MDL based Structural Interpretation of Images under Partial Occlusion

Sowmya Ramakrishnan and Peter Forte
School of Computer Science and Information Systems
Kingston University
Surrey, KT1 2EE, UK
sowmya.r@computer.org
pforte@kingston.ac.uk

Abstract

In this paper an information theoretic approach is provided for resolving border ambiguity under partial occlusion. The proposed framework allows structural interpretation of images prior to the application of domain specific knowledge. The central idea behind MDL based figure-ground grouping is merging those place-token candidates whose composed description length (whole) is better than sum of their individual description lengths (parts). The computational theory is illustrated by application to blocks-world images.

1 Introduction

The human visual system is able to attach depth perception to objects even in the absence of stereo data, making good decisions to distinguish foreground (figure) from background (ground). In addition particular shape completions can be predicted for occluded contours. However the majority of computer vision systems today lack this level of generality. Traditionally such systems have relied on use of domain specific knowledge and models in order to derive structural interpretation of images. In contrast, human visual perception is able to infer structure from the image without prior knowledge of the scene [20] [18] [23].

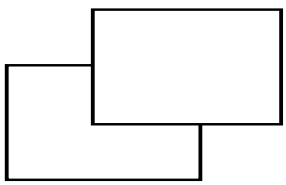


Figure 1: Simple line Drawing illustrating perception of relative depth under partial occlusion

Figure 1 shows an example of a simple 2D line drawing which a typical observer would interpret as 'a rectangle partially occluded by another rectangle' despite the fact

BMVC 2001 doi:10.5244/C.15.57

that other consistent figure-ground interpretations are possible in principle. Three of these are shown in figure 2(a),(b) and (c). All interpretations postulate two objects but with different assumptions about which is figure and which is ground. The human observer prefers interpretation (a) with a rectangular hidden contour rather than (b) or (c). This accords with Gestalt principles [13] and the tendency to avoid accidental configurations [18]. In the absence of such principles the scene in figure 1 is subject to unresolvable figure-ground ambiguity.

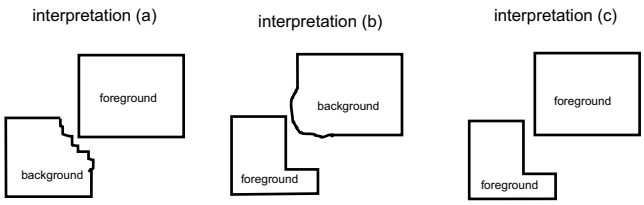


Figure 2: A few possible interpretations of figure 1

Figure 3 illustrates additional problems that arise in image data using an example from the blocks world. Once straight line edges have been segmented out from the image there exists difficulty in identifying which lines should be grouped to form major figural areas of the image. Furthermore, in the absence of domain knowledge, partitioning the figural set into groups to identify a specific foreground structure from its background is a challenging task because occlusion, illumination conditions, reflections, shadows etc., can cause image contours to get fragmented and displaced. In figure 3 it can be observed that the partitioning is ambiguous at the occlusion boundary between the blocks. Traditional cues like T-junctions cannot be used straight away to identify the foreground until gaps are perceptually closed. Recent figure-ground discrimination systems like [2] [19] perform well in terms of removing texture edgels and preserving edgels corresponding to smooth boundaries of large groups but the grouping is often imperfect at the point of occlusion. This example is analysed in section 4.

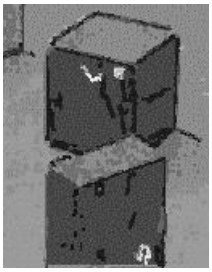


Figure 3: Image contours fragmented by partial occlusion resulting in border ambiguity

Often the figure-ground problem can be formulated as a 'border ownership' problem [21] (as in the Rubin's vase example, figure 4) and is inter-related with determination of salient regions. In general figure/ground segregation involves two aspects:

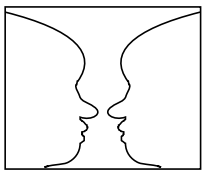


Figure 4: Rubin's vase: figure ground ambiguity

1. deciding to which region a boundary curve belongs (border ownership ambiguity),
2. deciding how the ground continues behind the figure (amodal completion).

It is the purpose of this paper to present a computational framework to resolve border ownership ambiguity based on the Minimum Description Length (MDL) principle.

2 Related past work

Although the initial contribution for perceptual analysis of figure/ground organization came from psychology [13] [12] [5], defining a computational basis for visual perception has become a computer vision research problem [20] [18] [23] [14].

Some recent 2D figure-ground analyses [8] [21] model the shape figure boundary as a source of heat diffusing inwardly. Smoothness over a pixel neighbourhood is the major constraint imposed to constrain the hypotheses to fit the diffusion field. Bayesian modelling [8] and a leaking energy model coupled with Markov random walks [21] are some strategies used to minimize an entropy criterion while fitting the diffusion field to local hypotheses.

The detection of salient image point sets through a saliency measure is closely related to figure-ground discrimination [2] [19] [7]. Each saliency measure is a function of a set of affinity values assigned to pairs of edges and all affinity measures incorporate Gestalt principles of good continuation and proximity in some form or other, e.g. cocircularity [9], closure [19]. Jacobs' [11] measure detects salient convex groups while Mahmud, Williams and Thornber [19] detect closed salient groups by first applying a local affinity measure proportional to the likelihood that a smooth contour passes through the given edges. This affinity measure is then used to compute a global saliency measure proportional to the relative number of closed contours that join a pair of edges.

In the intermediate level of organization, hypotheses of segments, arcs and points of interest are generated from salient chains and grouped according to parallelism, proximity, collinearity, continuity etc., to form graphs of geometric primitives [1] [18]. Smooth continuation is an important factor that promotes figure-ground separation and optimal discovery of object-like components in the scene corresponding to structurally salient parts of the image. This factor has been interpreted in terms of minimum energy curves in the context of surface organization [4], generic positioning or non-accidental alignment [22], contour smoothness under occlusion of opaque surfaces [22], and low curvature between edgels [2].

The role of convexity has been highlighted in determining border-ownership and figure-ground organization both in psychology [12] and in computer vision [21] [11] [4].

In real images contours are often fragmented by occlusion, shadows and low reflectance contrast. Contour closure is a strong indicator of sets of fragments that could have possibly originated from a common object [6].

Recently there has been renewed interest in discovering image structure in advance of applying domain knowledge in order to supply better structural descriptions going into recognition stage. Boundary based region segmentation incorporating region intensity and geometry [10], merging of color-based segmentation with depth information [3], cluster analysis to obtain coherent groups [7] are modern approaches in grouping / segmentation. However the problem of figure-ground grouping to disambiguate occlusion boundaries at intermediate level has not been directly addressed by any of the above works.

3 Simplicity and Structure

3.1 The MDL Principle

The Gestalt School [13] tried to express the intuitive notion of simplicity conveyed by Occam's razor through the principle of Prägnanz. Since the Gestalt psychologists, a number of researchers have attempted to define visual perception in terms of simplicity of description [5]. Leclerc emphasized the idea that simplicity of description is an important guiding principle in vision and used minimal-length-encoding to formally express this notion to perform image segmentation [14]. The MDL (Minimum Description Length) principle has since been applied in computer vision research for a wide variety of problems like image partitioning [16], edge [17] and part segmentation, data clustering, object recognition. The MDL principle [15] states that "the best theory to explain a set of data is the one which minimizes the sum of

- the length, in bits, of the description of the theory; and
- the length, in bits, of data when encoded with the help of the theory."

Following Li's notation [16] let ϕ = (ϕ_1, ϕ_2,..., ϕ_k) be a parameter vector of hypothesis or model ϕ with k components. Let $p(x/\phi)$ represent a parametrized class of probability functions that assigns a probability to any observation x=(x_1, x_2,..., x_n). Given a hypothesis space Φ, we want to select the hypothesis ϕ such that the length of the shortest encoding of data x together with hypothesis ϕ is minimal. This leads to the minimization of

$$L'(x,\phi) = L(x/\phi) + L(\phi) = -log_2 p(x/\phi) + L(\phi) \quad (1)$$

where $L(\phi)$ is a measure of the information contents in the model parameters. Any bit saving obtained by encoding the data with the model ϕ (i.e. decrease of data encoding length due to the use of the model) is to be traded off against the additional cost $L(\phi)$ of encoding the model parameters (called model overhead). Errors in fitting incur a further cost because they have to be additionally encoded. Any data outlying the model increases the bit cost as it requires to be separately encoded. Based on the principles of algorithmic information theory [15] the length of a compact description (MDL) of the data is taken as a valid objective measure for simplicity.

3.2 Computational theory

Following Marr [20], the visual perception process may be modelled as consisting of three tasks - selection, grouping and discrimination of tokens extracted from the image. These processes are applied at different levels of abstraction to obtain tokens of a higher level. In order to obtain a satisfactory grouping organization, the grouping process should progressively be able to identify the regularity relationships in the input data and accordingly package the input data so that only the package unit is subsequently visible to the ensuing grouping processes. When repeated at several levels, the grouping process should produce structures that increasingly correspond to object components present in the scene that originally created the image [18]. At the highest level, interpreting image structure involves joining or severing contour fragments to determine figure-ground arrangements of structurally salient parts of the image [23]. Saund [22] conceptualizes events as 'non-accidental' when grouped configurations could be described with fewer relative parameters than in the generic ungrouped case. Each new package created by the grouping process requires fewer relative parameters for description than the original input set.

The objective of a figure-ground segregation module is to be able to localize figural regions in the image with potential object-hood. In order to produce a meaningful semantic representation, which can be used by higher level reasoning processes, the primary task is to form sensible interpretations that capture overall image structure. A basic premise underlying most of past works interpreting image structure is that not only more generic and least accidental figure-ground interpretations agree well with human visual perception results [18] but also such interpretations are invariably ones of minimal cost [23].

Hypothesizing after Leclerc that the grouping process intends to combine place-tokens such that progressively simpler descriptions (models) are obtained, place-tokens (α_1, α_2,..., α_k) form a valid group Λ if their unified description length is less than the aggregate sum of their individual description lengths.

$$L'(\Lambda, \Phi) < L(\alpha_1, \phi_1) + L(\alpha_2, \phi_2) + \ldots + L(\alpha_k, \phi_k) \tag{2}$$

The above inequality expresses the idea that the group as a whole is a better explanation of the given data than ungrouped individual place-tokens by providing a simpler description. Thus we consider *figure-ground grouping as a bottom-up process that combines image place-tokens to achieve the simplest description*. The Gestalt concepts of proximity and continuity play a key role in restoring *configurational wholes* from *parts*. The comparison of the minimum description lengths of spatial models to establish that the information content of 'whole' is more compact than that of aggregation of 'parts' puts this intuitive notion on a firm quantitative footing.

Therefore the central idea behind grouping based on MDL is merging those place-token candidates whose composed description length is smaller than sum of their individual description lengths. A similar view on grouping has been taken by [17] and [7] but not applied towards resolving border ambiguity. Applying MDL to figure-ground grouping makes explicit the tradeoff between how structured the image is to how complex (succinct) its description. In general, figure-ground organization schemes construct and regularize an objective function by penalizing any deviation from desired smoothness or genericity based on local evidence subject to constraints identified at the required level of abstraction [2] [8] [19] [22]. MDL is a better regularization metric than choices of ad

hoc cost functions and more suitable to figure-ground problem because the trade off helps identify high-level structures in the image with simpler descriptions.

4 MDL based Structural Interpretation model

We now describe our approach to this problem and the results obtained from our implementation.

Our model consists of 4 stages. In the first stage boundary contour segments are recovered from proximal points based on Li's MDL based line/ curve fitting [16]. The basic MDL computation is performed as in equation (4) [given in Appendix] for each line in the structural hypothesis fitting a number of data points to the line. In the second stage collinear lines are grouped. The composed MDL is computed for collinear and non-collinear cases. Depending on the minimum composed MDL a collinear or non-collinear model is selected [17]. The third stage consists of initialising closed groups, which compete for figural status. The first step in this stage is checking for closure. For a given set of proximal segments with gaps, the closure property is validated using the measure of closure formulated by Elder and Zucker [6]. The next step is to close the gaps. As suggested by [1] elementary junctions could be detected from straight-line segments by finding 'actual' and 'virtual' intersections. To allow merging of fragmented edges, small gaps are filled in if a composed model gives a lower MDL as recommended in [17]. The final stage consists of evaluating the MDL of figure-ground hypotheses supported by local structural combinations put in mutual competition at the ambiguous border. A lower MDL value (measured in bits) indicates a preferred interpretation. Results for simple line drawings and image data are shown for ϵ = 1.0, δ= 0.0001. Figures 5 to 8 show the line drawings, images used in experiments and their structural interpretations subjected to MDL selection. Since offset points and outliers are an integral part of the MDL framework stable fitting is observed even without perfect line data. Although results have been shown for straight line segments the method is readily extendible to curvilinear segments. The description lengths computed for various interpretations for each figure are listed in Table 1. It can be observed that in each case the interpretation chosen as the winner by the MDL selection module is perceptually plausible. Apart from resolving border ambiguity (fig 5 and 8) the method also helps evaluate structural interpretations for choosing better groups and figural completions (fig 6 and 7).

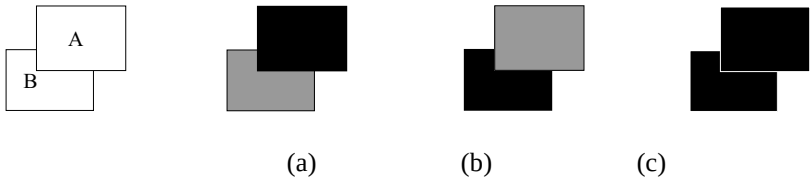
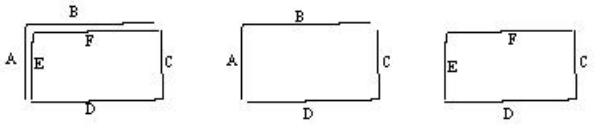


Figure 5: Structural interpretations for figure 1

For the line drawing example in figure 5 the ambiguous border may belong to either region A or region B. Which region is the foreground and which one the background depends on the ownership of the ambiguous border. Three structural interpretations figure 5(a), 5(b) and (c) are shown. In 5(a) region A owns the ambiguous border and is the foreground. In 5(b) region B owns the ambiguous border and is the foreground. In 5(c)

both A and B are placed coplanar and both of them are in the foreground. The MDL bit-value computed for these 3 interpretations are shown in Table 1. It can be observed that the MDL value for interpretation 5(a) is the least. This shows that our method picks out an interpretation that coincides with a perceptually plausible choice.

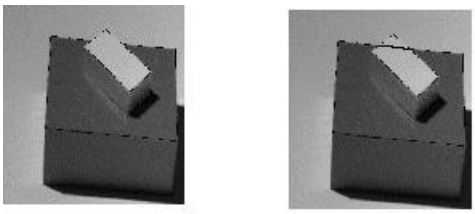
In the example shown in figure 6 it is illustrated how our MDL selection method can also successfully choose between ambiguous structural groups. A situation is shown where segments A and B compete with segments E and F to be grouped with segments C and D. The groups ABCD and EFCD are shown in figure 6(a) and 6(b) respectively. The MDL values computed for each of these grouping interpretations are given in Table 1. It can be seen that EFCD is the winner because it has a lower description length.



(a) (b)

Figure 6: Ambiguity in forming structural groups

In figure 7 structural interpretations are superimposed on a portion of an image to illustrate how our method can decide the suitability of figural completions. Figure 7(a) shows the outlines of the bigger and smaller blocks with figural completion for the bigger block undecided. In figure 7(b) the missing part of the background is completed and so the bigger block has a complete outline. The two structural interpretations are evaluated using MDL with the results being shown in Table 1. It can be seen that the interpretation 7(b) has a much lower description length compared to 7(a) and hence the winner. Here a natural choice of figural completion has been used. The true advantage of the method is that a selection can be made when a number of figural completion curves are postulated.



(a) (b)

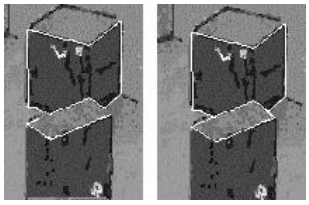
Figure 7: Structural interpretations superimposed on cropped portion of grayscale image (copyright of image: INRIA-Syntim image database)

In figure 8 two different interpretations of the image example given in figure 4 are shown. In interpretation 8(a) the lower block is the background while upper block is the foreground. The ambiguous border belongs to the upper block in this interpretation. In interpretation 8(b) the lower block is the foreground occluding the upper block. The ambiguous border belongs to the lower block in this interpretation. The description lengths

figure number	MDL in bits Interpretation (a)	MDL in bits Interpretation (b)	MDL in bits Interpretation (c)	winning interpretation
5	340.0	425.3	449.6	(a)
6	206.6	152.8		(b)
7	915.4	564.7		(b)
8	684.2	285.4		(b)

Table 1: MDL results and choice of winning interpretation

of these two interpretations have been shown in Table 1. It can be seen that interpretation 8(b) has a lower MDL compared to interpretation 8(a). Interpretation 8(b), the winner by our selection procedure, coincides with the perceptually reasonable choice.



(a) (b)

Figure 8: Interpretations superimposed on blocks image of figure 3

5 Conclusion

An approach towards resolving border ambiguity in images has been proposed through novel application of the MDL principle. The approach processes geometric information in advance of applying domain-specific knowledge. It also helps compare and select grouping configurations.

Inferring the structure of the physical domain and reconstruction of a larger fraction of the environment from image data is a crucial step for any bottom-up vision model attempting to carry out complex tasks such as learning and recognition, especially with the availability of very limited domain knowledge. The model proposed here is a step in this direction. However, as Marr [20] recognized, the intricacy of figure-ground segmentation arises from several subproblems and may not be derivable from a single underlying theory. Minimizing descriptional complexity may only be a single aspect of this multi-faceted problem, which still remains a challenge for the computer vision research community to explore further.

6 Appendix

This section gives details of Li's MDL based 2D shape description [16] [17] which is employed as the descriptive language by the structural interpretation model described in

this paper. Equations and notations have been reproduced as in [16] [17]. Li's 2D shape description uses two-part MDL coding consisting of the fitting model part and the outlier part, i.e., $\phi = [\phi_0(\lambda) + \phi_m(n - \lambda)]$ where ϕ_m is the model (line or ellipse) to be fitted to n data points, ϕ_0 is the outlier part and λ is the number of outliers. Assuming that a point is randomly located in the given range R_x x R_y, for a coordinate x or y having a uniform distribution in its range the description length of the variable is $lb(R_x/\epsilon)$ or $lb(R_y/\epsilon)$. The description coding length of a point for M x M image is

$$L_{pt}(R; x, y) = 2 * log_2 \frac{M}{\epsilon} \tag{3}$$

The total description length of data x using line segment model ϕ_{ls} is

$$L'(x, \phi_{ls}) = (L_{\sharp points} + L_{lineparameters} + L_{modelpoints} + L_{outliers})(x, \phi_{ls}) \tag{4}$$

in which

$$L_{lineparameters}(x, \phi) = 4 * log_2 \frac{M}{\epsilon} \tag{5}$$

for describing the two end points of a straight line segment,

$$L_{modelpoints}(x, \phi) = (N - \lambda) L_{offset}(\xi, \sigma) \tag{6}$$

with

$$L_{offset}(\xi, \sigma) = \frac{log_2 e}{2} . \frac{\xi^2}{\sigma^2} + log_2 \frac{\sigma}{\epsilon} + \frac{log_2(2\pi)}{2} \tag{7}$$

where $L_{offset}(\xi, \mu, \sigma)$ specifies that the coding length of the offset error of a data point with respect to the line is approximated with Gaussian distribution. This coding length is approximated by the expected coding length of the outcome of a centered Gaussian distribution $\xi \sim N(\mu = 0, \sigma^2)$ quantized with resolution ϵ and

$$L_{outliers} = 2\lambda log_2 \frac{M}{\epsilon} \tag{8}$$

where the λ points classified as outliers are modeled as random points having uniform distribution in the image domain.

7 References

[1] Alquier.L and Montesinos.P, Representation of linear structures using perceptual grouping, *IEEE Computer Society workshop on Perceptual Organization in Computer Vision*, June 1998.

[2] Amir.A and Lindenbaum.M, Ground from Figure discrimination, *Computer Vision and Image Understanding*, vol 76, no:1, pp 7-18, Oct 1999.

[3] Andrade-Cetto.J and Sanfeliu.A, Integration of perceptual grouping and depth, *IEEE International Conference on Pattern Recognition*, 2000.

[4] Boyer.K.L(ed), *Perceptual Organization for artificial vision systems*, Kluwer, 2000.

[5] Buffart.H, Leeuwenberg.E and Restle.F, Coding theory of visual pattern completion, *Journal of Experimental Psychology: Human Perception and Performance*, volume 7, no:2, April, 1981.

[6] Elder.J and Zucker.S, A measure of closure, *IR 93-2*, McGill Research centre for Intelligent Machines, McGill University, Montreal, 1994.
[7] Gdalyahu.Y, Shental.N, Weinshall.D, Perceptual grouping and segmentation by stochastic clustering, *CVPRIP 2000*, Atlantic city.
[8] Geiger.D, Kumaran.K and Parida.L, Visual Organization for figure/ground separation, *IEEE Conference on Computer Vision and Pattern Recognition* CVPR '96, San Francisco, 1996.
[9] Herault.L and Houraud.R, Figure - ground discrimination : a combinational Optimization approach, *IEEE transactions on Pattern Analysis and Machine Intelligence*, 15(9) , 899-914, September 1993.
[10] Hoogs.A, Mundy.J, An integrated boundary and region approach to perceptual grouping, *IEEE International Conference on Pattern Recognition*, 2000.
[11] Jacobs.D.W, Robust and Efficient Detection of Salient Convex Groups, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 18, No. 1, January 1996.
[12] Kanizsa,G, *Organization in Vision: Essays on Gestalt Perception* , Praeger, New York, 1979.
[13] Koffka,K, *Principles of Gestalt Psychology*, Harcourt Brace, NewYork, 1935.
[14] Leclerc.G.Y, Constructing simple stable descriptions for image partitioning, *International Journal of Computer Vision*, 3, 73-102, 1989.
[15] Li.M and Vitanyi.P, *An introduction to Kolmogorov Complexity and its applications*, Springer Verlag, NewYork, 1993.
[16] Li.M.X, Minimum description length based 2-D shape description, in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition 1993*, Berlin, Germany (H.-H. Nagel et al., Ed.), pp. 512-517, IEEE Computer Society press.
[17] Lindeberg, T. and Li, M. X, Segmentation and classification of edges using minimum description length approximation and complementary junction cues, *Computer Vision and Image Understanding*, 67(1):88-98,1997.
[18] Lowe.D.G, *Perceptual Organization and Visual Recognition*, Kluwer Academic, Boston, 1985.
[19] Mahmud.S, Thornber.K.K, Williams.L.R, Segmentation of salient closed contours from real images, *IEEE International Conference on Computer Vision*, Corfu, Greece,1999.
[20] Marr,D., *Vision*, Freeman Press, San Fransisco,1982.
[21] Pao.H, Geiger.D and Rubin.N, Measuring Convexity for Figure/Ground separation, *Proceedings of the International Conference on Computer Vision*, Sept 20 - 25, Corfu, Greece,1999.
[22] Saund.E, Perceptual Organization of Occluding Contours of Opaque Surfaces, *Computer Vision and Image Understanding*, vol 76, no:1, 70-82, Oct 1999.
[23] Witkin.A and Tenenbaum.J, On the role of Perceptual Organization, in: Pentland.A (ed), *From Pixels to Predicates, Recent Advances in Computational and Robotic Vision*, Ablex Publishing Corporation, Norwood, New Jersey, 1986.