Robust morphological scale-space trees

J. Andrew Bangham, Javier Ruiz Hidalgo, Richard Harvey
School of Information Systems
Tel: +44 1603 592847, Fax: +44 1603 593345,
Email: {ab,jrh,rwh}@sys.uea.ac.uk
Norwich, NR4 7TJ, UK.

Abstract

This paper derives a new tree representation of an image and shows how the tree may be derived from graph morphology and connected-set, alternating sequential, filters. The resulting *scale tree* forms a pyramid of increasing size objects where the nodes correspond to features of a particular scale. The tree structure itself may be made fairly insensitive to geometrical changes in the image. By parsing the tree and using attributes associated with the nodes, image processing operations such as filtering, segmentation and detection can be performed.

1 Motivation

A long-standing, ambitious, goal for a computer vision system is to extract a simple description in terms of meaningful objects, their positions and geometric relations to one another, from an image. Ideas of this type are currently being embodied in the proposed MPEG 4 standard which will facilitate the evolution of systems that allow interaction with the audio visual scene. An audio visual scene may be understood as a composition of primary audio visual objects, according to a script that describes their spatial and temporal relationship [1]. Coding an image, in such a fashion, should be an extremely useful step in the process of recognizing objects in the image. A first step to building an MPEG 4 coder might be to build a tree according to the *scale* of features where the relationship between objects is defined in terms of the inclusion of one object within, or occlusion by, another. That is, the topology of the image.

First we illustrate our meaning and show how such a *scale tree*, that represents the image topology, might be extracted from a greyscale image. Then it is considered how, in practice, a scale tree derived from a real image has to be simplified and finally it is shown how such a tree can be used for a simple image processing application, namely the separation of moving objects from a sequence.

2 Pyramid representations

The representation of an image as a scale-related pyramid is a well known theoretical development [2] and is sometimes used as a practical way of reducing computation [3]. All methods in common use, work the same way: the image signal bandwidth is reduced by filtering with a finite impulse response filter and the filtered signal is then down-sampled. Contenders for the filter include: Laplacian of Gaussians [2], Gabor filters [4], wavelets [5] or B-splines [6]. Discretized Gaussians have the particular advantage that they preserve scale-space causality [7]. However, the method described here does not rely on linear filtering at all: mathematical morphology provides the theoretical framework space. Mathematical morphology is the analysis of signals, particularly images, by shape. The subject was primarily developed by Serra [8] from work by Matheron [9] and other roots including Blum [10]. To preserve scale-space causality in two or more dimensions it is necessary to use connected set granulometries [11, 12, 13]. However, the granules extracted by this method can change significantly with small changes in the noise or clutter, in other words they are not robust and a tree predicated on granulometry could change significantly with small changes in the image. More recently it has been shown that connected set sieves, or alternating sequential filters, are more robust [14]. Like well known diffusion based filters [15, 16, 17] these systems preserve scale-space causality [18].

These processors are said to transform the signal to another domain, called granularity (each granule will be represented as a node in a scale tree), and such a transformation is invertible [19]. This means that, if a tree is built where each node is a granule, then the image may be rebuilt from the nodes and, if nodes are deleted, that the resultant, simpler, tree will be identical to the tree obtained from an image in which the corresponding granule had been deleted.

Although this paper deals only with two-dimensional images the sieve is defined as operating over a graph [20] so, in principle, operates on images defined in any finite number of dimensions. The graph is denoted $G = (V, E)$ where the set of vertices, V, are pixel labels and, E, the set of edges, represents the adjacencies. Defining $C_r(G)$ as the set of connected subsets of G with r elements allows the definition of $C_r(G, x)$ as those elements of $C_r(G)$ that contain x.

$$C_r(G, x) = \{\xi \in C_r(G) | x \in \xi\} \tag{1}$$

Morphological openings and closings, over a graph, may be defined as

$$\psi_r f(x) = \max_{\xi \in C_r(G,x)} \min_{u \in \xi} f(u) \tag{2}$$

$$\gamma_r f(x) = \min_{\xi \in C_r(G,x)} \max_{u \in \xi} f(u) \tag{3}$$

The effect of an opening of size one, ψ_1, is to remove all *maxima* of area one when working in two dimensions. In one-dimension it would remove run-lengths of length one. γ_1 would remove *minima* of size one. Applying ψ_2 to $\psi_1 f(x)$ will now remove all maxima of area two and so on. The $\mathcal{M}$ and $\mathcal{N}$ operators are defined as $\mathcal{M}^r = \gamma_r \psi_r$ and $\mathcal{N}^r = \psi_r \gamma_r$. Sieves, and filters in their class such as alternating sequential filters with flat structuring elements, depend on repeated application of such operators at increasing scale.

This cascade behavior is key, since each stage removes maxima or minima of a particular scale. The output at scale r is denoted by $f_r(x)$ with $f_1 = \mathcal{Q}^1 f = f$ and $f_{r+1} = \mathcal{Q}^{r+1} f_r$ where $\mathcal{Q}$ is one of the γ, ψ, $\mathcal{M}$ or $\mathcal{N}$ operators. The differences between successive stages of a sieve, called *granule functions*, $d_r = f_r - f_{r+1}$, contain non-zero regions, called *granules*, g_r's of only that scale. Each g_r is a connected set of r pixels. Illustrations of sieves and formal proofs of their properties appear elsewhere [21], and the generation of an 2D *scale tree* appears to be less than an $\mathcal{O}(n \log n)$ process. In practice we can generate the tree in less than a second on a Pentium PC.

3 Parsing real trees

The left of Figure 1 shows a small, simple, image. Nevertheless it is associated with a complicated tree (shown in the right of Figure 1). The challenge, particularly for pattern recognition, is to decide which parts of the tree are significant and retain them. Furthermore, if the tree is to be simplified then it should be simplified in a way that satisfies Marr's Principle of Least Commitment [22].

The conventional approach [21] has been to describe the image via *channels* which are the sums of granules over a fixed range of scales. Although such a simplification is useful it is often the case that objects appear in several channels. This is not surprising since shading causes intensity variations across the object and there is often blur due to imperfections in the imaging system. A solution to this problem is to track the object through scale and look for a peak in the scale-selection surface [7]. In terms of the tree representation this amounts to concatenating long chains of, non-branching, low contrast nodes into a single node.

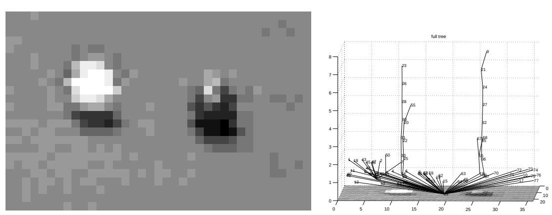

Figure 1: A 4 bit greyscale excerpt from the sequence used in Figure 5 (left) and its associated scale tree (right)

Figures 1, 2, 3 illustrate this process. Figure 1 shows a segment of an image from the sequence shown in Figure 5. The tree shown at the right has a large number of nodes associated with image detail and has some long chains associated with the objects (billiard balls in this case). For each branch the sequence $g_{s_i}, i = 1 \ldots N$ represents the granule amplitude at each node where N is the number of nodes in that branch. Each node has scale s_i. In Figure 4 we plot the rate of change of granule intensity $\Delta_i = g_{s_i}/(s_i - s_{i-1}), i > 2$ versus the index, i. i is the tree depth measured down the branch. The peak in Δ_i is, for an object, the node at which its rate of change of intensity with respect to scale is maximized. For example a perfect disc of area s would yield a sequence of Δ_i that is all zero except for one value at its true scale. For blurred images, such as shown in Figure 1, we plot in Figure 4 the scale selection measure and collapse chains onto the node with the maximum Δ_i. Figure 2 shows the

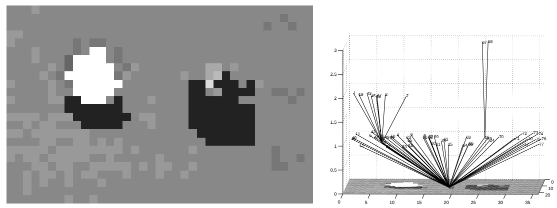

Figure 2: Image from Figure 1 after long chains have been collapsed

image and its tree after this operation. Granules at a number of scales have been removed, the tree is simplified and the objects have sharp edges. We emphasize that the use of connected sets means that there is no windowing effect. However there could still be too many branches.

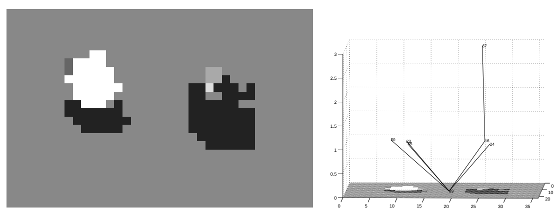

Figure 3: Image from Figure 2 after low contrast nodes have been removed.

Figure 3 shows the image after the tree has been pruned by removing all nodes with an amplitude that differs from their parent by less than two units. This heuristic uses a hard decision and it would be desirable to replace it with a more principled step based on probabilities.

4 Using the scale tree for motion segmentation

Figure 5 shows just one of many potential applications of the technique. Frames from a sequence taken from a noisy domestic television signal are shown. The original sequence (top row) shows a white cue ball hitting a black ball. The second row shows the result of using the scale tree to look for moving objects. In this motion filtering example the full tree is used without the collapsing or pruning step previously described. The moving objects have been extracted from the sequence. Initially each node is visited, starting at the root, and the flat-zone associated with each node is translated around the equivalent

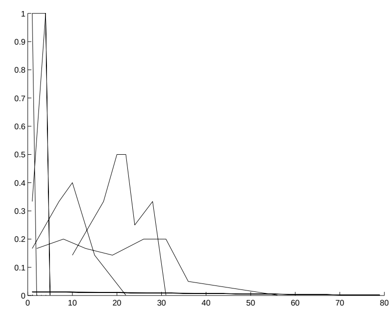

Figure 4: Graphs showing amplitude of granules as a function of scale.

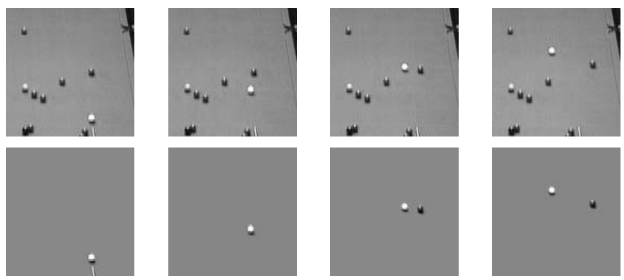

Figure 5: Top row, some frames from a 8 bit greyscale movie sequence, note the cue (bottom right of the first frame) strikes the white ball which in turn collides with the black ball. The second row shows the same frames from which the moving balls have been extracted with the help of a scale tree.

region of the next image to find the minimum absolute difference. This reflects the motion vector. The motion vector for a parent node tends to be a good estimate for the motion vector for its children and so the search can be fairly straightforward. A similar strategy has been reported using max and min trees [23]. However these structures are much less robust to noise.

5 Discussion

A new way to obtain a topological tree representation of an image has been presented. Each node is a granule and, as such, has at some stage in the processing of the image, been associated with an extremum and, therefore, is bounded by an edge. Each node has a set of attributes. In our current formulation this includes: granule amplitude, granule shape (coded using a combination of a bit map and pointers) and position (in x, y and scale). Reports suggest that the x, y and scale attributes are likely to be robustly estimated in the presence of noise and clutter [14]. The branching structure of the tree is determined by the way features lie within each other. It has been shown that this scale tree could, with certain images, represent the image in a manner that is consistent with an object tree, where the objects fit the definition required of a primary visual object sitting on a plane as envisaged in MPEG 4.

Although nodes in this new scale-tree do represent some, perhaps even many, of the objects in the image it is not expected that it accurately identifies all objects. We therefore see the tree as a first approximation to the correct object tree. It is an approximation that can be obtained from a single image. Currently we are working on strategies for modifying the structure in the light of further evidence, such as colour, motion vectors or stereo disparity. We are also exploring a more subtle approach using the original single image where shape-from-shading clues are used to modify the scale-tree and make it more closely represent an object tree.

Once the scale-tree is as close to an object tree as possible, the tree will be a very useful representation of the image. Not only can it be used for filtering, but it can also be used for object recognition. This can be done at two levels. (1) The tree structure itself codes object topology that is, to a large extent, independent of geometrical scaling, rotations and distortions and (2) a more detailed matching can be performed by also using attributes of the nodes.

The main potential problem with the system may lie in how well it scales to large and complex images, however, we are already processing real images and there is evidence that with a careful choice of node attributes, pointers in particular, that the approach will prove realistically fast.

References

[1] F. Pereira. MPEG 4: a new challenge for the representation of audio-visual information. In *Proc. Picture Coding Symposium 1996*, 1996.

[2] P.J. Burt and E.H. Adelson. The laplacian pyramid as compact image code. *IEEE Transactions on Communication*, COM-31:532–540, 1983.

[3] A. Lanitis, C.J. Taylor, and T.F. Cootes. A unified approach to coding and interpreting face images. In *Proceedings of the Fifth International Conference on Computer Vision*, pages 368–373, 1995.

[4] J. Daugman. Complete discrete 2-d gabor transforms by neural networks for image analysis and compression. *IEEE Trans. Acoustics Speech and Sig. Proc.*, 36:1169–1179, 1988.

[5] E.P.Simoncelli, W.T.Freeman, E.H.Adelson, and D.J.Heeger. Shiftable multi-scale transforms. *IEEE Trans. on Info. Theory*, 38(2):587–607, March 1992.

[6] M. Unser, A. Aldroubi, and M. Eden. A family of polynomial spline wavelet transforms. *IEEE Trans. on Sig. Proc.*, 30(2):141–162, Jan 1993.

[7] T. Lindeberg. *Scale-space theory in computer vision*. Kluwer, 1994.

[8] J. Serra. *Image Analysis and Mathematical Morphology vol 1.* Academic Press, New York, 1982.

[9] G.Matheron. *Random sets and integral geometry.* Wiley, 1975.

[10] H. Blum. Biological shape and visual science (part 1). *J. Theoret. Biol.*, pages 205–287, 1967.

[11] L. Vincent. Morphological area openings and closings of greyscale images. In *Workshop on Shape in Picture*. NATO, 1992.

[12] Corinne Vachier and Fernand Meyer. Extinction value: a new measurement of persistence. In Ionas Pitas, editor, *Proc. of 1995 IEEE Workshop on nonlin. sig. and image proc.*, volume 1, pages 254–257, June 1995.

[13] J. Serra and P. Salembier. Connected operators and pyramids. In *Proceedings of SPIE Conf. on Image Algebra and Mat. Morph., Volume 2030*, pages 65–67, 1993.

[14] R. Harvey, J.A. Bangham, and A. Bosson. Scale-space filters and their robustness. In *Proc. First Int. Conf. on Scale-space theory*, pages 341–344. Springer, 1997.

[15] J. J. Koenderink. The structure of images. *Biological Cybernetics*, 50:363–370, 1984.

[16] A. P. Witkin. Scale-space filtering. In *8th Int. Joint Conf. Artificial Intelligence*, pages 1019–1022. IEEE, 1983.

[17] P.Perona and J.Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Patt. Anal. Mach. Intell.*, 12(7):629–639, July 1990.

[18] J.A. Bangham, P.D. Ling, and R. Harvey. Nonlinear scale-space causality preserving filters. *IEEE Trans. Patt. Anal. Mach. Intell.*, 18:520–528, 1996.

[19] J.A. Bangham, P. Chardaire, C.J. Pye, and P.D. Ling. Multiscale nonlinear decomposition: the sieve decomposition theorem. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(5):529–539, 1996.

[20] H.J.A.M. Heijmans, P. Nacken, A. Toet, and L. Vincent. Graph morphology. *Journal of Visual Computing and Image Representation*, 3(1):24–38, March 1992.

[21] J.A.Bangham, R.Harvey, and P.D.Ling. Morphological scale-space preserving transforms in many dimensions. *J. Electronic Imaging*, 5(3):283–299, July 1996.

[22] D. Marr. *Vision.* W. H. Freeman and Company, New York, 1982.

[23] Phillippe Salembier, Albert Oliveras, and Luis Garrido. Motion connected operators for image sequences. In *Sig. Proc. VIII Proc. Eusipco-96*, volume 2, pages 1083–1086. Edizioni LINT, 1996.