Texture and Features

Nonlinear Scale-Space from n-Dimensional Sieves

J. Andrew Bangham, Richard Harvey, Paul D.Ling and Richard V. Aldridge

School of Information Systems, University of East Anglia, Norwich, NR4 7TJ, UK.

Abstract. The one-dimensional image analysis method know as the *sieve*[1] is extended to any finite dimensional image. It preserves all the usual scale-space properties but has some additional features that, we believe, make it more attractive than the diffusion-based methods. We present some simple examples of how it might be used.

1 Introduction

The use of scale-space for the analysis of images is well established and there is an interest in incorporating scale-space processors as part of high-level computer vision tasks.

Scale-space vision is usually associated with diffusion based systems [2, 3, 4] in which the image forms the initial conditions for a discretization of the diffusion equation $\nabla \cdot (D \nabla I) = I_t$. In linear scale-space processors the diffusion parameter, D, is a constant and, although the system has the desirable property that it does not introduce new extrema as scale increases (preserves *scale-space causality*), it blurs the edges of objects in the scene. Since edges are thought to be important for most high-level vision operations, a variant, in which D is made to vary as a function of $\|\nabla I\|$, is sometimes used.

It is said that scale-space should be "semantically meaningful," for which three properties have been enunciated [5], *causality* (above), *immediate localisation* and *piecewise smoothing*. The second is a requirement that the "region boundaries should be sharp and coincide with semantically meaningful boundaries at that resolution" and the third property is that "intra-region smoothing should occur preferentially over inter-region smoothing."

To these properties one might add some practical requirements: (i) The system should be *scale-calibrated*. At a particular scale one should see features of only that scale. This allows shapes to be measured accurately. (ii) The scale-space should be *manipulable*. One should be able to process the image in the scale-space domain and reconstruct it to produce an enhanced image.

This paper presents a system that has all the desirable properties of scale-space and also some additional, extremely powerful, features. The system has similarities with morphological image processing systems.

Mathematical morphology [6, 7] is based on the analysis of shape and has developed separately. However, a recent welcome development has been the effort to achieve some unification of scale-space and morphology [8, 9, 1, 10, 11]. We continue this trend with an analytical and experimental study of a type of decomposition called a *sieve* operating in two or more dimensions.

2 Properties of sieves

2.1 Definitions

An arbitrary array of pixels, or voxels, can be described by a connected graph [12] $G = (V, E)$ where V is the set of vertices and E is the set of pairs that describe the edges.

Definition 1. When G is a graph and $r \geq 1$, we let $\mathcal{C}_r(G)$ denote the set of connected subsets of G with r elements. When $x \in V$, we let $\mathcal{C}_r(G, x) = \{\, \xi \in \mathcal{C}_r(G) \mid x \in \xi \,\}$.

All subsequent operations take place over such sets. The structure of an element of $\mathcal{C}_r$ is determined by the adjacency of image pixels. In three dimensions the connected subsets of G would normally be six-connected.

Definition 1 allows a compact definition of an *opening*, ψ_r, and *closing*, γ_r, of size r.

Definition 2. For each integer, $r \geq 1$, the operators, $\psi_r, \gamma_r, \mathcal{M}^r, \mathcal{N}^r \colon \boldsymbol{Z}^V \to \boldsymbol{Z}^V$ are defined by

$$\psi_r f(x) = \max_{\xi \in \mathcal{C}_r(G,x)} \min_{u \in \xi} f(u), \gamma_r f(x) = \min_{\xi \in \mathcal{C}_r(G,x)} \max_{u \in \xi} f(u),$$

and $\mathcal{M}^r = \gamma_r \psi_r$, $\mathcal{N}^r = \psi_r \gamma_r$. ($\psi_r$ and γ_r are well defined since, for each $x \in V$, $\mathcal{C}_r(G, x)$ is nonempty [13].)

Thus $\mathcal{M}^r$ is an opening followed by a closing, both of size r and in any finite dimensional space.

We can now define the M- and N-sieves of a function, $f \in \boldsymbol{Z}^V$.

Definition 3. Suppose that $f \in \boldsymbol{Z}^V$.
(a) The M-sieve of f is the sequence, $(f_r)_{r=1}^{\infty}$, given by

$$f_1 = \mathcal{M}^1 f = f, \qquad f_{r+1} = \mathcal{M}^{r+1} f_r, \quad \text{for integers, } r \geq 1.$$

(b) The N-sieve of f is the sequence, $(f_r)_{r=1}^{\infty}$, given by

$$f_1 = \mathcal{N}^1 f = f, \qquad f_{r+1} = \mathcal{N}^{r+1} f_r, \quad \text{for integers, } r \geq 1.$$

M- and N-sieves are one type of alternating sequential filter [7]. We use the term sieve, partly for brevity, and partly because the properties we concentrate on are not unique to alternating sequential filters but are also to be found with, for example, a sequence of recursive median filters [14]. In addition, not all alternating sequential filters have the properties studied here.

Theorem 4. *If $(f_r)_{r=1}^{\infty}$ is the M- or the N-sieve of an $f \in \boldsymbol{Z}^V$, then, for each integer, $r \geq 1$, f_r is r-clean.*

Definition 5. For each integer, $r \geq 1$, we let $M^r = \mathcal{M}^r \cdots \mathcal{M}^2\mathcal{M}^1$, and $N^r = \mathcal{N}^r \cdots \mathcal{N}^2\mathcal{N}^1$. So the M-sieve (resp. N-sieve) of an $f \in \boldsymbol{Z}^V$ is $(M^r f)_{r=1}^{\infty}$ (resp. $(N^r f)_{r=1}^{\infty}$).

The term r-clean means that extremal level connected sets have r or more pixels [13]. The sieve has the effect of locating intensity extrema and "slicing off peaks and troughs" to produce *flat zones* [15] of r or more pixels hence r-clean, a function from which the flat zones have been "cleaned."

Since all the pixels within each extremal connected set have the same intensity, a simple graph reduction at each stage can lead to a fast algorithm [16]. The one-dimensional equivalent has already been reported [17]. It has order complexity p, where p is the number of pixels, for a complete decomposition ($r = p + 1$).

In one-dimension the flat zones created by the sieve have a length r. This means that the scale parameter can be used for the precise measurement of features [18]. For a two-dimensional image with regular pixelation the flat zones created by the sieve have a defined area and we refer to the sieve as an area decomposition c.f. [19]. In three dimensions we have a decomposition by volume.

2.2 Properties

One important property of the M- and N-sieves is that, for a particular edge, $\{x, y\} \in E$, as r increases, the change in image intensity, $f_r(y) - f_r(x)$, does not change sign and its absolute value never increases. In particular, if it vanishes for some r, it is then zero for all larger r. Formally, we have the following result.

Theorem 6. *Suppose that $(f_r)_{r=1}^{\infty}$ is the M- or the N-sieve of an $f \in \boldsymbol{Z}^V$, that $\{x, y\} \in E$ and put $\delta_r = f_r(y) - f_r(x)$, for each, r. Then $\delta_1 \geq \delta_2 \geq \cdots \geq 0$, or $\delta_1 \leq \delta_2 \leq \cdots \leq 0$.*

In other words n-dimensional sieves preserve scale-space causality.

At each stage of the M- or N-sieve one can examine either the filtered image (usually the case) or the difference between successive stages. These differences are called *granules* and are defined as follows.

Definition 7. When $f \in \boldsymbol{Z}^V$, the M-granule decomposition of f is the sequence of functions, $(d_r)_{r=1}^{\infty}$, defined by $d_r = f_r - f_{r+1}$, for each $r \geq 1$, and $\{d\}$ is the granularity domain. There is an equivalent definition for the N-granule decomposition.

Theorem 8 (Invertibility). *Suppose that $(d_r)_{r=1}^{\infty}$ is the M-granule decomposition of an $f \in \boldsymbol{Z}^V$ with finite support then the original image can be rebuilt from the granularity domain $f = \sum_{r=1}^{\infty} d_r$. Suppose that $(\hat{d}_r)_{r=1}^{\infty} \subseteq \boldsymbol{Z}^V$ are such that, for each integer, $r \geq 1$, $\hat{d}_r = 0$ on $[d_r = 0]$, $\hat{d}_r$ is constant and nonnegative on each positive granule of d_r and $\hat{d}_r$ is constant and nonpositive on each negative granule of d_r and let $\hat{f} = \sum_{r=1}^{\infty} \hat{d}_r$. Then $(\hat{d}_r)_{r=1}^{\infty}$ is the M-granule decomposition of $\hat{f}$.*

This last property is vital if one wishes to manipulate an image in the granularity domain. For example, a pattern recognising filter can be produced by: decomposition using M^r or N^r, selectively removing granules (filtering in the granularity domain); and rebuilding. Such a filter is idempotent.

3 Results

Figure 1(A) shows an image of a doll manually segmented from a larger image.

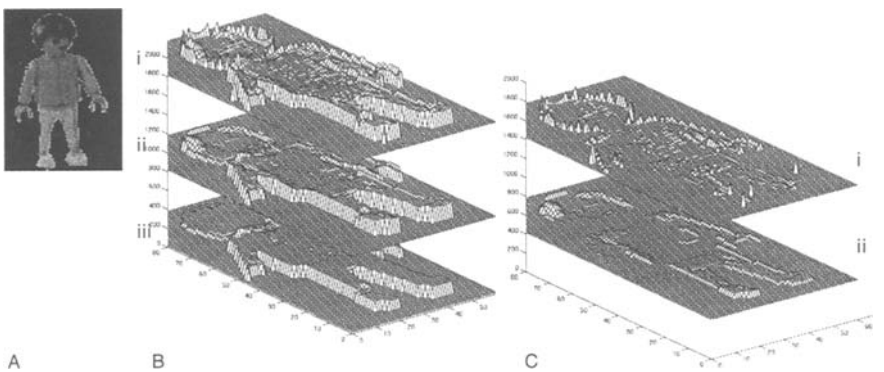

Fig. 1. (A) Shows the original image I; (B)(i) shows, in relief, the original image, B(ii) shows $M_4(I)$ and B(iii) shows $M_{25}(I)$. (C) shows the differences between B(i), B(ii) and B(iii)

It is shown as a topological relief in Figure 1(B). The spiky detail, particularly evident around the head, represents small scale regional extremes of area that are removed by sieving to scale 4, as in Figure 1(B)ii. The small extrema that are removed are shown in Figure 1(C)i (actually $M^4(I) - I$). Likewise the pale highlight on the (left side) of the hair, is a feature represented by a larger scale extremum that is removed with a larger, scale 25, area sieve. The differences between Figure 1(B)ii and (B)iii are shown in Figure 1(C)ii. No new features can be seen in Figure 1(B)iii and what remains is a set of large flat zones with edges in the same places as those in the original. This is consistent with Theorem 6 and Theorem 4. This localisation of features is an important advantage over, say, Gaussian scale-space which blurs the edges.

The process is essentially regional and is not equivalent to redefining the intensity map over the whole image. Nor does it alter the corners of objects in the way that a morphological or median filter with a rigid structuring element would be expected to.

However, there is a problem. In Figure 1(A) there are over 5,000 pixels and so, potentially, 5,000 different area scales. In practice, this is often far too many. Figure 2 (bottom right panel) shows an image of four pieces of string. The two

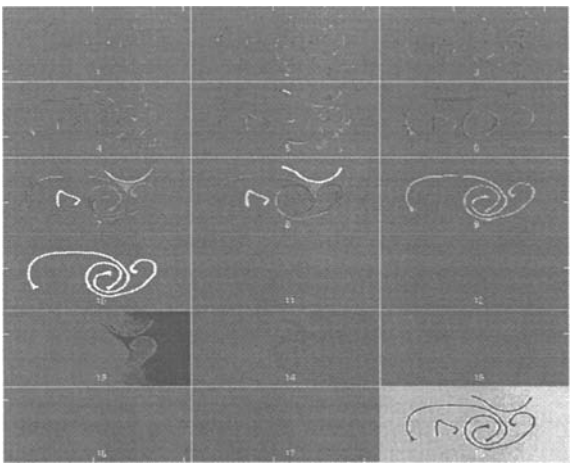

Fig. 2. Panel 18 shows the original image and panels 1–17 show an area decomposition

longer pieces are the same length. The image is decomposed into 17 area *channels*. Each channel is formed as the partial sum of granules from a range of scales, $c_i = \sum_{r=r_i}^{r=r_{i+1}} d_r$. Most of the activity due to the longer pieces of string is observed in channel 10, whilst that for the two smaller pieces can be seen in channels 7 and 8. Of course, since the pieces of string are not of uniform intensity, and there are shadows, activity is not confined to a single channel, rather it peaks at the appropriate scale.

This result would be difficult to achieve using a filter with a structuring element (for example, an alternating sequential filter by reconstruction [20] with square structuring element) because, a structuring element chosen to encompass the small bits of string would remove the more coiled of the long strings. The area sieve does not rely on a match between the shape of a structuring element and an object. The strings can be arranged in any shape (although, if the string forms loops it might be necessary to use a sieve that takes advantage of the sign of the local extremum, cf. [19]).

Figure 3 shows a more complicated example: a human face. Note the nostrils in channel 5, the eyes in channels 6 and 7, the mouth in channels 9 and 10 and the entire face in channel 15. The larger scale channels can be used for segmentation, for example, channel 15 could be used as a mask to pick out facial features from the background for, by Theorem 6, it accurately represents features in the original image. This image is one of a movie sequence in which the activity associated with the mouth moves from a peak in channel 9 to a peak in channel 10/11 and back.

Figure 4 shows a set of transverse scans of a human head generated by X-ray tomography. The sections start at the level of the nose and go up to the top of the head. A guide frame, which is opaque to X-rays, is attached to the head. The white, calcified bony tissue of the skull is readily identified and could be

Fig. 3. An area decomposition of an image (top) into a number of channels

segmented out by simple thresholding, however, we have chosen to use a very large scale volume sieve.

Figure 5(A) shows a rendered reconstruction of the volume channel containing between 25,600 and 51,200 voxels. The left cheekbone can be seen looping out at the bottom of the image, and the upper part of the nose is visible at the bottom left. The guide frame fastenings protrude from the skull.

The third ventricle is a complex shaped void within the brain. It appears as a darkened region circled in panel 16 of Figure 4 and runs up through the brain, bifurcating to form two "wings" that come together at the top. The primary segmentation is made by taking a volume channel encompassing 800 to 1600 voxels. This first step achieves a near perfect segmentation but some other features of the same volume, are also present. Since they do not penetrate to the same depth, they occupy a larger area of those slices in which they are present and so can be removed by area sieving each slice to remove areas larger than 100 pixels. This leaves some small regions, which are no longer connected through the slices, that can be removed by re-sieving through the original volume sieve. From Theorem 8 this leaves the third ventricle unchanged.

The final example shows how the sieve might be used to create compact feature vectors from an image. An isolated object in an image can be represented as a hierarchy of nested connected subsets of G. The pixels that form the graph

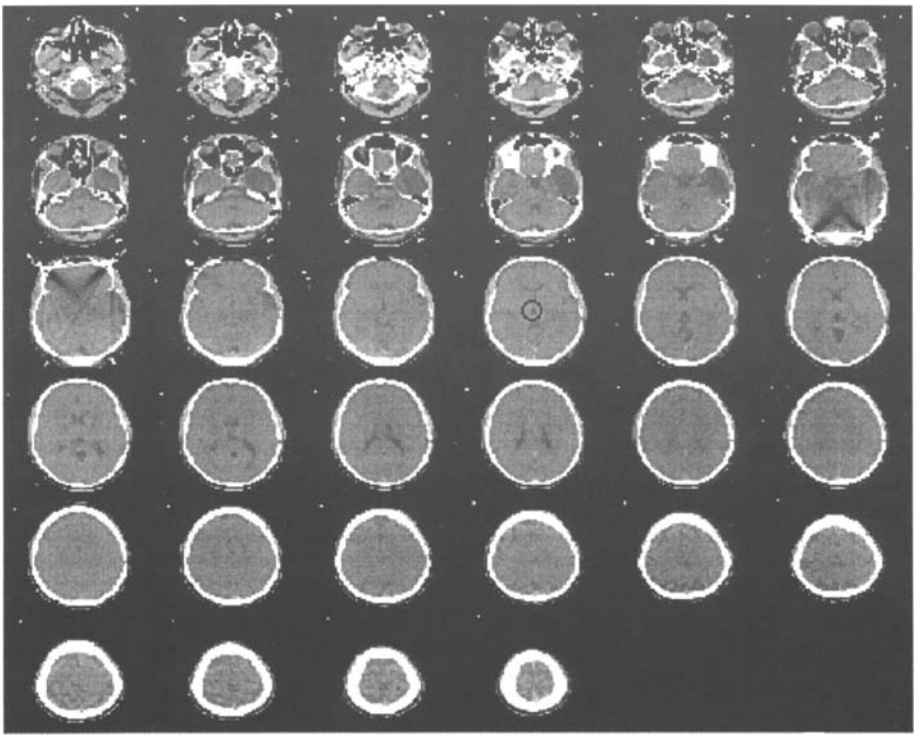

Fig. 4. Thirty-five Computer tomography scans through a human head. The third ventricle is circled in the sixteenth slice. Note thin connecting filaments are not rendered.

of a large granule may themselves be elements of smaller graphs that represent the finer scale features of that object. This is illustrated in Figure 6

In Figure 6 the top sketch shows an apparently meaningless collection of objects but in the lower sketch the shaded areas immediately identify two faces. The bar charts represent a count of the number granules at a particular scale enclosed within a particular object as a proportion of the total number of granules within the object. Both the faces in (B) contain the same distribution of area-granules, and so this feature vector can be used for rotation-independent pattern recognition; even though, in this case, the shaded regions have different shapes.

Figure 7(B) shows an area-channel obtained from the image in Figure 7(A) (a QuickTake image taken at a "freshers" party with a flash). Notice that the regions labelled 4, 8 and 11 are readily visible. This is because they form maxima relative to the background in their region.

Each of the regions in Figure 7(B) are well defined, for the edges correspond directly to intensity changes in the original image (a consequence of Theorem 6, but the faces are not necessarily identifiable from the outlines because the illumination is very uneven and the face yielding region 11 is partly occluded. The

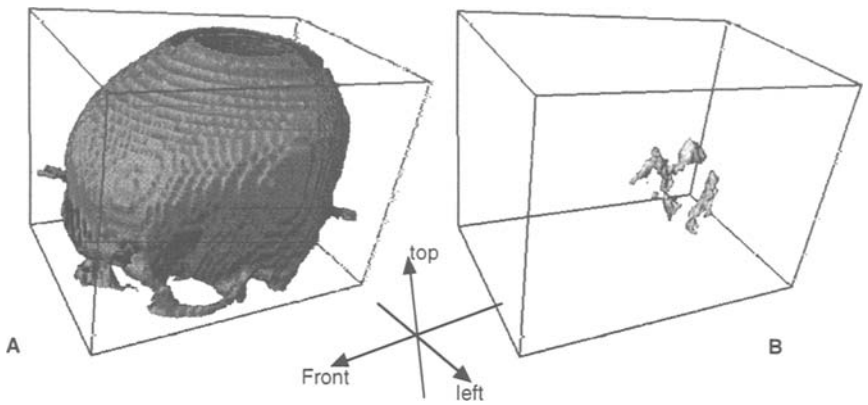

Fig. 5. (A) shows a rendered 3D image of the skull. (B) shows the third ventricle on the same co-ordinated system as (A)

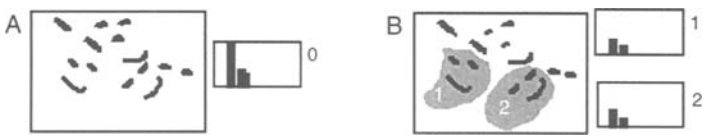

Fig. 6. (A) Shows a collection of shapes (black connected sets). In (B) the same shapes that have been grouped to make sense. Panel 0 shows granule density (described in the text) of A, Panels 1 and 2 show the density for features within sets 1 and 2.

set of segments do, however, form a starting point for developing a heuristic for finding the faces. Here we use the new approach, indicated in Figure 6, for obtaining a set of rotation-independent features for each region.

Each region is taken in turn, Figure 7(C), and examined to find the number of smaller area regions it contains at each scale. This is plotted on the ordinate, as a fraction of the total number of smaller area segments within each segment, as a function of scale, abscissa. Each of the plots is an intensity and rotation-independent signature of the regions in Figure 7(B) and, to the extent that plots 4, 8 and 11 are similar to each other and differ from the others, they represent a way of distinguishing faces from the background. In this experiment only regions 1 and 6 might be confused with faces on this criterion alone. A better feature vector might include other components (measurements made with lower dimensional sieves for example) that would allow more reliable recognition. Here, our purpose is only to illustrate the principle of using area decomposition to aid rotation independent recognition.

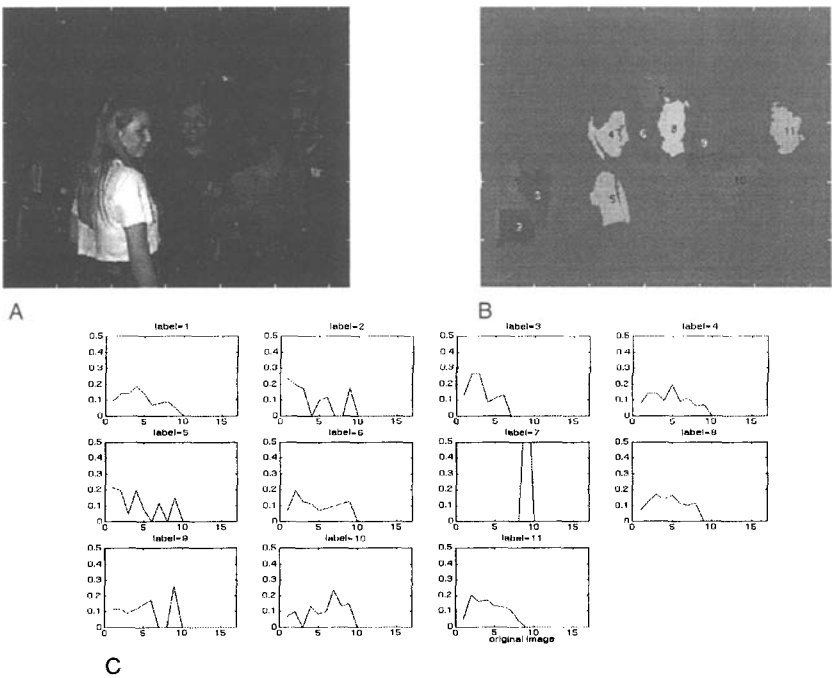

Fig. 7. (A) original image; (B) labelled single channel of original image; (C) granule densities for each of the regions in (B)

4 Conclusion

A nonlinear method of generating scale-space has been presented. It has the following features: it preserves scale-space causality; sharp-edged objects in the original image remain precisely localised; it operates on images defined in any finite dimension; it is scale-calibrated; the image may be processed in the scale-space domain; the decomposition appears to offer semantically meaningful feature vectors.

The method presented here is unusual, and may be unique, in having all these properties. Although the theory has been presented for M- and N-sieves our experimental evidence suggests that one can form an n-dimensional sieve using recursive median filters and it has the above properties. This has been proved in the one-dimensional case and such sieves have been shown to have good performance in noise [14]. The empirical evidence suggests that the n-dimensional sieve is also robust.

The one-dimensional version is extremely quick to compute and the n-dimensional version may also be efficiently coded although the order complexity has yet to be presented. Initial studies show that the n-dimensional sieve can form a valuable part of computer vision systems.

References

1. J. Andrew Bangham, Paul Ling, and Richard Harvey. Scale-space from nonlinear filters. In *Proc. First International Conference on Computer Vision*, pages 163–168, 1995.
2. J.J.Koenderink. The structure of images. *Biological Cybernetics*, 50:363–370, 1984.
3. Tony Lindeberg. *Scale-space theory in computer vision.* Kluwer Academic, Dordrecht, Netherlands, 1994.
4. Bart M. ter Harr Romeny, editor. *Geometry-driven diffusion in Computer vision.* Kluwer Academic, Dordrecht, Netherlands, 1994. ISBN 0-7923-3087-0.
5. Pietro Perona and Jitendra Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Patt. Anal. Mach. Intell.*, 12(7):629–639, July 1990.
6. G.Matheron. *Random sets and integral geometry.* Wiley, 1975.
7. J.Serra. *Image analysis and mathematical morphology Volume 2: Theoretical Advances*, volume 2. Academic Press, London, 1988. ISBN 0-12-637241-1.
8. Paul T. Jackway and Mohamed Deriche. Scale-space properties of the multiscale morphological dilation-erosion. In *Proc. 11th IAPR Conference on Pattern Recognition*, 1992.
9. R. van den Boomgaard and A. Smeulders. The morphological structure of images: the differential equations of morphological scale-space. *IEEE Trans. Patt. Anal. Mach. Intell.*, 16(11):1101–1113, November 1994.
10. Corinne Vachier and Fernand Meyer. Extinction value: a new measurement of persistence. In Ionas Pitas, editor, *Proc. 1995 IEEE Workshop on nonlinear signal and image processing*, volume 1, pages 254–257, JUNE 1995.
11. M. H. Chen and P. F. Yan. A multiscale approach based upon morphological filtering. *IEEE Trans. Patt. Anal. Mach. Intell.*, 11:694–700, 1989.
12. 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.
13. J. Andrew Bangham, Paul Ling, and Richard Harvey. Nonlinear scale-space in many dimensions. Internal report, University of East Anglia, 1995.
14. J. Andrew Bangham, Paul Ling, and Robert Young. Multiscale recursive medians, scale-space and transforms with applications to image processing. *IEEE Trans. Image Processing*, pages –, January 1996. Under review.
15. J.Serra and P.Salembier. Connected operators and pyramids. In *Proc. SPIE*, volume 2030, pages 65–76, 1994.
16. Luc Vincent. Morphological grayscale reconstruction in image analysis: applications and efficient algorithms. *IEEE Trans. Image Processing*, 2(2):176–201, April 1993.
17. J. A. Bangham, S. J. Impey, and F. W. D. Woodhams. A fast 1d sieve transform for multiscale signal decomposition. In *EUSIPCO*, 1994.
18. J. A. Bangham, T. G. Campbell, and M. Gabbouj. The quality of edge preservation by non-linear filters. In *Proc. IEEE workshop on Visual Signal Processing and Communication*, pages 37–39, 1992.
19. Luc Vincent. Grayscale area openings and closings, their efficent implementation and applications. In Jean Serra and Phillipe Salembier, editors, *Proc. international workshop on mathematical morphology and its applications to signal processing*, pages 22–27, May 1993.
20. P. Salembier and M. Kunt. Size sensitive multiresolution decomposition of images with rank order based filters. *Signal Processing*, 27:205–241, 1992.