j}$ 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:

$$q_{j,i} = q_{i,j} = \frac{1}{2}\left\{-K_1 \cdot |\mathcal{M}_{i,j}| \;+\; K_3 \cdot \chi^2(\mathcal{M}_{i,j}, \mathcal{R}_i \cap \mathcal{R}_j)\right\} \tag{6}$$

Intuitively, 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 (4) is, technically speaking, a *Quadratic Boolean Optimisation Problem*, as the solution space can be represented as corners of an M-dimensional hypercube. In [4], [1] and [6] this optimisation problem was tackled by using different greedy strategies, which we have found unsuitable to our minimisation because of 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. More details on the optimisation stage can be found in [7].

4 On the determination of the constants

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

In [7] 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 the experiments (Sec. 5) indicated as an optimal combination. In the case of K_3, the experiments show that the above equation slightly over estimates the value found to be optimal in the experiments ($\sigma = 1 \ldots 3$), probably because the model/codon noise distribution is not Gaussian.

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

5 Experimental Results

Figure 2 shown four examples of the MDL filtering method using the same set of parameters K_1, K_2, K_3 and K_4.

In each experiment, the original edge image, the initial set of part hypotheses and the final filtered set are shown on the left, centre and right figures, respectively.

It can be seen that the initial sets includes many poor hypotheses and multiple ambiguous interpretations of the edge data. In all the examples, the MDL method proposed here managed to filter down this redundant set and yield the correct part segmentation: the surviving part hypotheses are the minimal set of part models that most economically represent the edge image in the "language" of generic parts, in the very spirit of the MDL principle.

In [7], 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 have been identified and are reported in the following section.

6 Discussion

From recent literature it appeared that the MDL hypothesis competition method would be able to handle in principle several middle-level segmentation problems.

However, some problems have surfaced from experiments with different values of the constants K_1, K_2, K_3 and K_4 (as extensively discussed [7]), and in our opinion the considerations that follow bear a certain significance. Most of these problems can be attributed to the well known figure-ground ambiguity in edge images but some issues are more specific to the MDL method.

Figure 3 shows how, within the proposed framework, three parallel lines can give rise to an ambiguity in terms of their interpretation by part models. This situation arises often in images with multiple objects and it is even accentuated when parts of the lines are missing because of occlusion or poor edge detection.

In [7] it is shown that by changing the constants K_1, K_2 and K_4, the highest values of the bit saving S could be obtained in diverse situations. Cases B and C are theoretically equivalent [7] for the ideal situation shown in Fig. 3. Case D is slightly favoured over B and C (less non-supported contour portions) but unbalanced lengths and some cluttering could arbitrarily favour either. Case A could be favoured if the length of the segments is small and therefore the saving due to the description of the edge by the models cannot compensate for the model overhead itself (e.g., see the final segment of the index in the hand test image). Case E is very unlikely to happen because is too costly to encode.

Although the MDL method is in principle very stable, situations like the above can lead to some instabilities, as shown in [7]; ambiguities and instability of this kind have not been previously noted in related literature. The tuning the parameters on a simple test example in order to favour a particular solution would clearly be useless because in real conditions even a small percentage of missing contour would turn the balance towards alternative solutions.

Finally, there is the problem of scale. Although it has been advocated elsewhere [5, 4] that one of the main advantages of MDL is its scale-independence, practically we have found that bigger model hypotheses are slightly favoured over

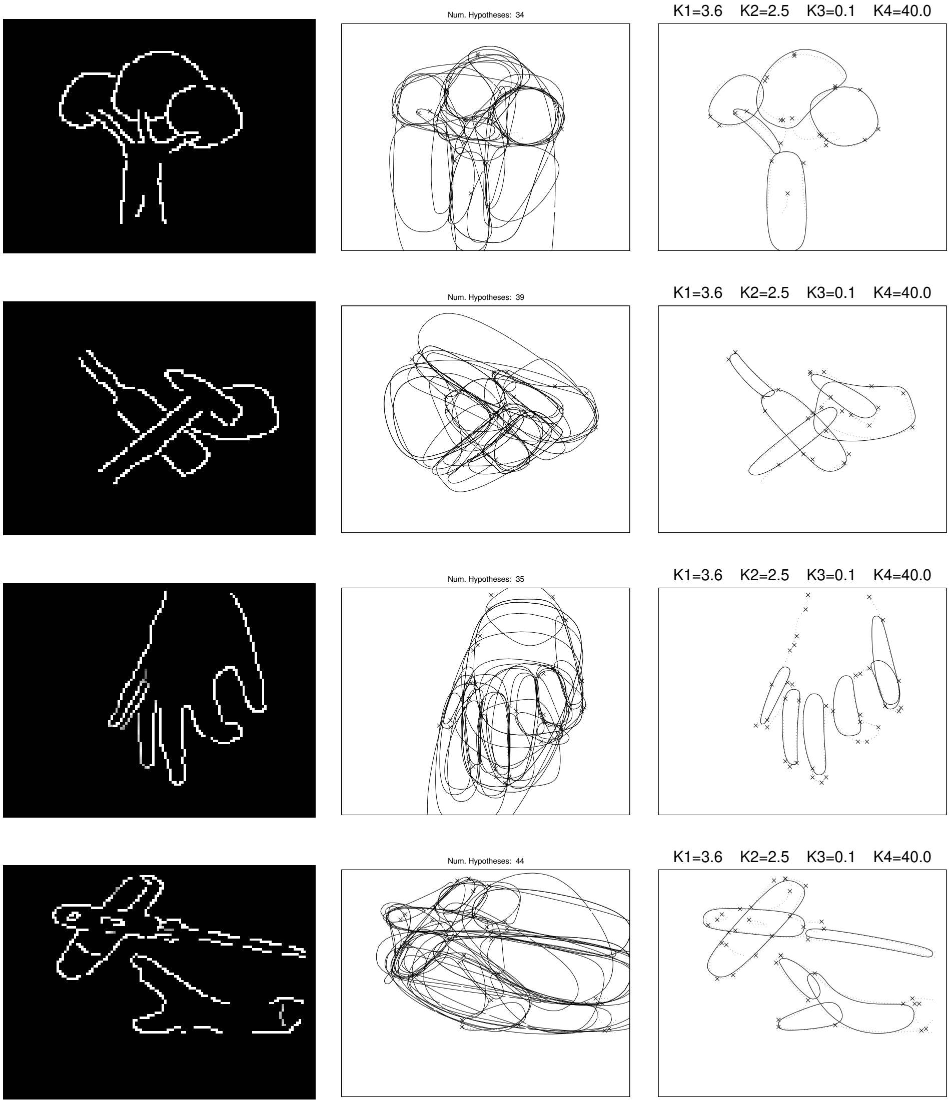

Figure 2: Four hypothesis filtering examples. Left: original edge image; Centre: redundant set of hypotheses; Right: selected hypotheses. The top two examples are from synthetic image representing a tree and a composite image of a beer bottle, a hammer and a round object beneath. The two other examples are from of a real grey-level image of a hand (128x128) and a screwdriver, marker and wooden stick (256x256) processed by an implementation of a Canny edge detector. See text for more details.

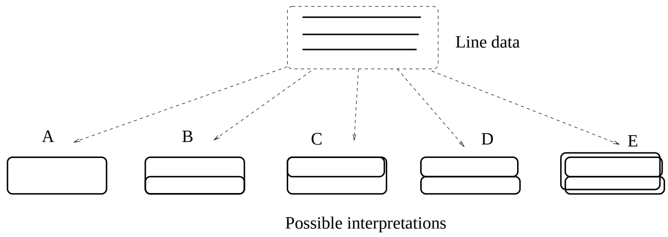

Figure 3: Taxonomy of possible MDL filtering results for the case of three parallel lines. A: only the bigger hypothesis is selected, which normally corresponds to the actual part outline. B, C: the bigger one plus either of the small ones are selected. D: both small hypotheses are selected. E: all three hypotheses are selected. Ambiguities often arise for cases B, C and D.

small ones (such as the last segment of the index finger in the hand example) because they lead to higher savings in bits. These scale problems have actually been considered assets in works such as [4] and [1] because bigger models were supposed to more economically describe the surface being segmented under their Gaussian-noise assumption. A simplistic solution could be to tie K_4 to the model scale but this not only would contradict one of the main assumptions of the MDL method but has no theoretical support; further work has to be done in this regard.

7 Conclusion and future work

In this paper, a principled approach is described that allows an initial set of part hypotheses produced from ordinary edge images to be filtered down to a few ones that have a high likelihood of corresponding to actual parts of objects. The method is inspired by recent work in surface segmentation by the Minimum Description-Length principle [4, 1].

A number of relevant contributions can be identified. First above all, the segmentation strategy proposed in [4] and [1] (that was shown to produce good results in 3D surface segmentation) is for the first time used here to perform part segmentation from ordinary unsegmented edge images. Together with the part-based grouping method in [7], the filtering method proposed here constitutes one of the handful of methods for part segmentation from unsegmented edge images to be found in the literature. Another contribution regards the use of a genetic algorithm to maximise the quadratic boolean cost function in Eqn. (4). In other works, this maximization was performed by greedy methods, which have been found impractical when the quality of the initial hypotheses is poor. Finally, some theoretical aspects of the MDL method have been investigated and discussed, e.g. the problem of ambiguities described in Sec. 6 which had not been pointed out in previous work. These problems are inherent to the use of sole edge information but could also arise whenever the data is incomplete, cluttered or the fitting residuals are simply too high and no noise model is available.

However, the work could improved upon in several aspects. Codons have been used as atomic entities both for generating part hypotheses and for finding additional image support, expressed in fact in terms of codons. This assumption, although yielding good results, is rather simplistic and perhaps image support should be directly found in the raw edge or gradient image, which would ease some problems in determining support regions. Another significant improvement could come from the integration of more information (such as regions, colour and depth) under the MDL framework. Thus far, no work has been done on merging different kinds of information under an MDL framework but this could be just incidental, since a clear derivation could be made from a Bayesian frame of mind. Such an integration could resolve some ambiguities, as shown for a simple case in [7]. Finally, an intriguing possibility would be to explore the use of a *multi-population* genetic algorithm in which equally high-scoring alternative solutions could be left to evolve, up to migrations, in parallel; a similar possibility was also investigated in [2].

Acknowledgements: Maurizio Pilu was partially sponsored by SGS-THOMSON Microelectronics. The authors whish to thank A.W. Fitzgibbon for useful discussions and for providing an genetic algorithm in-house implementation.

References

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

[2] A. Hill and C. Taylor. Model-based image interpretation using genetic algorithms. *Image and Vision Computing*, 10(5):295–300, June 1992.

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

[4] 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.

[5] E. 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, Aug. 1989.

[6] A. Pentland. Automatic extraction of deformable part models. *International Journal of Computer Vision*, (4):107–126, 1990.

[7] M. Pilu. *Part-based Grouping and Recogntion: A Model-Guided Approach.* PhD Thesis, Department of Artificial Intelligence, University of Edinburgh, Scotland, 1996. Forthcoming.

[8] M. Pilu, A. Fitzgibbon, and R. Fisher. Training PDM on models: The case of deformable superellipses. In *Proceedings of the British Machine Vision Conference*, Edinburgh, Sept. 1996. This volume.

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