MULTI-SCALE HIERARCHICAL SEGMENTATION

L.D. Griffin, G.P. Robinson, A.C.F. Colchester
Department of Neurology,
UMDS, Guy's Hospital, London SE1 9RT, England.

Abstract

In the absence of a priori information on scales of interest vision systems should initially process in a scale invariant manner. The fact that any signal can only be sampled discretely further constrains the initial processing. The paper argues that a representation satisfying these requirements is an hierarchical segmentation of scale-space. An algorithm is presented to compute such representations. The algorithm has been designed so that its operation is scale invariant in the following sense: the addition of finer scale information only ever adds to the computed representation and never changes what was discoverable from coarser scales. It is noted that such a scheme has benefits even when the scales of interest are known.

1 Introduction

A theory of image segmentation must address three questions: the nature of the input (section 1.1); the nature of the output of the process (section 1.2); and the method whereby the inputs are processed to produce the output (section 1.3).

1.1 Scale-Space and Image Measurement A unified theory of image measurement has been developed by Koenderink [1988 and 1992]. For the visual system to exhibit scale, shift and rotational invariance its measurements must take the form of derivatives of an aperture function. The aperture function ranges in size from the inner scale (pixels) to the outer scale (whole image). The condition that no detail should be generated as scale is increased dictates that the aperture function should be the isotropic Gaussian kernel. These measurements can be organized into a scale-space which can be visualized as a stack of images formed into a volume. Each horizontal slice of the scale-space is the original image blurred to the degree associated with the aperture function at that level.

1.2 Image Segmentation Outputs of most previous segmentation algorithms fall into the following categories -

- Division of the image plane by a series of closed loops [e.g. Marr and Hildreth 1980].
- A set of not necessarily connected edge fragments [e.g. Canny 1986].
- A partition of the image plane [e.g. Leclerc 1989].

None of the above representations are rich enough to express the relation of object/sub-object. Previously we have presented an algorithm for constructing hierarchical segmentations (HS) of grey-level images [Griffin et al. 1992b]. An improved version of this algorithm is described in section 2. An HS represents

BMVC 1993 doi:10.5244/C.7.29

image structure as a recursive partitioning of the image. However, even this representation is not rich enough to be scale-invariant. If objects changed their shape but maintained their topology over scale then an HS (of the image plane) would be sufficiently rich for scale-invariance. All that would have to be ensured would be that structure at the bottom of the hierarchy was the first to disappear as the observer receded from the scene. As it is, objects do change their topology during blurring. Consider a lightly leafed tree. At close range the tree is seen to be a complex connected mesh of leaves and branches. As the viewpoint moves away (equivalent to blurring the image), at some point, thin twigs attaching leaves to branches will no longer be resolvable and the leaves will appear to be floating in mid-air (the leaf objects are now disconnected). As the viewpoint moves still further the leaves re-merge and one is left with a tree-shaped connected blob. An important aspect of this phenomenon is that changes in topology under blurring are not accompanied by significant perceptual changes [Koenderink 1986].

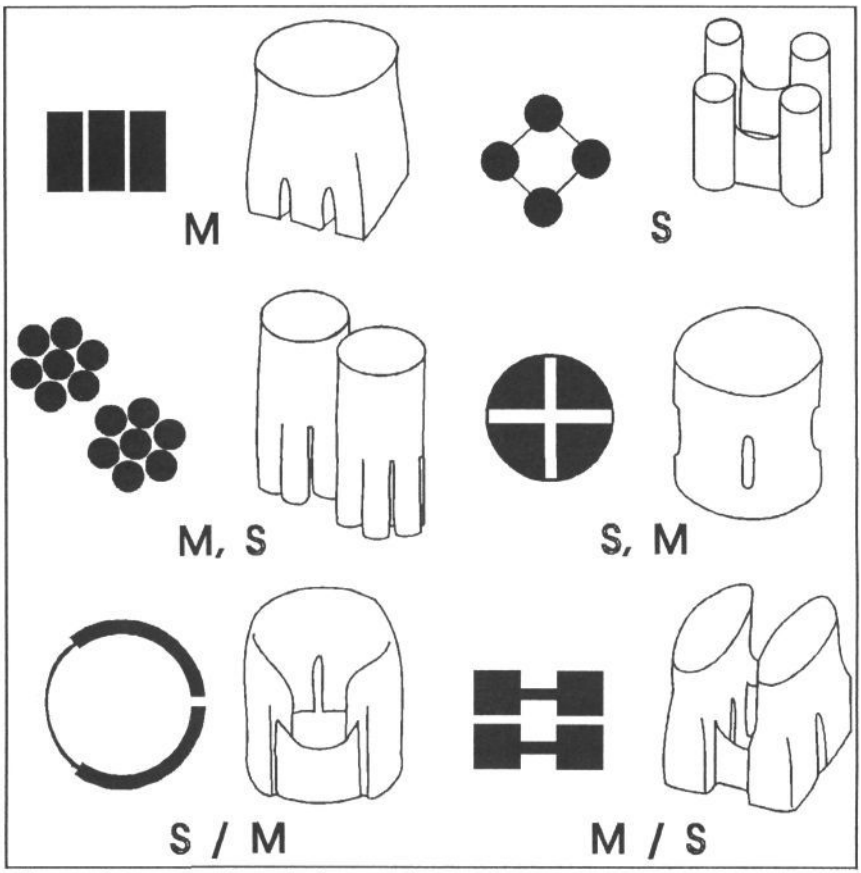

Figure 1 - Examples of changes in topology over scale.

Fig. 1 shows examples of the richness of this process. Various shapes and their outlines under blurring are shown. The outlines have been organized into surfaces

(enclosing 'shape-volumes') showing the structure of the shape in scale-space. The vertical direction is scale, and horizontal cross-sections are the shape at particular scales. The shapes show examples of merging (M), splitting (S) and various combinations of the two.

The dynamic shape theory of Koenderink [1986] shows how these shape volumes are to be defined for binary images. The algorithm presented in section 2 is a solution to the more general problem of grey-level images.

1.3 Edge Measures Edge measures which attempt to detect intensity discontinuities (as compared to, for instance, texture boundaries) are typically calculated from the response of one or more filters. Most filters [e.g. Canny 1986] are based on the Gaussian kernel and its derivatives (as advocated by Gaussian scale-space theory). Often these filters are applied at several scales and the results combined to emphasize those edges which exist over a range of scales. In section 2.2 the concept of stability over scale is examined within the framework of Gaussian scale-space theory and a novel edge measure is presented.

2 Single-Scale Hierarchical Segmentation

In our single-scale HS algorithm the image is represented as a graph, where, initially, nodes correspond to pixels and the links between nodes represent edges and correspond to pixel adjacencies (i.e. the cracks between pixels). The procedure iteratively groups nodes to form regions separated by edges. The clustering algorithm is detailed in section 2.1. In section 2.2 we derive a measure for the stability of edges over scale. This measure is used to modify the gradient to produce an edge strength which combines strength and stability.

2.1 Graph Merging The input image is represented as an undirected graph where the nodes represent image objects and the links the object adjacencies. Initially there is a node for each pixel in the original image and a link for each pair of adjacent pixels (4-way connectivity used). The graph is iteratively reduced by merging sets of adjacent nodes until only a single node remains. As the graph is reduced, node/sub-node relationships are recorded. Thus every object (apart from individual pixels) has a set of sub-objects and every node (apart from the whole image) has a parent object of which it is a sub-part. As the hierarchy is formed descriptive values (attributes) are accumulated within the nodes. These attributes are used in the calculation of an edge measure which guides the node-merging.

To select which nodes to merge, an edge measure (see section 2.2) is calculated for each link in the graph. The edge measure must satisfy two criteria: firstly, that it is high if the two nodes are 'dissimilar' and low if 'similar' (similar and dissimilar being defined by the particular edge measure used); and secondly that

it is calculated solely from consideration of the two nodes at either end of it. We require the merging technique to be independent of the edge measure used and invariant under linear transformations of the luminance, so we cannot use a threshold on the edge strength to discover which nodes should be merged. Instead we group in a similar manner to watershed techniques [Griffin et al. 1992a]. A pointer is set across the weakest edge of each object (i.e. towards the object with which it wants to merge first). This set of pointers groups the nodes into equivalence classes. Exactly one equivalence class is formed for each minimally weak edge. These occur where two objects mutually point at each other. Separating the equivalence classes is a network of ridges of high edge strength. Repeated application of this procedure produces a hierarchy, but not a very satisfactory one: there is a problem of 'interference' between regions with a different number of levels of structure.

The problem can be understood my means of an example. Consider an image with two neighbouring regions of different mean luminance. Imagine one region to be smooth (for instance the background) and the other textured. If we proceed as described, then the smooth region will quickly cluster together in only a few iterations, while the textured region may still be quite fragmented. At the next iteration the smooth region will merge with a portion of the textured region. The end result is that the hierarchy will not have an object corresponding to the entire textured region. The problem occurs in those equivalence classes that have an internal edge stronger than some external edge. The remedy is simple. The weakest external edge of each equivalence class is determined. Internal edges which are stronger than the weakest external edge of the equivalence class are noted and objects are removed so that offending internal edges become external. Of the two objects that could be removed the one with the stronger weakest edge is chosen. The value of the weakest external edge of an equivalence may change during this process; it may become stronger but never weaker, so the order of the removals is irrelevant. No equivalence class will be completely destroyed, as neither of the minimally-linked pair at the centre will ever be deleted. The equivalence classes are then merged and a new graph is formed from the resulting nodes and those nodes not taking part in any merge. The graph will always contain at least one minimally weak edge so the graph is reduced at each iteration and the procedure only halts once a single node (corresponding to the entire image) is all that remains.

The resulting hierarchical structure of nodes and child nodes (objects and sub-objects) descends all the way from the outer scale (the entire image) down to the inner scale (individual pixels). Since the hierarchy is of variable depth, different parts of the image will have a different number of levels of structure.

2.2 Edge Measure It has been noted that the occurrence of an edge at multiple scales contributes to its perceptual significance [Marr and Hildreth 1980]. Bischoff

and Caelli [1988] attempted to make this precise by defining the stability of an edge (in their case Laplacian zero-crossings) as the largest continuous range of scales over which the zero-crossing lies within some neighbourhood of the point. The area of the neighbourhood increases linearly with scale.

It can be shown [Koenderink 1988] that the paths of steepest ascent on the isophote surfaces of scale-space are given by (**L** is the luminance) -

$$\vec{s} = (-L_x L_\sigma , -L_y L_\sigma , L_x^2 + L_y^2)^T \qquad \text{(x,y,}\sigma\text{) system}$$

This is conveniently expressed in the (w,v,t) gauge co-ordinate system [Haar Romeny et al. 1991]. The w-direction is in the direction of the gradient, the v-direction is tangent to the isophote through the point and $t=2\sigma^{1/2}$ (the natural unit of length in scale space [Koenderink 1992]).

$$\vec{s} = (-\sqrt{\sigma}\, L_\sigma , 0 , L_w)^T \qquad \text{(w,v,t) system}$$

Assuming that we are willing to identify an edge through scale by means of these isophote projections we can use this expression to calculate the angle (Θ) between the tangent to the isophote projection and the σ=constant plane

$$\theta = \arctan(L_w / |\sqrt{\sigma}\, L_\sigma|)$$

This shall be referred to as the phase with respect to scale and shall be used to characterize the stability of an edge point. We note that unlike previous definitions this value is defined for all points of the image and not just for Laplacian zero-crossings.

We combine this angle with the gradient magnitude to produce an edge measure **E** (referred to as the modified gradient) which reflects both the strength of the edge (as given by the gradient magnitude) and its stability (as given by the phase with respect to scale). In scale-space we have the property that $L_\sigma = \nabla^2 L$ which allows this value to be calculated from derivatives within the image plane -

$$E = L_w . \arctan(L_w / |\sqrt{\sigma}\, \nabla^2 L|)$$

Figure 2 shows: a 1D blurred step edge; its gradient; and the modified gradient. This shows how the non-linearity of the modified gradient has produced a cusp at the point of maximum response (for a step edge) rather than a simple maximum.

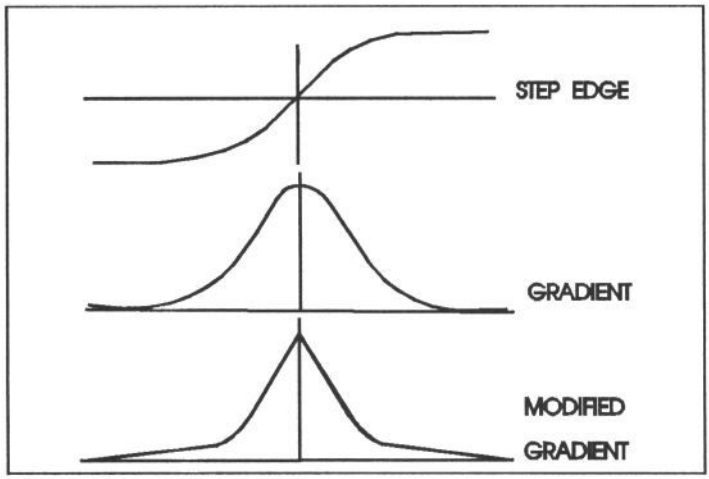

Figure 2

2.3 Single-Scale Image Measurement For the edge measure detailed in section 2.2 three attributes are required: mean luminance, mean Laplacian and area. The area attribute is needed for the recalculation of the luminance and Laplacian attributes after node merging. The initial nodes in the graph correspond to individual pixels and so have an area of 1 and a luminance inherited from the pixel. The Laplacian is calculated with the 9-point mask described in [Lindeberg 1990].

To calculate the strength of an edge between two regions we require values for the gradient and the Laplacian. We use the difference in mean luminance (between the two regions) and the average of the mean Laplacians (of the two regions) for these two values.

3 Multi-Scale Hierarchical Segmentation

In section 1.1 it was pointed out that since the topology of image objects can change over scale, a hierarchical segmentation of the image plane is not sufficiently rich to capture image structure in a scale-invariant manner. A representation which is capable of scale invariance is a hierarchical segmentation of scale-space. In such a representation, objects are connected volumes of scale-space. Cross-sections through these volumes give the shape of the object at that scale. As before, sub-objects are completely contained within their parent.

The establishment of the relations between images at different scales is referred to as the correspondence problem [Koenderink 1990 p502]. Previous attempts to solve it [Lifshitz 1987] have concentrated on establishing the isophote projections of the image points in a consistent manner. This has proved to be hard. One problem with it is the breakdown of nice causal behaviour at critical points of the image. Koenderink [1989] has shown that this anomaly does not occur if one

looks at areas instead of paths. The solution presented in this paper proceeds by calculating an area-based representation of the image at each scale and establishing a correspondence between the representations. The representation used is the hierarchical segmentation described in the previous section. The correspondence proceeds stepwise. Initially, the second coarsest scale representation ($\mathbf{R}_2$) is put into correspondence with the coarsest scale representation ($\mathbf{R}_1$). During this process $\mathbf{R}_2$ may be modified. Then $\mathbf{R}_3$ is put into correspondence with the $\mathbf{R}_1$+$\mathbf{R}_2$ structure with what ever modification is necessary, and so on.

The technique of establishing correspondence was motivated by 3 constraints -

C1 The correspondences generated should be consistent with the object/sub-object relationships already discovered at coarser scales (scale-invariance).

C2 All objects at a given scale should have a cause at the next finest scale (scale-space causality).

C3 Although the addition of information at a finer scale may reveal that two apparently distinct objects are in fact connected, no edge at a coarser scale should be removed by the addition of finer scale information (scale-invariance).

and two considerations -

C4 As much freedom as possible should be left to the segmentation generated within the scale (use your data).

C5 Causes should be close to effects (isophote projections).

3.1 Multi-Scale Image Measurements The requirement of scale invariance dictates that (i) scale should be sampled logarithmically (ii) the number of samples per unit area, at a given scale, is proportional to that scale. Together this gives scale-space the shape of an exponentially tapering tower (such as the Eiffel Tower) of which the quad-tree [Rosenfeld 1984] is an example.

Koenderink [1988] has argued that for 8-bit images a scale sampling ratio of 1.155 (approximately one fifth of an octave) is appropriate. This is a comparable figure to the results of psychophysical experiments [Caelli et al. 1983] which test the ability of observers to discriminate between images blurred by a small amount. A scale-space constructed with this sampling scheme and sampled at the Nyquist frequency within scale is a factor of 7.43 greater in size than the original image.

The set of samples so defined lacks a topology. The required topology has a within- and a between-scale component. Within scale the 4-way connectivity implicit in the grid-like arrangement of samples is used. The between-scale topology takes the form of a lattice structure. Each sample is connected to a set

of samples in the finer scale (potential causes) and a set of samples in the coarser scale (potential effects). This can be envisaged as a pair of cones pointing upwards and downwards from each point. The spread of these cones represents the fastest lateral movement possible by an image feature during blurring. The angle of the cones is determined by the precision of the image: the higher the precision the greater the possible lateral movement. These cones are not unlike the light cones of general relativity in that they represent the limits of causal linkage.

3.2 Combining Segmentations from Different Scales The scale-space segmentation is created from coarse to fine. Sequentially, each scale is attached to the bottom of the growing segmentation until the whole of scale-space from inner- to outer-scale is represented. In the following we sketch the process whereby a finer scale is attached.

Constraint **C1** guides the determination of cause/effect relationships between the objects of the finer scale (causes) and the partial scale-space segmentation so far discovered (effects). The two hierarchy roots are connected as cause (fine) and effect (coarse) and then the depth 1 sub-objects of each hierarchy are dealt with. A system of cause/effect relationships is established between these objects (see next paragraph). Since both merge and split events may occur these cause/effect relationships are potentially many-to-many. Some causes may be without effects but all effects will have a cause(s). The cause/effect relations partition the causes and effects into equivalence classes. The procedure then continues recursively within each equivalence class and so proceeds down the two hierarchies.

How are the cause/effect relations determined between the two sets of objects? We consider the set of pixels making up the effect objects. The between-scale topology of scale-space defines a set of pixels in the finer scale which are potential causes. (*) For each effect pixel at least one of the attached finer scale pixels will belong to a cause. Thus there is at least one cause/effect relationship implied by each effect pixel. We select one relationship for each effect pixel (guided by **C5**) and instantiate it. As a consequence of this, when we come to deal with the sub-causes and sub-effects statement (*) will still hold. This means that the process will continue all the way down the effect hierarchy and so **C2** will be satisfied.

It is however possible (likely) that constraint **C3** will be violated. This will occur if a common cause is found for two adjacent effects. Such events are detected and dealt with by modifying the finer scale hierarchy. Any offending causes responsible for a violation of **C3** are removed and replaced with their sub-parts (a pixel object will never cause such a violation, so there is no problem here). Then the cause/effect relations are re-determined between the effects and the now changed set of causes. This process of modification continues until cause/effects

are set up that do not violate C3. This is always possible, but in extreme situations may mean that the causes are fragmented right down to the pixel level.

The set of cause/effects thus found are consistent with C3 but may imply the merging of groups of causes. This occurs if they are adjacent and have a common effect. Such groups are detected and hierarchically clustered. The root of this clustering is then put into cause/effect relation with the common effect of the clustered objects. This step undoes many of the initial modifications that were necessary to ensure that C3 would be satisfied.

4 DISCUSSION AND CONCLUSIONS

We have presented single- and multi-scale hierarchical segmentation algorithms. The single-scale algorithm generates hierarchical segmentations of the image plane. The multi-scale algorithm links and modifies hierarchies from a range of scales into a hierarchical segmentation of scale-space. The process is scale invariant in the sense that the addition of finer scale information only ever adds to the representation and never changes what is already there.

The shift from a hierarchical segmentation of the image plane to a hierarchical segmentation of scale-space allows a richer set of image features to be represented. The elements of the segmentation are connected volumes of scale-space. The 'shape volumes' make explicit many features whose detection is regarded as crucial in shape representation.

- Objects which are not connected at some scale (e.g. the inner scale) can still be grouped into a single gestalt.
- Objects can be split into sub-parts at narrowings (if a split occurs).
- Spurs/dents can be identified and related, as protruding from or penetrating into, some more significant part.

It is this added benefit of also producing a structure suitable for shape processing that makes the scheme particularly attractive. The insistence of scale-invariance is thus seen to be worthwhile not only for vision systems with unknown inner-scale (such as mobile robots) but also for systems where the inner-scale can be precisely known (such as medical imaging and automated inspection systems).

Several image features cannot be represented by this scheme. Firstly, there are shadows and transparency, both cases of overlapping objects. This cannot be expressed in the current scheme. Secondly, there are dot patterns [Robinson et al. 1992]. In dot patterns groups of widely separated points are linked together; this could not be achieved by their being parts of some common object at a coarse scale, as the individual dots would disappear at an earlier stage of blurring.

Acknowledgements

This work was carried out as part of the SERC funded ISIS project.

References

Bischoff,W.F. & Caelli,T.M. (1988) Parsing scale-space and spatial stability analysis. *Comp. Vis. Graph. Image Process.* 42:192-205.

Caelli,T.M., Brettel,H., Rentschler,I., Hilz,R. (1981) Discrimination thresholds in the two-dimensional spatial frequency domain. *Vis. Res.* 23:129-133.

Canny,J. (1986) A computational approach to edge detection. *IEEE Trans. Patt. Anal. Mach. Intell.* PAMI-8:679-698.

Griffin,L.D., Colchester,A.C.F. and Robinson,G.P. (1992a) Scale and segmentation of grey-level images using maximum gradient paths. *Image and Vis. Comput.* 10(6):389-402.

Griffin,L.D., Colchester,A.C.F. and Robinson,G.P. (1992b) Structure-sensitive scale and the hierarchical segmentation of images. In: *Proc. VBC '92*, Robb,R.(ed.). pp24-32.

ter Haar Romeny,B.M., Florack,L.M.J., Koenderink,J.J. and Viergever,M.A. (1991) Scale-space: its natural operators and differential invariants. *Image and Vis. Comput.* 10(6):376-388.

Koenderink,J.J. (1986) Dynamic shape. *Biol. Cybern.* 53:383-396.

Koenderink,J.J. (1988) Image structure. In: *Mathematics and Computer Science in Medical Imaging*, Viergever, M.A., Todd-Pokropek,A. (eds.), Springer-Verlag, Berlin, pp67-104.

Koenderink,J.J. (1989) A hitherto unnoticed singularity of scale-space. *IEEE Trans. Patt. Anal. Mach. Intell.* PAMI-11(11):1222-1224.

Koenderink,J.J. (1990) *Solid Shape*. MIT Press.

Koenderink,J.J. (1992) Generic neighbourhood operators. *IEEE Trans. Patt. Anal. Mach. Intell.* PAMI-14(6):597-605.

Leclerc,Y.G. (1989) Constructing simple stable descriptions for image partitioning. *Int. J. Comp. Vis.* 3:73-102

Lindeberg,T. (1990) Scale-space for discrete signals. *IEEE Trans. Patt. Anal. Mach. Intell.* PAMI-12:234-254.

Lifshitz,L.M. (1987) Image segmentation via multiresolution extrema following. Univ. North Carolina, Dept. Comp. Sci. Tech. Rep. 87-012.

Marr,D.C. and Hildreth,E. (1980) Theory of edge detection. *Proc. Roy. Soc.* B-207:187-217

Robinson,G.P., Griffin,L.D. and Colchester,A.C.F. (1992) The Delaunay/Voronoi selection graph: a method for extracting shape information from 2D dot patterns with an extension to 3D. In: *Proc. BMVC*, Leeds, Sep. 1992.

Rosenfeld,A. (ed.) (1984) *Multiresolution image processing and analysis*. Springer, Berlin.