ROBUST AND FLEXIBLE MULTI-SCALE MEDIAL AXIS COMPUTATION

by

MING XU

A thesis submitted to the Faculty of Engineering
of The University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

School of Electronic and Electrical Engineering
Faculty of Engineering
The University of Birmingham
January 2001

Abstract

The principle of the multi-scale medial axis (MMA) is important in that any object is detected at a blurring scale proportional to the size of the object. Thus it provides a sound balance between noise removal and preserving detail. The robustness of the MMA has been reflected in many existing applications in object segmentation, recognition, description and registration. This thesis aims to improve the computational aspects of the MMA.

The MMA is obtained by computing ridges in a “medialness” scale-space derived from an image. In computing the medialness scale-space, we propose an edge-free medialness algorithm, the Concordance-based Medial Axis Transform (CMAT). It not only depends on the symmetry of the positions of boundaries, but also is related to the symmetry of the intensity contrasts at boundaries. Therefore it excludes spurious MMA branches arising from isolated boundaries. In addition, the localisation accuracy for the position and width of an object, as well as the robustness under noisy conditions, is preserved in the CMAT. In computing ridges in the medialness space, we propose the sliding window algorithm for extracting locally optimal scale ridges. It is simple and efficient in that it can readily separate the scale dimension from the search space but avoids the difficult task of constructing surfaces of connected maxima. It can extract a complete set of MMA for interfering objects in scale-space, e.g. embedded or adjacent objects. These algorithms are evaluated using a quantitative study of their performance for 1-D signals and qualitative testing on 2-D images.

to my wife, Huiju

Acknowledgements

I would like to thank my supervisor David Pycock for his leading me into this marvellous research area and the time spent on reviewing this thesis.

I would like to thank my examiners, Prof. Michael Brady in University of Oxford and Dr. Michael Spann, for their helpful remarks and discussion.

I am very grateful to the Committee of Vice-Chancellors and Principals (UK) and the School of Electronic and Electrical Engineering in University of Birmingham for their funding my study.

I would like to thank Prof. Stephen Pizer and his colleagues in University of North Carolina at Chapel Hill for their continuously providing us papers, theses, software as well as advice on the MMA, which have contributed to this research at different stages.

I am very grateful to Dr. Tim Ellis, who I am now working with in City University, for his encouragement for my thesis writing and reviewing a part of this thesis. I am also very grateful to James Black in City University for his reviewing a part of this thesis.

I would like to thank Dr. Sridhar Pammu for his maintaining the computer systems, Dr. Xinquan Shen for his advice on scientific research, and Dr. Mounther N. Salous for his help in using software and facilities.

Finally, I would like to thank my wife Huiju Zheng for her sacrifice and love, and thank my parents for their love and remote blessing.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

1-D	One dimension(al).
2-D	Two dimension(al).
3-D	Three dimension(al).
CA	Credit Attribution.
CMAT	Concordance-based Medial Axis Transform.
CT	Computer Tomography.
DoD	Distance of Distinguishability.
DoP	Distance of Peaks.
GRSM	Global Ridge Strength Measure.
HMAT	Hough-like Medial Axis Transform.
HMAT-2	An adaptation of the Hough-like Medial Axis Transform.
KLD	Koller Line Detector.
LoG	Laplacian of Gaussian.
LUT	Look-up-Table.
MAT	Medial Axis Transform.
MMA	Multi-scale Medial Axis.
MR	Magnetic Resonance.

Chapter 1 Introduction

1.1 The Importance of Multi-Scale Medial Axis

The solution of many problems in computer vision depends on the ability of image processing algorithms to adequately represent the shapes of objects in a scene. The classical approach to shape description in grey-level images has been to apply some sort of edge detection operation and represent objects based on their boundaries, e.g. using chain codes and B-splines. This approach has two difficulties. First, while boundary-based representations may be used to describe properties such as the length, local curvature and orientation of the boundary, they fail to directly capture more global properties of the object, such as the end-to-end length, length-to-width ratio, overall orientation, symmetry, etc. Secondly, edge detection is especially sensitive to intensity variations and noise. In a discrete image, edges cannot be reliably extracted without some notion of spatial scale over which to measure discontinuity (edge). The use of an unduly small scale emphasises fine detail including noise. The use of an unduly large scale distorts the form of detected objects and can lead to many details not being detected. Often any one image will involve structures at several scales. Therefore, it is not trivial to choose an appropriate scale.

A solution to the first difficulty above is Blum's medial axis transform (MAT) [BLU67]. In this scheme, a binary object is represented by the locus of centres of maximal disks inscribed within the object, together with the radii of these disks. Each medial axis point is associated with two or more boundary points where the maximal disk tangentially touches the boundary. In this way, the MAT establishes the relationship between two or more boundary sections separated by the width of the object, and can provide a greater level of global information, such as the overall length and orientation, symmetry, and changes of local width (narrowing or flaring). In addition, Blum *et al.* have shown that the spatial position of the axis, and the radius of the maximal disk encode the position, orientation and curvature of boundaries [BLU78]. These properties of the MAT express global and local shape properties.

A solution to the difficulty of selecting an appropriate scale is provided by scale-space theory [WIT83] [KOE84] and the mechanism of automatic scale selection [FRI93] [LIN94] (see Section 2.1 for details). The medial axis transform, combined with these theories, leads to the concept of the multi-scale medial axis (MMA) [CRO84] [FRI93] [PIZ94] [MOR94]. The MMA for an object is defined as a set of curves in scale-space. Each spatial position on a curve marks the middle of the object and is associated with a spatial scale that indicates the approximate width of the object. For a 2-D image, the MMA is obtained by computing the 1-D ridges in a 3-D “medialness” scale-space derived from the image.

The significance of the MMA is that the geometric measurement of the medial axis and local width is made at a scale proportional to the size of the object of interest. Because larger objects can naturally withstand a greater amount of blurring than smaller objects, the MMA method uses measurement scales according to the context across the image: it removes more noise and detail for larger objects, while removing less noise and retaining detail for smaller objects. This is in contrast with the use of a single scale in many existing methods. In addition, the multi-scale medial axes are invariant to spatial translation, rotation and scaling of the image plane as well as linear variation of image intensity [PIZ98]. They also have a low sensitivity to noise (spatially uncorrelated), blurring (compared to the object’s width), and shape variation [MOR98]. Therefore, the multi-scale medial axis provides a robust tool for shape representation.

1.2 Applications of Multi-Scale Medial Axis

The invariance of the multi-scale medial axis to spatial translation, rotation and scaling and its insensitivity to disturbances on the form (intensity plus shape) provide a good basis for various image analysis tasks, including segmentation, recognition and registration.

(1) Segmentation and Visualization

The MMA has been widely used in the detection of blood vessels in medical images [KOL95] [FEL97] [KRI98] [SAT98] and roads for remote-sensing [KOL95] [STE98]. The blood vessels and roads in these images are symmetric structures where the scale of symmetry varies with context. This characteristic is directly addressed by the MMA. The detected MMA can even help reconstructing the vessels and providing improved visualization.

Maximum Intensity Projection (MIP) has long been used to display 3-D vessel images (e.g. MR angiography) from a single view. For a given image plane, S, this method assigns each pixel, P, on S a value equal to the maximum of those pixels that have P as their common projection point on S (Fig. 1.1(a)). However, it contains no information about the relative depth of the vessels. One vessel can partially obscure another. On the other hand, an iso-surface representation of the initial image can account for the relative depth of the vessels, but the selection of an appropriate threshold is critical (Figs. 1.1(b) and (c)). Krissian *et al.* [KRI98] computed the multi-scale medial axes for these vessels and reconstructed the vessels using the scale (width) information of the MMA. The result is more complete and better connected interpretation (Fig. 1.1(d)).

(2) Description and Recognition

Fritsch *et al.* described an object in an MMA hierarchy ordered by scale [FRI95]. The large-scale MMA for the global figure is a parent to its child MMA corresponding to small-

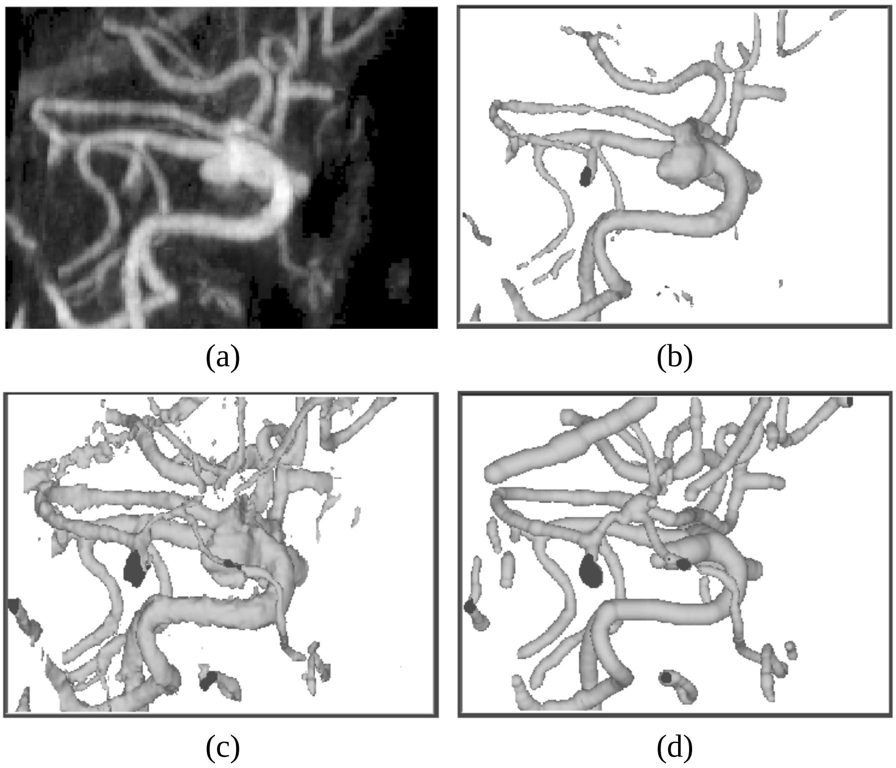


Fig. 1.1. The MMA used in vessel detection and visualization: (a) maximum intensity projection of a 3-D vessel image; (b) and (c): iso-surfaces of the image with high and low thresholds, respectively; (d) vessels reconstructed using the MMA. (Courtesy of [KRI98])

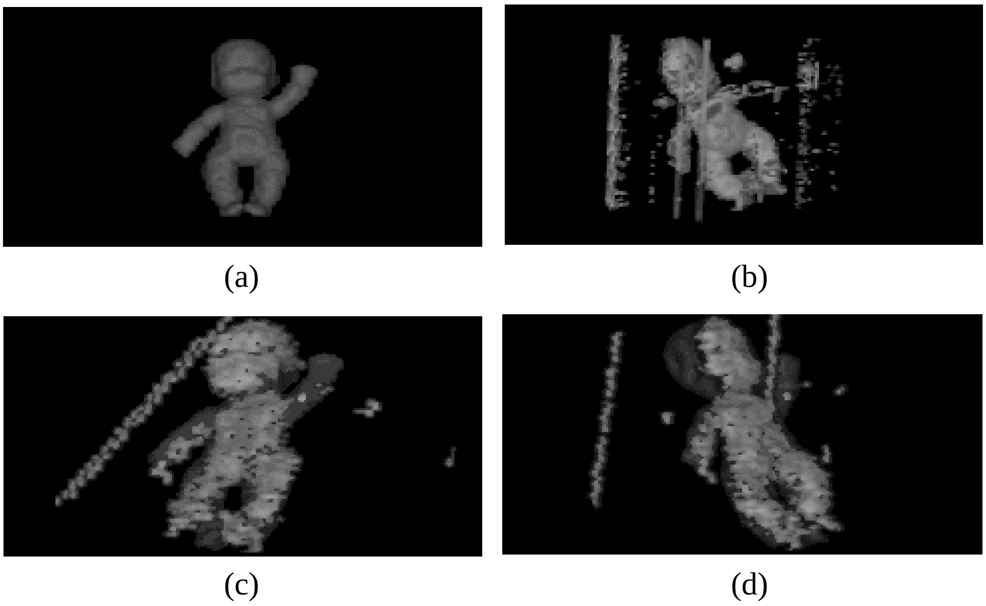
scale sub-figures, e.g. protrusions and indentations. This leads to a graph description of the object, the nodes of which contain statistical information related to the individual MMA, and the arcs of which contain information related to the sub-figure type (e.g. protrusion or indentation) and the relationship of the sub-figure MMA to its parent (e.g. relative position, size and orientation). From such a graph, for a given object, from a population of different images, one can generate an object model that can in turn be used to automatically recognize the same object in another image. Such models typically contain both means and variances in its nodes and arcs to provide a flexible model description.

(3) Registration and Fusion

Fritsch developed an automatic procedure for verifying treatment setup in radiotherapy via the registration of portal image pairs using the MMA [FRI93]. The spatial location of the

MMA was used to define a set of matching axes in a reference, portal image. The same set of labeled axes was identified in a treatment portal image and automatically registered. Axis correspondence was determined using moment-based shape measure computed on the set of all possible MMA pairs in both images. It was shown, in a series of simulated images, that this automatic procedure could determine setup errors of 1mm in translation and 1 degree in rotation [FRI93].

Data fusion by the registration of CT and ultrasound images using the MMA was demonstrated by Liu *et al.* [LIU94], see Fig. 1.2. The original ultrasound image misses the entire back of the doll, and one of the arms is at different positions in the two images (Figs. 1.2(a) and (b)). However, the robustness and richness of the MMA representation still allows the two images to be registered using the principal MMA branches for the torso, extracted from both images. The fused information from the more detailed CT image provides detail that is not present in the ultrasound image (Figs. 1.2(c) and (d)).



(a) (b)

(c) (d)

Fig. 1.2. Image fusion using MMA-based registration of 3-D images: (a) volume renderings of a CT image and (b) an ultrasound image for a baby doll, (c)(d) two views showing the fusion . (Courtesy of [LIU94])

1.3 Contributions of This Thesis

This thesis aims to improve the computational aspects of the MMA and makes the following contributions to the multi-scale medial axis method:

(1) We propose an edge-free medialness algorithm, the Concordance-based Medial Axis Transform (CMAT). The CMAT medialness not only depends on the symmetry of the positions of boundaries, but also the symmetry of the intensity contrasts at boundaries. Therefore, it excludes the spurious MMA branches that arise from isolated edges and appear in traditional medialness algorithms.

(2) We propose a sliding window algorithm for extracting locally optimal scale ridges in the medialness scale-space. This algorithm is simple and efficient being able to readily separate the scale dimension from the search space whilst avoiding the difficult task of constructing surfaces of connected medialness maxima. In contrast to the globally optimal scale ridge definition [FRI93], it can extract a complete set of MMA for assemblies of objects, e.g. embedded or adjacent objects.

(3) We set up a framework for comparing the performance of different medialness functions using the “operator radius” rather than the standard deviation of the Gaussian (or its derivative) as the scale parameter. This allows various medialness functions to be quantitatively compared. So far as the author is aware a detailed quantitative comparison of these medialness algorithms has not previously been reported.

(4) We analyze the abilities of scale-space and the globally optimal scale ridge definition [FRI93] to distinguish objects within an embedded object group. Based on this analysis, we give some indications on the selection of scale sampling rate, which is still an open problem and an active research area in scale-space theory.

1.4 Organization of This Thesis

In Chapter 2 the concepts of scale-space and medial axis transform are introduced. Then different schemes for robustly extracting medial axes are reviewed and the account of multi-scale medial axis method extended.

In Chapter 3 the computation of boundariness is introduced. The Concordance-based Medial Axis Transform is described and the related concordance property presented. The limits of the globally optimal scale ridge definition to distinguish embedded objects are demonstrated, and the sliding window algorithm for extracting locally optimal scale ridges in medialness scale-space introduced.

In Chapter 4 the performance of the CMAT medialness in the 1-D case is analyzed quantitatively and compared with, selected, traditional medialness operators. The performance of the CMAT on 2-D data sets and of the sliding window algorithm are also demonstrated.

In Chapter 5 the relationship of the CMAT medialness algorithm to existing edge-free algorithms and the relationship between the sliding window algorithm to existing scale-space ridge detection algorithms are discussed.

Finally, in Chapter 6, conclusions are drawn and suggestion is made for future work.

Chapter 2 Literature Review

2.1 Multi-Scale Analysis

2.1.1 The Nature of Scale

The problem of scale is faced in any physical measurement such as imaging. An inherent property of objects in the world and in images is that they only exist as meaningful entities over a range of scales [KOE84]. The result of a measurement process depends on the scale at which this measurement is made. A simple example of this is given by an image of trees. At a spatial scale of centimetres, they are perceived as branches and leaves. As the scale increases to a metre or so, the perception is of the outline of the trunk and canopy. As the scale further increases to kilometres, the forest of trees is perceived. The correct selection of an appropriate scale is not normally difficult to determine in this situation for a known task. However, the scale selection problem cannot be avoided when analyzing unfamiliar scenes. One such example is the detection of edges corresponding to the discontinuities of the image function. Edge detection is usually computed as a difference between pixels in some neighbourhood. There is seldom a sound reason for choosing a particular size of neighbourhood, since the “right” size depends on the sharpness of the edge under investigation. This size is in general unknown at the pre-processing stage and may vary across an image. Therefore, to avoid missing any useful information, we would like to treat the scale of the observation as a free parameter and analyze objects at a range of scales.

When an image is captured the finest scale is defined and the multi-scale process can only be simulated by convolving the original image with progressively larger blurring operators. For this multi-scale representation to behave in a reasonable way, some general constraints need to be imposed on the operators for the construction of scale-space.

One crucial constraint is that no new detail should be generated as scale increases [KOE84] [WIT83] [YUI86] [LIN92]. Koenderink introduced the concept of causality [KOE84], which means that new level surfaces must not be created in the scale-space representation as scale increases. Lindeberg determined that the operators should not increase the number of local extrema in any 1-D signal under convolution [LIN92]. Another crucial constraint for scale-space operators is that the scale-space representation must be invariant to spatial translation, rotation and scaling [ter91] [FLO92]. Each of these two constraints needs to be further combined with some other requirements, e.g. linearity and semi-group property. Although the general constraints for constructing scale-space representations are presented in a variety of ways by different authors, they each reach the same conclusion, that the Gaussian kernel and its derivatives are the uniquely appropriate blurring operator when considering a wide range of contexts in which little, if any, prior knowledge is available.

2.1.2 Linear Scale-Spaces and Diffusion

The linear scale-space of a signal, as introduced by Witkin [WIT83] and further developed in [KOE84] [BAB86] [YUI86] [LIN94], is an embedding of the original signal into a one-parameter family of derived signals constructed by convolution with Gaussian kernels of increasing width. Let $\mathbf{x} = (x_1, x_2, \cdots\cdots, x_N)$ denote a spatial position in N-dimensional space and $I(\mathbf{x})$ represents a given signal. Then the scale space representation, $L(\mathbf{x},\sigma)$, of the signal, $I(\mathbf{x})$, is defined as:

$$\begin{aligned} &L(\mathbf{x},\sigma) = I(\mathbf{x}) * G(\mathbf{x},\sigma) \\ &L(\mathbf{x},0) = I(\mathbf{x}) \end{aligned} \tag{2.1}$$

where σ is the scale parameter; $G(\mathbf{x},\sigma)$ is the Gaussian kernel with standard deviation σ:

$$G(\mathbf{x},\sigma)=\frac{1}{\left(2\pi\sigma^2\right)^{N/2}}\exp\left(-\frac{\mathbf{x}^T\mathbf{x}}{2\sigma^2}\right) \tag{2.2}$$

Under the progressive Gaussian blurring, it was found that [WIT83]:

- The derived signals become smoother with increasing scale (Fig. 2.1(a)). For a 1-D signal, new local extrema of any order spatial derivative cannot be created at a larger scale [LIN94].
- The location of each feature point (e.g. a local maximum, an edge or a zero-crossing of the derivative of the signal) changes continuously with scale. Therefore, it is possible to track and relate the feature points across scale.
- Feature points annihilate in pairs with increasing scale. The scale at which a feature point annihilates indicates the significance of the feature point.
- The location of each feature point at coarse scale is shifted from its fine scale location (Fig. 2.1(b)). The localisation of a feature point is most accurate at finest scale and can be traced from coarse to fine scale to combine the advantages of robustness and

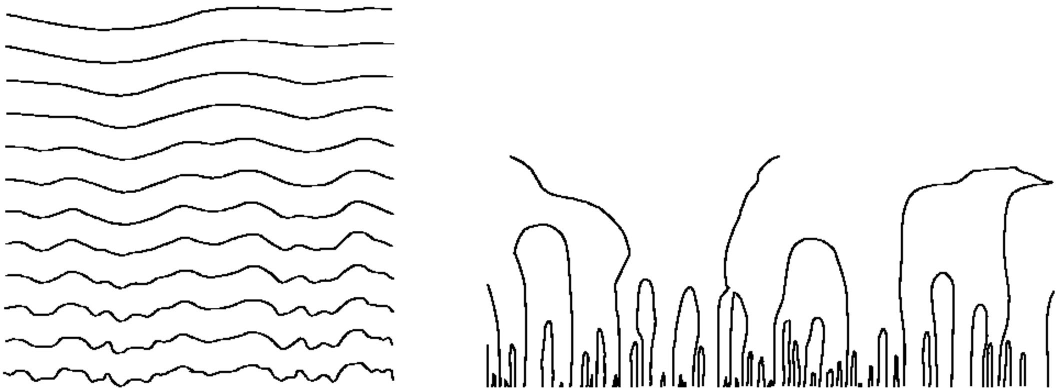

(a) (b)

Fig. 2.1. The scale-space representation of a signal (a) and the zero-crossings of its second derivative (b). (Courtesy of [WIT83])

accuracy [BER87].

In addition to convolving the signal with a set of Gaussian kernels of increasing size, another way to generate the scale-space is using the *diffusion equation* [KOE84]. The diffusion equation is the well-known physical equation that describes how a heat distribution, L, evolves over time t in a homogeneous medium with uniform conductivity c, given an initial heat distribution $I(\mathbf{x})$ [WID75]. The diffusion equation is expressed as:

$$\begin{aligned} \frac{\partial L(\mathbf{x},t)}{\partial t} &= c\nabla^2 L(\mathbf{x},t) \\ L(\mathbf{x},0) &= I(\mathbf{x}) \end{aligned} \tag{2.3}$$

At local maxima of L, $\nabla^2 L < 0$; thus $\partial L / \partial t < 0$ and L decreases. At a local minima of L, $\nabla^2 L > 0$; thus $\partial L / \partial t > 0$ and L increases. Therefore the diffusion equation has the effect of suppressing small local variations. It was found that computing the scale-space representation of a signal at scale σ corresponds to iteratively diffusing the signal using the diffusion equation for a time $t = \sigma^2 / 2$ [KOE84]. Each step of the iterative diffusion is analogous to convolving the scaled image, computed at the previous step, by a much smaller Gaussian kernel. Due to the semi-group property of the Gaussian kernels, the result of convolving a Gaussian kernel with a Gaussian kernel is another, larger Gaussian kernel, i.e.:

$$G(\mathbf{x},\sigma_1) * G(\mathbf{x},\sigma_2) = G(\mathbf{x},\sqrt{\sigma_1^2 + \sigma_2^2}) \tag{2.4}$$

2.1.3 Anisotropic Diffusion

Under the above uniform diffusion, the location of features, e.g. edges, at a coarse scale is displaced from their original location. To improve the localisation accuracy for feature points, it is desirable to restrict the diffusion of information around these feature points [PER90] [WHI93], using anisotropic diffusion.

By considering the conductivity, c in Eq. (2.3), to be a variable over space and time (scale), the anisotropic diffusion equation is defined as:

$$\frac{\partial L(\mathbf{x},t)}{\partial t} = \nabla \cdot \left(c(\mathbf{x},t) \nabla L(\mathbf{x},t) \right) \tag{2.5}$$

For the edge detection in scale-space, to encourage diffusion within a region in preference to diffusion across boundaries, the conductivity, c, can be chosen as a monotonically decreasing function of the gradient magnitude. This anisotropic diffusion process corresponds to applying a larger Gaussian kernel within a region and a smaller Gaussian kernel across boundaries.

The anisotropic diffusion is sometimes criticised because it seeks to obtain a robust feature detection at large scales by successively diffusing the signal, but it uses the estimates of these features, at each scale including small scale, to control the diffusion process. Therefore, the signal-to-noise ratio in the original image is critical to the success of the feature detection.

2.1.4 Feature Detection in Scale-Space

Although in principle the Gaussian kernel is all one needs to generate a scale-space, the result is only a set of scaled versions of a given image. This is highly insufficient for a complete local description of the image structure. For a vision system to be able to derive a variety of symbolic representations from images, it is required to combine the convolution output of the Gaussian derivative operators of different orders and at different scales into a more explicit description of image geometry. Florack, ter Haar Romeny *et al.* [FLO92] have constructed a complete hierarchical family of up to N-order derivatives of scale-space filters (N may approximate infinity), called *N-jets*, and shown that this family of derivatives is sufficient for a complete determination of the local image structure. On the other hand, given a weakly bounded input, the scale-space derivatives are infinitely differentiable and guaranteed to converge at any scale [LIN94], which facilitates the (possibly non-linear) combination of scale-space derivatives.

Of the many possible schemes for combining scale-space derivatives, an important class generates outputs that do not depend on the arbitrary choice of the image coordinate system, i.e. the output values are invariant to the rotation and translation of the image function. This ensures that geometric responses reflect image geometry only. A simple example of a geometric invariant is the first-order gradient magnitude, $\|\nabla L(\mathbf{x},\sigma)\|$, which captures edge-like properties [CAN86]. Another example is the second-order Laplacian, $\nabla^2 L(\mathbf{x},\sigma)$, which captures edges [MAR80], blobs [LIN94] and ridges [FRI93]. More complex combinations of scale-space derivatives for detecting corners, junctions and ridges can be found in [ter91] [FLO92] [LIN94] [LIN98].

2.1.5 Automatic Scale Selection

Although scale-space theory describes how information can best be combined across scale, it does not indicate how to select the most appropriate scale for further analyses. Fritsch, Pizer *et al.* [FRI93] [PIZ94] and Lindeberg [LIN94] have contributed to the development of a framework that incorporates the mechanism of automatic scale selection in scale-space feature detection. In scale selection, response from a feature detector (i.e. some combination of normalised Gaussian derivatives) is compared across scale. For each spatial position, the scale at which the detector assumes a local maximum response reflects the characteristic length of underlying image features observed from that position and is likely to be the “optimal” scale for the analysis of that position. Such “optimal” scales are usually varying across an image, e.g. increasing along the middle line of a flaring shape.

When the size (scale) and structure (feature type) of a feature detector is best fit to a local region of the form (intensity plus shape), the strongest response will be generated at this scale that is proportional to the characteristic length of the image feature of interest. This characteristic length has been proved to correspond to the radius of a “blob”, the width of an elongated object, the width of edge slope (diffuseness), or the wavelength of a sinusoidal signal [FRI93] [LIN98].

The significance of automatic scale selection is the adaptive scale selection within an image. Robustness is enhanced by using the largest degree of smoothing that does not sacrifice image structure. Scale selection allows the most appropriate scale of smoothing to automatically be used in each part of an image.

2.2 Medial Axis Transform (MAT)

2.2.1 Definitions

The medial axis of a planar shape is the locus of centres of maximal disks inscribed within that shape [BLU67], see Fig. 2.2(a). When each medial axis point is associated with the radius of its maximal disk, it is referred to as medial axis transform (MAT), because the shape can be recovered as the union of these maximal disks. In an equivalent definition of the MAT, known as the “prairie fire”, Blum considered a fire front initiated simultaneously on all the boundaries and propagating with a constant speed inside the shape. The medial axis is the locus of points where fire fronts originating from different boundary points meet and then quench, see Fig. 2.2(b). The time at which the fire front reaches the quenching point corresponds to the radius of the maximal disk [BLU67].

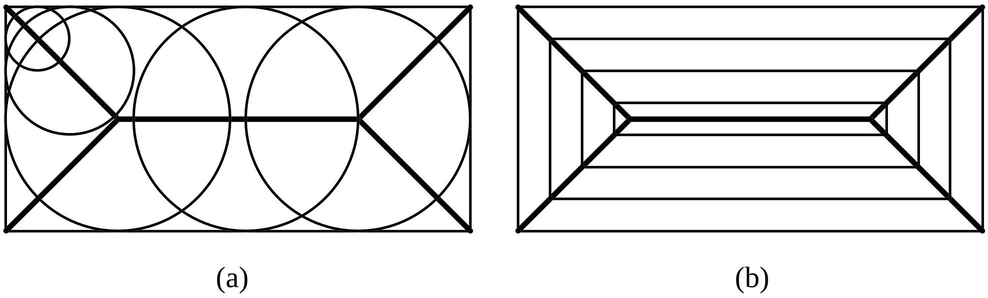

(a) (b)

Fig. 2.2. Medial axis definitions using (a) maximal disks and (b) prairie fire model.

Some equivalent definitions of the MAT can be found in [MON68]. In Blum’s later symmetric axis transform (SAT) [BLU73], the restriction that the maximal disks must lie entirely within the object was removed, and thus the SAT can capture additional symmetries. For example, the SAT of the above rectangle has a horizontal axis as long as the shape and an additional vertical axis as high as the shape. This concept is equivalent to the prairie fire that continues to propagate after meeting other fire fronts. Blum’s work on the MAT led to a number of other definitions focusing on the medial properties of objects. These include the smoothed local symmetries (SLS) [BRA84], which is the locus of the midpoint of the chord connecting the two tangent points of each maximal disk, and the process inferring symmetric axis (PISA) [LEY87], which is the locus of the midpoint of the arc connecting the two tangent points of each maximal disk. However, these transforms are not readily described in terms of object boundary [MOR95].

The MAT provides a direct encoding of local properties of object shape, such as boundary orientation and curvature, and of global properties, such as overall length and orientation [MOR95]. Another attractive property of the MAT is that the branching structure of an object is reflected by the branching of the axes. The graphical structure facilitates the decomposition of a complicated shape into simple parts. In this case, branching points of the medial axes or local width minima along the medial axes separate each connected medial axis tree into segments. The union of the maximal disks of each axis segment naturally constitutes a simple part of the shape.

2.2.2 MAT for Binary Shapes

Many attempts have been made to implement the MAT for segmented (binary) shapes in digital images, though they may be referred to as different names such as skeletonization and thinning. These methods often fail to preserve one or another fundamental property such as connectivity and Euclidean metric. Most of these algorithms can be classified as:

(1) Topological thinning

Most algorithms in the literature are based on topological thinning or morphological erosion. They repeatedly peel off the pixels on the contour of an object whenever the removal does not change the shape topology. This is an approximation of the “prairie fire” process.

Different sequential [HIL69] [ARC78] [DIL87] and parallel [CHI87] [GUO89] implementations exist. Because of the anisotropic nature of the commonly used rectangular grids, these algorithms tend to use a regular metric, e.g. city block and chessboard, rather than Euclidean metric, to preserve connectivity. While the connectivity of the resulting skeletons can be guaranteed, their geometric accuracy is limited and their representations are rotation dependent.

(2) Analytical computation

Another class of algorithm computes the symmetric axes from a polygonal approximation of a shape. Some methods find pairs of opposite boundary line segments and compute the midlines between them [BOO79] [SHA81]. Many other algorithms are based on the computation of a Voronoi diagram (VD) of the boundary line segments [MON69] [LEE82] [BRA92] [OGN95]. These algorithms split the polygonal shape into regions separated by the VD of the boundary line segments or its dual the Delaunay triangulation (DT), and delete the peripheral DT cells which are not relevant to the description of the shape. The final result is the Voronoi skeletons. Recursive deletion of DT cells may be avoided if an adequate measure for the “relevance” of each VD branch is chosen. The problem here is that a polygonal approximation is often insufficiently accurate for general shapes, and introduces numerous additional skeleton branches.

(3) Distance transform (DT)

A distance transform assigns each pixel within a shape a value equal to the distance to the nearest boundary point, according to some metric. This results in a distance map for that shape. The skeletons of the shape are the ridges of the distance map. Different metrics can be used to compute the distance transform. Skeletons based on the regular metric [ROS66] [ARC89], e.g. city block and chessboard, can be computed very quickly and are assured to be connected, but they are not accurate and not robust under rotations. Methods based on Euclidean or quasi-Euclidean metrics [DAN80] [HO86] [KLE87] [WRI93] are accurate, but disconnection of skeletal branches requires special treatment.

2.2.3 MAT for Grey-Level Shapes

After Blum's MAT definition for binary shapes, several generalizations of the MAT to grey-scale images were proposed in the early literature. According to the MAT definition that they are based on, these algorithms can be classified as:

(1) Methods based on the "maximal disks"

Ahuja *et al.* found maximal homogeneous disks in a grey-scale image. The set of centres, radii, and average grey levels of these disks defines a generalized MAT, called the spatial piecewise approximation by neighbourhoods (SPAN)[AHU78]. Pal *et al.* defined the fuzzy medial axis transformation (FMAT) [PAL92], which uses a union of maximal fuzzy disks to represent a grey-scale image. The membership value in a fuzzy disk is the minimum intensity among those points that have the same distance from the centre of this disk according to the regular metric.

(2) Methods based on the "prairie fire" model

Peleg *et al.* defined the min-max medial axis transform (MMMAT) [PEL81], in which min and max operations are iterated over a neighbourhood. This is analogous to the binary "shrinking" and "expanding" operations of binary mathematical morphology. Wang *et al.* proposed the GRADMAT [WAN82], which computes a score, for each point P, by accumulating edge response maxima around a circle centred at P and at a range of circle radii. The points with a high score lie midway between pairs of edges and were considered as MAT points. Arceli *et al.* proposed a parallel thinning algorithm that relies on an iterative erosion of grey-level images [ARC95]. He noted that the skeleton branches are located along ridges, or located centrally within a plateau of the image function. For the ridges, the sets of pixels with an increasing intensity level were successively removed until only the set of ridge pixels was left; For the plateaus, pixels were removed symmetrically.

2.3 MAT Significance Hierarchies

One problem of the MAT is that small variations in the boundary of a shape can greatly change the medial axis structure. For example, a small protrusion on the boundary will produce a long MAT branch and can distort the main branch from which the peripheral branches stem, see Fig. 2.3. A related problem is the lack of a measure for the relevance of MAT branches to the description of the shape. A hierarchical MAT representation is potential to describe the relative importance of each medial axis and to overcome these effects. The description of the detailed aspects of a shape provided in the lower levels of this hierarchy does not disturb the primary description at higher levels of the hierarchy [PIZ87]. Another advantage of the hierarchical MAT representation is that it supports top-down (large scale first) object description, recognition and matching. Several methods have been proposed to construct hierarchical MAT representations.

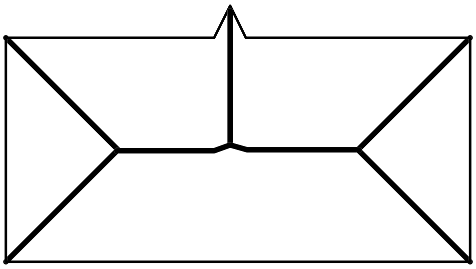


Fig. 2.3. The medial axes are sensitive to detailed boundary perturbations.

2.3.1 MAT Pruning for Binary Shapes

To systematically identify which branches of the MAT could be pruned it is necessary to assign each axis point a measure of "significance" for the representation of the global properties of the shape. Then less essential sections of the MAT can be removed. MAT pruning does not include resolution reduction and thus preserves the topological structure of a shape. A hierarchy of medial axes can also be constructed, based on this significance measure [OGN95].

An early significance measure is the propagation velocity of each symmetrical axis point in the prairie fire model [BLU73], which relies further on the velocity of the fire front (boundary) and the angle between meeting fire fronts. The propagation velocity of each axis point is low for a "flat" protrusion (in the boundary) that often results in a long axis branch but is less significant for the description of the shape. For example, letting v_f be the velocity of the fire front and θ be the inner angle between two fire fronts, the propagation velocity of the axis point for this angle is $v_a = v_f / \sin(\theta / 2)$. It is slow for an obtuse angle ($\pi / 2 < \theta < \pi$) and fast for an acute angle ($0 < \theta < \pi / 2$).

The effect of pruning on the MAT is similar to morphological opening. Clearly pruning an insignificant axis should not significantly change a shape description. A significance measure based on this principle was defined as the maximal thickness of the implied erosion [HO86] [BRA92]. The significance measure of a point A on an axis with endpoint E is: *R*(E)+*dist*(A, E)-*R*(A), where *R* is the radius of the maximal inscribed disk for an axis point and *dist* is the distance between points. This measure is low for a "flat" protrusion on the boundary, in which *R*(A) is much larger than *R*(E). The importance of this measure is that it increases monotonically from the tip of the axis inwards towards a limiting value. Therefore, a threshold can be applied without danger of disconnecting the MAT. A similar significance measure, defined as the erosion area in the MAT pruning, was described in [SHA98].

Another significance measure is the ratio of the length of the boundary unfolded by an axis segment to the length of the axis segment [BLU78]. For "flat" protrusions in the boundary, the axes are usually associated with small amount of boundary they unfold. Blum [BLU78] suggested a differential measure $\partial B(a)/\partial a$, where a is the axis arclength parameter and $B(a)$ the length of the boundary unfolded by the axis segment reaching axis point a. Axis significance is determined by the integration of the differential significance measure over a. Ogniewicz [OGN95] suggested a significance measure that is the length of boundary unfolded by an axis segment. This measure also increases monotonically inward along the axis and thus is called "insideness". Ogniewicz found that the axis points deep inside an object, usually with a high "insideness" measure, are less sensitive to boundary variations and more relevant to the description of significant shape properties, see Fig. 2.4.

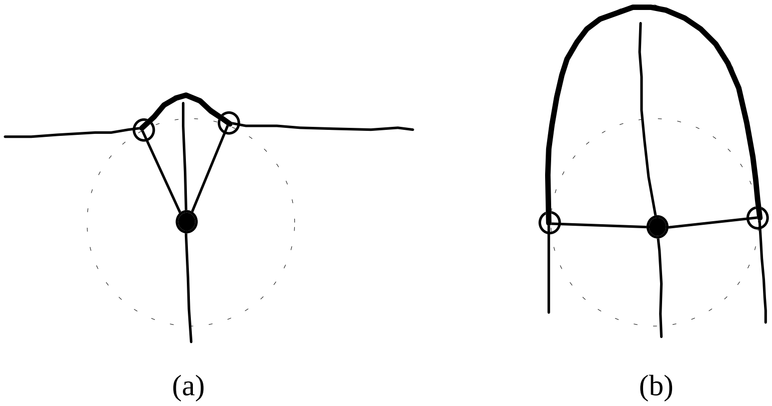

(a) (b)

Fig. 2.4. The boundary (thick line) unfolded by (a) an insignificant axis and (b) a significant axis.

2.3.2 Contour Smoothing for Binary Shapes

The following two categories of MAT hierarchies involve multiple resolution procedures. At lower resolution object descriptions are simplified. The importance of each axis branch is determined by the order of annihilation under successive resolution reduction. To obtain a multi-resolution object representation contour smoothing and region blurring have been used.

Contour smoothing may be applied to various boundary representations, such as curvature. Dill *et al.* iteratively deleted the contour of an object but retained points representing “significant” convexity [DIL87]. The “significance” of the convexity is determined by first smoothing the boundary curvature with a low-pass filter at multiple resolutions and then thresholding the curvature. The remaining points of “significant” convexity constitute a robust set of skeletons. Rom *et al.* divided the object boundary into segments separated by points of maximum curvature and repeatedly removed the smallest segments by fitting a spline to the local object boundary [ROM93]. The axis for each removed segment is computed separately and thus does not disturb the axis representation of the principal shape property.

A more systematic approach to contour smoothing is given by curve evolution [SET85] [MOK86] [KIM95], in which the velocity of fire fronts depends on a constant component (like Blum’s model) and a smoothing component proportional to curvature, see Fig. 2.5. Let *s* be the parameter of contour position then the contour representation is $C(s,t) = (x(s,t), y(s,t))$ and the curve evolution is defined as:

$$\begin{cases} \dfrac{\partial C(s,t)}{\partial t} = (\beta_0 + \beta_1 \kappa(s))\mathbf{N} \\ C(s,0) = C_0(s) \end{cases} \tag{2.6}$$

where *t* is the time duration (similar to scale), **N** is the inward normal to the curve, the sub-

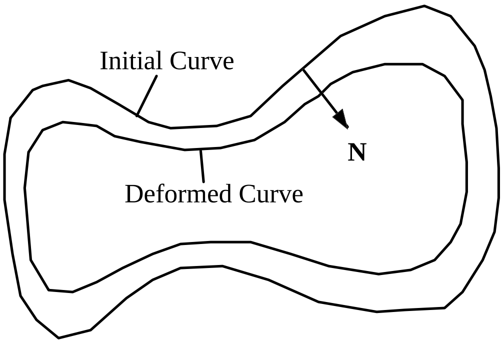


Fig. 2.5. Curve evolution.

script 0 denotes the initial curve prior to deformation, and $\kappa(s)$ is the curvature of the contour. Gage *et al.* proved that in the case of $\beta_0 = 0$, the curvature deformation of a shape, by Eq. (2.6), is a non-linear Gaussian smoothing process over the contour [GAG86]. It gives an anisotropic smoothing over the contour in the sense of Perona *et al.*'s work on anisotropic region blurring [PER90].

Kimia *et al.* [KIM95] began with a boundary contour and propagated the contour inward according to Eq. (2.6). When opposing boundaries meet, a "shock" occurs, the two sides form a common middle and annihilate. The locus of points at which shocks occur defines axes like the MAT. The amount of diffusion is determined by the ratio, β_1 / β_0.

Tari *et al.* [TAR97] computed a blurred version of the distance transform of a contour using linear diffusion. The level curves of this blurred distance transform are a smoothed version of the level curves of the original distance transform and a smoothed analogue of the successive shape outlines produced by prairie fire model. The medial axes are the loci of maximum curvature along level curves. Because the propagating velocity of the points on the blurred level curves was found to be the sum of the curvature and a constant [TAR97], this method is closely related to the curve evolution.

2.3.3 Region Blurring for Grey-Level Shapes

Region blurring involves applying a blurring filter to the intensity of an image and then obtaining a smoother boundary. Koenderink treated a binary shape as a characteristic function, e.g. letting the function be 0 outside the shape and 1 inside, then convolved it with an appropriate Gaussian [KOE86]. The result is a grey-scale representation. The contour of the "blurred" shape is considered as an intensity level curve of this image. The level curve may be chosen such that the area of the shape remains constant after each resolution reduction. The main criticism on region blurring is that topology is not preserved. However, Pizer *et al.* have shown that it is unnatural to insist that the topology be maintained under resolution reduction [PIZ87].

It is argued that region blurring provides a better representation than contour smoothing [KOE86]. In addition region blurring is directly applicable to grey-level images. On the contrary, the MAT pruning and contour smoothing approaches start with a segmented shape. The initial boundary contour, often measured at a small, fixed scale, gives an initial topology for the shape that is maintained through the MAT pruning and contour smoothing processes. Therefore, the result of the MAT pruning and contour smoothing is much influenced by the initial fixed-scale measurement, which is at odds with the original motivation for using multi-resolution analysis and a significance hierarchy. In addition, the selection of the initial scale is not trivial, because the prior knowledge about image scale is often not available and small-scale measurement is not resistant to noise.

An example of region blurring is the Intensity Axis of Symmetry (IAS) [GAU89], in which the symmetric axes are calculated for each intensity level curve of the blurred grey-scale image. The IAS is a set of branching sheets in 3-D space (2-D for spatial domain and 1-D for intensity levels). Under successive region blurring, sheets of medial axes annihilate into other sheets of medial axes, resulting in a progressive simplification of the IAS description.

Borgefors *et al.* also considered a binary shape as a characteristic function [BOR98]. He computed a grey-level pyramid from a binary image. Each level of the pyramid is obtained by applying a 3×3 blurring mask to the next higher resolution level and then sub-sampling the blurred result. At each resolution level, a distance transform is used to generate the skeletons at that resolution level.

The result of the above algorithms is a set of MAT branches (or sheets) at each blurring level. This is an over complex representation of the shape properties. More seriously it is not clear how to determine an appropriate blurring level at which to compute the MAT for a specific shape. Nor is it clear what the relative significance is for MAT branches when they are computed at several blurring levels. One solution to these issues is to use a multi-scale medial axis (MMA) representation. An MMA representation provides a principled approach to computing MAT branches across scale from a grey-level image. An appropriate scale for a specific shape property can be determined if the significance measure of the MAT is “optimal” across scale.

An earlier work on multi-scale medial axis is Crowley’s graphical representation of the MAT in a resolution pyramid [CRO84]. Crowley *et al.* computed the difference of low-pass transforms (DOLP), more commonly known as the Difference of Gaussian (DoG), at progressively lower resolutions for a grey-level image. At positions where the radially symmetric central lobe of a DOLP filter is a best fit to the form (shape plus intensity), the response of the DOLP is high. This is analogous to the “maximal disk” definition of the MAT for binary shapes. The peaks (ridges) in the DOLP response are detected and linked with adjacent peaks (ridges) at neighbouring positions and resolution levels. This results in a tree-like MAT representation in the resolution pyramid, which is insensitive to fine detail and noise. However, the shortcoming of this method is that successively resampling the image space means that the result is not quantitatively reliable. In addition the DOLP computation and ridges detection procedure are adhoc and rotation dependent. A more systematic approach to multi-scale medial axis computation is to use differential geometry to detect ridges in a scale-space representation [PIZ94] [FRI93] [MOR94] [LIN98] of an image. This is introduced in Section 2.4.

2.4 Multi-Scale Medial Axis

The multi-scale medial axes (MMA) for an object in a 2-D image are a set of 1-D curves in 3-D scale-space. For each MMA point, the spatial coordinate, (x, y), indicates the middle position of the object; the scale parameter, σ, specifies the approximate width of the object at that position. The MMA curves are obtained by first computing a measure called "medialness" over scale-space and then detecting scale-space ridges in the medialness manifold. The medialness function, $M(x, y, \sigma)$, is defined as the degree to which a position (x, y) resembles an object middle when examined at a particular scale, σ. At a given scale σ, the medialness function gives the strongest response for those objects with a specific radius, r, in proportion to σ. For variable width objects, the medialness values are relatively high along a track through the middle of the object and going up and down in scale proportional to the local object width. This track of high medialness, i.e. the ridges in medialness, is the MMA (see Fig. 2.6). In the following section, the medialness and ridge definitions in scale-space are reviewed.

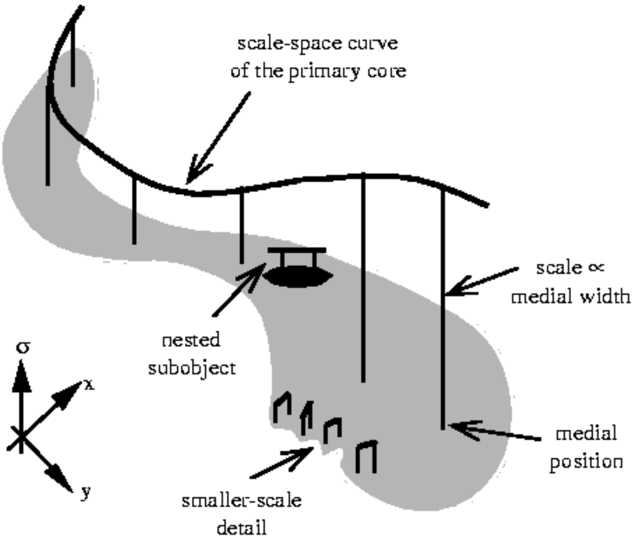


Fig. 2.6. The MMA curves in scale-space. (Courtesy of [MOR95])

2.4.1 Medialness Computation

2.4.1.1 Traditional medialness operators

There are many definitions for the medialness function, each of which is suitable for a particular class of image or object. A thorough review of traditional medialness operators can be found in [PIZ98]. To make medialness responses that are invariant to translation, rotation and zoom, these medialness operators are usually based on normalised Gaussian derivatives of image intensity, $\sigma^k D^k L$, where $L(\mathbf{x},\sigma) = G(\mathbf{x},\sigma) * I(\mathbf{x})$. Medialness operators can be either local or multi-local.

(1) Local medialness operators

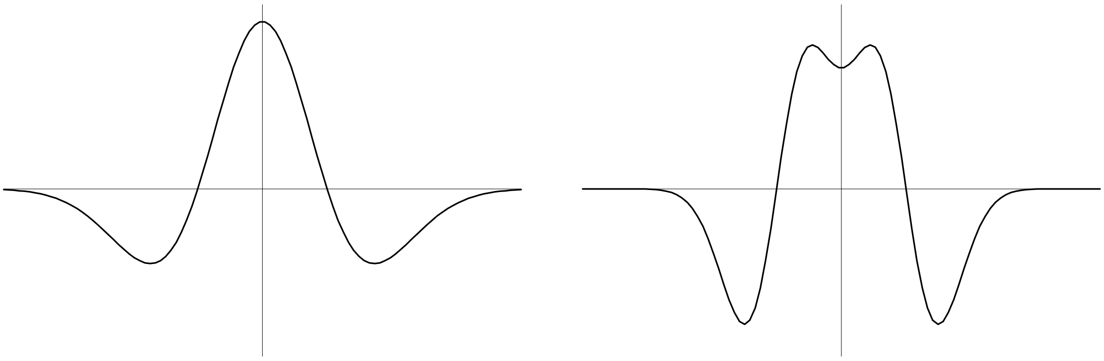
Local operators depend only on local spatial derivatives of the blurred image intensity at each potential axis point. If the local derivative operation is combined with the Gaussian blurring to form a convolution kernel, it resembles Blum's "maximal disk" definition for the MAT of a binary shape. For a 2-D grey-level image, when the radially symmetric central lobe of the convolution kernel of a local operator is best fit to the object (shape plus intensity), the kernel centre is an axis point.

A linear local operator is the normalised Laplacian of a Gaussian (LoG) [FRI92] [FRI93]:

$$\begin{aligned} K(x,y,\sigma) &= -\sigma^2 \nabla^2 G(x,y,\sigma) \\ &= -\frac{1}{2\pi\sigma^2}\left(2 - \frac{x^2+y^2}{\sigma^2}\right)\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \end{aligned} \tag{2.7}$$

The profile of this operator in 1-D is shown in Fig. 2.7(a). Another linear operator is the Difference of Low-Pass (DOLP) transforms or DoG (Difference of Gaussian) [CRO84], which is a difference of two Gaussians with different sizes and an approximation to the LoG operator.

In addition, some non-linear, data-dependent operators have been used. This allows the



(a) (b)

Fig. 2.7. The 1-D profiles of medialness operators: (a) the LoG and (b) HMAT.

choice of medialness operators that respond only to involute pairs[1] and are selective to image properties at border involutes. For example, the orientation of the operator can be adjusted to respond preferentially to parallel boundaries or the polarity of intensity change at boundaries.

A non-linear (orientation dependent) local medialness function is [ter91] [FRI93] [LIN98]:

$$M(x,y,\sigma) = -\sigma^2 L_{pp} \tag{2.8}$$

where L_{pp} and L_{qq} are the maximal and minimal principal curvatures of $L(x,y,\sigma)$ with principal directions ***p*** and ***q***, respectively. These principal curvatures and directions correspond to the eigenvalues and eigenvectors of the Hessian matrix of second derivatives of $L(x,y,\sigma)$:

$$\begin{pmatrix} L_{xx} & L_{xy} \\ L_{xy} & L_{yy} \end{pmatrix} \tag{2.9}$$

Therefore, unlike the isotropic convolution of the LoG (Eq. (2.7)), this medialness function applies the LoG convolution only in the direction that maximizes the medialness function. It is particularly effective for objects with parallel sides and uniform internal intensity. It

---

[1] The maximal disk of each MAT point is often tangent to the boundary of a shape at two distinct points. These two related points are called involutes.

is less sensitive to intensity variations along the object axis than the linear, local LoG medialness, because these two kinds of medialness functions are related to L_{pp} (principal) and $L_{xx}+L_{yy}=L_{pp}+L_{qq}$ (isotropic) measures, respectively. The isotropic LoG operator involves applying second derivatives along an object axis in direction $\boldsymbol{q}$.

A shortcoming of the medialness definitions in Eqs. (2.7) and (2.8) is that they also give strong responses for “blob” structures, where L_{pp} and L_{qq} are of large and roughly equal values. Lindeberg proposed two blob-free medialness functions [LIN98]:

$$M(x,y,\sigma)=\sigma^{4\gamma}(L_{pp}^2-L_{qq}^2)^2 \tag{2.10}$$

$$M(x,y,\sigma)=\sigma^{2\gamma}(L_{pp}-L_{qq})^2 \tag{2.11}$$

which have high values only when the principal curvatures are significantly different, i.e. for elongated structures.

A non-linear (polarity dependent), local medialness function, which ignores the difference between a bright object on a darker background and a dark object on a brighter background, can be found in [ter91] [FRI93] and is defined as:

$$M(x,y,\sigma)=-\sigma^2\left(L_{pp}^2+L_{qq}^2\right)^{1/2} \tag{2.12}$$

(2) Multi-local medialness operators

Multi-local operators generate medialness by querying derivatives of the blurred image intensity, e.g. the gradient magnitude, at a distance from each potential axis point. They are inspired by the observation that the maximal disk of each MAT point is tangential to an object boundary at two or more places. When they are applied to a 2-D image, the edge responses around a circle, where the gradient of the blurred image intensity is towards the circle centre, are accumulated. A medial axis point is a local maximum of medialness response with respect to circle radius and position.

A linear multi-local medialness function is the Hough-like Medial Axis Transform (HMAT) [MOR91] [MOR95]. It is the integration of directional boundariness around a circle with centre coinciding with the operator centre and radius proportional to scale. The directional boundariness is defined as:

$$B(x,y,\sigma,\theta) = -\sigma \mathbf{u}(\theta) \cdot \nabla L(x,y,\sigma) \tag{2.13}$$

where $\mathbf{u}(\theta) = (\cos\theta, \sin\theta)$. The HMAT medialness is defined by:

$$M(x,y,\sigma) = \int_0^{2\pi} B((x,y) + r\mathbf{u}(\theta), \sigma, \theta) d\theta \tag{2.14}$$

where $r = k\sigma$ for some constant radius-to-scale ratio *k*. Therefore, the HMAT medialness operator is defined by:

$$K(x,y,\sigma) = \sigma \int_0^{2\pi} \left[-\mathbf{u}(\theta) \cdot \nabla G((x,y) + r\mathbf{u}(\theta), \sigma) \right] d\theta \tag{2.15}$$

The profile of this operator in 1-D is shown in Fig. 2.7(b).

A non-linear (orientation dependent), multi-local medialness function was studied in [PUF95] and defined as:

$$M(x,y,\sigma) = -\sigma \mathbf{u}(\theta) \cdot \nabla L((x,y) + r\mathbf{u}(\theta); \sigma) + \sigma \mathbf{u}(\theta) \cdot \nabla L((x,y) - r\mathbf{u}(\theta); \sigma) \tag{2.16}$$

where the angle θ is selected to maximize the medialness function over all possible angles. This medialness function is especially effective for objects with parallel boundaries. A similar medialness function can be found in [KOL95] where θ is selected as the principal direction that maximizes the second derivative of the function $L(x,y,\sigma)$.

A non-linear (polarity dependent), multi-local medialness function can be obtained by integration of the absolute value of the boundariness around a circle centred at each putative axis point [MOR94] [MOR95] using:

$$M(x,y,\sigma)=\int_0^{2\pi}\left|B((x,y)+r\mathbf{u}(\theta),\sigma,\theta)\right|d\theta \tag{2.17}$$

This non-linear function is useful if an object's intensity near the boundary is brighter in some locations than the background and in some places darker than the background. Conversely, the linear HMAT kernel is useful only for objects with the same polarity of boundary transitions.

(3) Comparison between local and multi-local medialness operators

Multi-local operators are suitable for a greater variety of object appearances than local operators. The ratio between operator radius and the blurring scale, called radius-to-scale ratio, $k = r/\sigma$, can be tuned to suit image properties such as noise and pre-existing blur. For example, if the image is noisy one might decrease *k* and use larger blurring scales for a given object radius. In noise-free environments where one desires as much precision as possible, a larger value for *k* might be used. As a result, multi-local operators usually have low weights (sometimes zero) around operator centres (see Fig. 2.7(b)) and are less sensitive to variations in image intensity, e.g. embedded objects and correlated noise, within objects.

The adaptive nature of multi-local medialness operators is also reflected in the pairing between boundary transitions of opposite polarities. For example, they can pair boundaries that are brighter than the background on one side and darker than the background on the other, see Eq. (2.17). Local medialness operators can only pair boundaries with the same transition polarity, for either bright objects on dark backgrounds or dark objects on bright backgrounds.

Due to the adaptive nature of multi-local operators, the edge-free medialness functions, reviewed in Section 2.4.1.2, are based on multi-local operators. However, the use of multi-local operators is computationally expensive. Multi-local operators involve the integration and dot product operation over all the angles around a "fuzzy" disk, which cannot be decomposed into separable operators in *x* and *y* dimensions. On the contrary, local operators based on

Gaussian derivatives can do so and the computational cost can be reduced from $O(\mathrm{M}^2)$ to $O(\mathrm{M})$, where M is the width of the operator [HUE86].

A partial remedy for multi-local operators, described by Bharath *et al.* [BHA99], used a polar separable function to represent the radial and angular weightings for the HMAT operator. The angular weighting for each boundary point is approximated by the application of angular steering techniques [FRE91]. The boundariness (gradient magnitude), $B(x, y, \sigma)$, is multiplied pointwise by the steering weights computed from the gradient angle field, $\theta_B(x, y, \sigma)$, and the resulting scalar responses are each convolved with their corresponding filter masks (bases). The output of these filters is the summation of a truncated series and found to be sufficient to approximate the HMAT medialness. This scheme greatly reduces the computational cost. [BHA99].

2.4.1.2 Edge-free medialness operators

One shortcoming of the traditional medialness operators is that they are not only sensitive to symmetries, but also respond to edges. This can disturb the extraction of the MMA. The edge responses may be partially distinguished from symmetry responses under an appropriate scale-normalization scheme. One such scheme is multiplication by the *k*-th power of scale for medialness operators based on the *k*-th derivative of a Gaussian, i.e. $\sigma^k D^k L$. In this case the response for an isolated step-edge can have a constant peak amplitude at a range of scales [FRI93]. Such a medialness response forms a “fuzzy sheet” in scale-space, extending from the step edge to proportionally larger scales (Fig. 2.8(a)).

On the other hand, the response of a normalized operator to a symmetric structure varies across scale and has its peak value at a scale proportional to the object width. Fig. 2.8(a) shows the cross-section of the medialness for two parallel edges of an object. This creates two “fuzzy sheets” of medialness that intersect to produce an enhanced medialness in the middle of the object. For points on these sheets, the direction, $\mathbf{e}_1$ (refer to Fig. 2.10), in which the magnitude of the second derivative is greatest, is through the sheet; the direction, $\mathbf{e}_3$, in which the magnitude of the second derivative is least, is parallel to the boundaries and points into the picture. The direction, orthogonal to both, is $\mathbf{e}_2$. The maximum medialness across scale, as a function of position, is shown in Fig. 2.8(b). The resulting curve is relatively flat until near the middle of the object. The strong peak in the middle of the object is the desired symmetry response. The MMA is readily detected as the peak of this curve.

Due to discrete spatial sampling and noise in the original image, there are always small fluctuations in the edge response peak computed at each scale, which is, thereafter reflected as small fluctuations along $\mathbf{e}_2$ and in the curve shown in Fig. 2.8(b). These fluctuations satisfy the definition of a ridge (using either the optimal scale ridge or height ridge definition) in scale-space and can lead to the detection of an MMA branch along the direction $\mathbf{e}_3$ [MOR95], see

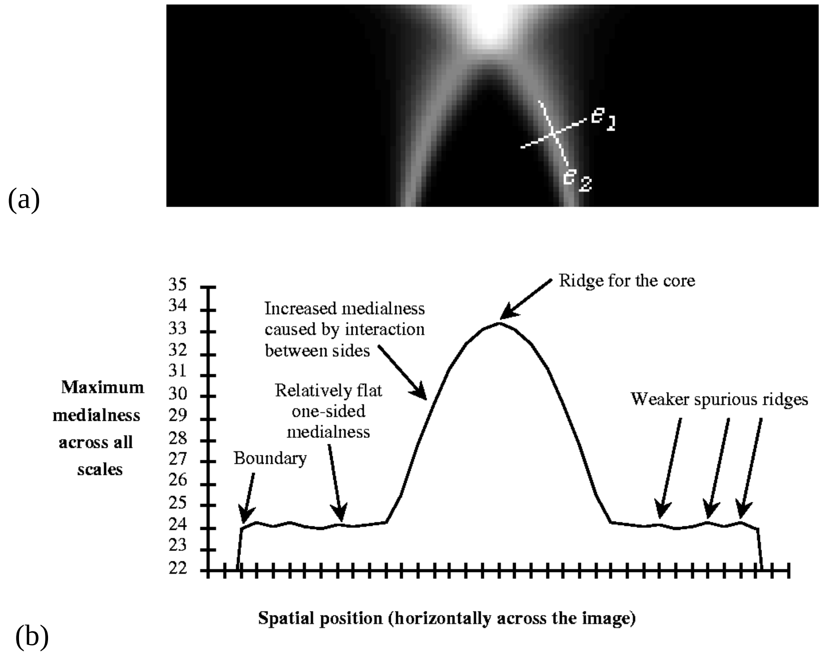


Fig. 2.8. The medialness scale-space for two parallel edges: (a) a cross-section and (b) the maxima across scale as a function of position (Courtesy of [MOR95]).

Section 2.4.2.2.

It is clear that ridge detection alone is not sufficient for multi-scale medial axis detection; some measure of ridge strength is desirable to differentiate a symmetry from artifacts of discrete sampling and noise. This is similar to one motivation for Canny's edge detector [CAN86], in which the amplitude of the edge response (gradient) at the maximum provides a good estimate of edge strength and a hysteresis thresholding scheme can be used on those edge candidates selected by non-maximum suppression. This gives a better separation between signal and noise, or strong edges and weaker edges.

Because a ridge strength measure is closely related to the medialness operator being used, existing ridge strength measures can be classified as either local or multi-local measurement.

(1) Local measurement of ridge strength

The estimation of ridge strength may be based on a local measure at a potential ridge point. One obvious measure is the medialness response itself, $M(x, y, \sigma)$. Lindeberg defined a ridge saliency measure as the integration of the medialness response along each ridge [LIN98]. Therefore, long connected medial axis branches arising from high contrast objects are attributed a greater ridge saliency. However, his medialness function, $-\sigma^2 L_{pp}(x, y, \sigma)$, gives a strong emphasis to edges, which are prone to forming small creases from scale to scale, especially at small scales. Thus, some spurious medial axes, usually close and parallel to boundaries, can be generated.

Another local measure of ridge strength is the cross-ridge second derivative, $M_{pp}(x, y, \sigma)$ (ridge "sharpness"). It considers only local properties close to the "top" of the ridge. There is no sense of the overall height or breadth of the ridge as measured across the contributing region of boundariness. Therefore, it cannot best differentiate small "hills"(fluctuations of edge responses) and a large mountain (symmetrical interaction of edge responses) on a plateau (edge responses). It is clear that a global measure of ridge strength, which considers the relative local and medium range importance of the ridge, is required.

(2) Multi-local measurement of ridge strength

Multi-local ridge strength measures are often based on multi-local medialness operators. There are two types of multi-local ridge strength measures. The first type is to compare the medialness response at a potential ridge with the edge responses that contribute to the ridge and are at a distance from the ridge.

For the HMAT operator, the medialness at a ridge point is the sum of the edge responses from contributing boundary points. Therefore, any elevation in medialness above the edge response of one boundary point must be the result of interaction with other boundary points. Morse defined a global ridge strength measure (GRSM) [MOR95] as the ratio of the total medialness at a potential ridge point (the altitude of the mountain) to the maximum edge response (the altitude of the plateau) that contributes to the ridge point:

$$R(x,y,\sigma) = \frac{M(x,y,\sigma)}{\max_{\theta}\left\{B((x,y)+r\mathbf{u}(\theta),\sigma,\theta)\right\}} \tag{2.18}$$

The GRSM value arising from the interaction of two boundaries is greater than that arising from an isolated boundary. Thus it can differentiate medial and edge responses. Because this measure is invariant to intensity, it ignores the difference between bright and dim objects. The sensitivity to dim objects must be traded off against the sensitivity to small noise and computational errors.

Morse also suggested an iterative process, called credit attribution [MOR95], to enhance the HMAT medialness response for structural symmetry and suppress edge responses. In the credit attribution algorithm, the initial medialness, at $t = 0$, is computed in the same way as the HMAT. At iteration *t*, for each boundary point at (*x*, *y*), the (circular) locus of all medialness points, at $(x, y) - r\mathbf{u}(\theta)$, that the boundary point (*x*, *y*) contributes to is examined and the medialness on this locus is summed. Then the medialness is re-accumulated, with each boundariness contribution being weighted by the ratio of the medialness at the medial point and the medialness sum around each boundary point, i.e.:

$$M_t(x, y, \sigma) = \int_0^{2\pi} B((x, y) + r\mathbf{u}(\theta), \sigma, \theta) W_t((x, y) + r\mathbf{u}(\theta), \sigma, \theta) d\theta \tag{2.19}$$

$$W_t(x, y, \sigma, \theta) = \begin{cases} 1 & t = 0 \\ \dfrac{M_{t-1}((x, y) - r\mathbf{u}(\theta), \sigma)}{\int_0^{2\pi} M_{t-1}((x, y) - r\mathbf{u}(\varphi), \sigma) d\varphi} & t > 0 \end{cases} \tag{2.20}$$

This process is repeated until the desired level of sharpening is achieved or the maximum computation is exceeded. The credit attribution algorithm assumes that each boundary point has a constant sum of contributions to the medialness space. If one of its surrounding locations, at $(x, y) - r\mathbf{u}(\theta)$, has a greater initial medialness than the other surrounding locations, i.e. with an enhanced $M_{t-1}((x, y) - r\mathbf{u}(\theta), \sigma)$, the boundariness contribution in that direction θ will obtain a greater weight in the next iteration. In this way, the boundariness contribution becomes greatest when the contributions interact.

The second type of multi-local ridge strength measures is to compare a pair of edge responses contributing to a potential ridge [WAN82] [SUB93] [KOL95]. These measures have been widely used in the detection of ridges in 1-D signals and curvilinear structures in 2-D images. Like using the medialness function in Eq. (2.16), Koller computed a pair of shifted edge responses that were steered in the principal direction, θ, that maximizes the second derivative of the blurred image $L(x, y, \sigma)$, i.e.:

$$\begin{aligned} R_L(x,y,\sigma) &= -\sigma\mathbf{u}(\theta)\cdot\nabla L((x,y)+r\mathbf{u}(\theta);\sigma) \\ R_R(x,y,\sigma) &= \sigma\mathbf{u}(\theta)\cdot\nabla L((x,y)-r\mathbf{u}(\theta);\sigma) \end{aligned} \tag{2.21}$$

He noted that the final medialness response must be large, if both edge responses are large, and zero, if either edge response is zero. Therefore, rather than using the sum of the two edge responses to produce medialness as in Eq. (2.16), either the minimum or geometric mean of the two edge responses was used in computing the output response:

$$\begin{aligned} M(x,y,\sigma) &= \min\{R_L(x,y,\sigma), R_R(x,y,\sigma)\} \\ M(x,y,\sigma) &= \sqrt{R_L(x,y,\sigma)R_R(x,y,\sigma)} \end{aligned} \tag{2.22}$$

As a result, the medialness response of a single edge is suppressed and the remaining response can be directly used as a measure of ridge strength.

2.4.1.3 Preview of our medialness function

In Section 3.2 we propose a Concordance-based Medial Axis Transform (CMAT) to compute edge-free medialness responses. The initial medialness is computed in the same way as for the HMAT, using Eq. (2.14). At a second, non-iterative stage, the medialness is re-accumulated using Eq. (2.19) but the directional boundariness is weighted by a function, *f*, of the ratio of the boundariness value to the total initial medialness, i.e.:

$$W_t(x,y,\sigma,\theta) = \begin{cases} 1 & t=0 \\ f\left[\dfrac{B((x,y)+r\mathbf{u}(\theta),\sigma,\theta)}{M_0(x,y,\sigma)}\right] & t=1 \end{cases} \tag{2.23}$$

Boundariness is introduced in Section 3.1. The computation of the CMAT medialness response from the boundariness is described in Section 3.2.1. In Section 3.2.2 the concordance property, which expresses the symmetry of boundariness strength, is discussed. In Section 3.2.3, alternative definitions for computing the CMAT, using different forms of the function *f*, are discussed. In Section 4.1, we present quantitative performances of the CMAT algorithm and compare them with those of selected traditional medialness operators. The results of the applications of the CMAT to synthetic and medical images are presented and compared in Section 4.2. Finally, the relationship of the CMAT with existing edge-free medialness algorithms is discussed in Section 5.1.

2.4.2 Detection of Medialness Ridges

2.4.2.1 Ridge definitions in 2-D space

Definitions of ridges can be found in the mathematical literature dating back for more than a century [KOE94]. Since each ridge finder is correct in some sense but may fail in specific cases, the literature does not converge to a unique description of ridges. Instead, most of the numerous definitions fall into one of the following four categories.

(1) Water flow definition

This method is based on the global drainage pattern of rainfall on a terrain map [GAU93] [KOE94] [GRI91]. If the image intensity is viewed as height, the image gradient can be used to predict the direction of drainage in an image. By following the image gradient downhill from each point in the image, the set of points that drain to the same local intensity minimum can be identified and is called a watershed region of the image. The lines that separate watershed regions correspond to intensity ridges and valleys. From this the ridges can be identified.

The water flow definition has not been widely used for a practical reason: there is no local operator to detect watershed ridges [EBE94a]. The ridges have to be determined by a nonlocal process, in which a ridge point may be relocated by the disturbance on the drainage paths, even though no change in intensity has occurred around this ridge point [EBE94a]. Therefore, it is more widespread to regard the 2-D image intensity as a surface in 3-D space and extract the ridges using differential geometry. These definitions are local and involve measurement of curvatures associated with surfaces in general.

(2) Level curvature definition

This method defines a ridge as the locus of positions that are local maxima of the isophote (level curve) curvature of a surface [EBE94a], see Fig. 2.9(a). This definition arises from the notion that a person walking at a constant altitude around a mountain terrain might label the mountain ridges as those points where he experiences a maximal change in direction. The isophote curvature of a function, $f(x, y)$, may be written as:

$$\kappa(x,y) = \frac{-f_{xx}f_y^2 + 2f_x f_y f_{xy} - f_{yy}f_x^2}{\left(f_x^2 + f_y^2\right)^{3/2}} \tag{2.24}$$

A position, (x, y), can be identified as a ridge point, if $\kappa(x,y)$ is a local maxima in the tangent direction of the level curves:

$$\left(\frac{f_y}{\sqrt{f_x^2 + f_y^2}}, \quad \frac{-f_x}{\sqrt{f_x^2 + f_y^2}} \right) \tag{2.25}$$

Maintz *et al.* noted that the intensity profile along a line perpendicular to a ridge is relatively concave [MAI96]. They defined ridge points as those with minimum (negatively maximum) second derivative in the direction normal to the gradient direction. The second directional derivative they obtained is $\kappa(x,y)\sqrt{f_x^2 + f_y^2}$. Thus this method is also based on isophote curvature.

The main limitation of isophote curvature definition is that it fails to detect horizontal ridges. An example is the function, $f(x,y) = 1 - x^2$, which has two parallel isophote lines and a horizontal ridge. This is also reflected in the level curvature expression in Eq. (2.24), which involves division by the gradient magnitude that is zero along a horizontal ridge.

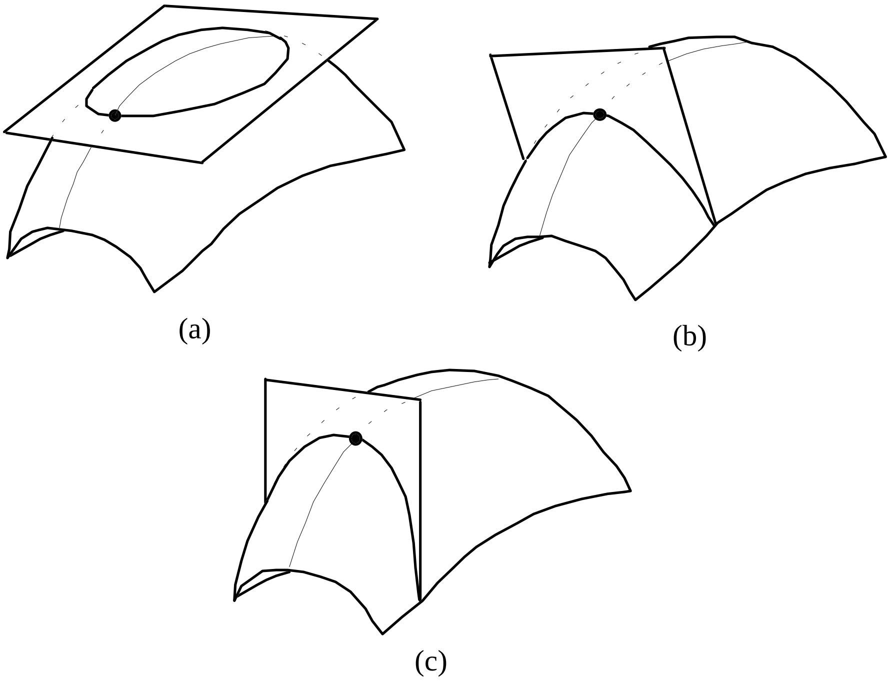


Fig. 2.9. Ridge definitions on a 2-D surface and associated section planes: (a) level curvature, (b) principal curvature, and (c) height ridge.

Thirion *et al.* extended the level curvature definition to 3-D images [THI92] to extract crest lines. For a 3-D image, the manifold with an iso-value is typically a 2-D surface. Thirion's crest lines are defined as those points on a level surface, whose principal curvature is a local maximum in the corresponding principal direction (the definitions of the principal curvature and direction are given below in (3)). Fidrich further extended Thirion's work to 4-D images (and 4-D scale-spaces of 3-D images) [FID96].

(3) Principal curvature definition

At a point on a surface, the surface may bend in quite different ways for varying directions. This can be represented by the normal curvature, which is the curvature of the curve formed by the intersection of the surface and one normal section plane. The normal section plane is spanned by the normal vector and a tangent vector of the surface. The maximum and minimum values of the normal curvature of the surface at a point are the *principal curvatures*.

The corresponding tangent directions of the surface are the *principal directions* [ONE66]. A point is defined as the ridge point when one of its two principal curvatures of the surface assumes a local maximum in the corresponding principal direction [EBE94a] [KOE91] [MON95], see Fig. 2.9 (b).

For a ridge with a flat top and steep slopes, this definition has the problem that the true ridge may be lost and, instead, two separate ridges of principal curvature, symmetrical to the true ridge, will be found [EBE94a] [STE98]. These are intuitively spurious ridges. For example, the principal curvature ridge for the surface, $f(x,y)=1-x^2$, is at $x=0$; Those for the surface, $f(x,y)=1-x^4$, are at $x=\pm 1/\sqrt{6}$.

(4) Height definition

A height ridge point is defined as a local maximum of the image intensity in the direction that minimizes (negatively maximizes) the second directional derivative of the intensity [HAR83] [EBE94a] [KOL95], see Fig. 2.9(c). The detection of height ridge points can resort to solving the eigenvectors of the Hessian matrix of second derivatives of the intensity. For a 2-D function $f(x,y)$, let $\lambda_1 \le \lambda_2$ denote the eigenvalues of the Hessian matrix of second derivatives

$$\begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix} \tag{2.26}$$

with corresponding eigenvectors $\mathbf{e}_1$ and $\mathbf{e}_2$, respectively. A point is defined to be on a ridge of f, if there is a maximum of f in the direction of (negatively) greatest curvature:

$$\begin{aligned} &\lambda_1 < 0 \\ &\mathbf{e}_1 \cdot \nabla f = 0 \end{aligned} \tag{2.27}$$

For a 3-D function $f(x,y,z)$, let $\lambda_1 \le \lambda_2 \le \lambda_3$ denote the eigenvalues of the Hessian matrix of second derivatives:

$$\begin{pmatrix} f_{xx} & f_{xy} & f_{xz} \\ f_{xy} & f_{yy} & f_{yz} \\ f_{xz} & f_{yz} & f_{zz} \end{pmatrix} \tag{2.28}$$

with corresponding eigenvectors $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$, respectively. A point is determined to be a 1-D ridge point, if $\lambda_2 < 0$ and $\mathbf{e}_i \cdot \nabla f = 0$ for i=1, 2, see Fig. 2.10. Also, a point is determined to be a 1-D valley point if $\lambda_2 > 0$ and $\mathbf{e}_i \cdot \nabla f = 0$ for i=2, 3.

Eberly *et al.* showed that the height ridge definition gives an intuitively better result than the level and principal curvature definitions [EBE94a]. However, this definition may lead to the entire area of a radially symmetric surface being classified as ridge points [HAR92], as in the case of a cone defined by $f(x, y) = \sqrt{x^2 + y^2}$. Haralick has suggested a post-processing step to eliminate these "ridge" points [HAR92].

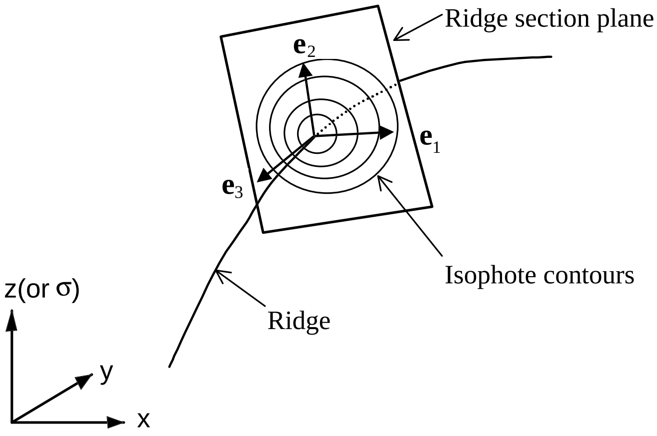


Fig. 2.10. A 1-D height ridge in 3-D space (or 3-D scale-space) with the three eigenvectors of the Hessian matrix.

2.4.2.2 Ridge detection in scale-space

There are many reports of methods for the multi-scale detection of ridges. The method presented in this thesis is inspired by, and has the closest connection with the following algorithms that detect ridges in 2+1 dimensional scale-space and involve automatic scale selection.

(1) Crowley's peak and ridge linking

A discrete definition of ridges in a multi-resolution pyramid (analogous to a scale-space) was given in [CRO84], where the underlying function is the image intensity convolved with the difference of low-pass (DOLP) kernel. Ridges with major (P-paths) and minor (L-paths) extensions in the scale dimension are extracted separately. P-paths are identified by locating peaks (P-nodes) at each resolution level and linking them with those P-nodes at adjacent positions and in adjacent levels. L-paths are identified by locating ridges (R-nodes) at each resolution level. The R-nodes that have larger DOLP values than those R-nodes at neighbouring locations and in adjacent resolution levels are selected as L-nodes. These L-nodes are further linked with the largest adjacent L-nodes with the same direction in adjacent resolution levels to form ridge paths (L-paths) in the 3-D pyramid space. This process results in a graphical representation of scale-space ridges. However, this ridge definition is not rotation-invariant and has mathematical weakness [PIZ98] [EBE94a]. For example, the R-nodes are defined as local directional maxima in any of the four directions associated with the 8-neighbourhood of the pixel. A more strict ridge detector should depend on all the 8-neighbourhood, like in the height ridge definition [HAR83].

(2)Maximum convexity ridges

The maximum convexity ridge definition interprets the joint contribution of spatial space and scale to the identification of potential ridge points. It is an extension of the height ridge definition to the 3-D scale-space [MOR94] [EBE94b], see Fig. 2.10. A 1-D ridge point is defined as the local maximum in both the directions, $\mathbf{e}_1$ and $\mathbf{e}_2$, corresponding to the two negatively greatest eigenvalues of the Hessian matrix. Therefore, the ridge point is a local maximum in the plane spanned by the two eigenvectors, $\mathbf{e}_1$ and $\mathbf{e}_2$, normal to the ridge; the ridge is in the direction of the third eigenvector, $\mathbf{e}_3$.

Because one dimension of a scale-space is the scale parameter, which needs to be specifically treated, each spatial and scale change must be defined in terms of the relative scale. Differentiation must be performed with respect to dx/σ and $d\sigma/\sigma$, respectively, under Riemannian geometry in scale-space [EBE94b]. A modified Hessian matrix that reflects the interdependence of spatial space and scale is required. The Riemannian geometry is hyperbolic when the radius-to-scale ratio is equal to 1. Under hyperbolic geometry, the shortest path (distance) between two points in scale-space is not a straight line but a circular arc. Due to the complexity of the non-Euclidean geometry, the maximum convexity ridge definition has found limited use.

(3) Optimal scale ridges

In the optimal scale ridge definition, the contributions of space and scale to the identification of potential ridge points are considered independently [FRI92][PIZ94]. The maximal response over scale at each position is located and projected onto the image plane (i.e. the scale dimension is determined as one distinguished direction, $\mathbf{v}_{1,}$ in Fig. 2.11), forming an optimal scale response in 2-D spatial space. Then scale-space ridges are considered as the spatial ridges in the 2-D optimal scale response (i.e. the distinguished directions, $\mathbf{v}_2$ and $\mathbf{v}_3$, are normal to and consistent with the direction of the spatial ridge, respectively, see Fig. 2.11). A benefit of the optimal scale ridge definition is that the search space which must be considered is reduced by one dimension (the scale dimension) and ridges can be detected in 2-D using

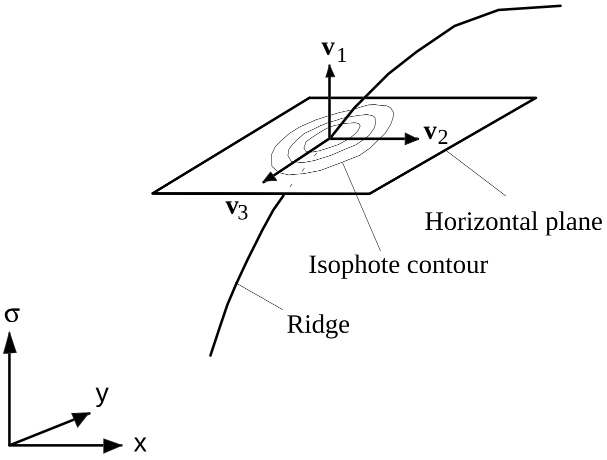


Fig. 2.11. A 1-D optimal scale (or optimal position) ridge in 3-D scale-space with the three distinguished directions.

Euclidean geometry and existing techniques as those described in Section 2.4.2.1.

The original definition of optimal scale ridges locates all local maxima over scale at each position. These maxima in scale-space may then be partitioned into connected subsets. For each connected subset, the locally maximal response forms a sub-image, whose spatial ridges constitute a part of the MMA. However, for complicated images, determining the connectivity of the subsets of maxima is problematic [FRI93]. It is necessary to take account of when the boundaries of a subset surface are encountered and when it is necessary to jump to another subset surface. Therefore, a procedure for computing ridges on discontinuous sub-images has not been developed [FRI93].

As an alternative, a simplified strategy for optimal scale ridge detection has been proposed [FRI93], in which only the global maximum through scale at each position is considered and the ridge search is conducted over a single image of the globally optimal scale response. Although this method can capture a significant portion of the MMA, it can restrict single points to belong to separate MMA branches at different scales, as in the case of embedded objects. Despite this disadvantage, this simplified strategy of the optimal scale ridge detection has been widely used to detect blood vessels in medical images [KOL95] [FEL97] [KRI98] [SAT98] and roads in remotely sensed images [KOL95]. Indeed it is appropriate that a point should not be the centre of several vessels or roads of different sizes.

(4) Lindeberg's ridge surface

Lindeberg proposed a scale-space ridge definition [LIN98], in which the contributions of space and scale to the identification of potential ridge points are also considered independently. In this method, the spatial ridges of the 2-D medialness response at each scale are extracted. Such ridges at a range of scales constitute ridge surfaces in the 3-D scale-space. Scale-space ridges are made up of those points, on the ridge surfaces, which are local maxima in a scale-increasing direction. A theoretically reasonable direction is one in the tangent plane of the ridge surface. Such an approach was proposed in [LIN98] but no implementation details and results were presented. Constructing connected ridge surfaces and determining their tangent planes in the scale-space is not a trivial task, especially for complex images. Alternatively, the scale axis can be selected as the scale-increasing direction. In this way, constructing ridge surfaces can be avoided by searching for those points that are both spatial ridges at a single scale (i.e. determining the distinguished directions, $\mathbf{v}_2$ and $\mathbf{v}_3$, in Fig. 2.11) and maxima over scale (i.e. considering the scale axis as the distinguished direction, $\mathbf{v}_{1,}$ in Fig. 2.11) [LIN98]. However, this approach tends to merge many edges as ridges, because the medialness responses of isolated edges, resulting from a traditional medialness operator, satisfy the ridge definition in this approach.

(5) Other developments based on maximal convexity ridges

Because computing medialness everywhere can be computationally very expensive and generally the MMA curves are sparse in scale-space, it is desirable to localize calculations to regions which contain MMA. A fast ridge extraction algorithm, *ridge flow and traversal*, was proposed in [PIZ98]. It finds an approximation to a ridge by user-interaction, moves to the ridge by looking for the position at which the gradients in the first two eigenvector directions vanish (i.e. $\mathbf{e}_i \cdot \nabla f = 0$ for i=1, 2), and then follows the ridge by approximating ridge tangent in $\mathbf{e}_3$ and root finding of $\mathbf{e}_i \cdot \nabla f = 0$ for i=1, 2. By following the ridge direction, it increases the chance that the ridge will not be lost. Fritsch *et al.* used this technique to detect ridges in 2-D optimal scale responses (“stimulated cores”) [FRI95]. Furst *et al.* used marching cores, which combine this technique with the marching-cubes approach [LOR87], to detect ridges in higher dimensional scale-spaces [FUR96b] [FUR98].

Because the MMA, unlike the medial axis, is a disconnected collection of curves without branching, Damon suggested the use of “connector curves” to continue the ridge and valley curves and fill in the gaps [DAM99]. These “connector curves” are obtained by relaxing one of the conditions for the ridge-valley curves. Let f with the eigenvalues λ_i and eigenvectors $\mathbf{e}_i$ of the Hessian matrix be as in Eq. (2.28). A point is defined as an r-connector point if $\mathbf{e}_i \cdot \nabla f = 0$ for i=1, 2 but $\lambda_2 > 0$; and a v-connector point if $\mathbf{e}_i \cdot \nabla f = 0$ for i=2, 3 but $\lambda_2 < 0$. A ridge, valley, or connector curve can only terminate when either the Hessian of f is singular (zero eigenvalue) or a partial umbilic point occurs (identical eigenvalues). Therefore, when a ridge ends, a local search for the beginning of a connector curve can be conducted to continue the ridge. By filling in the gaps between ridges, small disturbances of an image may change (e.g. split) the structure of a ridge but will not change the structure of ridge-valley-connector set. Furst *et al.* extended the “connector curves” to finding 1-D ridges in a 4-D scale-space [FUR96a].

2.4.2.3 Preview of our ridge detection scheme

In Section 3.3 we propose a sliding window algorithm to extract a complete set of optimal scale ridges. It differs from Fritsch’s simplified strategy of the optimal scale ridge detection in that it searches for each local maximum over scale, rather than the global maximum, at each location. At each scale in this algorithm, the “global” medialness maximum over scale (for individual locations), within a small scale range (sliding window), forms a “global” optimal scale response for that scale. A pixel is determined as a scale-space ridge point at the current scale, if it is both a ridge point in the optimal scale response at the current scale and a local maximum over scale. The implementation presented here avoids the non-trivial task of constructing sub-surfaces of connected maxima in the optimal scale ridge and Lindeberg’s ridge definitions.

To illustrate how the globally optimal scale ridge definition (Fritsch’s simplified strategy) fails in distinguishing embedded objects and the importance of locally optimal scale ridges, the behaviour of an embedded pulse model in scale-space is analyzed in Section 3.3.2. Using this pulse model, we also show how scale sampling rate and the width of the sliding window can be selected. The sliding window algorithm is described in Section 3.3.3. In Section 3.3.4, the ridge detection and linking procedures in each 2-D optimal scale response, used in the sliding window algorithm, are introduced. The applications of the sliding window algorithm to both synthetic and medical images are demonstrated in Section 4.3. The relationship between the sliding window algorithm and existing scale-space ridge definitions is discussed in Section 5.2.

Chapter 3 Theory and Computation

3.1 Boundariness

3.1.1 Scale-Space Boundariness

The scale-space for an original image, $I(\mathbf{x})$, is defined as:

$$L(\mathbf{x},\sigma) = I(\mathbf{x}) * G(\mathbf{x},\sigma) \tag{3.1}$$

where $\mathbf{x} = (x_1, \ldots, x_n) \in R^n$ denotes spatial position in an n-dimensional superplane (R^n represents an n-dimensional real domain); G is the n-dimensional Gaussian kernel with unit volume:

$$G(\mathbf{x},\sigma) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\mathbf{x}^T\mathbf{x}}{2\sigma^2} \right) \tag{3.2}$$

and $\sigma \in R^+$ is the scale and the standard deviation of the Gaussian (R^+ represents a one-dimensional positive real domain).

Because this CMAT medialness computation depends on the accumulation of boundariness responses in the scale-space for an image, some operators must applied to find points with boundary-like properties in the scale-space. However, it is not necessary to generate the image scale-space explicitly and then apply boundary-detection operation to this scale-space. If the boundary-detection operation is linear and shift invariant, e.g. differentiation in the gradient direction, considering that convolution is associative, the scale-space boundariness response can be written as:

$$\mathbf{B}_0(\mathbf{x},\sigma) = \nabla[I(\mathbf{x}) * G(\mathbf{x},\sigma)] = I(\mathbf{x}) * \nabla G(\mathbf{x},\sigma) \tag{3.3}$$

The formulation $I(\mathbf{x}) * \nabla G(\mathbf{x},\sigma)$ avoids the regularization problem involved in computing derivatives of a discrete function, and is necessary to compute the derivatives of the image [ter91]. The corresponding scale-space boundariness response in the 1-D spatial domain is:

$$B_0(x,\sigma) = I(x) * G_x(x,\sigma) \tag{3.4}$$

3.1.2 Size-Invariant Boundariness

Let a 1-D unit step function, $U(x)$, be defined as:

$$U(x) = \begin{cases} 1 & x > 0 \\ 0 & otherwise \end{cases} \tag{3.5}$$

The scale-space boundariness response for $U(x)$ is:

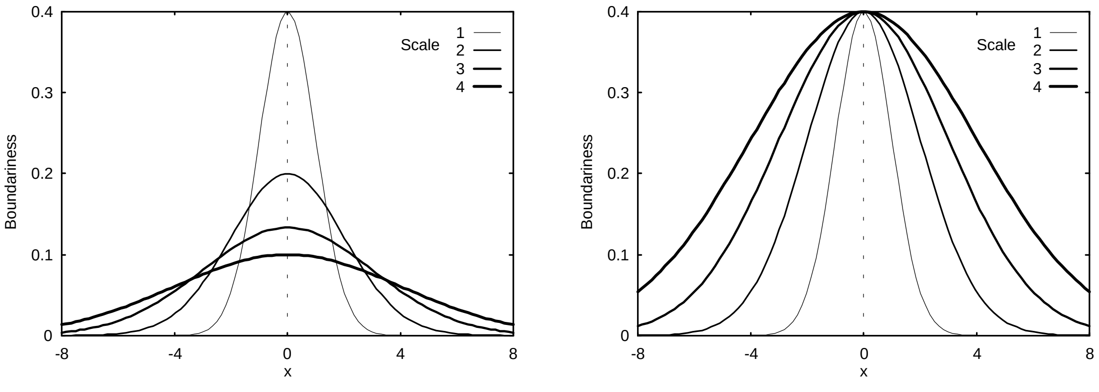
$$\begin{aligned} B_{0,U}(x,\sigma) = U(x) * G_x(x,\sigma) = \frac{dU(x)}{dx} * G(x,\sigma) \\ = \delta(x) * G(x,\sigma) = G(x,\sigma) \end{aligned} \tag{3.6}$$

as illustrated in Fig. 3.1(a). The magnitude of this response is scale-dependent. Therefore any medialness response, arising from the accumulation of such boundariness responses, may be also scale-dependent. As a result, when the medialness response is tracked through scale-space, an enhancement of medialness may result from either the interaction between boundariness responses, due to structural symmetry, or a change in the magnitude of boundariness responses themselves, due to scale changes. Therefore, the scale-space boundariness response must be scale-normalised to maintain a constant peak magnitude for a single step edge. In the 1-D spatial domain, the scale-space boundariness response for the input signal, $I(x)$, becomes:

$$B(x,\sigma) = I(x) * \sigma G_x(x,\sigma) \tag{3.7}$$

The scale-space boundariness response for a unit step edge becomes:

$$\begin{aligned} B_U(x,\sigma) = \sigma G(x,\sigma) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2} \right) \\ = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{X^2}{2} \right) \end{aligned} \tag{3.8}$$



(a) (b)

Fig. 3.1. Boundariness at several scales for the unit step function, (a) before and (b) after scale-normalisation.

as shown in Fig. 3.1(b), where $X = x/\sigma$ is the dimensionless "natural spatial coordinate" [ter91]. The new boundariness response for a step edge has a constant peak magnitude through scale. The boundariness response is also size-invariant, in the sense that, if we introduce the scaling transformations $x \to \lambda x$ and $\sigma \to \lambda\sigma$, then:

$$B_U(\lambda x, \lambda\sigma) = B_U(x, \sigma) \tag{3.9}$$

A position λx from the step edge will obtain the same boundariness value at a scale $\lambda\sigma$, as a position x from the edge at a scale σ.

The size-invariant boundariness response for the n-dimensional spatial domain is:

$$\mathbf{B}(\mathbf{x},\sigma) = I(\mathbf{x}) * \sigma\nabla G(\mathbf{x},\sigma) \tag{3.10}$$

It can be jointly represented by the gradient magnitude:

$$b(\mathbf{x},\sigma) = \left\|\mathbf{B}(\mathbf{x},\sigma)\right\| \tag{3.11}$$

and a unit gradient vector:

$$\hat{\mathbf{n}}(\mathbf{x},\sigma) = \frac{\mathbf{B}(\mathbf{x},\sigma)}{\left\|\mathbf{B}(\mathbf{x},\sigma)\right\|} \tag{3.12}$$

This definition of boundariness response is used in the following text.

3.2 Concordance-based Medial Axis Transform

3.2.1 The Definition

3.2.1.1 The definition for continuous co-ordinates

For each point $\mathbf{x}$ in boundariness space, the initial contribution, $b(\mathbf{x},\sigma)$, to medialness space is made at:

$$\mathbf{y} = \mathbf{x} + r\hat{\mathbf{n}}(\mathbf{x},\sigma) \tag{3.13}$$

where $\mathbf{y} \in R^n$ is the n-dimensional spatial position in medialness space, and $r = k\sigma$ (k is the ratio of radius to scale, see Fig. 3.2).

The integration of initial contributions in medialness space provides the initial medialness response. Written as a Radon Transform [GEL66], this is:

$$m_0(\mathbf{y},\sigma) = \int_{R^n} b(\mathbf{x},\sigma)\delta(\mathbf{y} - \mathbf{x} - r\hat{\mathbf{n}}(\mathbf{x},\sigma))d\mathbf{x} \tag{3.14}$$

where the Dirac function $\delta(\mathbf{x})$ is defined by $\int_{R^n} \delta(\mathbf{x})d\mathbf{x} = 1$ and $\delta(\mathbf{x}) = 0$ when $\mathbf{x} \neq \mathbf{0}$. In 2-D

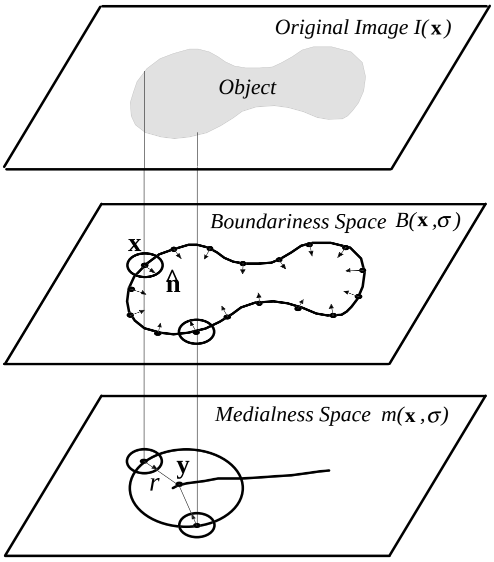


Fig. 3.2. Medialness is the integration of boundariness contributions.

space, the initial medialness at **y** is the integration of directional boundariness around a circle centred at **y** with radius *r*; the orientation of the boundariness being towards the centre of the circle, see Fig. 3.2. This definition is close to Blum's MAT [BLU67] and the HMAT [MOR94].

The initial medialness has high values for symmetric structures but also responds to edge structures, because the boundariness response, $b(\mathbf{x},\sigma)$, for an isolated edge results in the medialness response, $m_0(\mathbf{y},\sigma) = b(\mathbf{x},\sigma)$. To obtain a medialness response that is sensitive only to symmetric edges, the contribution to medialness from boundariness is constrained by a weighting function, $p(\mathbf{x},\sigma)$, called *contribution confidence*. The requirements for contribution confidence are:

(1) The medialness response should not be greater than the initial medialness response, i.e. $0 \le p(\mathbf{x},\sigma) \le 1$.

(2) A single boundary point produces no medialness response, i.e.:

$p(\mathbf{x},\sigma) = 0$ *if* $m_0(\mathbf{y},\sigma) = b(\mathbf{x},\sigma)$.

(3) Given the boundariness contribution, $b(\mathbf{x},\sigma)$, the greater the initial medialness, $m_0(\mathbf{y},\sigma)$, the more possible it is for **x** to contribute to a true symmetric structure. Therefore $p(\mathbf{x},\sigma)$ increases with $m_0(\mathbf{y},\sigma)$, or decreases with $b(\mathbf{x},\sigma)/m_0(\mathbf{y},\sigma)$.

The natural definition of contribution confidence, used here, is:

$$p(\mathbf{x},\sigma) = 1 - \frac{b(\mathbf{x},\sigma)}{m_0(\mathbf{y},\sigma)} \tag{3.15}$$

A discussion of alternative definitions of contribution confidence can be found in Section 3.2.3. The measure of contribution confidence can be considered to be an estimate of the extent to which one boundariness point contributes to existing evidence of medialness. Using this confidence measure the boundariness contribution to true medialness is:

$$\mathbf{C}(\mathbf{x},\sigma)=\mathbf{B}(\mathbf{x},\sigma)\,p(\mathbf{x},\sigma) \tag{3.16}$$

The magnitude of such contribution is:

$$c(\mathbf{x},\sigma)=\left\|\mathbf{C}(\mathbf{x},\sigma)\right\| \tag{3.17}$$

The medialness response is the integration of boundariness contributions to true medialness:

$$m(\mathbf{y},\sigma)=\int_{R^n} c(\mathbf{x},\sigma)\delta(\mathbf{y}-\mathbf{x}-r\hat{\mathbf{n}}(\mathbf{x},\sigma))d\mathbf{x} \tag{3.18}$$

The estimation of confidence in Eq. (3.15) provides the “concordance” property of the Concordance-based Medial Axis Transform (CMAT), by which the medial response is high only when multiple boundary points jointly provide evidence of a symmetric structure. With this formulation an isolated boundary cannot produce a medial response. This concordance property is analysed in Section 3.2.2.

The formula of $\hat{\mathbf{n}}(\mathbf{x},\sigma)$, defined in Eq. (3.12) and used in Eqs. (3.13), (3.14) and (3.18), determines the medialness responses for dark objects (low grey-level value) on a light background (high grey-level value). The same equations can be used to compute the medialness response for the inverse phase (light object on a dark background) by inverting the sign of the unit vector $\hat{\mathbf{n}}(\mathbf{x},\sigma)$ such that:

$$\hat{\mathbf{n}}(\mathbf{x},\sigma)=-\frac{\mathbf{B}(\mathbf{x},\sigma)}{\left\|\mathbf{B}(\mathbf{x},\sigma)\right\|} \tag{3.19}$$

It is relatively simple to combine the computation of both sets of medialness either by summing absolute or signed values of medialness.

3.2.1.2 Computation in discrete space

For a discretely sampled space, boundariness contributions to medialness space are typically distributed between grids. In order to improve the localisation of the medial axis, the real-values of the co-ordinates of boundariness contributions are preserved throughout the computation of medialness.

Let $\mathbf{m} \in I^n$ and $\mathbf{n} \in I^n$ be the discrete co-ordinates of points in n-dimensional boundariness and medialness spaces, respectively (I^n represents n-dimensional integer domain). Let $\mathbf{y} = (y_1, \ldots, y_N) \in R^n$ be the continuous co-ordinates of points in medialness space. Now we define a pseudo version of the discrete Delta function with real argument components:

$$\delta_p(\mathbf{y}) = \begin{cases} 1, & -1/2 \le y_i < 1/2, \quad i = 1, \ldots, n \\ 0, & \mathit{otherwise} \end{cases} \tag{3.20}$$

To estimate the contribution confidence of each boundariness point **m**, the initial medialness at the corresponding position, **y,** in Fig. 3.2 must be known. This contribution is the summation of initial boundariness contributions over the unit volume centred at **y,** i.e. :

$$m_0(\mathbf{y}, \sigma) = \sum_{\mathbf{m}} b(\mathbf{m}, \sigma) \delta_p(\mathbf{y} - \mathbf{m} - r\hat{\mathbf{n}}(\mathbf{m}, \sigma)) \tag{3.21}$$

Likewise the boundariness contribution, $b(\mathbf{m},\sigma)$, can be thought of as the accumulation of boundariness over a unit volume centred at **m**. The contribution confidence, $p(\mathbf{m},\sigma)$, and the boundariness contribution, $c(\mathbf{m},\sigma)$, are computed using Eqs. (3.15) and (3.17), respectively. With discrete sampling, there are few boundariness points that contribute to medialness value at exactly the same position. To compensate for the sparseness of these contributions, a weighted summation over a volume is used in the computation of the medialness response:

$$M(\mathbf{n}, \sigma) = \sum_{\mathbf{m}} c(\mathbf{m}, \sigma) G(\mathbf{n} - \mathbf{m} - r\hat{\mathbf{n}}(\mathbf{m}, \sigma), s) \tag{3.22}$$

Here G is the Gaussian kernel with the standard deviation s. This is similar to the procedure adopted in other medialness algorithms [MOR94]. To construct the scale-space of the original image we can either solve the diffusion equation or apply Gaussian filter kernels of increasing size. The Gaussian filter is a linear diffusion process by which the response at a point in the original image is distributed over an area, the size of which changes linearly with scale. The standard deviation, s, of the Gaussian weighting function in Eq. (3.22) is chosen as $s = \lambda\sigma$ (λ is a proportionality constant). The amplitude of the Gaussian function is 1% of the peak value at a displacement of more than three times its standard deviation. Therefore the radius of the volume influenced by a point is limited to $3s$. This is illustrated for the 2-D case in Fig. 3.3. Point B in boundariness space contributes to a circular area centred at A in medialness space. If we consider the weighted accumulation over an angular range θ, we have:

$$\theta = 2\sin^{-1}(3\lambda / k) \tag{3.23}$$

and θ is kept constant across scale. The value of λ can be chosen to suit image properties such as noise. One may use a large λ for noisy images and a small λ for noise-free images. Currently λ is set to 0.5 in our implementation.

The position argument in the Gaussian function of Eq. (3.22) is real-valued therefore it is

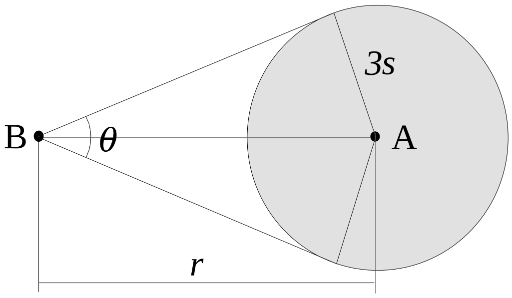


Fig. 3.3. Boundariness space (point B) contributes to medialness space centred at point A. All similar contributions within a radius of $3s$ of A are summed using a Gaussian weighting.

not practical to construct a Gaussian look-up table indexed by position to accelerate computation. Instead, each boundariness contribution is distributed linearly between the four nearest neighbouring grid points, in proportion to the distance to each neighbour, using:

$$m(\mathbf{n},\sigma) = \sum_{\mathbf{m}} c(\mathbf{m},\sigma) W(\|\mathbf{n} - \mathbf{m} - r\hat{\mathbf{n}}(\mathbf{m},\sigma)\|) \tag{3.24}$$

Where the distance weighting function W is defined as $W(d) = \max\{1 - d, 0\},\ d \geq 0$. This definition means that, if $d \leq 1$, $W(d)=1-d$; otherwise $W(d)=0$. Other weighting functions, such as a small Gaussian, are possible. The Gaussian weighting kernel is then applied to compute the final medialness response:

$$M(\mathbf{n},\sigma) = m(\mathbf{n},\sigma) * G(\mathbf{n},s) \tag{3.25}$$

3.2.1.3 Algorithmic description of CMAT

In the computation of the CMAT a look up table (LUT) is used to identify the correspondence between medialness and boundariness points. An outline description of the CMAT algorithm is given in Fig. 3.4 and the computation of the LUT is described in the following section. The LUT contains groups of boundariness points. The groups are identified in a second LUT.

```
1. For each scale $\sigma$
    1.1. For each discrete boundariness point compute the boundariness value using Eq. (3.10).
    1.2. For each discrete point in boundariness space compute the location of its contribution in medialness space using Eq. (3.13).
    1.3. Generate the LUT association between each discrete medialness point and the group of contributing discrete boundariness points (Fig. 3.6).
    1.4. For each discrete boundariness point
        1.4.1 Accumulate the initial boundariness contributions at the location of its contribution using the LUT (Fig. 3.7) and Eq. (3.21).
        1.4.2 Compute the contribution confidence and "true" contribution for each boundariness point using Eqs. (3.15) and (3.17).
    1.5. For each discrete medialness point accumulate the closest boundariness "true" contributions, using the LUT (Fig. 3.7) and Eq. (3.24).
    1.6. Convolve the result of step 1.5 with the Gaussian smoothing kernel, Eq. (3.25).
```

Fig. 3.4. Outline of CMAT algorithm

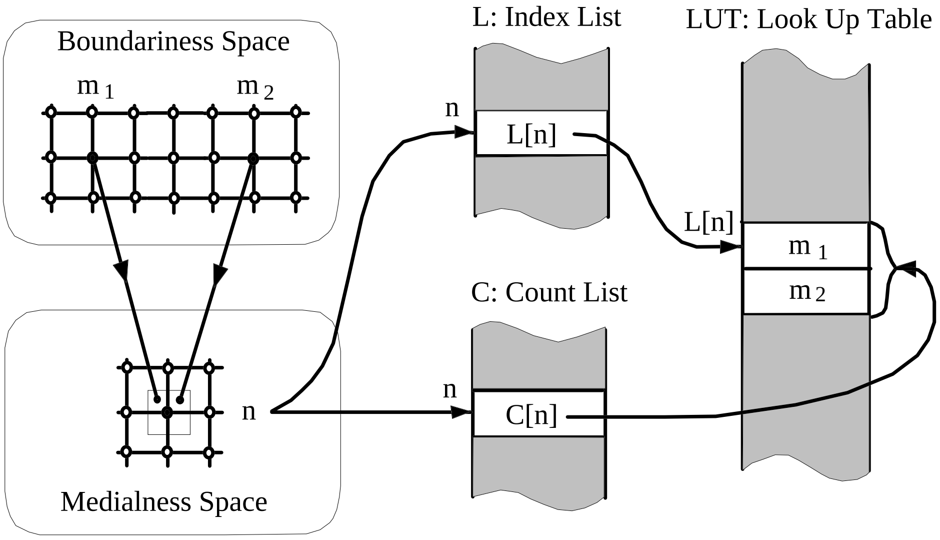


Fig. 3.5. The LUT for associating medialness and boundariness points.

Although medialness response is the accumulation of co-ordinated boundariness contributions, only a few boundariness points are relevant to the computation of each medialness value. A look-up table (LUT) is used to identify the boundariness points that satisfy $\delta_p(\mathbf{n}-\mathbf{m}-r\hat{\mathbf{n}}(\mathbf{m},\sigma))=1$, and are therefore relevant at each scale.

Suppose that there are N points in boundariness and medialness space. Then an N-element LUT (see Fig. 3.5), organised into N groups and indexed by medialness position n (n=1,…,N) will describe the association between medialness and boundariness points. A group, n, is the set of the boundariness points whose contributions are closest to medialness grid point n. The content of the LUT is the position, m, of the boundariness points. The start position and count of each group are recorded in arrays L and C, respectively. If group n is empty, its start position coincides with that of the next medialness point. The creation of the LUT is described in Fig. 3.6 and a detailed description of how the LUT is read is given in Fig. 3.7. The CMAT medialness responses for 2-D images are shown in Section 4.2.1.

m	the grid point co-ordinates in boundariness space; m=1, …, N
n	the grid point co-ordinates in medialness space; n=1, …, N
L[n]	the start address of group n in the LUT
C[n]	the counter of group n in the LUT
LUT[i]	the content of LUT indexed by i; i=0,…,N-1
Y[m]	the co-ordinates of medialness point associated with boundariness point m

```
1. For each discrete point, n (n=1,…N), in medialness space, set group counter
   C[n]=0.
2. For each discrete point, m, in boundariness space
    2.1  Compute the real-valued position, y, of its contribution in medialness space
         using Eq. (3.13).
    2.2  Find the grid point, n, in medialness space that is closest to y.
    2.3  Y[m]=n;
    2.4  C[n]=C[n]+1.
3. L[1]=0.
4. For successive discrete points, n, in medialness space; n=2, …, N,
    L[n]=L[n-1]+C[n-1].
5. For each discrete point, m, in boundariness space
    5.1  n=Y[m].
    5.2  Fill in boundariness position, m, into the LUT: LUT[L[n]+C[n]-1]=m.
    5.3  C[n]=C[n]-1.
```

Fig. 3.6. Algorithm for creation of LUT association between medialness and boundariness points.

```
1. For each discrete point, n, in the supporting region of medialness space:
    1.1. Compute the size of group n using  C[n]=L[n+1]-L[n],
    1.2. Repeat
        1.2.1 Read the boundariness co-ordinates, m, of a member of group n from the
              LUT as:  m=LUT[L[n]+C[n]-1].
        1.2.2 Accumulate the boundariness contribution from m.
              In step 1.4.1 of the CMAT algorithm (Fig. 3.4) use Eq. (3.21).
              In step 1.5 of the CMAT algorithm (Fig. 3.4) use Eq. (3.24).
        1.2.3 C[n]=C[n]-1.
        Until C[n]=0.
```

Fig. 3.7. Reading the LUT.

3.2.1.4 The HMAT-2 transform

This transform is an adaptation of the HMAT. It is the same as the CMAT with the concordance criteria omitted and differs from the HMAT in using a circular region for distributing boundariness contributions, rather than an arc [XU98]. It is presented here to demonstrate the role of the concordance criteria. The HMAT-2 medialness response is computed as a weighted summation of initial boundariness contributions over a region, that is:

$$M_0(\mathbf{n},\sigma) = \sum_{\mathbf{m}} b(\mathbf{m},\sigma) G(\mathbf{n} - \mathbf{m} - r\hat{\mathbf{n}}(\mathbf{m},\sigma), s) \tag{3.26}$$

The algorithm for the HMAT-2 is the same as that for the CMAT when step 1.4 is excluded and the *contribution confidence,* in step 1.5, is set to 1. The principle of the HMAT-2 algorithm is the same as that in the initial medialness computation. However, HMAT-2 is defined on discrete co-ordinates, while initial medialness is defined on continuous co-ordinates. Also, HMAT-2 includes a smoothing operation not present in the initial medialness computation. These differences all arise from the shift from continuous to discrete space. The HMAT-2 medialness responses for 2-D images are shown in Section 4.2.1.

3.2.2 The Concordance Property

3.2.2.1 Concordance factor

Suppose N points contribute to a medial point A; that the boundariness responses at these points are $b_i, i = 1, \ldots, N$; the sum of these boundariness values is S and the average is $\bar{b}$. The medialness estimate at point A, without considering the relation among b_i, (as computed in the HMAT) is:

$$M_{HMAT}(\mathbf{x}_{\mathrm{A}}, \cdot) = S \tag{3.27}$$

For the CMAT, the confidence of contributions to a true medial structure is $1 - \frac{b_i}{S}$, for each point b_i . Therefore the medialness response is:

$$\begin{aligned} M_{CMAT}(\mathbf{x}_{\mathrm{A}}, \cdot) &= \sum_{i=1}^{N} b_i (1 - \frac{b_i}{S}) \\ &= S - \frac{1}{S} \sum_{i=1}^{N} b_i^2 \end{aligned} \tag{3.28}$$

Examining $\sum_{i=1}^{N} b_i^2$, we have,

$$\begin{aligned} \sum_{i=1}^{N} b_i^2 &= \sum_{i=1}^{N} (b_i - \bar{b})^2 + N\bar{b}^2 \\ &= \sum_{i=1}^{N} (b_i - \bar{b})^2 + \frac{S^2}{N} \end{aligned} \tag{3.29}$$

Because $\sum_{i=1}^{N}(b_i - \bar{b})^2 \geq 0$,

$$\sum_{i=1}^{N} b_i^2 \geq \frac{S^2}{N} \tag{3.30}$$

$\sum_{i=1}^{N} b_i^2$ has a minimum value, $\frac{S^2}{N}$, when:

$$b_i = \bar{b} = \frac{S}{N}, \quad i = 1, \ldots, N \tag{3.31}$$

On the other hand when $b_i \geq 0$, $i = 1, \ldots, N$, it is obviously true that:

$$\sum_{i=1}^{N} b_i^2 \leq S^2 = \left(\sum_{i=1}^{N} b_i\right)^2 \tag{3.32}$$

$\sum_{i=1}^{N} b_i^2$ has a maximum value, S^2, when there is one and only one non-zero boundariness contribution, i.e.:

$$\begin{cases} b_k = S \\ b_i = 0, \quad i = 1, \ldots, N \ and \ i \neq k \end{cases} \tag{3.33}$$

Therefore $\sum_{i=1}^{N} b_i^2$ is constrained to the range:

$$\frac{S^2}{N} \leq \sum_{i=1}^{N} b_i^2 \leq S^2 \tag{3.34}$$

Combining Eqs. (3.29) and (3.34), we have:

$$0 \leq \sum_{i=1}^{N}(b_i - \bar{b})^2 \leq S^2\left(1 - \frac{1}{N}\right) \tag{3.35}$$

Let:

$$c = \begin{cases} 1 - \dfrac{\sum_{i=1}^{N} (b_i - \bar{b})^2}{S^2 \left(1 - \dfrac{1}{N}\right)} & N > 1 \\ 0 & N = 1 \end{cases} \tag{3.36}$$

Then the medialness response of the CMAT in Eq. (3.28) can be written as:

$$\begin{aligned} M_{CMAT}(\mathbf{x}_{\mathrm{A}}, \cdot) &= S - \frac{1}{S}\left[\sum_{i=1}^{N} (b_i - \bar{b})^2 + \frac{S^2}{N}\right] \\ &= \left(1 - \frac{1}{N}\right) Sc \end{aligned} \tag{3.37}$$

Note that *c* can also be written as:

$$c = \begin{cases} 1 - \dfrac{\sum_{i=1}^{N} (b_i - \bar{b})^2 \Big/ (N-1)}{S^2 / N} & N > 1 \\ 0 & N = 1 \end{cases} \tag{3.38}$$

The item $\sum_{i=1}^{N} (b_i - \bar{b})^2 \Big/ (N-1)$ is the sample variance, a measure of the dispersion, of the *N* observations, b_i, normalised by S^2/N, the largest value of the dispersion (see Eq. (3.35)). Therefore *c* measures the dispersion or variability of the contributions. The smaller the amount of variability, the greater the value of *c* and the more concordance there is among the contributions. Thus *c* is called the concordance coefficient. We note that $0 \le c \le 1$. When $c = 0$ Eq. (3.33) is satisfied, which is equivalent to there being only one boundary point casting its contribution to point A and there being no concordance. When $c = 1$ Eq. (3.31) is satisfied, which corresponds to *N* equal-valued contributions combining to give the maximal degree of concordance.

The CMAT medialness response in Eq. (3.37) is proportional to the sum of boundariness responses and to the concordance coefficient among the boundariness responses. In addition when $c = 1$ the medialness responses of the CMAT are smaller than those of HMAT by a factor $1-(1/N)$. This factor is related to the number of boundary points involved in a symmetry. Therefore the CMAT medialness response at end points and branch points is enhanced. A similar behaviour is present in the HMAT medialness response. Morse [MOR91] pointed out that such enhancement is beneficial because it more clearly identifies the end and branch points that best define the axis.

3.2.2.2 Boundariness response versus concordance

The medialness response of the CMAT is proportional to both the sum of boundariness responses and the concordance between boundariness responses. It is therefore important to consider the interaction of these two factors. Consider the idealised situation, shown in Fig. 3.8, of boundariness responses for two step edges that differ in magnitude. A scale-invariant boundariness operator, $\sigma\nabla G(x,\sigma)$, is used, and its response for a single step edge has a Gaussian-like profile with a constant peak magnitude through scale (see size-invariant boundariness in Section 3.1.2). The boundariness responses at A and B are the local maxima with the magnitudes of *V* and *U* respectively (*U*>*V*). C is the position at which the boundary response due to the edge at B is equal to the maximum response of the edge at A. Suppose that positions E and D are the midpoints of AC and AB respectively, and that the half distances of AC and AB are $k\sigma_1$ and $k\sigma_2$, respectively. Here *k* is the ratio of radius to scale.

Since: $$M_{HMAT}(\mathbf{x}_E,\sigma_1) = B(\mathbf{x}_A,\sigma_1) + B(\mathbf{x}_C,\sigma_1) = 2V$$

$$M_{HMAT}(\mathbf{x}_D,\sigma_2) = B(\mathbf{x}_A,\sigma_2) + B(\mathbf{x}_B,\sigma_2) = U + V$$

Then: $$M_{HMAT}(\mathbf{x}_D,\sigma_2) > M_{HMAT}(\mathbf{x}_E,\sigma_1).$$

By inspection it is clear that the HMAT will produce a maximal medialness response at posi-

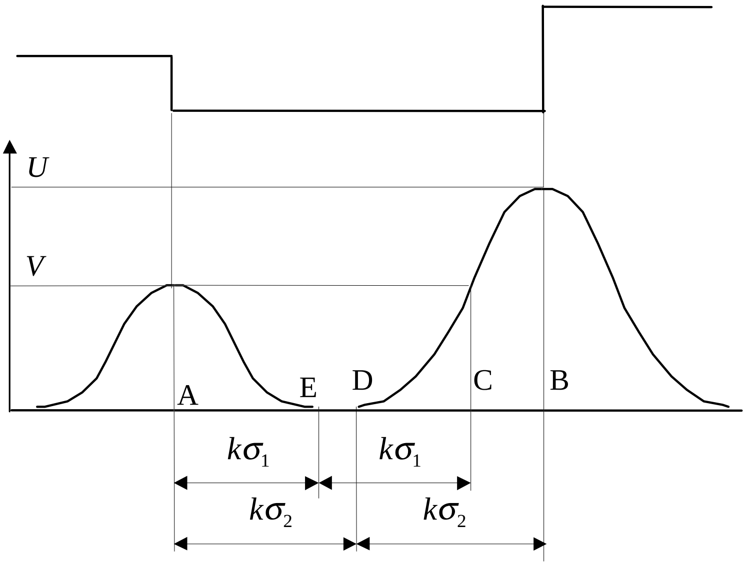


Fig. 3.8. Interaction of boundariness strength and concordance in estimating medialness response.

tion D, the medial position between the peaks of boundariness response, and give an accurate identification of medial position and object width. For the CMAT the concordance between the boundariness responses at positions A and B (which determine the medialness response at D) is weaker than the concordance between the boundariness responses at positions A and C (which determine the medialness response at E). However, the boundariness sums at D and E bear the inverse relationship therefore it is difficult to compare, in a simple manner, the relationship between the medialness responses at D and E.

In order to estimate how concordance influences the way in which boundariness contributes to medialness we analyze incremental changes in CMAT medialness. Suppose, as in the previous section, that N points in boundariness space contribute to a medial point; that the boundariness response at these points is $b_i, i = 1, \ldots, N$; and that the sum of these responses is S. Then the CMAT medialness response is defined by Eq. (3.28).

If any one boundariness response, b_k , is increased by Δb_k , then the summation is increased to $S + \Delta b_k$ and the increment in the CMAT medialness response is:

$$\Delta M_{CMAT}(\cdot,\cdot) = \frac{\sum_{i=1}^{N} b_i^2 - 2Sb_k + S^2}{S} \cdot \frac{\Delta b_k}{S + \Delta b_k} \tag{3.39}$$

Let:
$$T = \frac{\sum_{i=1}^{N} b_i^2 - 2Sb_k + S^2}{S} \tag{3.40}$$

using $\sum_{i=1}^{N} b_i^2 = b_k^2 + \sum_{\substack{i=1 \\ i \neq k}}^{N} b_i^2$, then:

$$T = \frac{\sum_{\substack{i=1 \\ i \neq k}}^{N} b_i^2 + (S - b_k)^2}{S} \tag{3.41}$$

Note that item *T* is not related to the increment Δb_k. If we treat all initial boundary responses as constant then *T* is also a constant. Note also that $T = 0$ only when $b_i = 0, i = 1, \ldots, N \text{ and } i \neq k$, this corresponds to one boundariness point contributing to the medialness value. In this case, no CMAT medialness response will be produced. In the case of multiple boundariness elements, $T > 0$, will there be an increase of boundariness response resulting in an increase in medialness response.

With the HMAT, the medialness increment due to increased boundariness response, Δb_k, is:

$$\Delta M_{HMAT}(\cdot,\cdot) = \Delta b_k \tag{3.42}$$

If the original boundariness responses are treated as constants (and so is their sum *S*), the increment of the HMAT and CMAT medialness responses change with the ratio:

$$\beta = \frac{\Delta b_k}{S} \tag{3.43}$$

Here β may be considered as the relative increment in boundariness. With this definition, the

HMAT and CMAT medialness increment can be written as:

$$\Delta M_{HMAT}(\cdot,\cdot) = S\beta \tag{3.44}$$

$$\Delta M_{CMAT}(\cdot,\cdot) = T\frac{\beta}{1+\beta} \tag{3.45}$$

Therefore the medialness increments for the HMAT and CMAT depend on the behaviour of the coefficients β and $\frac{\beta}{1+\beta}$, respectively. The relationship between β and $\frac{\beta}{1+\beta}$ is shown in Fig. 3.9. In summary:

1. The medialness of the CMAT, like that of the HMAT, increases monotonically with boundariness increment**.** The CMAT forms a maximal response at positions midway between peaks of boundariness responses and the magnitude of the medialness peak changes monotonically with the magnitude of the contributing boundariness peaks. Therefore medial position and object width can be identified as accurately using the CMAT algorithm as the HMAT algorithm.
2. While the HMAT medialness increases linearly with boundariness increment, the increase in CMAT medialness is less than that for the HMAT and approaches the limiting value of T as $\beta \to \infty$. For example, when $\beta = 1$, according to the definition in Eq. (3.43), $\Delta b_k = S$; In this case one boundariness response, b_k, becomes much larger than any other and therefore violates the previously computed concordance level; When $\beta = 1$ the coefficient $\beta/(1+\beta)$, in the CMAT medialness increment of Eq. (3.45) is decreased to half the value of the coefficient, β, used in the equation for the increase in HMAT medialness. This means that weak concordance among boundariness responses restrains the increase in the CMAT medialness response.

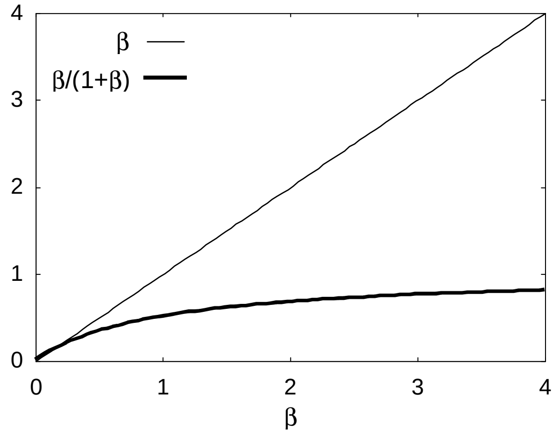


Fig. 3.9. Increment of HMAT and CMAT medialness related to boundariness increment: β and $\beta/(1+\beta)$ versus β.

3.2.3 Alternative Definitions of Contribution Confidence

In Eq. (3.15), we defined the contribution confidence function, $p(\mathbf{x},\sigma)$, in the form of $f(x) = 1 - x$, where $x = b(\mathbf{x},\sigma)/m_0(\mathbf{y},\sigma)$. In general, $p(\mathbf{x},\sigma)$ is defined as a function of the ratio $b(\mathbf{x},\sigma)/m_0(\mathbf{y},\sigma)$, i. e. :

$$p(\mathbf{x},\sigma) = f[b(\mathbf{x},\sigma)/m_0(\mathbf{y},\sigma)] \tag{3.46}$$

The requirements for the function $f(x)$ are:

$$\begin{aligned} &(1) \quad 0 \le f(x) \le 1, \quad 0 < x \le 1 \\ &(2) \quad f(1) = 0 \\ &(3) \quad f'(x) < 0 \end{aligned} \tag{3.47}$$

Condition (1) means that the medialness response should not be greater than the initial medialness response. Condition (2) means that a single boundary point produces no medialness response. Condition (3) means that given a boundariness contribution, the greater the initial medialness, the more possible it is for this boundariness point to contribute to a true symmet-

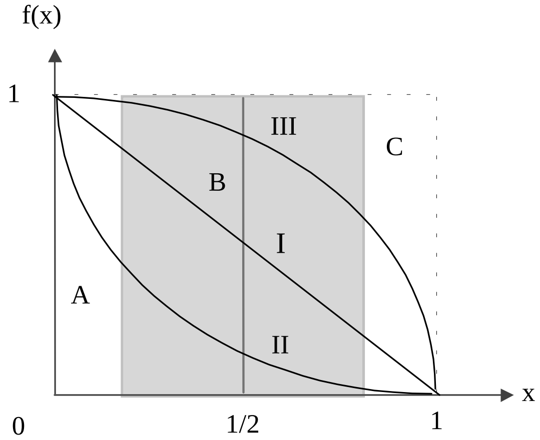


Fig. 3.10. Contribution confidence functions that strictly satisfy requirements of Eq. (3.47).

ric structure. These conditions constrain the values of *x* and $f(x)$ to the area between the coordinate axes and the dotted lines, shown in Fig. 3.10. This area can be divided into three regions according to the value of *x*:

Region A **(** $x \ll 1$ **):** This region corresponds to situations in which boundary points cluster around an end-point, or a weak boundariness response that contributes to the initial medialness jointly with other much stronger boundariness responses.

Region C **(** $x \to 1$ **):** This region corresponds to the situation in which a strong boundariness response contributes to the initial medialness jointly with other much weaker boundariness responses.

Region B: This region, centred at $x = 1/2$, corresponds to the situation in which two compatible boundaries, parallel to each other, contribute to the initial medialness.

The contribution confidence function defined in Eq. (3.15) corresponds to the function $f(x) = 1 - x$ represented by curve I in Fig. 3.10. Varying the shape of this function, without violating the requirements of Eq. (3.47), will emphasise (or de-emphasise) end-point medialness with respect to the medialness arising from parallel boundaries. Curve II suppresses and curve III enhances the relative medialness arising from parallel boundaries.

If the last requirement of Eq. (3.47), $f'(x) < 0$, is weakened a relationship for $f(x)$ can be defined to give greater emphasis to the medialness arising from parallel boundaries and less emphasis to the medialness arising from endpoints. Curve IV in Fig. 3.11 is one such function. The function of curve IV is:

$$f(x) = \begin{cases} 2x & if\, 0 < x \leq 1/2 \\ 2(1-x) & if\, 1/2 < x \leq 1 \end{cases} \tag{3.48}$$

To understand how this arises, consider the medialness for the “tube contour” shown in Fig. 3.12. In this figure the boundariness at each point on the contour is b and there are N points on the half-circular arc ($N >> 2$). Thus the medialness at points O (N contributions) and B (2 contributions, letting N=2), using Eq. (3.15), are:

$$\begin{aligned} M_I(\mathbf{x}_O,\cdot) &= Nb\left(1-\frac{1}{N}\right) = (N-1)b \\ M_I(\mathbf{x}_B,\cdot) &= b \end{aligned} \tag{3.49}$$

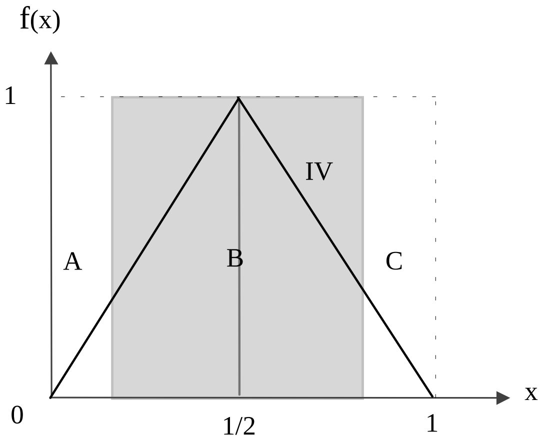


Fig. 3.11. A contribution confidence function that satisfies the weakened requirements of Eq. (3.47).

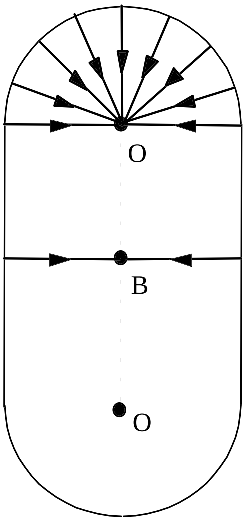


Fig. 3.12. Medialness accumulation from end-points and parallel boundaries.

Using Eq. (3.48) the medialness at points O and B becomes:

$$\begin{aligned} M_{IV}(\mathbf{x}_O,\cdot) &= Nb\frac{2}{N} = 2b \\ M_{IV}(\mathbf{x}_B,\cdot) &= 2b \end{aligned} \tag{3.50}$$

The resulting uniform-response medialness is no longer related to the number of contributions and the medialness along the middle line of parallel boundaries is greatly enhanced. The difference in the medialness resulting from curves I and IV is similar in nature to the difference between the “mean deviation from flatness” and the “principle deviation from flatness” [FRI93][ter91], computation of medialness. The “mean deviation from flatness” is based on the use of an isotropic convolution of a Laplacian with the image and the “principal deviation from flatness” is based on the application of a Laplacian convolution in the direction which maximises the response.

The analyses and results presented in this thesis are based on the contribution confidence $p(\mathbf{x},\sigma)$ defined in Eq. (3.15), in the form of $f(x) = 1 - x$ (curve I in Fig. 3.10). The CMAT medialness responses using other definitions of contribution confidence are shown in Section 4.2.3.

3.3 Sliding Window Algorithm for Ridge Detection

3.3.1 Introduction to the MMA for single objects

A symmetric pulse can be denoted by:

$$P(x) = \begin{cases} 1 & |x| < W/2 \\ 0 & otherwise \end{cases} \tag{3.51}$$

as shown in Fig. 3.13. This profile corresponds to a bright “object” on a darker background. *W*/2 is the pulse half-width. Note that:

$$P(x) = U(x + W/2) - U(x - W/2) \tag{3.52}$$

where *U*(*x*) is a unit step function defined in Eq. (3.5).

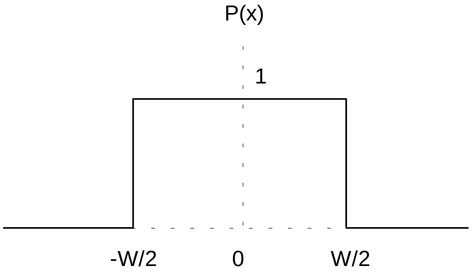


Fig. 3.13. A 1-D perfectly symmetric pulse.

The LoG medialness response of *P*(*x*), through scale-space, is:

$$\begin{aligned} M_P(x,\sigma) &= -\sigma^2 G_{xx}(x,\sigma) * P(x) \\ &= -\sigma^2 [G_x(x + W/2,\sigma) - G_x(x - W/2,\sigma)] \end{aligned} \tag{3.53}$$

as shown in Fig. 3.14. The scale parameter is replaced by the operator radius, $r=\sigma$ (see Section 4.1.1 for reasons), and the medialness response is normalised by its global maximum value. The LoG medialness forms a global maximum at the position of the pulse centre and at the operator radius of the pulse half-width [FRI93], which reflects the symmetry of the pulse. Therefore the location and size of the pulse (object) can be identified by searching for the

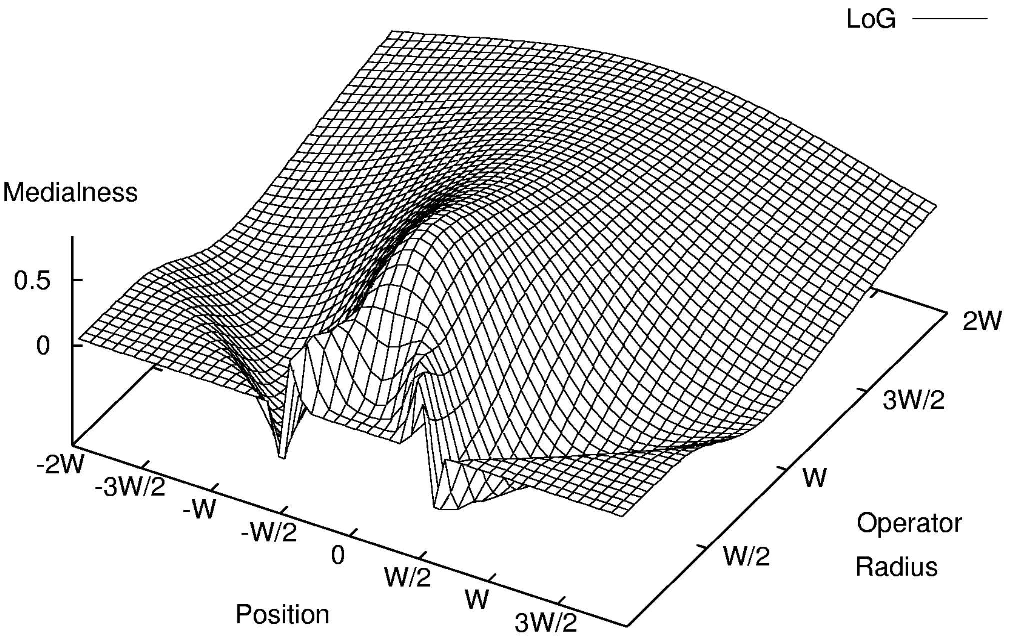
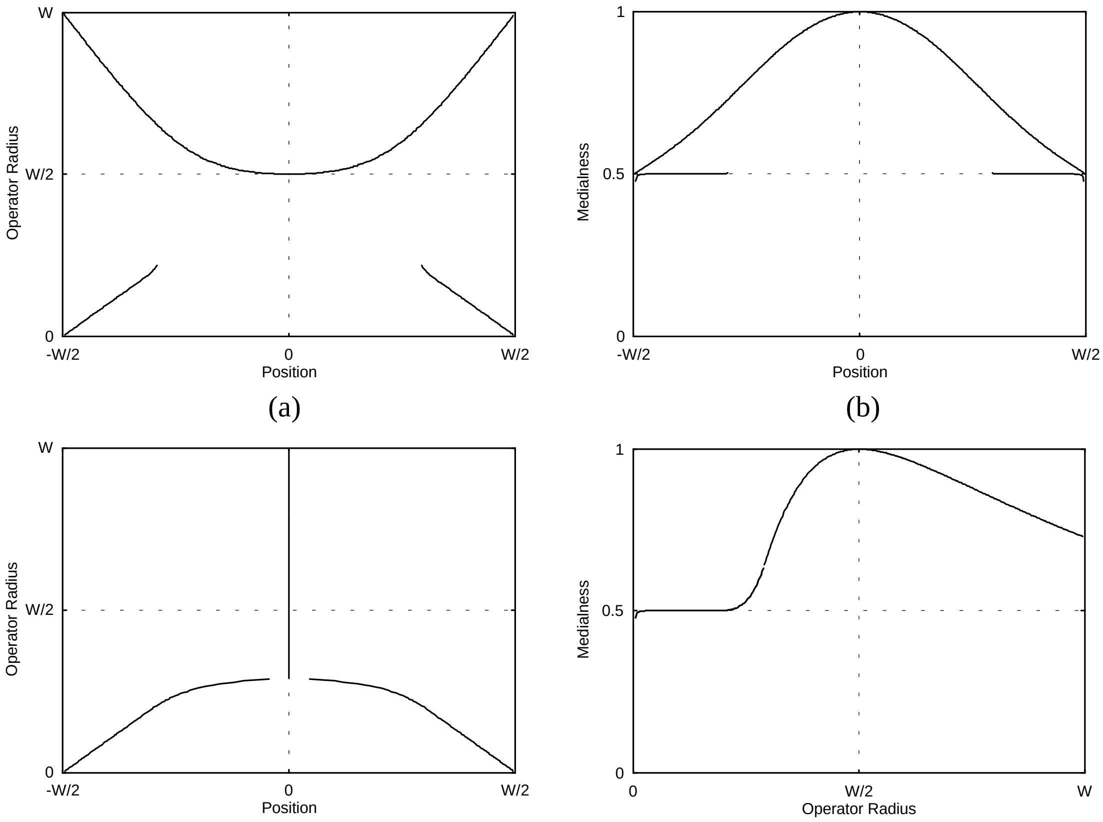


Fig. 3.14. The LoG medialness response through scale for the symmetric pulse shown in Fig. 3.13.

global maximum in the medialness scale-space.

Given the medialness response over scale and space, there are several efficient schemes that can determine the pulse centre and width. The optimal scale definition [FRI93] considers the local maxima with respect to scale at each position. The operator radii (scales) and responses of these maxima, as a function of position, are called *optimal radii (scales)* and *optimal radius (scale) response*, as shown in Fig. 3.15(a) and (b), respectively. The pulse centre always appears as a spatial maximum of the optimal radius response (Fig. 3.15(b)). The associated optimal radius of the pulse centre coincides with the pulse half-width (Fig. 3.15(a)).

The optimal position definition considers the local maxima with respect to position at each scale. The positions and responses of these maxima, as a function of operator radius (scale), are called *optimal positions* and *optimal position response*, as shown in Fig. 3.15(c) and (d), respectively. At the operator radius equal to the pulse half-width, there always is a maximum in the optimal position response (Fig. 3.15(d)). The associated optimal position of



(a) (b)

(c) (d)

Fig. 3.15. The loci of medialness maxima for the symmetric pulse: (a) The optimal radii and (b) optimal radius response, as a function of position, (c) the optimal positions and (d) optimal position response, as a function of operator radius.

this operator radius coincides with the position of the pulse centre (Fig. 3.15(c)).

It is noted that the loci of optimal scales and optimal positions are disjoint curves. This is because the summation of two approaching boundariness lobes may result in a single lobe of medialness, even when the boundariness lobes peak away from each other at a large distance. Numerical search techniques are required to locate the optimal-scale maxima and optimal-position maxima for the pulse. They cannot be determined analytically due to the presence of transcendental functions [FRI93].

For a 2-D elongated object, there are a range of local object centres and radii, corresponding to the centres and radii of locally maximal inscribed disks, called medial axis [BLU67]. For the optimal (radius) scale definition, the optimal radius response becomes a 2-D surface; The locations and associated optimal radii of the ridge points in such a response indicate the locations of the medial axes and the local widths [FRI93]. For the optimal position definition, each medial axis point is both a spatial ridge point of the medialness response at a scale and a local maximum over scale [LIN98].

3.3.2 The MMA of Embedded Objects

3.3.2.1 The embedded object model

Both the position and width of an isolated object can be correctly identified in the MMA representation. However, in the case of multiple objects existing within a close distance, the medialness response of one object begins to interfere with those of the other objects. This can alter the behaviour and localisation accuracy of the MMA and make ridge extraction using Fritsch’s simplified version of the optimal scale definition difficult.

The interference between multiple objects can occur for embedded objects. In this case, some locations may have an optimal scale for each object in the embedded grouping. The ability of the MMA to distinguish the response of each embedded objects has not been demonstrated [FRI93]. To illustrate how relative size, height, and position of an embedded object affect medialness response and MMA representations, an embedded pulse model is constructed using two overlapping pulses with adjustable relative width, height, and bias, as shown in Fig. 3.16, and defined as.

$$E(x) = P(x) + hP\left(\frac{x - bW/2}{e}\right) \tag{3.54}$$

Where $P(x)$ is a unit pulse of width W and centred at x=0, as defined in Eq. (3.51); Parameter e is the ratio of the width of the embedded pulse to that of the outer pulse ($0 < e \leq 1$); $e \to 0$ corresponds to only the outer pulse existing; when e=1 the embedded pulse coincides with the

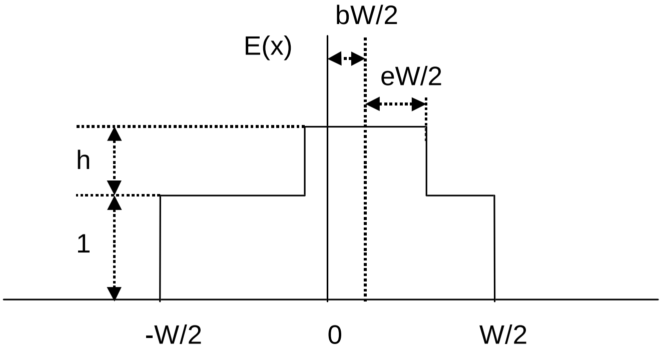


Fig. 3.16. An embedded object model.

outer pulse and the group is equivalent to a single pulse. The parameter *h* is the ratio of the height of the embedded pulse to that of the outer pulse ($0 \le h \le \infty$). Setting the parameter, *h*, to $h = 0$ corresponds to only the outer pulse existing and, as $h \to \infty$, corresponds to only the embedded pulse existing. The parameter *b* is the ratio of the offset of the pulse centres with respect to the width of the outer pulse ($-1 + e \le b \le 1 - e$). The value, $b = 0$, corresponds to two concentric pulses and when $b = \pm(1 - e)$ the right (left) edges of the embedded and outer pulses coincide.

To compute the LoG medialness response of this embedded pulse model, the function *E*(*x*) can be decomposed into unit step functions, *U*(*x*), defined in Eq. (3.5). Considering $dU(x)/dx = \delta(x)$, we have:

$$E(x) = U(x + W/2) - U(x - W/2) + hU[x - (b - e)W/2] - hU[x - (b + e)W/2] \quad (3.55)$$

$$\begin{aligned} M_E(x,\sigma) &= -\sigma^2 G_{xx}(x,\sigma) * E(x) \\ &= -\sigma^2 G_x(x,\sigma) * E_x(x) \\ &= -\sigma^2 G_x(x,\sigma) * \{\delta(x + W/2) - \delta(x - W/2) + h\delta[x - (b - e)W/2] \\ &\quad - h\delta[x - (b + e)W/2]\} \\ &= -\sigma^2 \{G_x(x + W/2, \sigma) - G_x(x - W/2, \sigma) + hG_x[x - (b - e)W/2, \sigma] \\ &\quad - hG_x[x - (b + e)W/2, \sigma]\} \end{aligned} \quad (3.56)$$

where $$G_x(x,\sigma) = -\frac{x}{\sqrt{2\pi}\sigma^3}\exp\left(-\frac{x^2}{2\sigma^2}\right) \tag{3.57}$$

The medialness response of an embedded pulse group (*e*=0.2, *h*=1, and *b*=0) is shown in Fig. 3.17. Operator radius is substituted for scale, $r=\sigma$, and the response is normalised by the global maximum value of the medialness response for an isolated, symmetric pulse, of unit height. Because of the linearity of the LoG operator, the medialness response of overlapping pulses is equivalent to the overlapping responses of the individual pulses. There are clearly two response peaks, one at a small radius and the other at a large radius, and two pairs of edge responses, one close to and the other relatively far from the pulse centre. Intuitively, the peak at the smaller value of radius and the edge responses closer to the pulse centre are attributed to the embedded pulse. The peak at the larger value of radius and the edge responses relatively far from the pulse centre are due to the outer pulse. In the following sections, we examine how

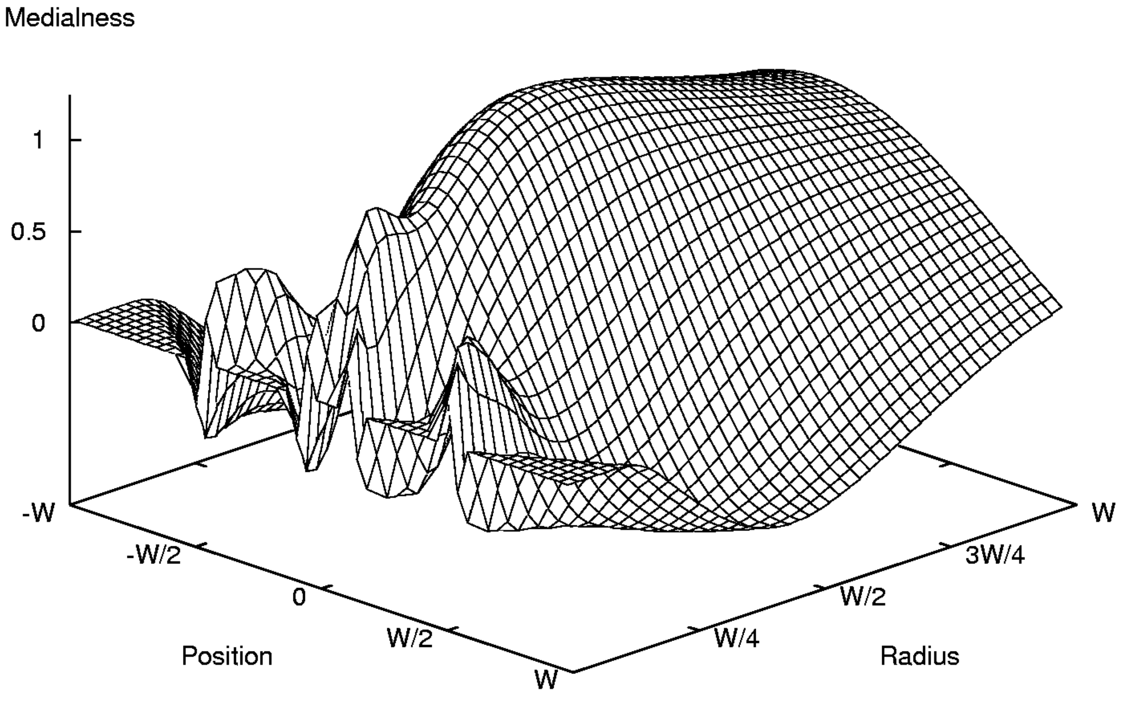


Fig. 3.17. The medialness response through radius for an embedded pulse.

the relative width, height, and position of an embedded pulse, with respect to the outer pulse, can influence medialness responses and the MMA representation.

3.3.2.2 Effect of relative object width

Fig. 3.18 shows the optimal radii and optimal radius responses, as a function of position, for embedded pulses of varying width. Each embedded pulse has an equal height contrast with the outer pulse (h=1), and their centres coincide (b=0). Fig. 3.19 shows the optimal positions and optimal position responses, as a function of operator radius, for the same set of pulses. The response traces through scale, at the centre of the embedded and outer pulses (x=0), when both pulses exist in isolation, are also illustrated in the right column of Fig. 3.19 and labelled as “Embedded (isolated)” and “Outer (isolated)”, respectively.

The LoG medialness response for a unit step edge is:

$$\begin{aligned} R_U(x,\sigma) &= -\sigma^2 G_{xx}(x,\sigma) * U(x) = -\sigma^2 G_x(x,\sigma) * \delta(x) \\ &= -\sigma^2 G_x(x,\sigma) = \frac{x}{\sqrt{2\pi}\sigma}\exp\left(-\frac{x^2}{2\sigma^2}\right) \end{aligned} \tag{3.58}$$

Then the LoG response trace through scale at a distance, d, from the unit step edge is:

$$R_U(d,\sigma) = \frac{d}{\sqrt{2\pi}\sigma}\exp\left(-\frac{d^2}{2\sigma^2}\right) \tag{3.59}$$

The LoG medialness response at the centre of a symmetric pulse is the linear summation of the LoG responses for both edges of the pulse. The response trace at the centre of the outer pulse is double that at a distance $d = W/2$ from a unit step edge; The response trace at the centre of the embedded pulse is double that at a distance $d = eW/2$ from a step edge of height h. Therefore, at the middle position, x=0, of the embedded and outer pulses, when they exist in isolation, the two response traces can be expressed as:

$$
\begin{aligned}
M_O(0,r) &= 2R_U(W/2,r) = \frac{W}{\sqrt{2\pi}r}\exp\left(-\frac{W^2}{8r^2}\right) \\
M_E(0,r) &= 2hR_U(eW/2,r) = \frac{heW}{\sqrt{2\pi}r}\exp\left(-\frac{e^2W^2}{8r^2}\right)
\end{aligned}
\tag{3.60}
$$

Due to the size-invariance of the LoG response to a step edge, the LoG response trace in the centre of a pulse is also size–invariant: $M_O(0,r)$ and $M_E(0,r)$ have the same maximum value (when h=1) at radii, $r=W/2$ and $r=eW/2$, respectively. Because of the linearity of the LoG operator, the response trace for the grouping of the two pulses, at x=0, is:

$$
\begin{aligned}
M_G(0,r) &= M_O(0,r) + M_E(0,r) \\
&= \frac{W}{\sqrt{2\pi}r}\exp\left(-\frac{W^2}{8r^2}\right) + \frac{heW}{\sqrt{2\pi}r}\exp\left(-\frac{e^2W^2}{8r^2}\right)
\end{aligned}
\tag{3.61}
$$

Despite the presence of edge responses at small radii, the optimal position response coincides with $M_G(0,r)$. It is also noted that $M_O(0,r)$ (also for the LoG response trace for any pulse) increases at 0.25 of its optimal radius, i.e. $r>W/8$.

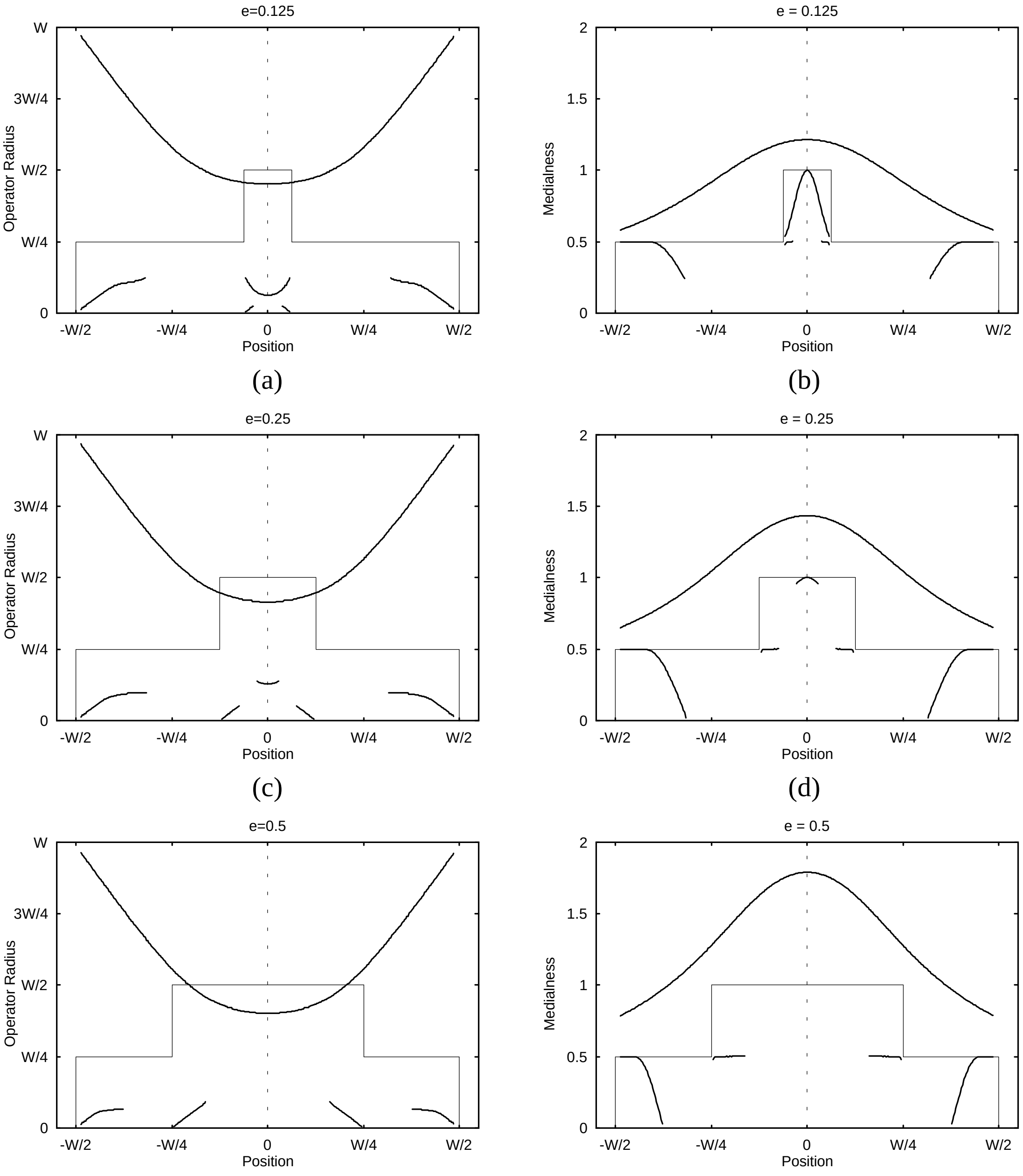
When the embedded pulse is much narrower than the outer pulse, e=0.125 in Fig. 3.18(a)-(b), the loci of optimal radii and optimal radius responses are almost the union of those for the embedded and outer pulses in isolation, except for the enhanced response of the outer pulse. In Fig. 3.19(b), the peak of the response trace for the embedded pulse is not influenced by the response trace for the outer pulse, because $e<0.25$ (the response trace of the outer pulse appears at a radius, $r>W/8$, larger than the optimal radius of the embedded pulse, $r=eW/2=W/16$). Therefore the embedded pulse forms a separate peak, at a radius of $r=eW/2$ and with unit magnitude, in the optimal position response, for the pulse grouping. At a radius, $r>W/8$, the optimal position response for the pulse grouping overlaps the response trace for the outer pulse and the long tail of the response trace for the embedded pulse. Therefore, the outer pulse forms another separate peak, at a radius of $r<W/2$ and with a greater magnitude, in the optimal

position response for the pulse grouping. Although both the embedded and outer pulses are distinguishable in scale-space, the outer pulse is always measured with a narrow width, contrary to the accurate representation of the embedded pulse. In addition, even with the same height contrast (h=1), the outer pulse always corresponds to a higher peak than the embedded pulse does in the optimal position response for the grouping (see also Fig. 3.18(b)). If the simplified strategy of the optimal scale definition is followed, only the outer pulse will be reflected in the MMA representation.

As the embedded pulse becomes wider, its response trace will be shifted closer to that of the outer pulse. When e=0.25, as shown in Fig. 3.19(d), the peak of the response for the embedded pulse begins to interfere with the response of the outer pulse, and becomes almost indistinguishable (see also the shortened loci in Fig. 3.18(c)(d)). Therefore the ability of scale-space analysis to discern embedded objects is surprisingly limited: it can only identify those embedded objects with at most 1/4 the width of an outer object, provided that h=1 (applied for $h \in [1/4, 2]$). This also implies that, for each position, no two peaks of response trace can exist at similar radii. For an object of width W, either embedded in a larger object or embedding a smaller object, no second peak in the central response trace is formed at radii, $r \in (W/8, 2W)$, for h=1.

When e=0.5, as shown in Fig. 3.19(f), the response trace for the embedded pulse is closer to that of the outer pulse, and the summation of both response traces results in a single lobe. The loci and responses look like those for a single pulse (Fig. 3.18 (e)(f) and Fig. 3.19(e)(f)). In this case (e>0.25), scale-space analysis tends to consider the grouping of the pulses as a single pulse, rather than as one pulse embedded in a wider pulse. This “new” pulse has an intermediate half-width between $eW/2$ and $W/2$, and an equivalent height contrast between 1 and h+1, both of which depend on the value of e and h. In general, a wider embedded pulse (greater *e*) increases both the equivalent radius and height of the new pulse; A greater magnitude embedded pulse (greater *h*) increases the equivalent height but decreases the equivalent half-width.

Due to the perfect symmetry of the pulse group when *b*=0, the location of the MMA coincides with the pulse centre. However, the embedding relation affects the half-width estimate associated with the MMA. Fig. 3.20 shows the radii of the response trace peaks, at the centre of the pulse group, as a function of the relative width, *e*, of the embedded pulse. The locus of the embedded pulse disappears at *e*=0.25, while that of the outer pulse is continuous through e (this may be reversed when *h* is of large value). In effect the response of the embedded pulse is absorbed in that of the outer pulse, rather than both being combined into a “new” pulse. Note that the loci of the embedded pulse when grouped or in isolation coincide, and thus the width of the embedded pulse is correctly reflected in the MMA representation. However, the outer pulse tends to be underestimated. Only at *e*=0 and *e*=1, corresponding to a single pulse, can the size of the outer pulse be correctly identified. In Fig. 3.20 (*h*=1), the smallest radius for the outer pulse (*r*=0.4W) occurs at *e*=0.4. It is also noted that, for the same location, no two local maxima over radius (scale) exist at a similar radius (scale dimension).

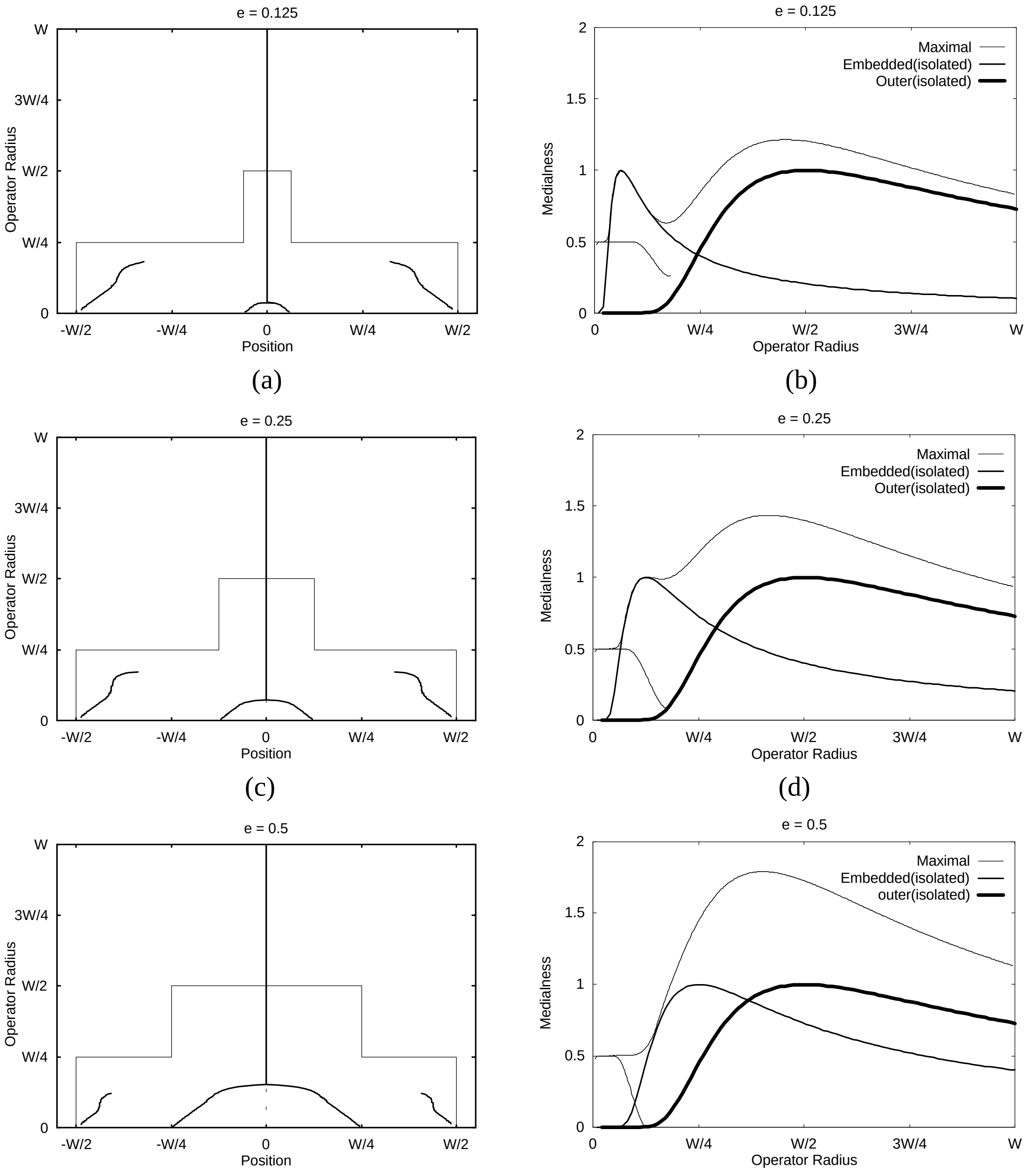


(a) (b)

(c) (d)

(e) (f)

Fig. 3.18. The optimal scale maxima for embedded pulses of varying width: the radii (left column) and responses (right column), as a function of position.



(a) (b)

(c) (d)

(e) (f)

Fig. 3.19. The optimal position maxima for embedded pulses with varying width: the positions (left column) and responses (right column), as a function of operator radius. The response traces at the centre of the embedded and outer pulses, when they exist in isolation, are also illustrated at right.

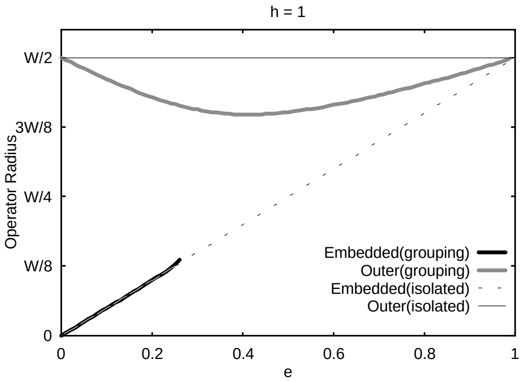


Fig. 3.20. The estimated radii, as a function of relative pulse width, e, for the embedded and outer pulses.

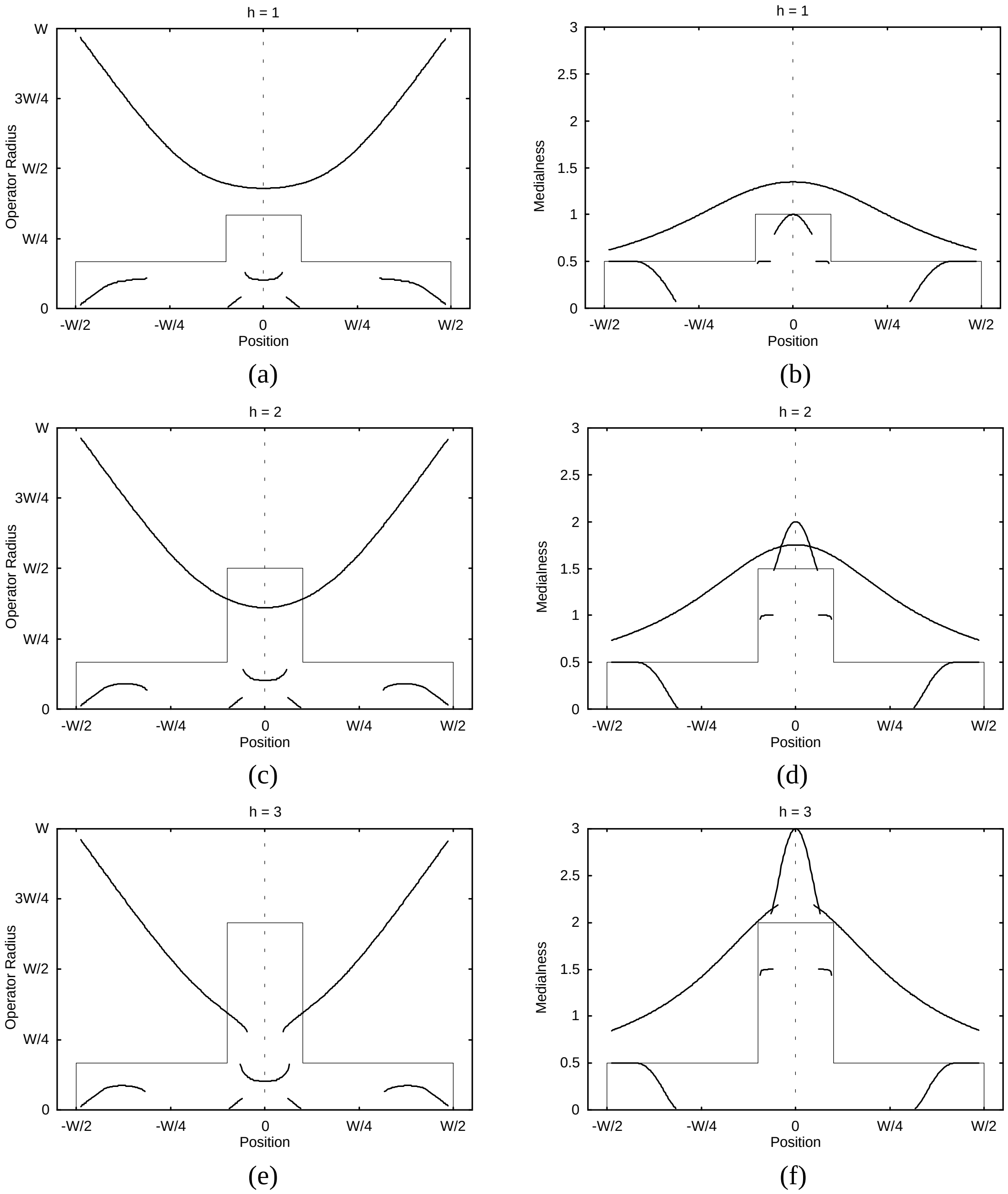
3.3.2.3 *Effect of relative object height*

Fig. 3.21 shows the optimal radii and optimal radius responses, as a function of position, for embedded pulses of varying relatively height, *h*. Each embedded pulse has a relative width, *e*=0.2, and its centre coincides with that of the outer pulse (*b*=0). Fig. 3.22 shows the optimal positions and optimal position responses, as a function of operator radius, for the same set of embedded pulses.

Because our selection of *e*<0.25, the peak of the response trace for the embedded pulse is not influenced by the response of the outer pulse (right column in Fig. 3.22). The loci and responses due to the embedded pulse show no structural change: the peak in the optimal position response is at a radius of *r*=0.2*W*/2, and the peak response increases linearly with relative pulse height, *h*. The optimal position response, due to the outer pulse, is the summation of the responses for both the outer and embedded pulses. The response peak due to the outer pulse is shifted to a smaller radius and disappears when the relative height, *h*=3, as shown in Fig. 3.22(f). As a result, the locus of optimal radii for the outer pulse breaks up at the pulse centre (Fig. 3.21(e)). This behaviour indicates the poor performance of scale-space analysis in discerning the presence of an embedded object group: only those outer objects with at least 1/3 the height contrast of an embedded object are detected. If the simplified optimal scale definition [FRI93] is used, the MMA representation of embedded pulses becomes even worse: at a relative height, h=1, only a response for the outer pulse is reflected in the globally optimal scale response. At relative height of h=2 and h=3 only the response for the embedded pulse is reflected in the globally optimal scale response.

Because Fritsch's scheme [FRI93] seeks only the global maximum over scale at each location, its MMA representation switches between the local maxima, over scale, generated by the embedded and outer pulses. At a relative pulse height of h=1.6, the local maximum of the embedded pulse dominates that of the outer pulse. When the relative pulse height h is slightly higher or lower than 1.6, the pulse grouping behaves as a single pulse with a width of $0.2W$ or $0.775W$. This is problematic since similar pulses can produce dramatic changes in the MMA representation.

Due to the perfect symmetry of the pulse grouping (b=0), the location of the MMA is kept at the pulse centre. Fig. 3.23 shows the radii of the local maxima of the response trace, at the pulse group centre, as a function of the relative height of the embedded pulse. The locus of the embedded pulse is continuous through h, while that of the outer pulse disappears at a relative height of h=2.7 when the estimated radius, $r=0.572W/2$, is at its smallest. The width of the embedded pulse is correctly reflected as the relative height of the embedded pulse varies (it coincides with the locus of the embedded pulse existing in isolation). However, the outer pulse width is underestimated. Only at a relative pulse height, h=0, corresponding to a single pulse, is the size of the outer pulse correctly estimated. It is noted again that, for each location, the local maxima over radius (scale) exist far away from each other in radius (scale dimension).

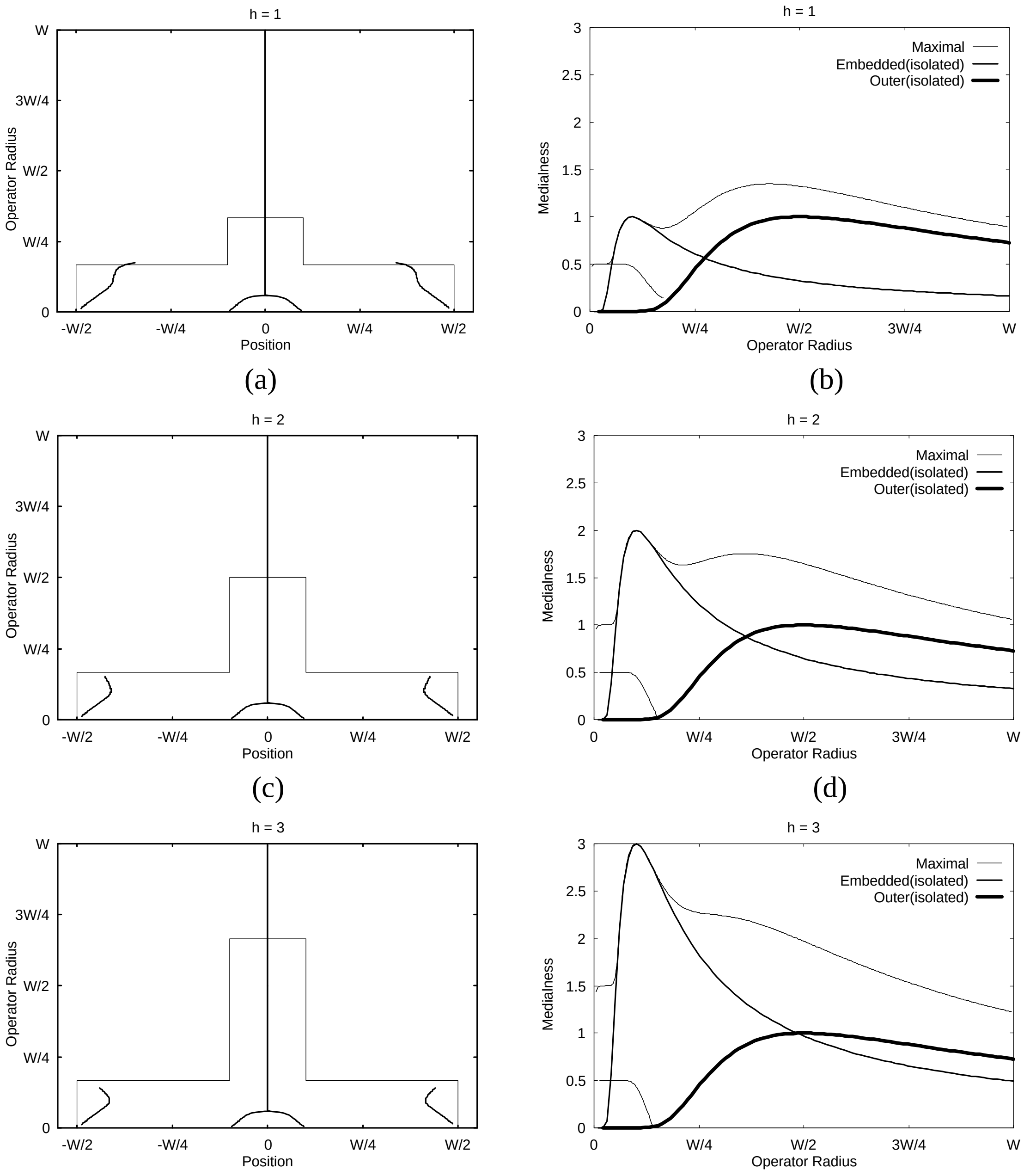


(a) (b)

(c) (d)

(e) (f)

Fig. 3.21. The optimal scale maxima for embedded pulses of varying height: the radii (left column) and responses (right column), as a function of position,.



(a) (b)

(c) (d)

(e) (f)

Fig. 3.22. The optimal position maxima for embedded pulses with varying height: the positions (left column) and responses (right column), as a function of operator radius. The response traces at the centres of the embedded and outer pulses, when they exist in isolation, are also illustrated at right.

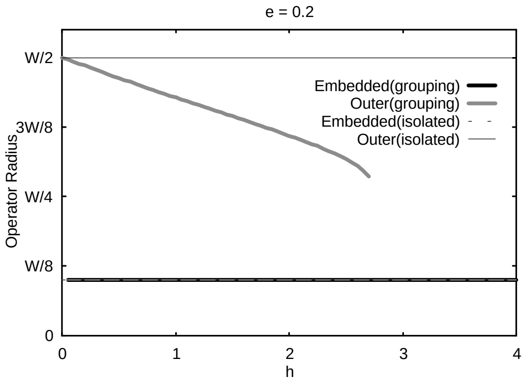


Fig. 3.23. The estimated radii, as a function of relative pulse height, h, for the embedded and outer pulses.

3.3.2.4 Effect of relative position

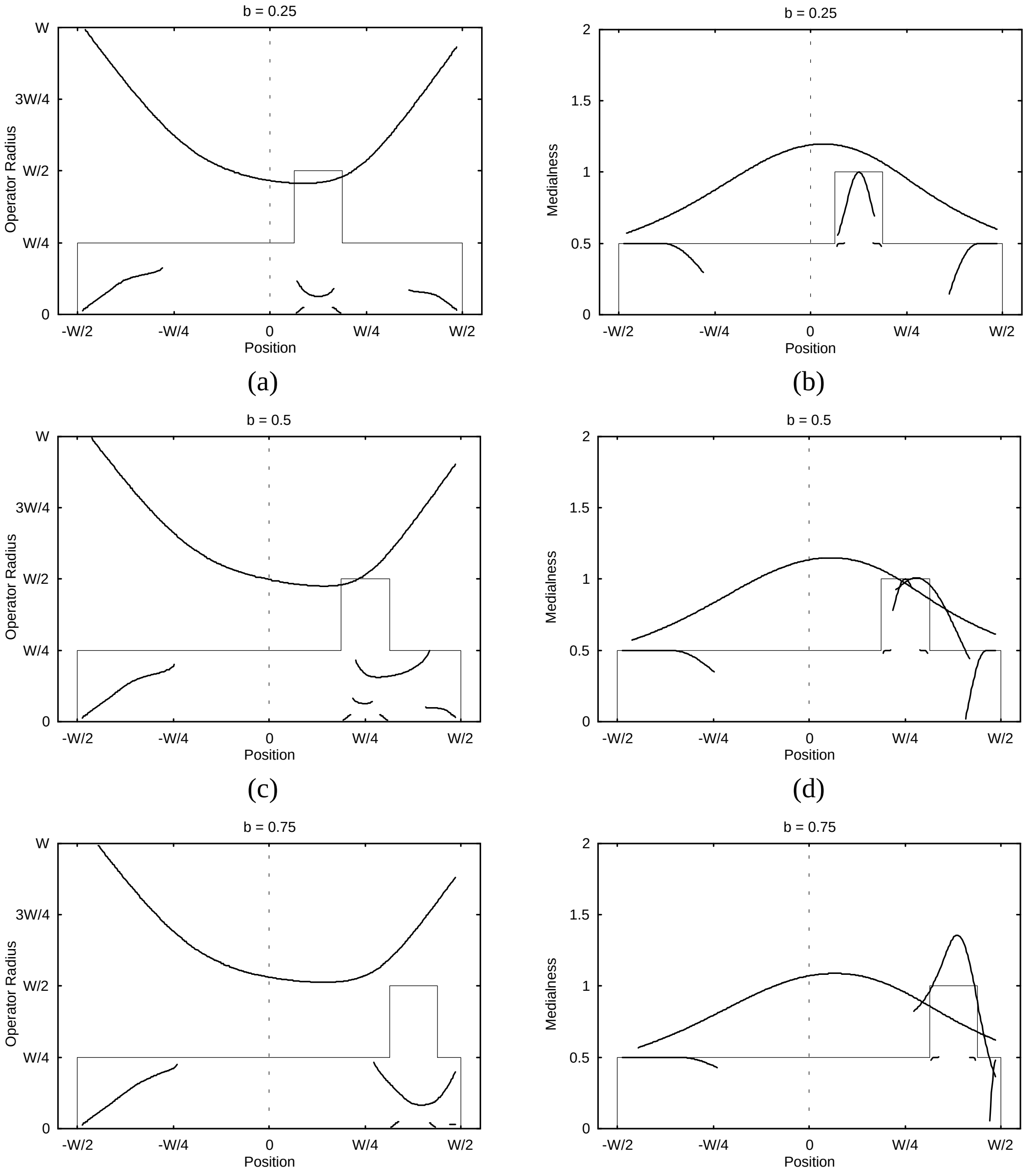
Fig. 3.24 shows the optimal radii and optimal radius responses, as a function of position, for embedded pulses of varying positional bias, *b*. Each embedded pulse has a relative width, *e*=0.125, and a relative height contrast, *h*=1. Fig. 3.25 shows the optimal positions and optimal position responses, as a function of operator radius, for the same set of pulses. In the right column of Fig. 3.25, the curve, labelled as “Embedded (isolated)”, represents the response trace at the centre, $x=bW/2$, of the embedded pulse existing in isolation. The curve, labelled as “Outer (isolated)”, represents the response trace at the centre, $x=0$, of the outer pulse existing in isolation. The curve labelled as “Spurious (isolated)”, if present, represents the response trace at the centre, $x=(b-e+1)W/4$, of a putative unit pulse bounded by the left edge of the embedded pulse and the right edge of the outer pulse. Because these response traces are for different positions, the optimal position response is no longer the simple summation of these response traces.

When the pulse offset, *b*=0.25, both the loci and responses for the outer pulse are shifted to those for the embedded pulse by the asymmetry of the grouping. The outer pulse is found at a position right of its centre and with a smaller radius. On the other hand, both the location and radius of the embedded pulse can be correctly identified (Fig. 3.24 (b) and Fig. 3.25(b)). In Fig. 3.24(b), the maximum of the optimal radius response for the embedded pulse is lower in magnitude than the response at the same location for the outer pulse, although the two pulses have the same height contrast. Therefore the medialness response at the centre of a narrower pulse (or a smaller object in 2-D case) can be weaker than the medialness response away from the centre of a wider pulse (or a larger object in 2-D case). If we detect only the globally optimal radius for each location, as in Fritsch's strategy, the MMA representation of narrower pulses (smaller objects) will be lost. Thus, large scale features tend to dominate small scale features, unless the smaller scale features have a much higher contrast, as shown in Section 3.3.2.2.

When the relative pulse offset, b=0.5, an additional locus and response lobe, similar to those for an intermediate size pulse, appear. In Fig. 3.25 (d), the middle peak of the optimal position response is fit to the response trace at the centre of a putative "pulse" formed by the left edge of the embedded pulse and the right edge of the outer pulse. Intuitively, this is a spurious pulse. It looks as though the embedded pulse is first embedded in the spurious pulse and then both are further embedded in the outer pulse. For the embedded pulse, the maximum of the optimal radius response remains at a position of $x=bW/2$ and the maximum of the optimal position response remains at a radius of $r=eW/2$. For the outer pulse, spatial localisation deviates further from the pulse centre.

When the relative offset of the pulse centre, b=0.75, the width of the embedded pulse approximates that of the spurious pulse and their response traces are close to each other, as shown in Fig. 3.25(f). The response of the embedded pulse is absorbed into that of the spurious pulse and disappears. Because the right edge of the spurious pulse has become the combination of the right edges of both the embedded and outer pulses, the spurious pulse has an enhanced response and is considered to have a greater equivalent height contrast. This happens in Section 3.3.2.2, when e=0.5.

Fig. 3.26 shows the estimated positions and radii, for a range of offset positions, b, for the embedded, outer, and spurious pulses. The loci for the outer pulse is continuous through b, while that for the embedded pulse seems to be replaced by the emerging loci of the spurious pulse at b>0.5. For the embedded pulse, both the position and radius are correctly identified, fit to the positions, $x=bW/2$, and the radii, $r=eW/2$. For the outer pulse, the detected position is always shifted to the side with the embedded pulse and the bias increases with pulse centre offset, b; The detected radius also increases with pulse offset b and is underestimated (overestimated) when b<0.6 (b>0.6), respectively. For the spurious pulse, the detected radius is similar to the radius, $r=(1-b+e)W/4$, of the putative pulse bounded by the left edge (at $x=(b-e)W/2$) of the embedded pulse and the right edge (at $x=W/2$) of the outer pulse (Fig. 3.26(b)). However, the detected position of the spurious pulse is much closer to the centre (at $x=bW/2$) of the embedded pulse, rather than the centre (at $x=(1+b-e)W/4$) of the putative pulse (Fig. 3.26(a)). If the locus for the spurious pulse is considered as the continuation of that for the embedded pulse with pulse offset b increasing, the location of the embedded pulse can be measured more accurately than the radius when the offset, b>0.5.

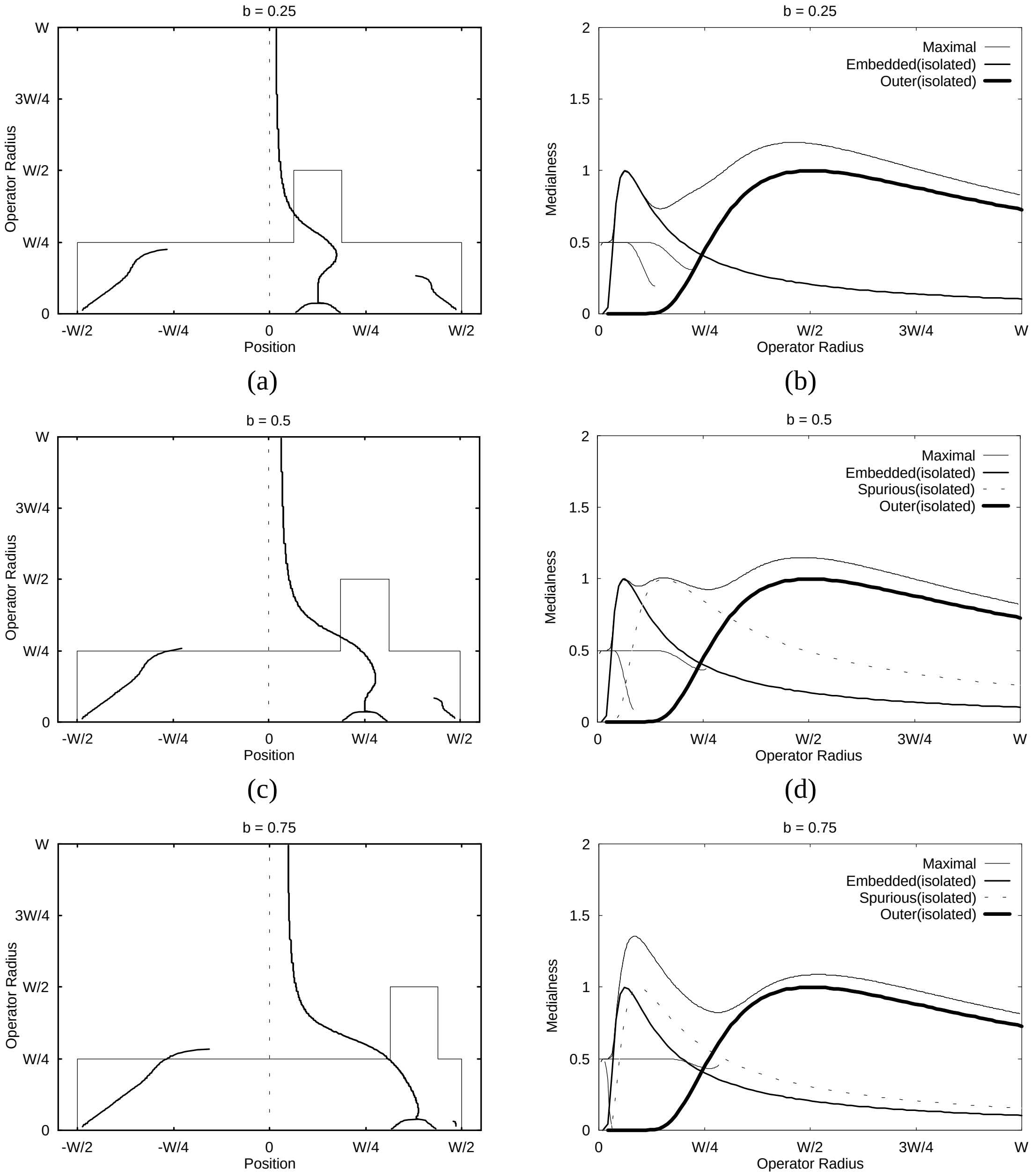


(a) (b)

(c) (d)

(e) (f)

Fig. 3.24. The optimal scale maxima for embedded pulses of varying position: the radii (left column) and responses (right column), as a function of position,.

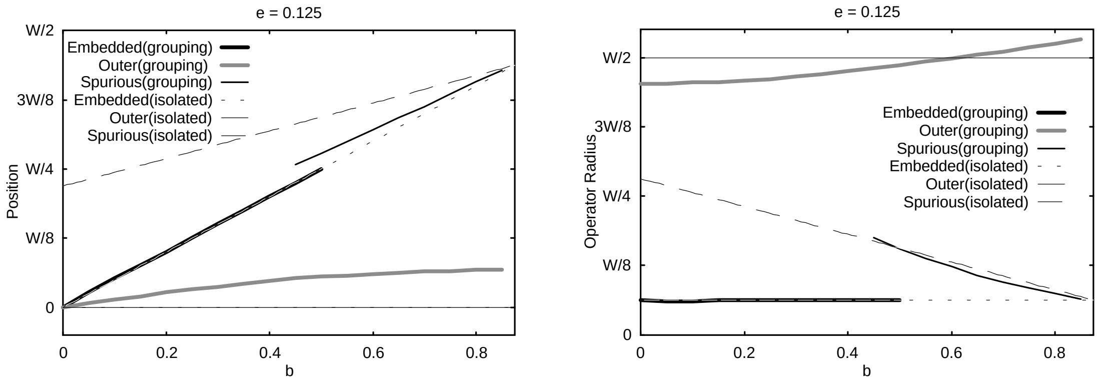


(a) (b)

(c) (d)

(e) (f)

Fig. 3.25. The optimal position maxima for embedded pulses with varying position: the positions (left column) and responses (right column), as a function of operator radius. The response traces at the centre of the embedded, putative, and outer pulses, when they exist in isolation, are also illustrated at right. The putative pulse is a unit pulse bounded by the left edge of the embedded pulse and the right edge of the outer pulse.



(a) (b)

Fig. 3.26. The estimated positions (a) and radii (b), as a function of pulse centre offset, b, for the embedded, outer, and spurious pulses.

3.3.2.5 The distinguishability of embedded objects

In Section 3.3.2.2, it was shown that, as the width of the embedded pulse increases, the peak of the response trace for the embedded pulse is lost at some distance from that for the outer pulse. In Section 3.3.2.3, it was also shown that as the height of the embedded pulse increases, the response peak for the outer pulse is lost at some distance from that for the embedded pulse. The distance between peaks of the response trace at the pulse centre, which we term the Distance of Peaks (DoP)[2], has been examined for a range of values of relative pulse width, *e*, and relative pulse height, *h*. Fig. 3.27 shows the isophote contours of the DoP through *e* and *h*. Let the radius of the embedded pulse be r_e, and then:

$$r_e = eW / 2$$

For those *e* and *h* values to the right of the contour for DoP=0.1r_e, there is only one peak in the response trace for the pulse group and thus DoP=0. In Fig. 3.27, the isophote contours for DoP=0.1r_e, 0.5r_e, and r_e coincide, except at the tips of the these contours, around *e*=0.3 and *h*=2.5. Therefore, for almost the full range of *e* and *h* values, DoP>0.5r_e or even DoP>r_e.

---

[2] DoP=0 if only one peak is found.

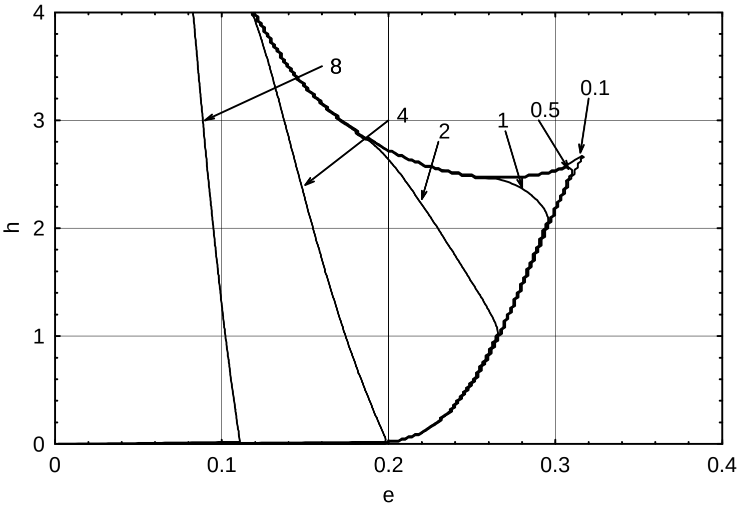


Fig. 3.27. The isophote contours of the DoP, as a function of e and h, corresponding to the DoP values of $0.1r_e$~$8r_e$.

This DoP minima are comparable to the radius of the embedded pulse, r_e. It seems that, as long as two peaks exist in a continuous response trace and the embedded pulse in isolation can be identified, that the DoP and both the pulses in a grouping are distinguishable.

However, under a discrete sampling of radius (scale dimension), either the embedded or outer pulse can be lost, even if both the peaks are distinguishable in a continuous response trace and the corresponding DoP is of high value. Fig. 3.28(a) and (b) show the undersampling cases for the response traces when $e \rightarrow 0.25$ and when $h \rightarrow 2.7$, respectively. Let r_A be the operator radius at the peak of the weaker response lobe and r_B be the operator radius at which the stronger response lobe has the same response as that at r_A. When the sampling interval in radius, Δr, is not sufficiently small in comparison with $| r_A - r_B |$, the peak of either the embedded pulse (Fig. 3.28(a)) or the outer pulse (Fig. 3.28(b)) can be lost in the discrete representation of the medialness scale-space. The distance, $| r_A - r_B |$, is critical to the distinguishability of embedded objects in a discrete representation and is here named the Distance of Distinguishability (DoD).

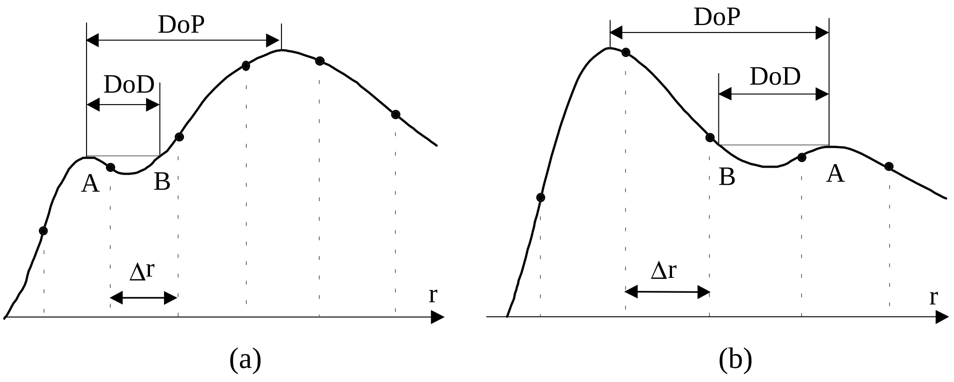


Fig. 3.28. The undersampling cases of the response traces (a) when e→0.25 and (b) when h→2.7.

We now turn to the selection of an appropriate sampling interval, Δr, with respect to the DoD. From the Shannon Sampling Theorem [OPE83], the response trace should be sampled at a rate:

$$\Delta r \leq DoD/2 \tag{3.62}$$

to avoid aliasing. Eq. (3.62) simply means that the sampling interval for radius should be chosen to be less than half the size of the smallest detail of interest in the response trace.

Unfortunately, the lossless sampling condition in Eq. (3.62) cannot be always satisfied because the DoD can approach zero with the variation of the relative width, *e*, and relative height, *h*, of the embedded pulse. Fig. 3.29 shows the isophote contours of the DoD for a range of values of *e* and *h*. Unlike the isophote contours of the DoP, those of the DoD do not coincide. Instead, the DoD value changes gradually and approaches zero.

There is a well established theory for spatial sampling that can be used to determine the minimum sampling rate required for a band-limited signal. To find the scale-space equivalent of the Nyquist Rate is an important, unanswered question in scale-space theory [MOR95]. In the case of the medialness scale-space for embedded objects, we find that 1/DoD in scale sampling plays the same role as the highest frequency, F_{max}, in Nyquist sampling, and that the sampling interval, Δr, should satisfy $\Delta r < DoD/2$ for a lossless discrete representation. In

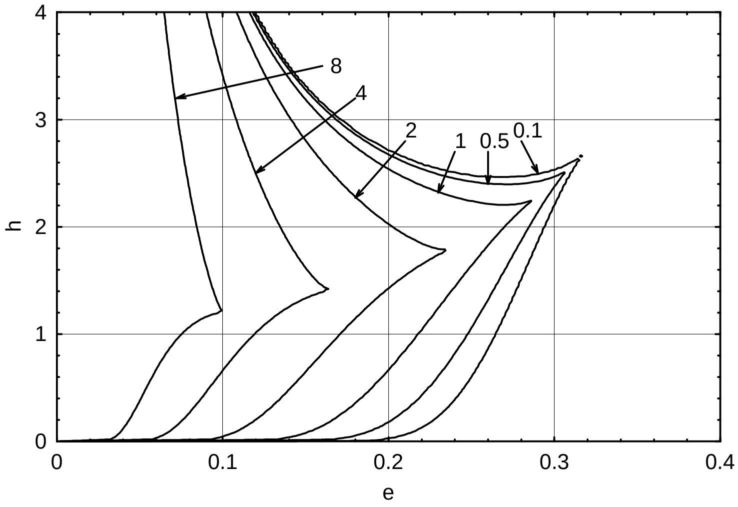


Fig. 3.29. The isophote contours of the DoD, as a function of e and h, corresponding to the DoD values of $0.1r_e$~$8r_e$.

spatial sampling, we can usually specify the sampling rate according to some prior knowledge of the signal domain, e.g. F_{max}=3KHz for a speech and F_{max}=5MHz for television video. However, similar prior knowledge is not available in scale sampling. It is known that the DoD is determined by the relative size, *e*, and contrast, *h*, of the objects in an embedded group (Fig. 3.29). The relative size and contrast are not known in a general analysis. Therefore from the viewpoint of embedded object analysis, it is difficult to find the minimum sampling rate for scale dimension. Instead, we have to discretely represent an embedded object group at some loss of information. The denser the scale sampling, the less the loss.

To identify an embedded object with radius r_e, at a scale resolution, DoD, the scale sampling rate, Δr, must satisfy Eq. (3.62). For a scale range ΔR, the number of scale slices that must be computed is: $\Delta R / \Delta r \geq 2\Delta R / DoD$. For an $N \times N$ pixel image, the size of memory required to record the floating point values of medialness scale-space is $8N^2 \Delta R / DoD$ bytes. Therefore both the computational cost and memory requirement is inversely proportional to the given scale resolution, DoD. If we try to capture the embedded pulse with the “smallest”

DoD (corresponding to the “highest” resolution), the sampling interval must be very small and the requirement for memory and processor speed is high[3]. For example, if we try to identify an embedded object with a radius, r_e=2 pixels (a width of 4 pixels), at a high scale resolution, DoD=0.1r_e, in a 256×256 pixel image at radii from 1 to 64 pixels, we need to compute 640 scale slices and save the data into a memory of at least 160 megabytes.

An alternative strategy for selecting the sampling interval is to make a compromise between the scale resolution, DoD, and the data volume that is inversely proportional to the DoD. If a scale resolution, DoD=r_e, is used, the memory required for the scale-space data now becomes 16 megabytes and only 64 scale slices need to be computed. What does the selection, DoD=r_e, mean in practice? Using a sampling interval, Δr=1 pixel, as in [FRI93], from Eq. (3.62), we have $r_e = DoD \geq 2\Delta r = 2$ pixels, and we can identify embedded objects with a width of at least 4 pixels and ignore tiny embedded objects.

---

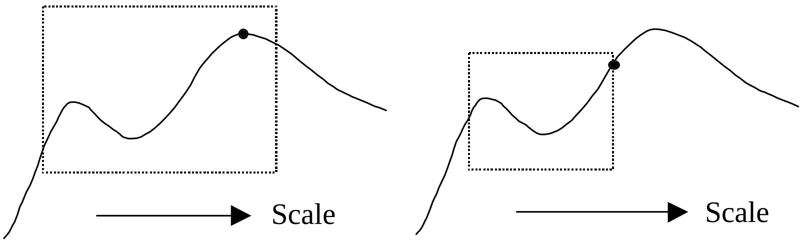
[3] The computational cost for the convolution of an image with a Gaussian-based operator increases linearly with scale by utilizing the separable nature of Gaussian-based operators.

3.3.3 Sliding Window Algorithm

3.3.3.1 The necessity for using a smaller range of scales

Though the definition of the globally optimal scale ridge [FRI93] can reflect a large portion of optimal scale ridges, there are strong reasons to consider the scale-space within smaller ranges of scales, rather than a full scale range:

- Ridge extraction is a local operation to check the geometry around a pixel. In the detection of scale-space ridges, this is reflected by the reliance on spatial and scale locality. In existing ridge detection algorithms, only the 3×3 neighborhood in 2-D spatial spaces [HAR83] [FRI93] and the 3×3×3 neighbourhoods in 3-D spatial spaces (or 2+1 dimensional scale-spaces) [MOR93] [MOR95] around each potential ridge point were examined.
- From the discussion in Section 3.3.2.5, we know that considering a smaller scale range each time can improve scale distinguishability. If the scale range used is greater than the Distance of Peaks (DoP), then the global maximum within this scale range is the greater magnitude peak (Fig. 3.30(a)), and the object corresponding to the weak peak is lost in the MMA representation. Even if the scale range used is less than the DoP but greater than the Distance of Distinguishability (DoD), the resulting global maximum within this scale range may occur at one end of the scale range, rather than the response peak due to an object (Fig. 3.30(b)). This object may still be excluded from the MMA representation. Therefore using a smaller scale range each time can more readily avoid interference between the medialness responses of different sized objects and a loss of MMA representations.



(a) (b)

Fig. 3.30. An object (the smaller peak) is lost in the MMA representation, when the scale range (dot line) is (a) greater than the DoP or (b) greater than the DoD. The black circles represent the global maxima within the scale range.

Suppose that Ω is the scale range being considered, and R is the set of globally optimal scale ridge points within Ω; Ω_i is a smaller scale range, $\Omega_i \in \Omega$; and R_i is the set of globally optimal scale ridge points within Ω_i. From the discussion above, it is intuitively true that, if $\Omega = \bigcup_i \Omega_i$, then $R \in \bigcup_i R_i$. A set of consecutive, smaller scale ranges would be an advantage in extracting a complete set of ridges. If the smallest range of scale is used (i.e. 3 samples), the response trace takes one of the following forms [4] (Fig. 3.31):

1) The response trace increases monotonically with scale. The global maximum occurs at the end scale (Fig. 3.31(a)).
2) The response trace decreases monotonically with scale. The global maximum occurs at the start scale (Fig. 3.31(b)).
3) The response trace has a local minimum. The global maximum occurs at either the start or the end scale (Fig. 3.31(c)).
4) The response trace has a local maximum (the global maximum too) at neither the start nor the end scale (Fig. 3.31(d)).

---

[4] The scale sampling rate is assumed to be sufficiently dense enough to represent the response trace by Shannon’s sampling criteria.

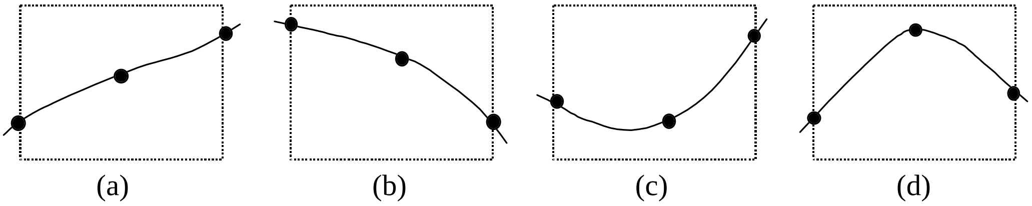


Fig. 3.31. The four types of response traces within a 3-sample sliding window: (a) increasing monotonically, (b) decreasing monotonically, (c) local minimum, and (d) local maximum.

It is obvious that the points in case (4) are candidates for the scale-space ridge and should be further examined to find whether they are ridge points of the globally optimal scale response (maximum projection onto the image plane) within that scale range.

3.3.3.2 Sliding window algorithm

Suppose that S_L and S_H are the smallest and largest scale indices, respectively, and $S_H \geq S_L + 2$; Each sliding window contains L scale-slices, and $3 \leq L \leq S_H - S_L + 1$. A sliding window corresponds to a small scale range. To improve the scale distinguishability, at each iteration (also indexed by scale), the sliding windows are moved by one scale index along scale dimension. Therefore, the sliding windows at two successive iterations partly overlap. In this way, we can identify the Distance of Distinguishability (DoD) as long as the discrete sampling of scale can reflect the DoD.

The index range of the sliding window is given by S_L+i to $S_L+i+L-1$ at iteration i. At each iteration the global maximum of the L scale slices, at each spatial position, forms a 2-D optimal scale response. The ridge points of this optimal scale response are identified as scale-space ridge points as long as the global maxima occur at neither the start nor the end scale of the sliding window. A detailed description of ridge detection in a 2-D space is given in Section 3.3.4.1.

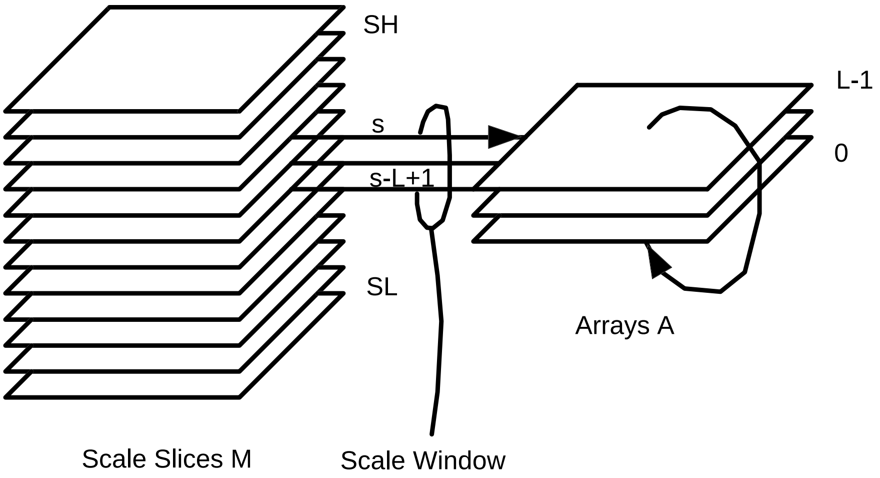


Fig. 3.32. The sliding windows extracted from scale slices.

In practice, it is not necessary to store medialness responses at each scale. Only *L* floating point arrays and one integer array are needed. The integer array, Sopt(x, y), records the optimal scale. The *L* floating point arrays, A(x, y, k), marked with a sequence number $k \in [0, L\text{-}1]$, record the most recent *L* scale slices, i.e. the *L* scale slices within the current sliding window. To avoid shifting the slice sequence within the *L* arrays, at a new iteration, the arrays, A(x, y, k), are treated as a circular buffer, see Fig. 3.32. The medialness response at scale *s* is always stored in array A(x, y, k) at $k=(s\text{-}S_L)$ *modulo L*. A detailed description of the Sliding Window algorithm is given in Fig. 3.33. The buffer, Mopt(x, y), in Fig. 3.33 can be eliminated by writing the maximum medialness response into array A(x, y, k) at $k=(s\text{-}S_L\text{-}L\text{+}1)$ *modulo L*, because the content of this array is the medialness response at the least recent scale and this response is not used in subsequent computations. The medialness response at the new scale is substituted at the next iteration.

S_L	the lower limit of scale.
S_H	the upper limit of scale.
$M(x, y; s)$	the medialness response at scale s.
$A(x, y; k)$	the floating point array, k, recording medialness at scale s; $k=(s-S_L)\%L$.
$M_{opt}(x, y)$	the maximum of $A(x, y; k)$ across k, within a sliding window.
$S_{opt}(x, y)$	the scale of the maximum of $A(x, y; k)$ across k, within a sliding window.

```
1  For $s=S_L$ to $S_L+L-2$
   $A(x, y; (s-S_L)\%L)=M(x, y; s)$
2  For $s=S_L+L-1$ to $S_H$
   2.1  $A(x, y; (s-S_L)\%L)=M(x, y; s)$
   2.2  For each scale, $ss \in [s-L+1, s]$, in a sliding window
        2.2.1  $M_{opt}(x, y)=\max_{ss} \{A(x, y; (ss-S_L)\%L)\}$
        2.2.2  $S_{opt}(x, y)=\arg\max_{ss} \{A(x, y; (ss-S_L)\%L)\}$
   2.3  If $(x, y)$ is a ridge point of $M_{opt}(x, y)$
        2.3.1  If $S_{opt}(x, y) \in (s-L+1, s)$
               2.3.1.1  $(x, y, S_{opt}(x, y))$ is labeled as an MMA point.
```

Fig. 3.33. The description of the sliding window algorithm. % denotes the modulo operation.

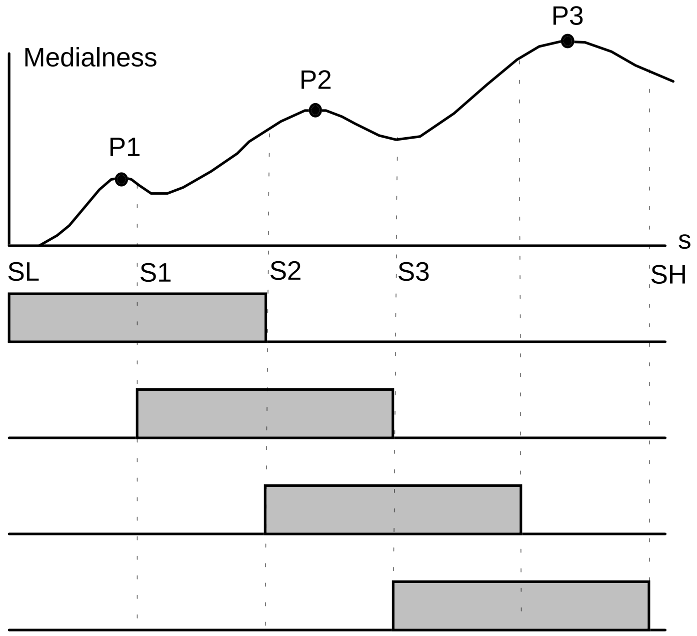


Fig. 3.34. The influence of a large sliding window on ridge extraction.

The selection of a large scale range, *L*, for each sliding window has a negative influence on the ridge extraction process. It might result in some ridge points being missed in the initial scale slices. However, this negative influence does not spread to the remaining scale slices due to the sliding nature of the scale window. This is illustrated in Fig. 3.34, in which there are three peaks in the response trace, P_1, P_2, and P_3, corresponding to three different-sized objects in an embedded group. When the sliding window is at the initial scale, $s=S_L$, the peak P_1 is not the global maximum within the sliding window and then is lost in the MMA representation. The peak P_2 is detected when the sliding window is at the scale, $s=S_1$, though the peak P_2 may be lost when the scale window is at the scale, $s=S_2$. Finally, the peak P_3 is detected when the sliding window is at the scale, $s=S_3$. In practice, a small scale range, $L=3$, is used for each sliding window.

The applications of the sliding window algorithm to 2-D images are shown in 4.3.1.

3.3.4 Ridge Segmentation in 2-D Space

3.3.4.1 Detection of ridge points

Here the height ridge definition [HAR83] [EBE94a] has been used, which considers a ridge point as a local maximum (of the underlying function) in the direction that maximizes the magnitude of the second derivative. This direction can be determined by calculating the eigenvectors and eigenvalues of the Hessian matrix of second derivatives [EBE94a], see Section 2.4.2.1 for details.

For 2-D space, solving the eigenvectors and eigenvalues of the Hessian matrix can be avoided and an analytical solution used [HAR83][5]. The polar to Cartesian coordinate transform defined by:

$$x = \rho \cos\alpha, \quad y = \rho \sin\alpha \tag{3.63}$$

is used to generate polar invariants for ridge detection.

The first derivative of function f in the direction α is denoted by:

$$\begin{aligned} f_\rho(x,y) &= f_x(x,y)\frac{\partial x}{\partial \rho} + f_y(x,y)\frac{\partial y}{\partial \rho} \\ &= f_x(x,y)\cos\alpha + f_y(x,y)\sin\alpha \end{aligned} \tag{3.64}$$

The second derivative of function f in the direction α is:

$$\begin{aligned} f_{\rho\rho}(x,y) &= f_{xx}(x,y)\cos^2\alpha + 2f_{xy}(x,y)\sin\alpha\cos\alpha + f_{yy}(x,y)\sin^2\alpha \\ &= \frac{1}{2}(f_{xx} + f_{yy}) + \frac{1}{2}(f_{xx} - f_{yy})\cos 2\alpha + f_{xy}\sin 2\alpha \end{aligned} \tag{3.65}$$

The direction α that maximizes $f_{\rho\rho}$ can be determined by differentiating $f_{\rho\rho}$ with respect to α, setting the derivative to zero, and solving for α:

---

[5] [HAR82] used an unusual definition of polar coordinates and did not give an explicit solution for α

$$f_{\rho\rho\alpha}(x,y) = -(f_{xx} - f_{yy})\sin 2\alpha + 2 f_{xy}\cos 2\alpha \tag{3.66}$$

Setting $f_{\rho\rho\alpha}(x,y) = 0$, we can find the two principal directions, one maximizing and another minimizing f_{pp}:

$$\tan 2\alpha = 2 f_{xy} /(f_{xx} - f_{yy}) \tag{3.67}$$

$$2\alpha = \begin{cases} \tan^{-1}(2 f_{xy} /(f_{xx} - f_{yy})) + 2l\pi, & \left|f_{xx} - f_{yy}\right| > 0 \\ \pi/2 + 2l\pi, & otherwise \end{cases} \quad (l \in I) \tag{3.68a}$$

or

$$2\alpha = \begin{cases} \tan^{-1}(2 f_{xy} /(f_{xx} - f_{yy})) + \pi + 2l\pi, & \left|f_{xx} - f_{yy}\right| > 0 \\ 3\pi/2 + 2l\pi, & otherwise \end{cases} \quad (l \in I) \tag{3.68b}$$

Thus

$$\alpha = \begin{cases} (1/2)\tan^{-1}(2 f_{xy} /(f_{xx} - f_{yy})) + l\pi, & \left|f_{xx} - f_{yy}\right| > 0 \\ \pi/4 + l\pi, & otherwise \end{cases} \quad (l \in I) \tag{3.69a}$$

or

$$\alpha = \begin{cases} (1/2)\tan^{-1}(2 f_{xy} /(f_{xx} - f_{yy})) + \pi/2 + l\pi, & \left|f_{xx} - f_{yy}\right| > 0 \\ 3\pi/4 + l\pi, & otherwise \end{cases} \quad (l \in I) \tag{3.69b}$$

It is noted that detecting local maxima along direction α and $\alpha+l\pi$ is the same. Therefore:

$$\alpha = \begin{cases} (1/2)\tan^{-1}(2 f_{xy} /(f_{xx} - f_{yy})), & \left|f_{xx} - f_{yy}\right| > 0 \\ \pi/4, & otherwise \end{cases} \quad (l \in I) \tag{3.70a}$$

or

$$\alpha = \begin{cases} (1/2)\tan^{-1}(2 f_{xy} /(f_{xx} - f_{yy})) + \pi/2, & \left|f_{xx} - f_{yy}\right| > 0 \\ 3\pi/4, & otherwise \end{cases} \quad (l \in I) \tag{3.70b}$$

The two solutions of α in Eqs. (3.68)-(3.70), which differ from each other by $\pi/2$, give the directions that maximize (α is the normal of a potential valley or the tangent of a potential ridge) and minimize (α is the normal of a potential ridge or the tangent of a potential valley) the directional second derivative $f_{\rho\rho}$, respectively. We need only to select the one direction that has smaller value (actually the smallest value) of $f_{\rho\rho}$. If for such a value of α we have:

$$f_{\rho\rho}(x,y) < 0 \quad \mathit{and} \quad f_{\rho}(x,y) = 0 \tag{3.71}$$

then a ridge point is detected at (*x*, *y*).

In discretely sampled space, the differentiation of f is performed numerically. Suppose $\sigma(x,y)$ and $f(x,y)$ are the optimal scale and the optimal scale response at (*x*, *y*), within a sliding window. Because the responses, $f(x,y)$, for different positions are obtained at different scales and one of the properties of a scale change is that the unit of measurement changes [ter91] [MOR93], all first derivatives computed using numerical differentiation between adjacent pixels are scaled by $\sigma(x,y)$, and all second derivatives are scaled by $\sigma^2(x,y)$. The first and second partial derivatives of f are:

$$\begin{aligned} f_x(x,y) &= \sigma(x,y)\frac{f(x+h,y)-f(x-h,y)}{2h} \\ f_y(x,y) &= \sigma(x,y)\frac{f(x,y+h)-f(x,y-h)}{2h} \\ f_{xx}(x,y) &= \sigma^2(x,y)\frac{f(x+h,y)-2f(x,y)+f(x-h,y)}{h^2} \\ f_{yy}(x,y) &= \sigma^2(x,y)\frac{f(x,y+h)-2f(x,y)+f(x,y-h)}{h^2} \\ f_{xy}(x,y) &= \sigma^2(x,y)\frac{f(x+h,y+h)-f(x+h,y-h)-f(x-h,y+h)+f(x-h,y-h)}{4h^2} \end{aligned} \tag{3.72}$$

where we set $h = 1$.

The ridge definition given in Eq. (3.71) is based on the gradient at direction α being zero. However, because the gradient changes rapidly at or near a ridge, it is rarely exactly zero at a specific pixel but instead changes rapidly from a positive to a negative value. In a discretely sampled function, such as the medialness response, it is better to test for a relative maximum in the direction, α, normal to the potential ridge. Fig. 3.35(a) illustrates a ridge and its normal. Three methods have been used to find a pair of appropriate points, in the direction α, to be compared with the central pixel (x, y).

(1) Nearest direction from the 8 neighbours

This method uses two of the 8-connected neighbours of the central pixel to approximate the candidates (Fig. 3.35(b)). Because α and $\alpha+\pi$ correspond to the effectively same direction for detecting directional maxima, we consider the direction in the range [0, π] as α. The pair of candidates are:

$$\begin{array}{ll}
(x+1, y) \;\; and \;\; (x-1, y), & if \;\; 0 \le \alpha < \pi/8 \;\; or \;\; 7\pi/8 \le \alpha < \pi \\
(x+1, y+1) \;\; and \;\; (x-1, y-1), & if \;\; \pi/8 \le \alpha < 3\pi/8 \\
(x, y+1) \;\; and \;\; (x, y-1), & if \;\; 3\pi/8 \le \alpha < 5\pi/8 \\
(x-1, y+1) \;\; and \;\; (x+1, y-1), & if \;\; 5\pi/8 \le \alpha < 7\pi/8
\end{array}$$

In 3-D space, this method corresponds to selecting two points, in opposite directions, from 26 neighbours of the central voxel.

(2) Bilinear interpolation amongst 4 neighbours

This method assumes that the pair of points are equi-distant from the central pixel and that the function is locally linear, see Fig. 3.35(c). Therefore at $(x+\cos\alpha,\, y+\sin\alpha)$ the interpolated value of the function is:

$$\begin{aligned}
&l = round(x+\cos\alpha), \qquad a = x+\cos\alpha - l \\
&k = round(y+\sin\alpha), \qquad b = y+\sin\alpha - k \\
&f(x+\cos\alpha, y+\sin\alpha) = (1-a)(1-b)\,f(l,k) + a(1-b)\,f(l+1,k) \\
&\qquad\qquad + (1-a)bf(l,k+1) + abf(l+1,k+1)
\end{aligned} \tag{3.73}$$

Another point can also be obtained by substituting $\alpha+\pi$ for α in the above equation.

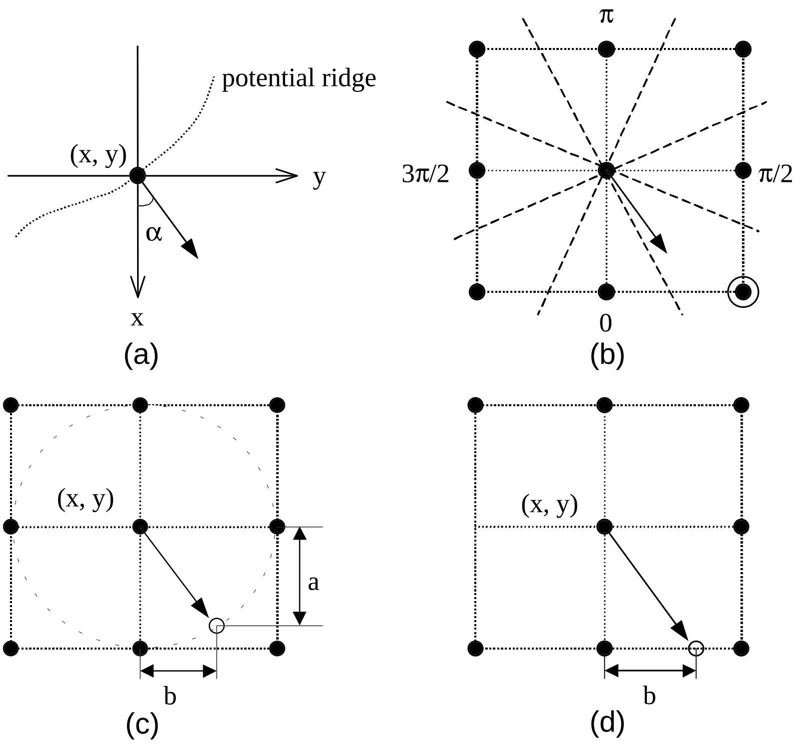


Fig. 3.35. Directional maximum detection: (a) a potential ridge and its normal, and the acquisition of a pair of points using (b) nearest direction in 8-neighbours, (c) bilinear interpolation in 4 neighbours and (d) linear interpolation in 2 neighbours. Black and white circles represent pixel grids and selected points, respectively.

When this method is extended to 3-D space, it corresponds to bilinear interpolation amongst the 8 vertices of a cube.

(3) Linear interpolation in 2 neighbours

This method assumes that the pair of points are in the direction α and the function is linear between neighbouring pixels. The location of the candidate point is found by projecting a ray thorough the central pixel (x, y) in the direction α until the ray intersects an edge of the square formed by the central pixel and its 8-connected neighbours (Fig. 3.35(d)). The value of this point can be linearly interpolated using the two end pixels of the line segment.

The second point can also be given by substituting $\alpha+\pi$ for α in above equations.

When this method is extended to 3-D space, it corresponds to projecting a ray from the central voxel until the ray intersects the face of the cube formed by the central voxel and its immediate neighbours. The value of the function is bilinearly interpolated using the 4 vertices of the square.

Generally, the nearest direction method is simplest to implement, whilst the interpolation methods in (2) and (3) more accurately reflect the ridge definition in Eq. (3.71).

The result of applying the above ridge detection to the optimal scale responses in the sliding window algorithm is a set of ridge points in 3-D scale-space. The position of each point is recorded with the scale (σ), local orientation (α), and medialness response $f(x, y, \sigma)$, etc.[6]. The position (x, y) indicates the location of the MMA, σ is proportional to the local width of the object, α reflects the local orientation of the object, $f(x, y, \sigma)$ is proportional to the intensity contrast of the object and, if the CMAT algorithm was used, also indicates the significance of a ridge point. It is potential for this rich set of attributes to be used by higher level processes for ridge linking and the like.

3.3.4.2 Linking ridge points

A ridge linking process is required to collect connected ridge points into a line, like the pearls of a necklace. Ridge points caused by noise and small disturbances, which correspond to short, and separate ridge lines (points), can also be removed at the same time.

We assign each pixel (x, y) a count, $C(x, y)$, to record how many ridge points exist at this point through scale. Each element of the second matrix, corresponding to a pixel in the image

---

[6] For medialness scale-space produced by linear operators, another attribute can be added to reflect the significance of a ridge point. Some potential options are the second derivative in the direction α (f_{pp}) and the global ridge strength measure [MOR95].

plane, contains $C(x, y)$ structures each recording the attributes (such as the position, scale, orientation and response) of a ridge point and a status flag. A status flag set to zero marks that the associated ridge point has been "processed" and that no further checking is needed.

The linking process first searches for the start of each ridge branch using a mask, shown in Fig. 3.36. Here the convention described in [GON92] is used. P_0 is the central pixel in the 3×3 mask and the potential start point of a ridge. $P_0=1$ (grey) if there is a ridge point at P_0, i.e. $C(P_0)>0$. P_i is the 8-neighbours of P_0, i=1,…,8. $P_i=1$ (grey), if $C(P_i)>0$ and $\sigma(\mathrm{P}_0)-2 \le \sigma(\mathrm{P}_\mathrm{i}) \le \sigma(\mathrm{P}_0)+2$. P_0 is determined as the start of a ridge line if the following conditions are satisfied: (1) $\mathrm{P}_0=1$; (2) $1 \le N(\mathrm{P}_0) \le 2$; (3) $S(\mathrm{P}_0)=1$, where $N(\mathrm{P}_0)=\sum_{i=1}^{8}\mathrm{P}_\mathrm{i}$ and $S(\mathrm{P}_0)$ is the number of 0-1 transitions in the ordered sequence of P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_1. Therefore a start point has one non-zero neighbour or two neighbours that are 4-connected to each other. Once a start point of a ridge branch is found, a new list for this ridge branch is created and the start point is stored as the first element of this list. The flag of P_0 is then set to zero to avoid it being re-checked. The only neighbour ($N(P_0)=1$) or one of the two adjacent neighbours ($N(P_0)=2$), which has stronger medialness response, is selected as the successor of the start point P_0 and forwarded to ridge following sub-process. On the other hand, if $P_0=1$ and $N(P_0)=0$, then P_0 is an isolated ridge point; it should be excluded from the linking process and its flag is set to zero.

In the ridge following sub-process, the ridge point selected as a successor is added to the list of the ridge branch being tracked. The ridge direction at this ridge point is either $\beta=\alpha+\pi/2$ or $\beta=\alpha-\pi/2$. this ambiguity can be removed by selecting a value for β that satisfies:

$$\Delta x \cdot \cos\beta + \Delta y \cdot \sin\beta > 0$$

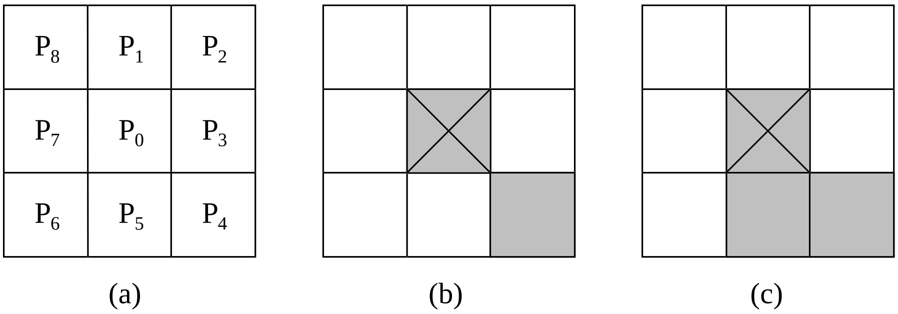


Fig. 3.36. Ridge starting point detection: (a) neighbourhood labelling, (b) a start point (marked by ×) when $N(P_0)=1$ and (c) a start point when $N(P_0)=2$.

where Δx and Δy are the relative coordinate increments of the current ridge point from its predecessor ($\Delta x, \Delta y \in \{-1, 0, 1\}$). The potential successor of the current ridge point is one of its 8-connected ridge points that satisfy:

(1) $\mathrm{flag}(\mathrm{P}_{\mathrm{potential}}) = 1$,

(2) $\beta_{\mathrm{current}} - \pi / 4 \le \beta_{\mathrm{potential}} \le \beta_{\mathrm{current}} + \pi / 4$,

(3) $\sigma(\mathrm{P}_{\mathrm{current}}) - 2 \le \sigma(\mathrm{P}_{\mathrm{potential}}) \le \sigma(\mathrm{P}_{\mathrm{current}}) + 2$.

It is possible that more than one 8-neighbour of the current ridge point satisfy the above conditions. However, only the one with the greatest medialness response is considered as the successor of the current ridge point, because a ridge point is the maximum in its normal direction. This successor is forwarded to the next iteration of ridge following as the new current ridge point. The flag of the current ridge point is set to zero. Those ridge points in the 8-neighbourhood of the current ridge point, which are not selected as the successor, arise mainly from the multiple responses when the ridge is 2-3 pixels wide, see Fig. 3.37(a). Therefore, their flags are set to zero if they have similar orientations to the selected successor. If no potential successor exists, the current ridge point is determined as the terminal point of the current ridge branch and the ridge following sub-process stops. The list of the ridge points is stored in a chain list if it contains at least 5 connected ridge points; otherwise it is deleted. In this way, short ridge lines due to noise and other disturbances are suppressed. The ridge linking process will resume scanning for the start point of a new ridge branch at the start point of the previous ridge.

In order to overcome small breaks (1 pixel width) in a ridge, the ridge following subprocess described above can be extended to consider a 5×5 neighbourhood of the current ridge pixel, if no successor can be found in the 3×3 neighbourhood. The successor is selected using the same criteria as that for a 3×3 neighbourhood. Once a successor at a gap of 1 pixel is found, a new ridge point is created to fill in the gap between the current ridge point and its successor, see Fig. 3.37(b). The attributes of this ridge point are linearly interpolated from those of the current ridge point and its successor, and are then added to the list of the current ridge branch. The successor at a gap is forwarded to the next iteration of ridge following as a new current ridge point. If no potential successor exists in the 5×5 neighbourhood, then the current ridge point is made a terminal point of the current ridge branch and the ridge following for this ridge stops.

The method of selecting a ridge successor, described above, treats junctions in a similar way to normal ridge points. At a junction, it will select the branch with the greatest response

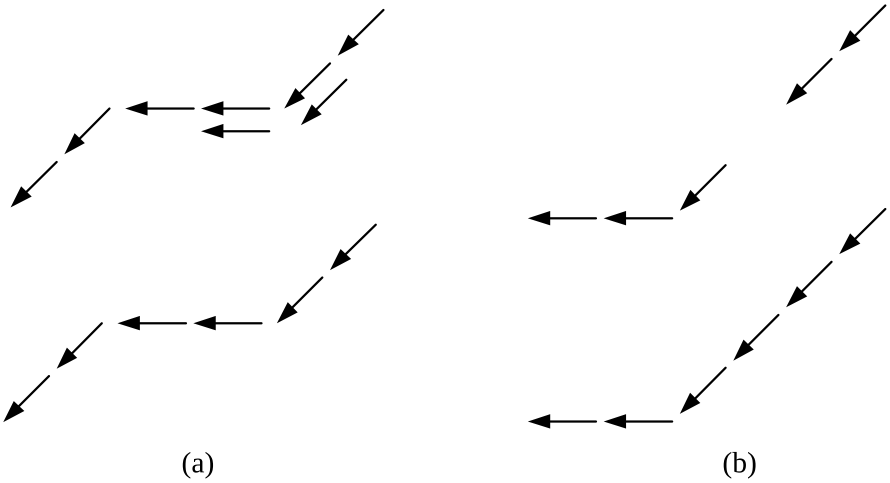

(a) (b)

Fig. 3.37. (a) Elimination of multiple ridges in parallel and (b) filling ridge breaks.

to follow without detecting the junction. The remaining branches will be followed in subsequent searches, after the main branch has been followed and labeled as “processed”. Junctions can be detected as a post-processing task of ridge linking, in which a body point in one ridge line can be declared as a junction point if it is a 8-neighbour of the start or terminal point of another ridge line. If both the start and terminal points of one ridge line are 8-neighbours of two body points of another ridge line, then the ridge has split into two branches and re-joined. If the start and terminal points of a ridge line are 8-neighbours, then the ridge is a closed loop.

At the post-processing stage, it is possible to connect adjacent ridge lines and to bridge large breaks in a long ridge line. If the terminal point of a ridge line is roughly in the direction of the start point of another ridge line but spatially separated, the two ridge lines may be combined and a set of new ridge points may be created to fill the gap between these two ridges.

The results of ridge linking for a variety of 2-D images are shown in Sections 4.3.1 and 4.3.2.

Chapter 4 Results

4.1 Performance Evaluation of the CMAT for 1-D Signals

4.1.1 Symmetric Pulses

To obtain quantitative estimates of the performance of the CMAT algorithm we have applied the LoG, HMAT and CMAT medialness algorithms to a variety of 1-D object profiles, such as the symmetric pulse shown in Fig. 3.13. The medialness responses for 2-D objects are not quantitatively analyzed here, because they can be influenced by more aspects of shape properties, such as the narrowing, flaring, bending, and limited length of shapes.

To give an objective and effective comparison among algorithms, both the response and scale have been normalised. The medialness function of each algorithm is normalised by the relevant global maximum, through scale and space, for a symmetric pulse with unit height. Each medialness response is sensitive to both polarities of boundary transition. For the HMAT and CMAT algorithms, the simple summation of signed medialness functions for both polarities of boundary transition is used, contrary to the summation of absolute values in [XU99].

The LoG medialness operator can be expressed as:

$$K(\mathbf{x},\sigma) = -\sigma^2 \nabla^2 G(\mathbf{x},\sigma) = \frac{1}{(2\pi\sigma^2)^{n/2}} \left(n - \frac{\mathbf{x}^T \mathbf{x}}{\sigma^2} \right) \exp\left(-\frac{\mathbf{x}^T \mathbf{x}}{2\sigma^2} \right) \tag{4.1}$$

where n is the number of dimensions and $\mathbf{x}$ is an n-dimensional vector for representing the spatial coordinates. Setting $K(\mathbf{x},\sigma) = 0$, we obtain $|\mathbf{x}| = \sqrt{n}\sigma$. The positive lobe of the LoG operator is within a distance of $\sqrt{n}\sigma$ from the centre of this operator. At a given scale, σ, the LoG operator gives the strongest response for objects with a radius of $\sqrt{n}\sigma$, because the positive central lobe matches the size of these objects; The HMAT and CMAT give the strongest response for objects with a radius of $k\sigma$. To compare these responses over scale, the radius of the operator is used as a base parameter. The radius of the LoG operator is defined as $r = \sqrt{n}\sigma$. The radius of the HMAT and CMAT operators is defined as $r = k\sigma$. These medialness operators may respond optimally to an object at different scales (the standard deviation of the Gaussian or its derivatives), σ, but should respond optimally at the same operator radius, r. In the following analysis the radius-to-scale ratio that determines the operator radius in the HMAT and CMAT, k, is set at 2. The Gaussian weighting for the CMAT in Eq. (3.22) and the equivalent distance weighting for the HMAT are not used, because the sampling interval of operator radius is set to be a multiple of the positional sampling interval and therefore, in the 1-D case each boundariness contribution is located at discrete positions in medialness space.

4.1.1.1 Medialness through scale

Figs. 4.1(a)-(c) show the LoG, HMAT and CMAT medialness responses through scale-space for the 1-D object profile of Fig. 3.13. Each medialness response forms a global maximum at the position of the pulse centre and at the operator radius of the pulse half-width, which reflects the symmetry of the pulse. However, for this single symmetrical structure, only the CMAT outputs a single response. The medialness surfaces of the LoG and HMAT are mixtures of boundariness and medialness properties and thus give a less clear description.

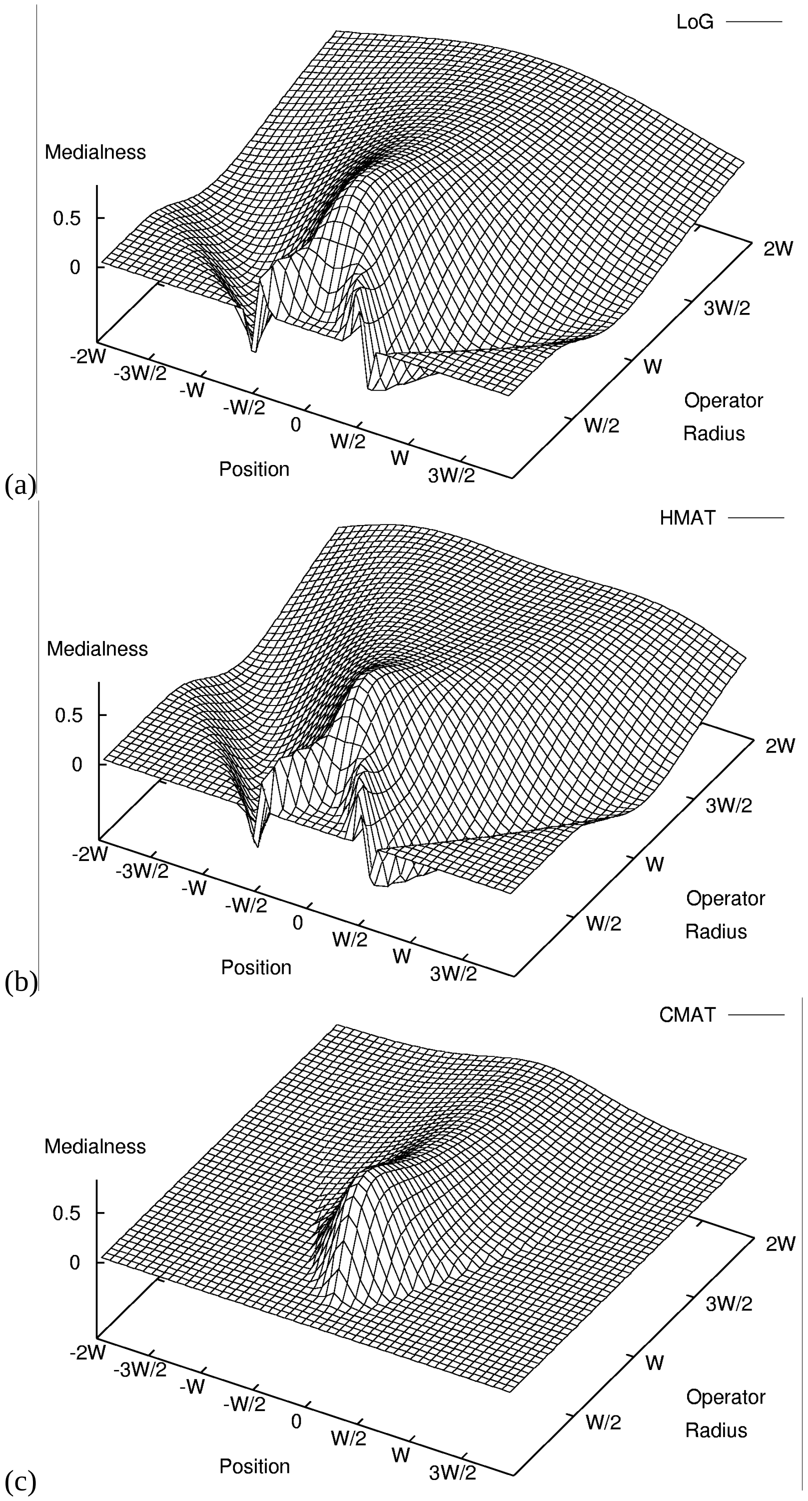
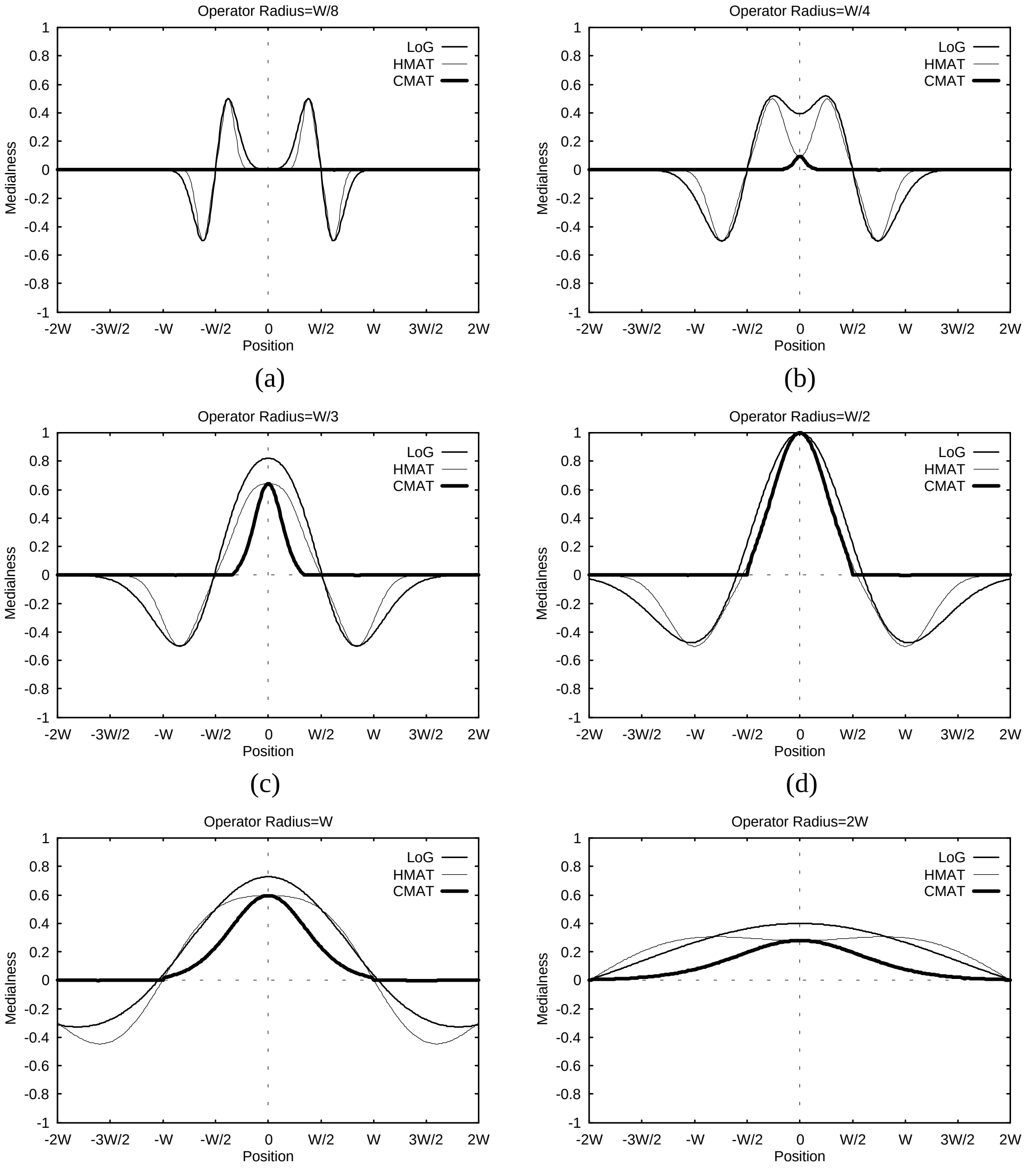


Fig. 4.1. Medialness responses through scale, for a symmetric pulse and generated from (a) LoG, (b) HMAT, and (c) CMAT.

4.1.1.2 Medialness at selected scales

The LoG, HMAT and CMAT medialness responses for a symmetric pulse at several radii are shown in Fig. 4.2. Fig. 4.2(a) shows that at a small radius of $r=W/8$, far from the pulse half-width, the medialness responses of the LoG and HMAT are already strong. These responses are caused by the edges. There is no interaction between the pair of edge responses for this operator radius, therefore, the CMAT produces no response. As the operator radius increases, $r=W/4$ in Fig. 4.2(b), there exists a minor interaction between the edge responses of the LoG and HMAT, and the CMAT begins to produce a small medialness response. At an operator radius of $r=W/3$, the interaction between edge responses of the LoG and HMAT becomes significant and the summation of edge responses combines into a single positive lobe, as shown in Fig. 4.2(c). The CMAT medialness response continues to increase as the operator radius increases the degree of interaction between edge responses. When the operator radius matches the pulse half-width, $r=W/2$ in Fig. 4.2(d), each algorithm produces its globally maximal response at the object centre and the curve of the CMAT response coincides with the main lobe of the HMAT response. The LoG and HMAT responses have two additional side-lobes. When the operator radius is larger than the pulse half-width, $r=W$ in Fig. 4.2(e), the medialness response of each algorithm begins to collapse. At $r=2W$ in Fig. 4.2(f), the HMAT response splits into two individual peaks again, violating an important premise of scale-space responses, that local extrema should not be generated as scale increases [LIN94]. The CMAT response is always equal to the HMAT response at the object centre. This is a property of the symmetric pulse used in this evaluation. Responses for more generic pulses are provided in Section 4.1.2.



(a) (b)

(c) (d)

(e) (f)

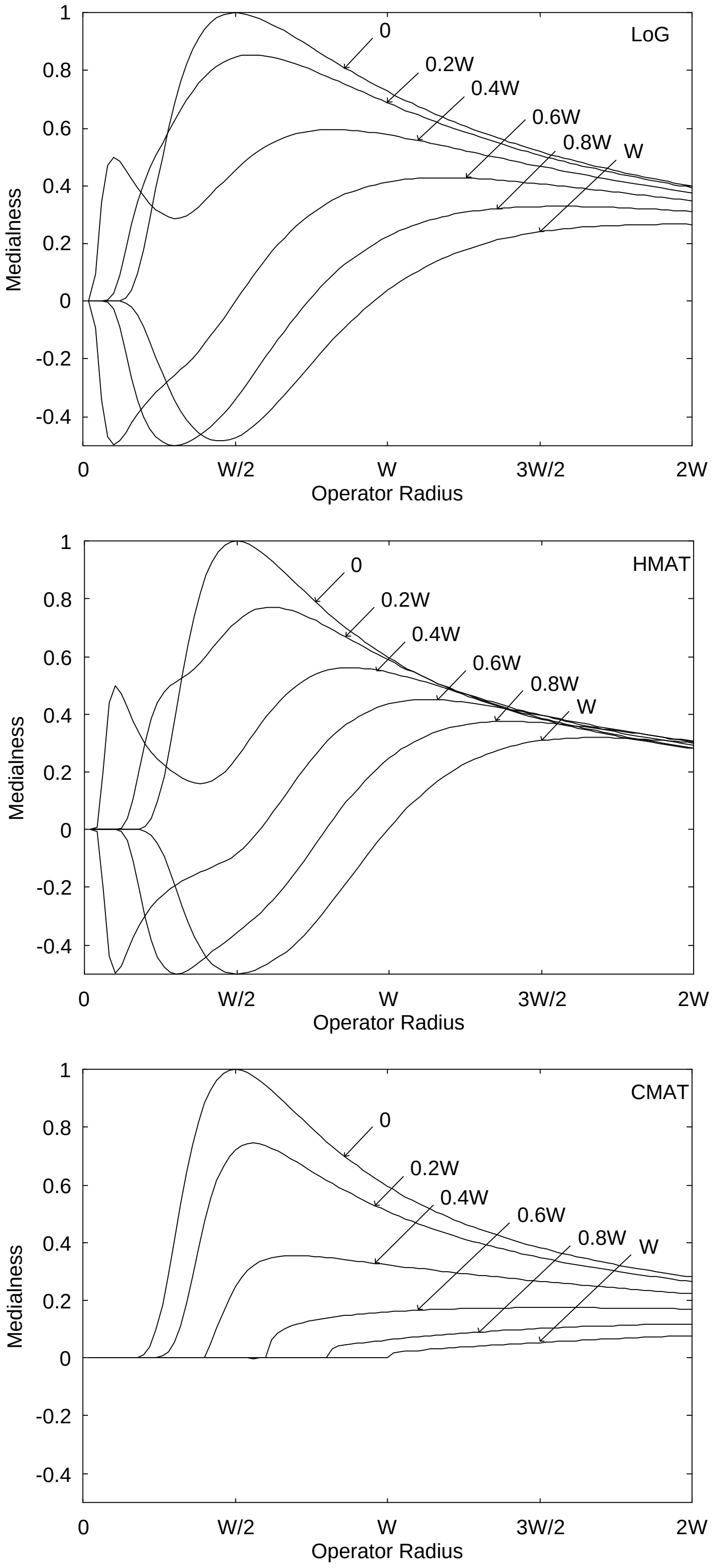
Fig. 4.2. Medialness responses at several scales, for a symmetric pulse and generated from the LoG, HMAT, and CMAT.

4.1.1.3 Medialness at selected positions

The optimal scale ridge definition [FRI93] is a method for determining the pulse centre and width from the medialness scale-space, which considers the medialness maxima with respect to scale at each position along an object. Fig. 4.3 shows the LoG, HMAT, and CMAT medialness responses over scale at several positions. Each plot represents a different displacement from the centre of the pulse.

Each algorithm produces its global maximal response at the object centre, $x=0$, and at an operator radius of $r=W/2$. For positions away from the pulse centre, each response collapses and exhibits a maximum at a larger radius than the pulse half-width. Such a radius for the CMAT is much closer to the pulse half-width than those for the LoG and HMAT. For positions near to the edge of the pulse, $x=0.4W$, the LoG and HMAT responses exhibit two peaks. The peak at smaller radius, which is negative at positions beyond the object ($|x|>W/2$), is not related to symmetry but edgeness. The CMAT response has only a single peak, which, like the second peak (at a larger radius) of the LoG and HMAT, correctly reflects the symmetry.

At a range of radii, $r>W/2$, the curves of the LoG and CMAT are always enveloped in those for positions closer to the pulse centre. This indicates that, for the LoG and CMAT, at $r>W/2$, the pulse centre continues to be the location of the only spatial maximum of the medialness response at a single radius. This is not the case for the HMAT. At $r>W/2$, the response may dominate those for positions closer to the pulse centre. This reflects from another aspect that the single spatial maximum at $r=W/2$ splits into two spatial maxima at larger operator radii (see Fig. 4.2(f)).



(a)

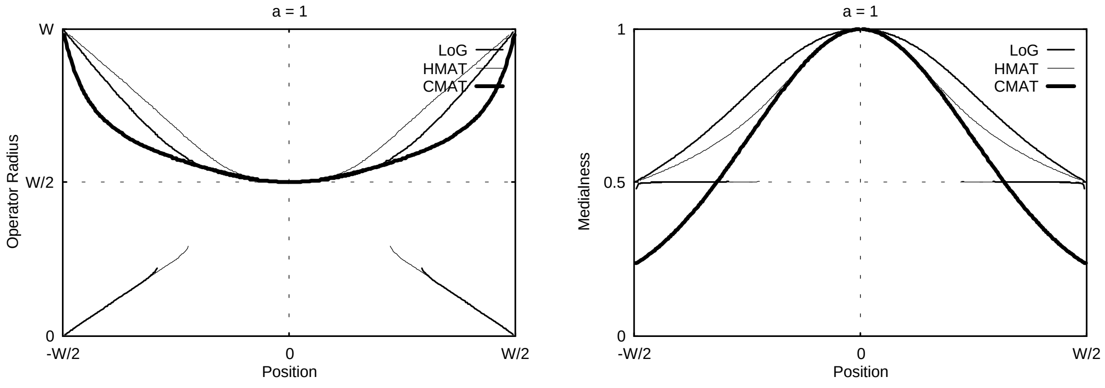
(b)

(c)

Fig. 4.3. Medialness responses through scale for several positions (labelled with each curve), for a symmetric pulse and generated from (a) LoG, (b) HMAT), and (c) CMAT.

4.1.1.4 Optimal scale maxima

Fig. 4.4(a) shows the optimal radii, as a function of position, for the symmetric pulse of Fig. 3.13 and generated from the LoG, HMAT and CMAT. The loci of the optimal radii for the LoG and HMAT contain three separate curves, two straight lines, which, at small radii, arise mostly from the response of a single edge. The loci of the CMAT is a single curve equivalent to those of the LoG and HMAT at larger radii, responding to the symmetry. At positions away from the pulse centre, each locus of symmetry occurs at a larger radius than the pulse half-width. The accuracy with which the half-width is estimated is, best first: CMAT, LoG, and HMAT. The estimation error of the CMAT is about one half that of the HMAT for most positions within the pulse.



(a) (b)

Fig. 4.4. The optimal scale maxima for the symmetric pulse: (a) the radii and (b) responses, as a function of position, generated from the LoG, HMAT and CMAT.

Despite the differences between the algorithms, the loci of optimal radii for each operator converge at three points: x=0 and r=W/2, and x=±W/2 and r=W (see Fig. 4.4(a)). The locus of the optimal radii at x=0 arises from the evenly-weighted contributions of both edges, for a radius of r=W/2. At x=W/2, the LoG medialness is the linear summation of their response from both edges, see Fig. 4.5(a):

$$M(W/2,r) = M_L(W/2,r) + M_R(W/2,r)$$

The LoG response to the right edge is always a zero-crossing at $x=W/2$, i.e. $M_R(W/2,r) = 0$, while $M_L(W/2,r)$ has a maximum at radius $r=W$, see Fig 4.5(a). Therefore the LoG has a maximum over radius at $r=W$. The analysis for the HMAT is the same as that for the LoG.

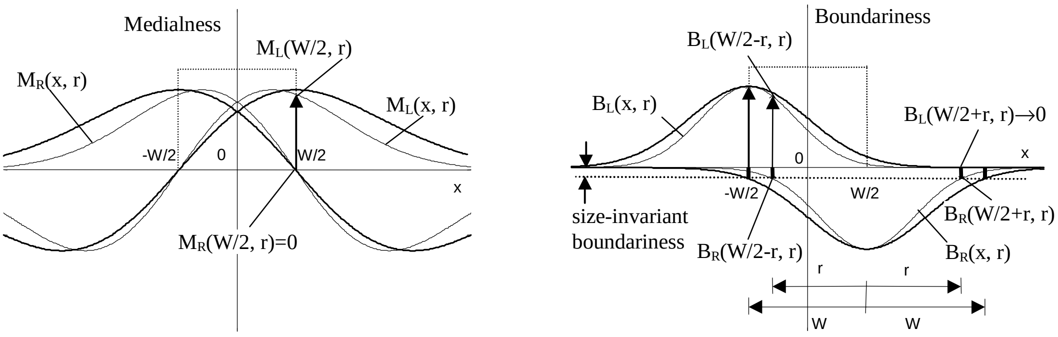
The CMAT medialness at $x=W/2$ is the **concordant** summation of the boundariness response for the pulse, $B(x,r)$, at $x=W/2\pm r$, i.e:

$$M(W/2,r) = \frac{2B(W/2-r,r)\cdot B(W/2+r,r)}{B(W/2-r,r)+B(W/2+r,r)}$$

The boundariness response for the pulse, $B(x,r)$, is the **linear** summation of the boundariness responses for the left and right step edges, $B_L(x, r)$ and $B_R(x, r)$ (see Fig. 4.5(b)), i.e.:

$$B(W/2-r,r) = \left|B_L(W/2-r,r) + B_R(W/2-r,r)\right|$$
$$B(W/2+r,r) = \left|B_L(W/2+r,r) + B_R(W/2+r,r)\right|$$

Because of the zoom-invariance to a step edge (see Section 3.1.2), the boundariness responses due to the right edge, at a distance r from the edge at $x=W/2$, $B_R(W/2-r, r)$ and $B_R(W/2+r, r)$,



.

(a) (b)

Fig. 4.5: The generation of an optimal scale maximum at the pulse border: (a) The LoG medialness and (b) the CMAT boundariness, due to the left (subscript L) and ridge (subscript R) edges of the pulse, at operator radius W (thick lines) and r (thin lines).

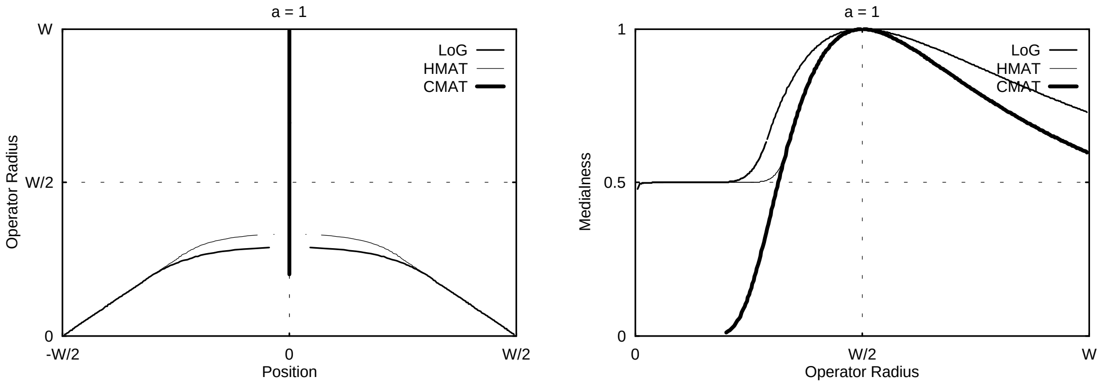
are invariant to r (short thick lines in Fig. 4.5(b)). At a radius close to $r=W$, $B_L(W/2+r, r)$ falls to zero because at a displacement of $3r/2$ (i.e. 3σ due to $r=2\sigma$), the boundariness response, due to the left edge, is negligible. On the other hand, $B_L(W/2-r, r)$ has its maximum at radius $r=W$. Therefore at a radius $r=W$, $B(W/2-r, r)$ has its local maximum and $B(W/2+r, r)$ has a small, constant value. At $x=W/2$ and $r=W$, the summation of the boundariness contributions, $B(W/2-r, r)$ and $B(W/2+r, r)$, are maximal but the degree of concordance is minimal. According to the conclusion in Section 3.2.2.2, the CMAT medialness, $M(W/2, r)$ is a maximum at $r=W$, because it is boundariness strength, rather than concordance degree, that dominates the CMAT medialness.

The optimal radius responses, for the LoG, HMAT and CMAT, are shown in Fig. 4.4(b). For each algorithm the object centre appears as a spatial maximum of the optimal radius response. The associated optimal radius, $r=W/2$ from Fig. 4.4(a), at the object centre indicates the pulse half-width. Therefore pulse centre and width can be efficiently identified from the 2-D medialness function for a 1-D symmetric pulse. The sequence of the selectivity to spatial position (best first) is: CMAT, HMAT and LoG. The sharp optimal radius response of the CMAT is an advantage for spatial maximum (ridge) detection.

4.1.1.5 Optimal position maxima

The optimal positions for the LoG, HMAT, and CMAT, as a function of operator radius, for the symmetric pulse are shown in Fig. 4.6(a). Unlike the loci of optimal radii, the locus of optimal positions is not continuous. Each algorithm exhibits a locus of optimal positions at the pulse center and at a range of radii, reflecting the symmetry of the pulse. For the LoG and HMAT there exist two additional branches at radii $r<W/2$, corresponding to edge responses.

The optimal position responses, as a function of operator radius, for the symmetric pulse



(a) (b)

Fig. 4.6. The optimal position maxima for the symmetric pulse: (a) the positions and (b) responses, as a function of operator radius, generated from the LoG, HMAT and CMAT.

are shown in Fig. 4.6(b). Each algorithm exhibits a maximum at the operator radius of pulse half-width. The associated optimal position, x=0 from Fig. 4.6(a), indicates the pulse centre. Therefore, the pulse center and width can be identified from the 2-D medialness function for a 1-D symmetric pulse using the optimal position definition.

Despite the separation of the optimal position loci for the LoG and HMAT (Fig. 4.6(a)), there is a continuous optimal position response for each algorithm. The CMAT response matches the HMAT's when boundariness responses have a major interaction. This results from a specific case: a symmetric pulse. Provided that more than one boundary response contributes to the medialness response, then Eq. (3.37):

$$M_{CMAT}(\mathbf{x}_{\mathrm{A}},\cdot) = \left(1 - \frac{1}{N}\right) Sc$$

may be applied. At the object centre, c=1 for a symmetric pulse. Therefore, the CMAT response, like the HMAT's, depends only on the sum of boundariness responses, S.

4.1.2 Asymmetric Pulses

In the previous section the medialness response for a symmetric pulse was considered. To observe the more general behaviour of these algorithms, we can allow the symmetry of a pulse to be less ideal. Because the medialness from any algorithm is directly related to boundariness responses, the ratio of the height contrast for the right edge to that for the left edge, *a*, is considered as an index for degree of symmetry. An asymmetric pulse is defined as (Fig. 4.7):

$$P_a(x) = \begin{cases} 0 & x < -W/2 \\ 1 & -W/2 \le x \le W/2 \\ 1-a & x > W/2 \qquad (0 \le a \le 1) \end{cases} \tag{4.2}$$

which is the difference of two step functions, $P_a(x) = U(x+W/2) - aU(x-W/2)$. Setting a=1 corresponds to a symmetric pulse and a=0 corresponds to a step function.

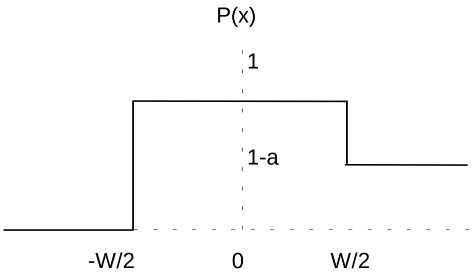


Fig. 4.7. A 1-D asymmetrical object.

4.1.2.1 Medialness through scale

Figs. 4.8(a)-(c) show the LoG, HMAT, and CMAT medialness responses through scale-space for an asymmetrical pulse, $P_a(x)$, with a=0.5. Each medialness response still forms a global maximum around the position of the object centre and at the operator radius of the object half-width, but the peak response becomes smaller than those, shown in Figs. 4.1(a)-(c), for a symmetric pulse, *P*(*x*). The LoG and HMAT responses are obviously shifted towards the more prominent edge.

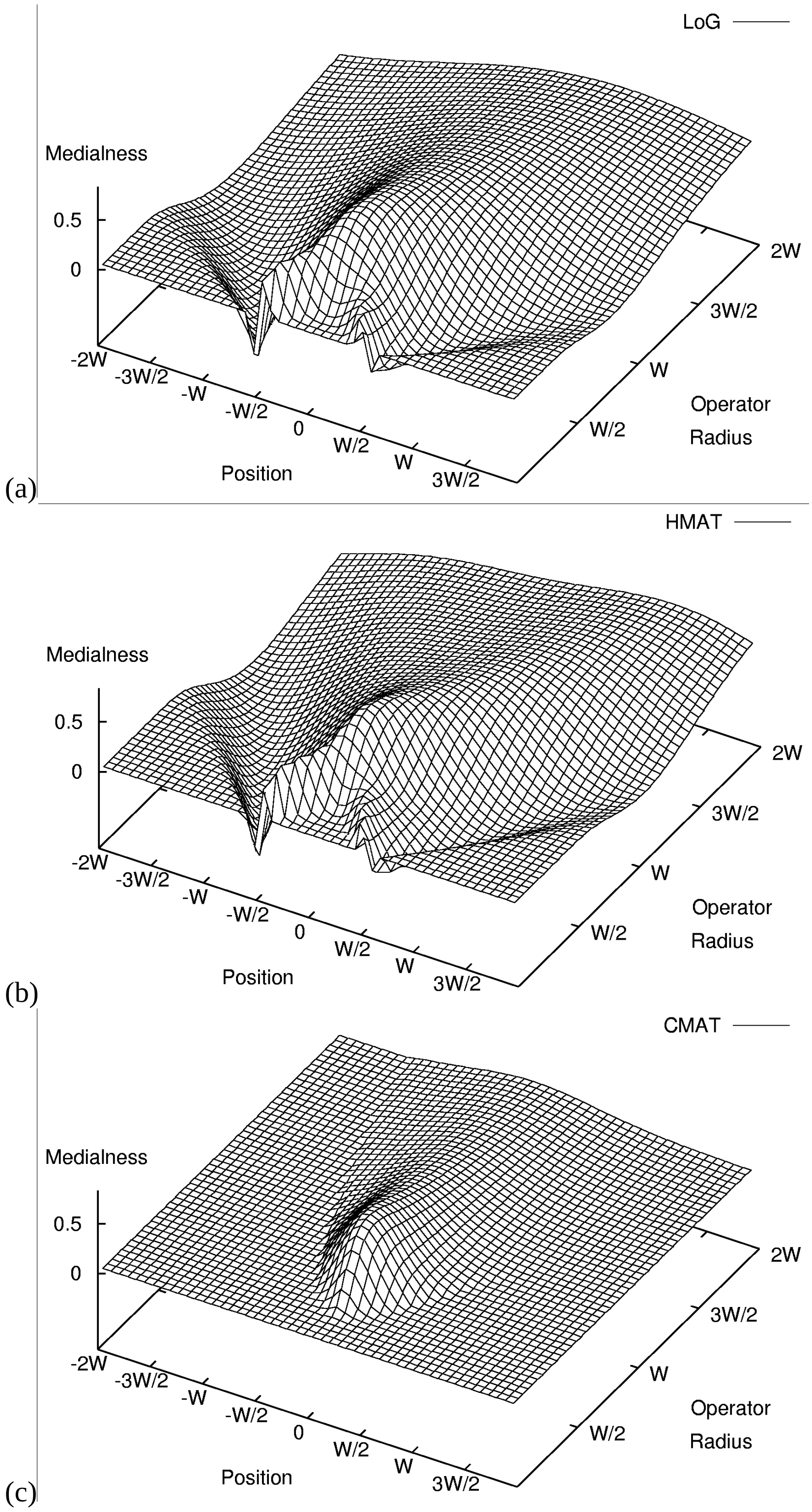
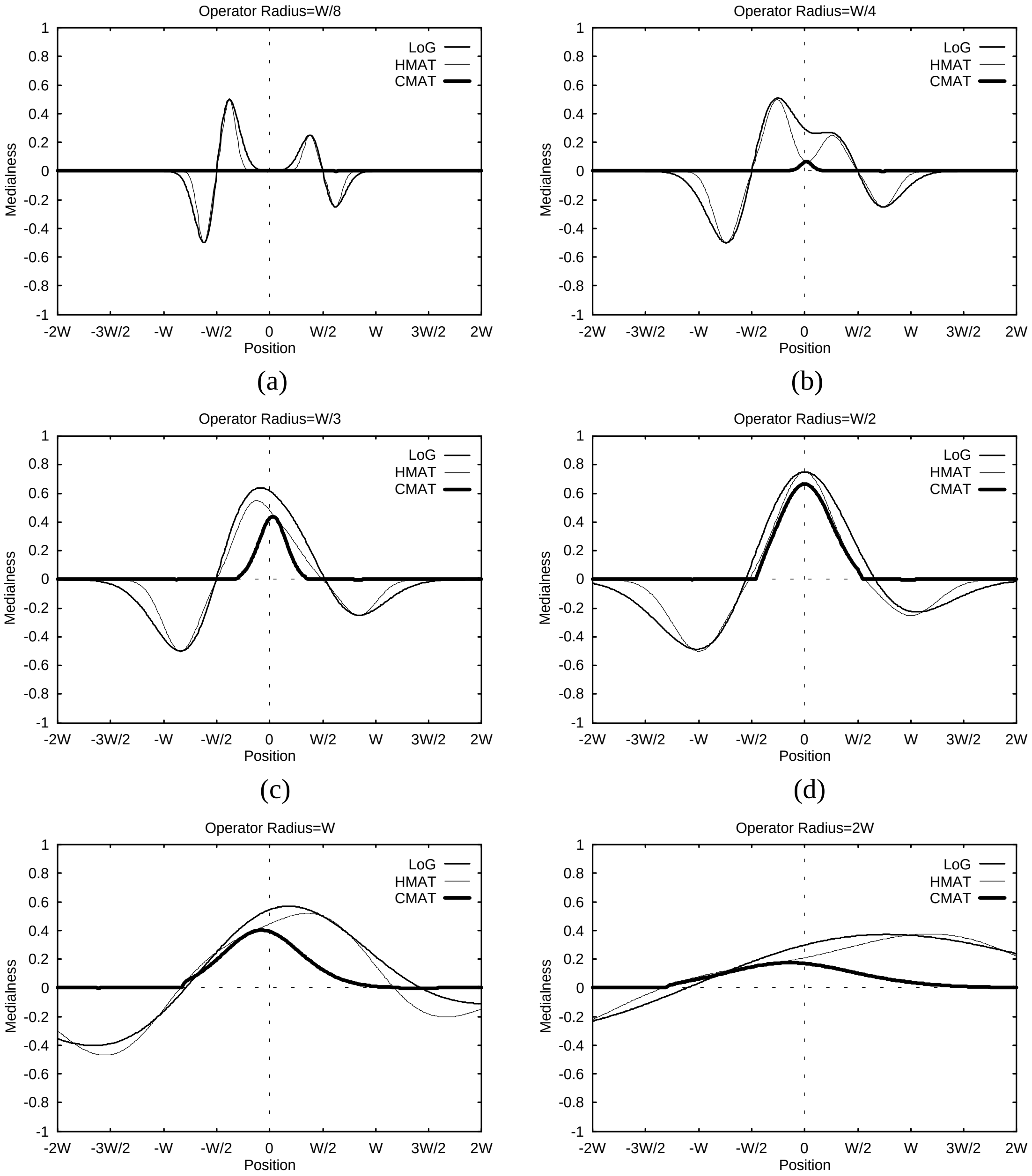


Fig. 4.8. Medialness response through scale, for an asymmetric pulse (a=0.5) and generated from (a) LoG, (b) HMAT, and (c) CMAT.

4.1.2.2 Medialness at selected scales

The LoG, HMAT, and CMAT medialness responses at several operator radii for the asymmetric pulse (a=0.5) are shown in Fig. 4.9. At radii, r=W/4 and r=W/3 (Figs. 4.9(b)-(c)), the edge responses of the LoG and HMAT combine into a single positive lobe which peaks away from the pulse centre; The peak of the CMAT response is much closer to the pulse centre. When the operator radius matches the pulse half-width, r=W/2 in Fig. 4.9(d), the peaks of the edge responses for the LoG and the HMAT are overlaid at the pulse centre, producing a global maximum of medialness. This maximum value is 75% that in Fig. 4.2(d) for a symmetric pulse, because the magnitude of one edge response is halved (see Fig. 4.9(a)). At r=W/2, the CMAT response at the pulse centre comes from the peak values of the boundariness responses at both edges. Because the CMAT medialness is dominated by boundariness strength (see Section 3.2.2.2), a global maximum is formed at the pulse centre, though the concordance degree between boundariness responses is not perfect (c<1). This smaller concordance factor has decreased the CMAT response a little, compared to the LoG and HMAT responses (Fig. 4.9(d)). At radii larger than the pulse half-width, r=W and r=2W in Figs. 4.9(e)-(f), each medialness response collapses and forms a maximum away the pulse centre. The peak of the CMAT response is much closer to the pulse centre than those of the others, which indicates that the CMAT has the best estimation of the pulse centre at a single operator radii.

From Figs. 4.2 and 4.9, the response of the CMAT is always enveloped in that of the HMAT, and they coincide with each other only at the radius equal to the half-width of a symmetric pulse (Fig. 4.2(d)). In effect the concordance facet of the CMAT selects that part of the HMAT response which arises from the interaction of boundary responses.



(a) (b)

(c) (d)

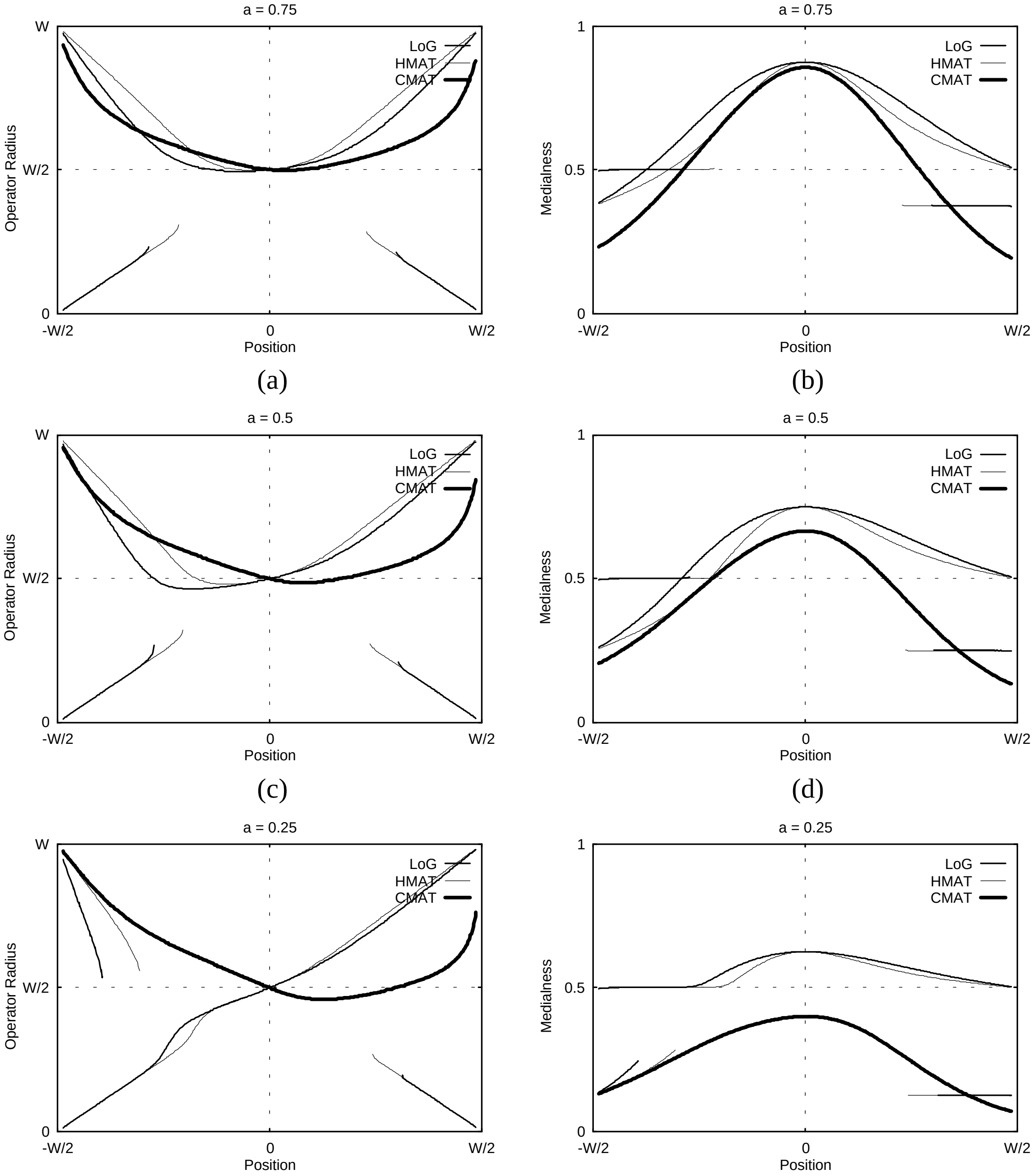
(e) (f)

Fig. 4.9. Medialness responses at several scales, for an asymmetric pulse (a=0.5) and generated from the LoG, HMAT and CMAT.

4.1.2.3 Optimal scale maxima

Figs. 4.10(a)-(f) show the optimal radii and optimal radius responses, as a function of position, for asymmetric pulses with varying degrees of symmetry. For the LoG and HMAT, the loci of optimal radii and associated responses, at a=0.75 and a=0.5, are similar to those for a symmetric pulse (a=1, see Fig. 4.4). They are only shifted lower on the left and higher on the right by the asymmetry of the pulse. A major change in the LoG and HMAT loci of optimal radii occurs at a=0.25, when the locus at large radii breaks up and combines with an edge response locus at small radii. The CMAT locus of optimal radii has no structural change with the variation in degree of symmetry. It is only distorted by the asymmetry. As a result, contrary to the case for the symmetric pulse, each algorithm may exhibit an optimal radius smaller than the pulse half-width, for positions away from the pulse centre. From Fig. 4.10, for each algorithm, no matter how the degree of symmetry changes, the locus of optimal radii still converges at the three points: x=0 and r=W/2, and $x=\pm W/2$ and $r=W$.

As the degree of symmetry worsens, the LoG and HMAT responses collapse because one of the boundariness responses weakens; The CMAT response collapses faster due to the reduced concordance. Once no concordance exists between boundariness responses, that is the asymmetric pulse evolves into a step edge (a=0), the CMAT response and optimal radius locus will disappear. The spatial maximum of each optimal radius response always appears at the pulse centre and the associated optimal radius matches the pulse half-width. Therefore, for each medialness algorithm, the representation of the pulse, i.e. the location of centre and width, is independent of asymmetry using the optimal radius definition.



(a) (b)

(c) (d)

(e) (f)

Fig. 4.10. The optimal scale maxima for pulses with varying degrees of symmetry: (left column) the radii and (right column) responses, as a function of positions, generated from the LoG, HMAT and CMAT.

4.1.2.4 Optimal position maxima

Figs. 4.11(a)-(f) show the optimal positions and optimal position responses, as a function of operator radius, for pulses with varying degrees of symmetry. For the LoG and HMAT, the loci of optimal positions have a major structural change, compared with those for a perfectly symmetric pulse. The principal locus at the pulse centre combines with an edge response locus to form a continuous locus through radius, when the pulse symmetry is perturbed. The CMAT loci show no structural change, compared with the response to a symmetric pulse. The asymmetry shifts the vertical locus in an opposite direction to the that of the LoG and HMAT.

At a radius other than the pulse half-width, the sequence of the accuracy with which the position of the pulse centre is found is (best first): CMAT, LoG, and HMAT. For a=0.5 as in the previous example (Figs. 4.8 and 4.9), the estimation bias of the pulse centre at $r=W$ is 36% of the pulse half-width for the LoG, 72% of the pulse half-width for the HMAT, and 16% of the pulse half-width for the CMAT. The HMAT medialness response more readily moves to one side, due to differences in boundary strength, than the LoG medialness. This result is contrary to the observation in [MOR94] concerning the behaviour of local (LoG) and multi-local (HMAT) medialness operators.

From Fig. 4.11, for each algorithm, no matter how the degree of symmetry of the pulse changes, the maximum of the optimal position response always appears at an operator radius equal to the pulse half-width, and the associated optimal position is kept at the pulse centre. Therefore, for each medialness algorithm, the representation of a pulse, i.e. the pulse centre and width, is independent of asymmetry using the optimal position ridge definition. It is noted that, unlike the case for the symmetric pulse, the optimal position response of the CMAT no longer matches that of the HMAT, due to a lessened degree of concordance between edge responses.

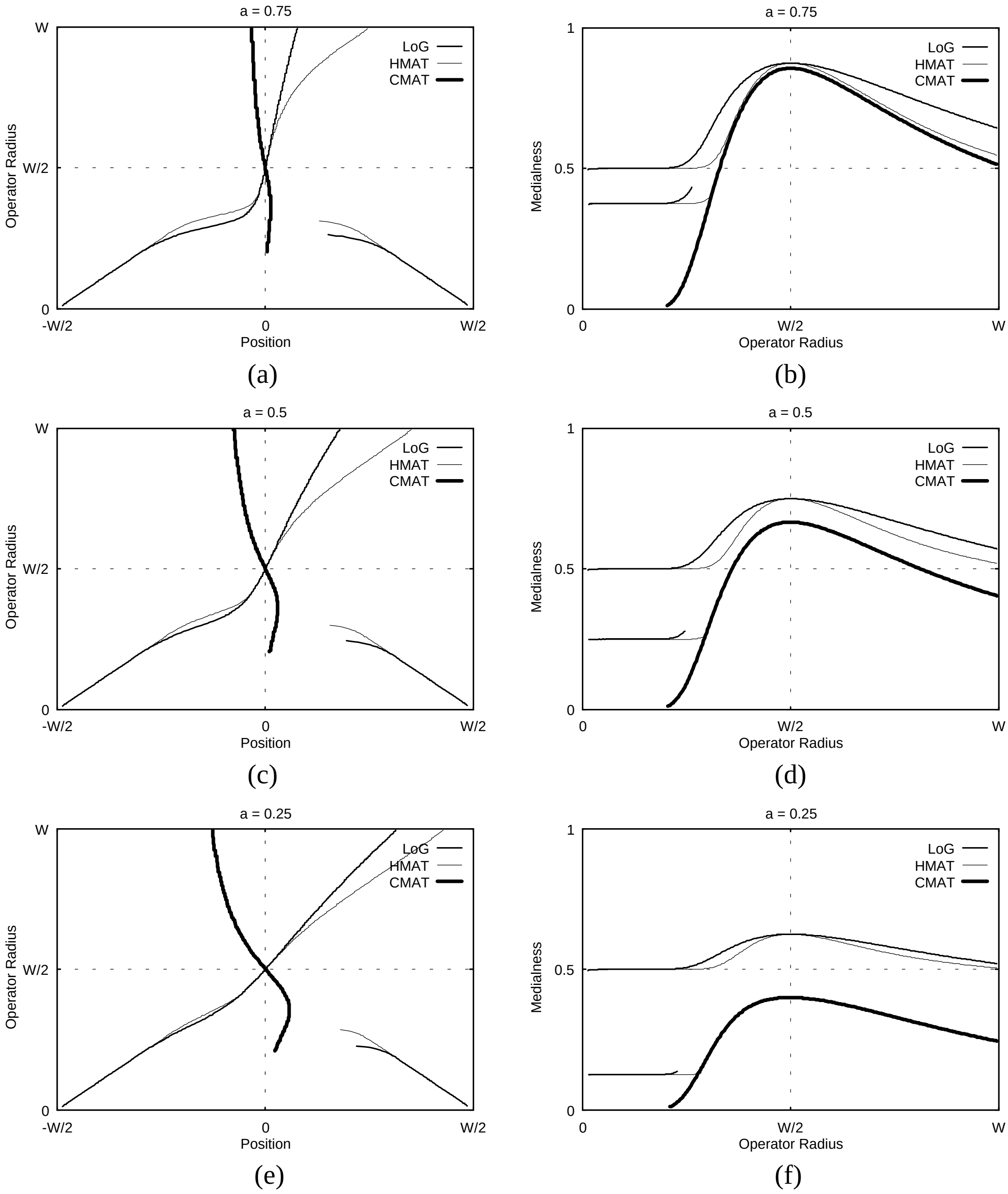


Fig. 4.11. The optimal position maxima for pulses with varying degrees of symmetry: (left column) the positions and (right column) responses, as a function of operator radius, generated from the LoG, HMAT and CMAT.

4.1.3 Noisy Pulses

One motivation for scale-space analysis is to separate the representation of large-scale structures from that of fine-scale detail, such as noise. To investigate how noise can affect the

CMAT medialness response and consequently the representation of an object, we add zero-mean, Gaussian distributed noise to a symmetric pulse. The noise level is defined by the ratio, *u*, of the standard deviation of the noise to the pulse height. Fig. 4.12 shows a noisy pulse with *u*=0.5.

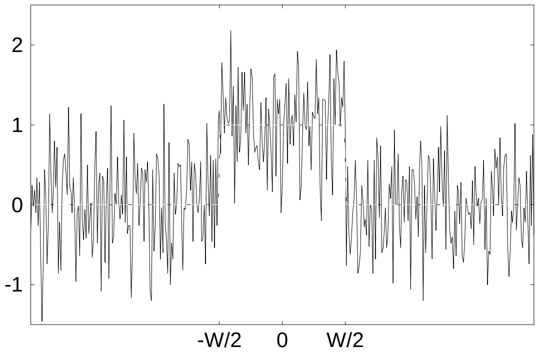


Fig. 4.12. A symmetric pulse with zero-mean Gaussian noise added (u=0.5).

4.1.3.1 Medialness through scale

The medialness responses through radii, for the noisy pulse in Fig. 4.12 and generated from the LoG, HMAT, and CMAT, are shown in Figs. 4.13(a)-(c). Despite the noise response being concentrated at a fine scale, each algorithm continued to form a global maximum, for the symmetry of the pulse, near the position of the pulse centre and at a radius close to pulse half-width. To correctly extract the representation of a pulse, the response must be separated from that of noise at an operator radius equal to the pulse half-width. Both increasing the signal-to-noise ratio and increasing the pulse width can enlarge the separation between the response for the pulse and that for the noise.

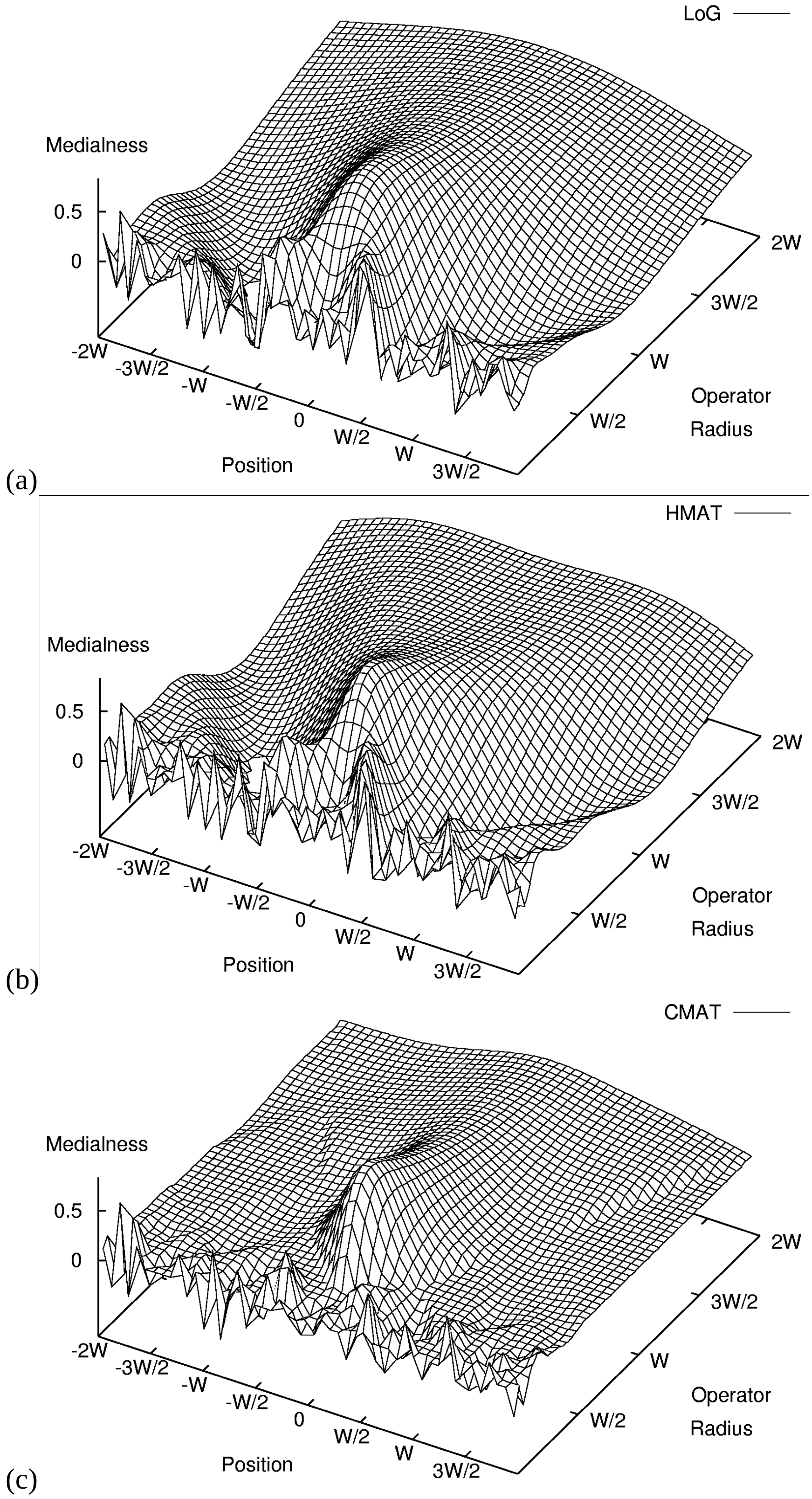


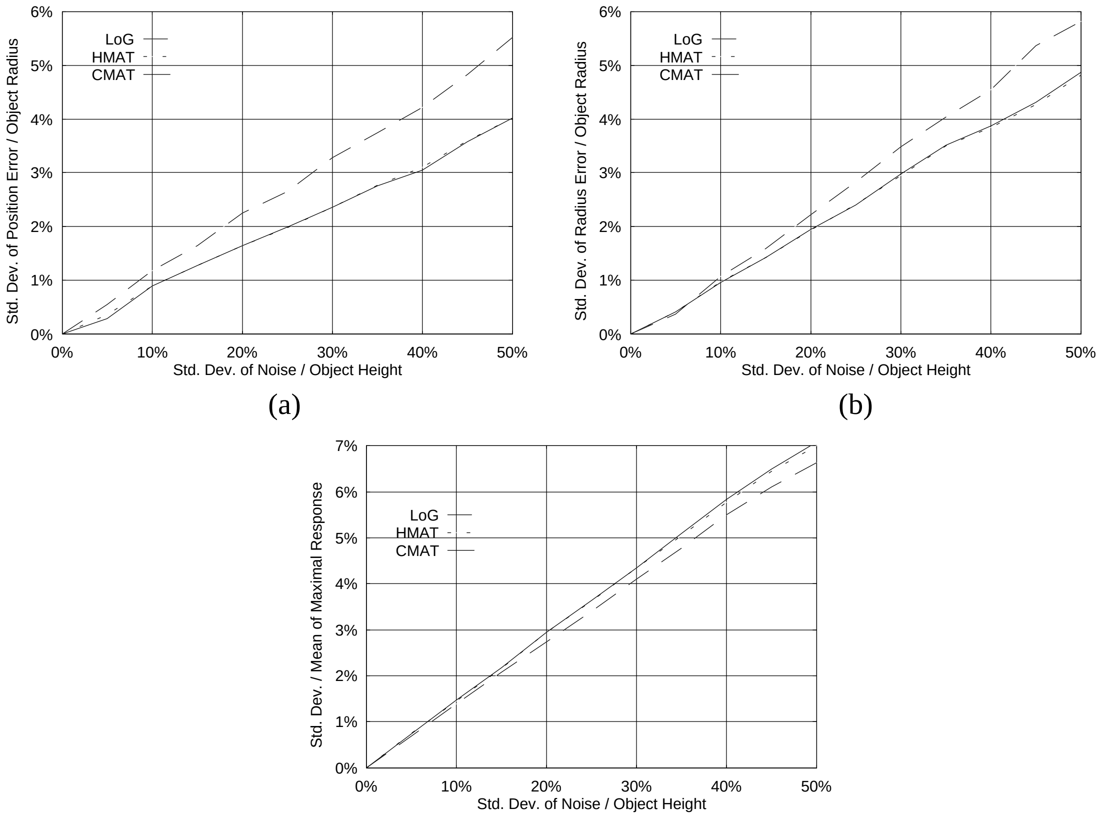
Fig. 4.13. Medialness responses through scale, for a noisy pulse (u=0.5) and generated from (a) LoG, (b) HMAT, and (c) CMAT.

4.1.3.2 Localisation of medialness maxima

A Monte Carlo simulation was performed to analyse the quantitative influence of noise on medialness response. The noise level, *u*, was increased from 0 to 50% in steps of 5%. At each noise level, one thousand noisy sample pulses were generated and input to the LoG , HMAT, and CMAT algorithms. The global maximum of medialness, at a position between [-*W*/4, *W*/4] and at a radius between [*W*/3, 3*W*/2], was taken as the estimated pulse centre and half-width. The estimated pulse centre, width, and response for noisy pulses are compared with those for a noise-free pulse.

It was found that the mean position, radius, and maximum response of the detected pulse centre were not changed by the addition of zero-mean Gaussian noise. However, the standard deviation of these quantities each increased linearly with noise level (see Figs. 4.14(a)-(c)). The curves for the CMAT always coincided with those for the HMAT, which implies again that these algorithms perform similarly in determining the position and half-width of a symmetric pulse. The HMAT and CMAT produce a 27% less estimation error, for pulse centre, and 15% less estimation error, for pulse half-width, than the LoG. This is consistent with Morse's conclusion [MOR94] that local operators, like the LoG, are sensitive to variations within an object because they require integration over the entire width of the object.

It is not surprising that the standard deviations of the maximum responses, of the HMAT and CMAT, are greater than that of the LoG. The LoG is a Gaussian (with standard deviation of $\sigma = r/\sqrt{n} = r$, for 1-D space) convolved with a second-derivative operator. The HMAT and CMAT involve the combination of boundariness responses which computation is based on a Gaussian (with a standard deviation of $\sigma = r/k$, i.e. $r/2$ in our test) convolved with a first-derivative operator. Thus, at a given operator radius, *r*, the degree of smoothing (σ) for



(a) (b)

(c)

Fig. 4.14. The standard deviation of estimation, with respect to W/2, for (a) pulse centre and (b) pulse half-width, and (c) the standard deviation of the medialness response, with respect to the global maximum response for a noise-free pulse. Each graph is plotted as a function of noise level and generated from the LoG (long dash), HMAT (short dash) and CMAT (solid).

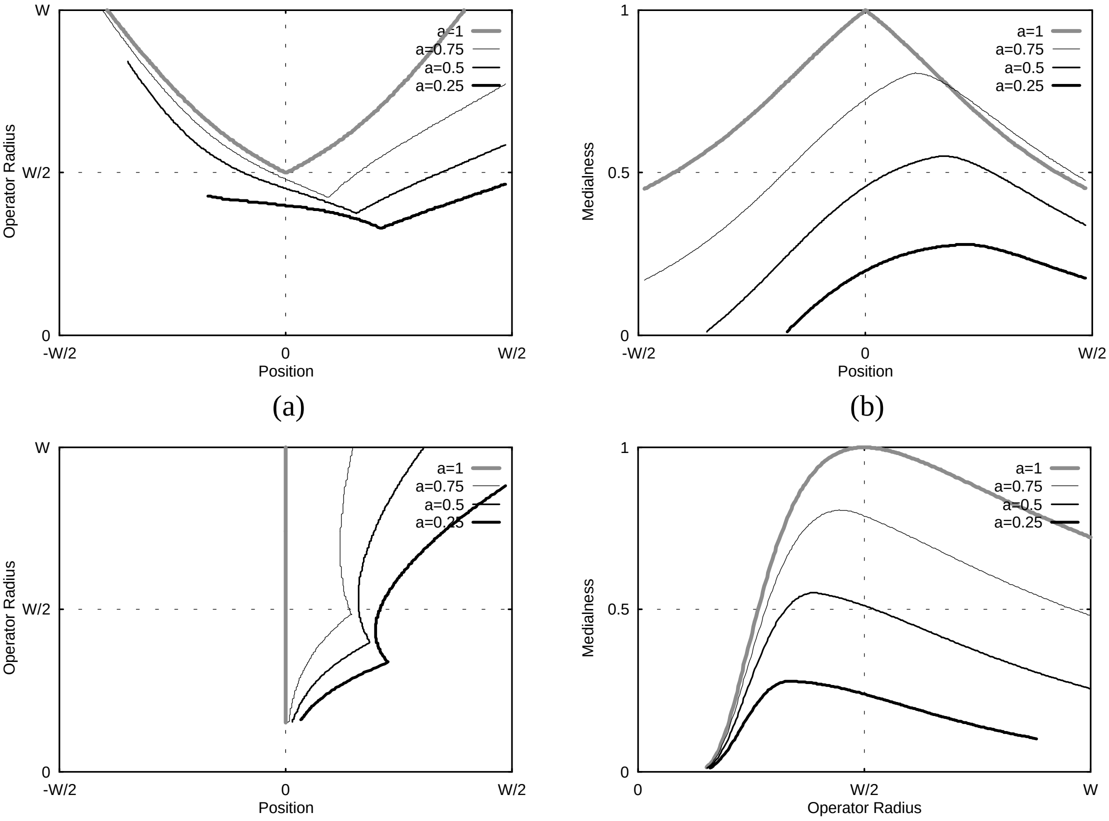
the LoG doubles that for the HMAT and CMAT. However, the variation of the medialness maximum response, in the presence of noise, differs by only 6%. The performance of the LoG is compromised by the second-derivative operator that is more sensitive to noise than the first-derivative operator.

4.1.4 Koller Line Detector

Koller *et al.* [KOL95] analyzed a symmetric pulse, like that shown in Fig. 3.13, and showed that the shift of edge response is determined by $s=\sigma$ and that a global medialness maximum appears at a scale of $\sigma=0.83356W/2$. To give an objective and efficient comparison with the CMAT, the operator radius, *r*, is taken to be equivalent to the scale parameter, σ, with a relation, r=1.19967σ and the medialness response normalised by the global maximum value for a symmetric pulse with unit height.

4.1.4.1 KLD minimum operator ($s=\sigma$)

Figs. 4.15(a) and (b) show the optimal radii and optimal radius responses, as a function of position, generated from the KLD minimum operator ($s=\sigma$) and for pulses with varying degrees of symmetry. For each pulse, there is only one locus reflecting the symmetry and edge responses are completely suppressed, as with the CMAT. As the degree of symmetry decreases the response becomes smaller, which is similar to that observed with the CMAT.



(a) (b)

(c) (d)

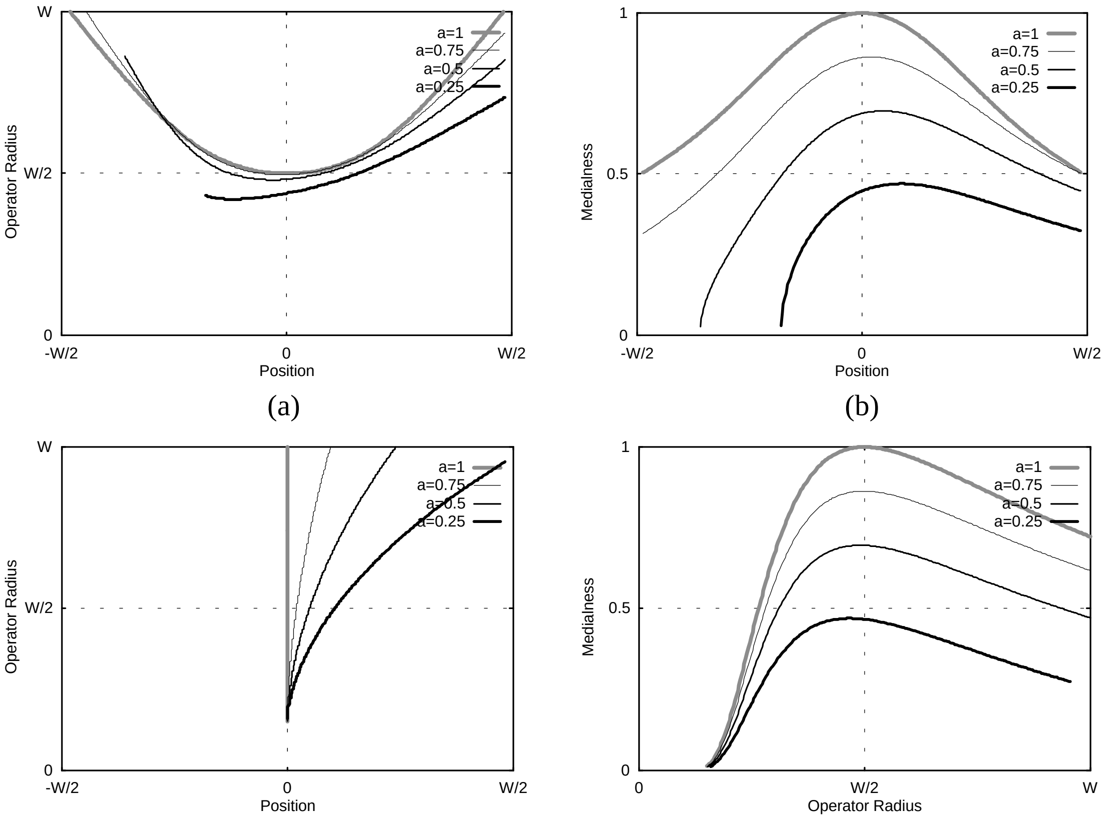
Fig. 4.15. Medialness maxima for the KLD minimum operator (s=σ): (a) optimal radii and (b) optimal radius responses, as a function of position; (c) optimal positions and (d) optimal position responses, as a function of operator radius, for pulses with varying degrees of symmetry.

However, for the KLD, the spatial maximum of the optimal radius response shift to emphasise the right side of the pulse centre, which means that the KLD minimum operator ($s=\sigma$) does not accurately localise the centre of an object with imperfect symmetry. In addition, the “sharp peaks” of the optimal radius response described by Koller [KOL95] occur only for a symmetric pulse ($a=1$).

Figs. 4.15(c) and (d) show the optimal positions and optimal position responses, as a function of radius, for pulses with varying degrees of symmetry. Again only one locus for each pulse is present. When the degree of symmetry is reduced, the locus of optimal positions is shifted towards the weaker boundary and the maximum of the optimal position response appears at an operator radius smaller than the pulse half-width. This means that the KLD minimum operator ($s=\sigma$) cannot accurately identify the width of an object with imperfect symmetry. The sudden turnings in the loci of optimal positions result from selecting the smallest edge response and abruptly switch of emphasis from one to the other.

4.1.4.2 KLD geometric mean operator ($s=\sigma$)

If we replace the minimum operation for shifted edge responses with the geometric mean ($s=\sigma$), the optimal radius maxima and the optimal position maxima are as shown in Fig. 4.16. These results are similar to those obtained using the minimum operation: the spatial maximum of the optimal radius response is shifted to the weaker boundary (Fig. 4.16(b)), and the maximum of the optimal position response is shifted to an operator radius smaller than the pulse half-width (Fig. 4.16(d)). Though the biases are smaller than those observed using the minimum operation, the KLD geometric mean ($s=\sigma$) cannot accurately identify the centre and width of a pulse with imperfect symmetry.



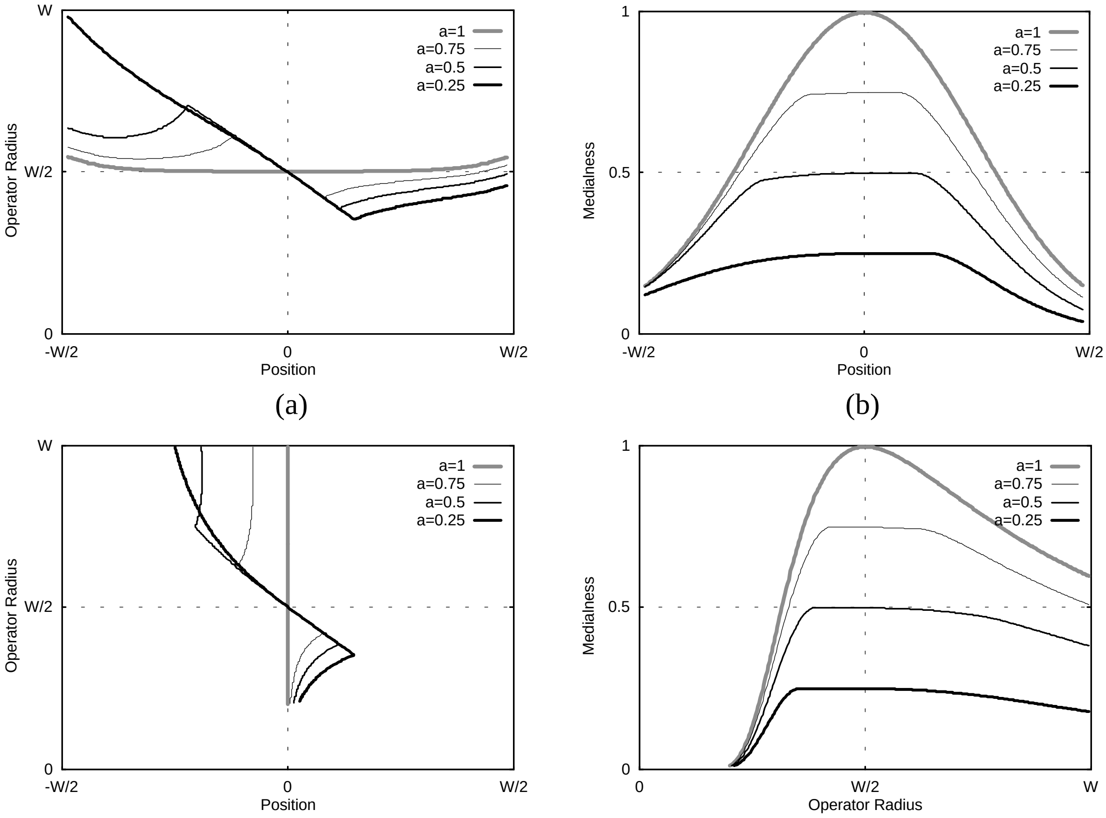
(a) (b) (c) (d)

Fig. 4.16. Medialness maxima for the KLD geometric mean operator (s=σ): (a) optimal radii and (b) optimal radius responses, as a function of position; (c) optimal positions and (d) optimal position responses, as a function of operator radius, for pulses with varying degrees of symmetry.

4.1.4.3 KLD minimum operator (s=2σ)

If we increase the shift of edge responses, relative to scale, to $s=2\sigma$, the results for the KLD minimum and geometric mean operators are as shown in Figs. 4.17 and Fig. 4.18, respectively.

In Fig. 4.17, the change in the shift-to-scale ratio greatly alters both the loci of optimal radii and optimal positions. The change in shape results from the repeated change (of the minimum operation) between the edge responses used to compute “medialness”. It is noted that both the loci converge at the pulse centre and at an operator radius equal to the pulse half-width. This comes close to the result of the CMAT, LoG, and HMAT, shown in Figs. 4.10 and



(a) (b)

(c) (d)

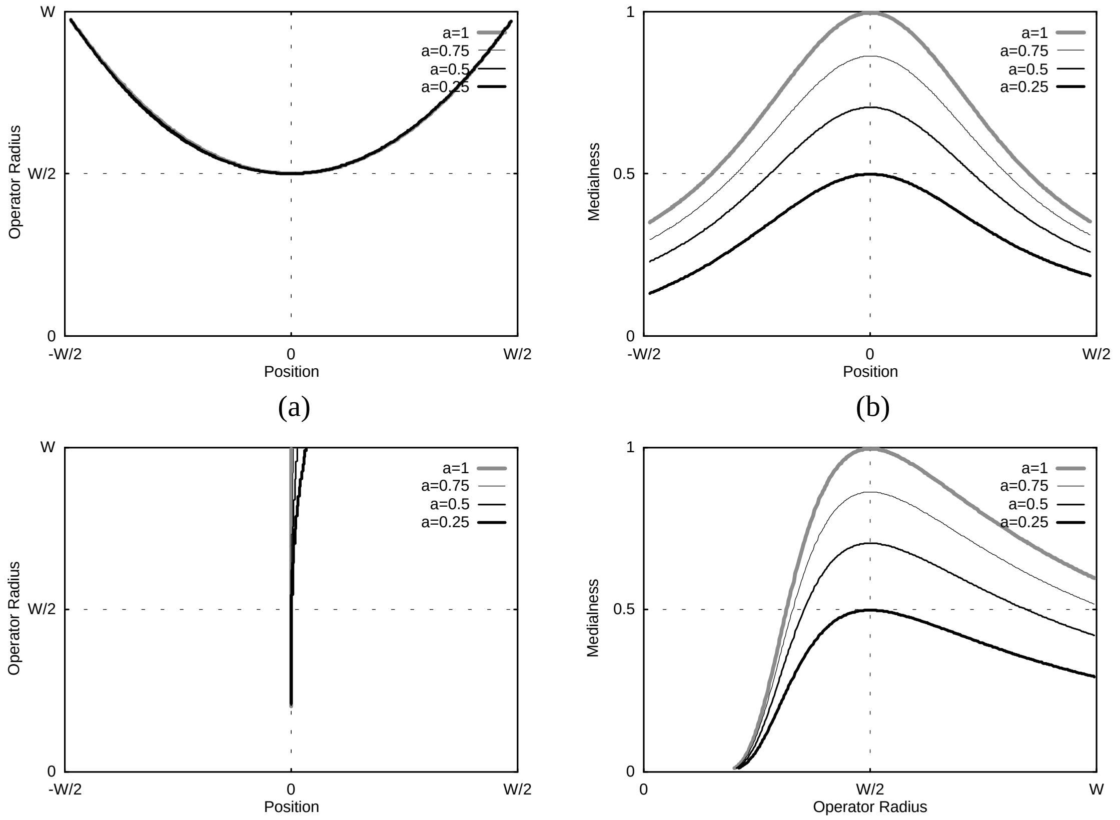
Fig. 4.17. Medialness maxima for the KLD minimum operator (s=2σ): (a) optimal radii, and (b) optimal radius responses, as a function of position; (c) optimal positions and (d) optimal position responses, as a function of operator radius, for pulses with varying degrees of symmetry.

4.11 and might be considered to imply that the localisation accuracy of pulse centre and half-width estimation is not affected by degree of symmetry. However, both the optimal radius response (Fig. 4.17(b)) and optimal position response (Fig. 4.17(d)) have wide flat tops, which makes the identification of pulse centre and half-width impractical for asymmetric pulses. On the other hand, the optimal radii for a symmetric pulse almost stay at an operator radius of $r=W/2$ ($a=1$ in Fig. 4.17(a)), which means the estimation of the width of a symmetric pulse is less critical for the KLD minimum operator ($s=2\sigma$), at positions off the pulse centre.

4.1.4.4 KLD geometric mean operator (s=2σ)

In Fig. 4.18(a), the loci of optimal radii coincide, which indicates that the KLD geometric mean operator (s=2σ) is robust to the variations in degree of symmetry. This is also reinforced by the balanced strength on both the sides of each optimal radius response (Fig. 4.18(b)). Each optimal radius response has its spatial maxima at the pulse centre and the associated optimal radius is equal to the pulse half-width. Therefore, we can accurately identify the pulse centre and half-width, using the optimal scale (radius) definition on the medialness computed by the KLD geometric mean operator (s=2σ). In Fig. 4.18(c), the optimal position maxima are displaced further from the pulse centre as degree of symmetry decreases. However, they still converge at the position of the pulse centre and the operator radius of the pulse half-width. The spatial displacements at large scales are much smaller than those of the CMAT, LoG, and HMAT (see Fig. 4.11). Therefore the KLD geometric mean operator (s=2σ) has the best estimation of the pulse centre at any single scale.

This property has a potential and useful application on decreasing the sampling rate of scale (radius). If the width of a line being detected has a limited range and we had prior knowledge of scale (e.g. roads in remote-sensing images), it would be possible to approximately localise the object centre or medial axis by computing medialness at one scale (radius). In Fig. 4.18(d), the maximum of the optimal position response occurs at an operator radius equal to the pulse half-width, and the corresponding position is the pulse centre. Therefore we can accurately identify the pulse centre and half-width using the optimal position definition.



(a) (b) (c) (d)

Fig. 4.18. Medialness maxima for the KLD geometric mean operator (s=2σ): (a) optimal radii and (b) optimal radius responses, as a function of position; (c) optimal positions and (d) optimal position responses, as a function of operator radius, for pulses with varying degrees of symmetry.

4.2 Applications of the CMAT to 2-D Images

4.2.1 The CMAT Medialness Responses

To assess the significance of the theories presented in Section 3.2, we evaluated the CMAT algorithm using synthetic and natural images. We compared these results with those obtained using our own implementations of the LoG, HMAT and credit attribution algorithms. The linear HMAT was used for comparison throughout this thesis. The results for the HMAT-2, the initial medialness response of the CMAT, are provided here to demonstrate how the improved CMAT responses arise. The size of each image is 128×128 pixels, except those in Figs. 4.25 and 4.26 which are 256×256 pixels. Each medialness algorithm was implemented to be sensitive to both polarities of boundary transition. The medialness images show the absolute value of medialness response. Therefore, in the medialness results for the LoG operator the zero-crossings, which reflect boundary information, are replaced by minima with zero value. As in the 1-D case, we use operator radius as a common parameter of scale for all the operators. Here the radius-to-scale ratio that determines operator radius, k, is set at 2 and the constant, which determines the distribution extent of boundariness contributions for overcoming the sparseness in discrete sampling, λ, is set at 0.5.

In the case of a 2-D image, the medialness response can be thought of as the density at each point in a 2+1 dimensional scale-space. However, no graphical visualisation is possible to display such a response completely using a single 2-D view. The medialness scale-space has to be visualised at a selected scale or in a selected section plane in 3-D scale-space..

Fig. 4.19 shows the medialness response, for a rectangle with a sawtooth edge, at three selected scales. For the LoG (first row) and HMAT (second row) operators, edge responses appear as lines framing the shape; the brightest regions correspond to medialness responses due to symmetry. At small operator radii of 5 and 10 pixels, the linear operators reflect more edgeness than symmetry. At an operator radius of 20 pixels, the linear operators produce of a central response for symmetry, the ridge of which provides a precise description of the overall shape, and an outer edge response. The response of the credit attribution algorithm (third row) makes the medialness response for symmetry more prominent; The outer edge responses present in Figs. 4.19(d) and (g) are eliminated. However, the credit attribution algorithm has limited impact on the way that isolated edges generate a “medialness” response (at radii of 5 and 10 pixels). This is not surprising because the enhancement of the credit attribution depends on the competition of the boundariness contributions on both the sides of each edge (the parallel lines in Figs. 4.19(e) and (h)); for an isolated edge or at an operator radius much smaller than the local width, the boundariness contributions on both sides of an edge have the same initial strength (HMAT medialness) and become two “winners”. The response of the HMAT-2 algorithm (fourth row) is similar to that of the HMAT (second row) but the responses for symmetry are more strongly emphasised. This is due to the angular range for weighting, θ in Eq. (3.23), being smaller than the effective angle ($\theta = \pi$) of the circular arc used in the HMAT algorithm [MOR91]. In the response of the CMAT algorithm (fifth row), only responses for symmetry are generated at all radii. The hierarchy and robustness of multi-scale analysis are demonstrated in the CMAT results: the triangular sawteeth and end corners are reflected at small radii (Fig. 4.19(n)), the rectangular shape is reflected at medium radii (Fig. 4.19(o)), and the elongated shape is reflected at large radii (Fig. 4.19(p)). The medialness of larger scale features is little affected by fine detail.

Fig. 4.20 shows two sections through the medialness response for the rectangle with a sawtooth edge (shown in Fig. 4.20(a)) to give an impression of the 3-D nature of medialness, in scale-space. Fig. 4.20(b) is a diagrammatic representation of the 3-D medialness scale-space, as a stack of 2-D slices at progressively larger scales. Fig. 4.20(c) shows a section through the 3-D scale space at plane C and Fig. 4.20(d) shows a section at plane D. Operator radius (scale) is linearly sampled from 1 to 64 pixels in steps of 1 pixel.

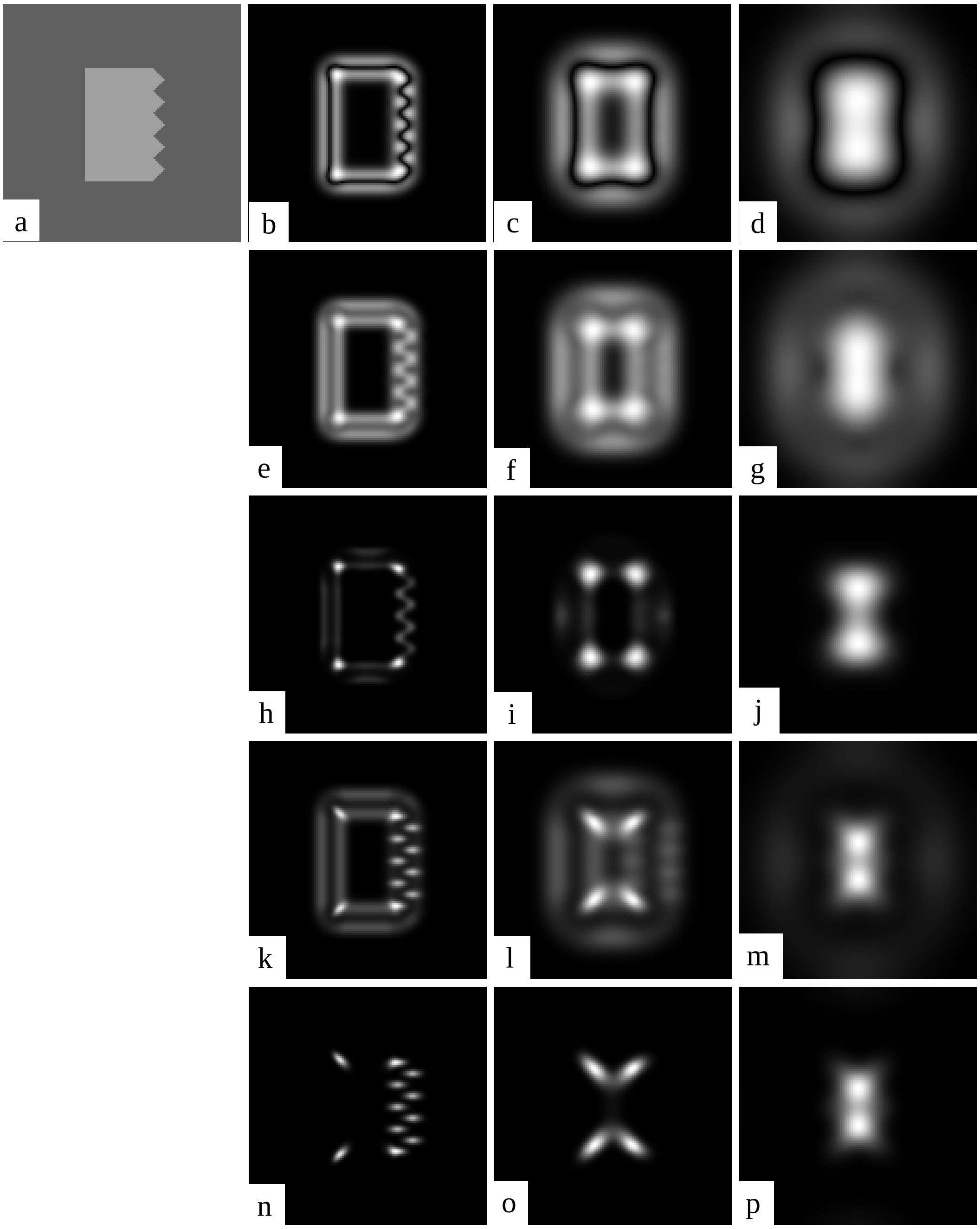
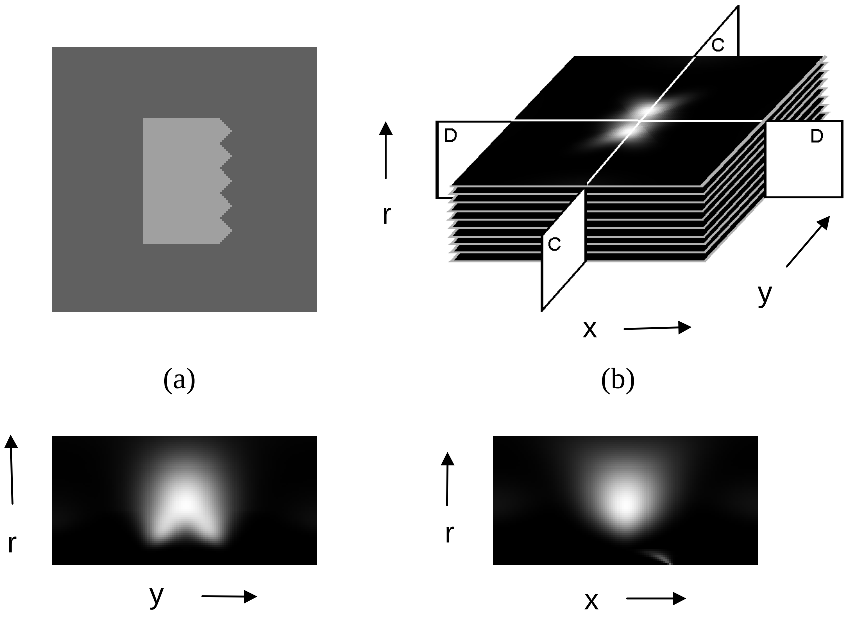


Fig. 4.19. The medialness response at selected scales for a rectangle with a sawtooth edge: (a) Original image; (b)-(d), LoG ; (e)-(g), HMAT; (h)-(j), credit attribution algorithm after 5 iterations; (k)-(m), HMAT-2; (n)-(o), CMAT. The columns from left to right are at operator radii of 5, 10 and 20 pixels, respectively. The half-width of the shape is 20 pixels.



(a) (b)

(c) (d)

Fig. 4.20. Sections of the CMAT medialness in scale-space for a rectangle with a saw-tooth edge: (a) original image, (b) diagrammatic representation of scale-space, (c) section on plane C, and (d) section on plane D.

Note that the shape in Fig. 4.20(a) is symmetric along both the vertical and horizontal axes, that is reflected in Figs. 4.20(c) and (d) by the strong medialness response. The medialness response in plane C is influenced by the extent of the shape. The branching of the medialness response in Fig. 4.20(c) corresponds to the symmetry of the ends of the shape in Fig. 4.20(a). This appears as two highlights in Fig. 4.19(p). For the plane D, Fig. 4.20(d), there is, in effect, no end region to modify the medialness response. The response in Fig. 4.20(d) can be regarded as a grey-level visualisation of the response shown in Fig. 4.1(c). At the bottom of Fig. 4.20(d) (small scale) there is a small “highlight” which arises from the middle sawtooth on the right of the shape in Fig. 4.20(a).

Fig. 4.21 shows the medialness responses, for a teardrop shape, at three selected scales. The radius of the circular arc at the bottom of the teardrop is 24 pixels. Therefore the radius at which the strongest medialness response should occur for this circular arc is 24 pixels. However, we find that the LoG (first row), HMAT (second row) and credit attribution algorithms (third row) begin to produce medialness responses for the circular arc, in the lower part of the teardrop shape, at the smallest radius used, 6 pixels, see Figs. 4.21(b)(e)(h). Further, these responses are displaced from the centre of the circular arc by a relatively large distance and resemble boundary responses. With the CMAT algorithm, the medialness response for the circular arc appears at an operator radius greater than 9 pixels and at a position close to the centre (fifth row). This is consistent with the medialness performance demonstrated in Section 4.1.1 whereby the CMAT produced a medialness response only at positions near to the pulse centre and at radii close to the pulse half-width.

Fig. 4.22 shows two sections through the medialness response for the teardrop shape of Fig. 4.22(a), to give an impression of the scale-space medialness for this shape. Fig. 4.22(b) is a diagrammatic representation of the 3-D medialness scale-space, as a stack of 2-D slices at successive scales. Fig. 4.22(c) shows a section through the 3-D scale-space at plane C and Fig. 4.22(d) shows a section at plane D. The shape in Fig. 4.22(a) displays several symmetries which vary in scale along the y axis. This can clearly be seen in Figs. 4.22(c) and (d).

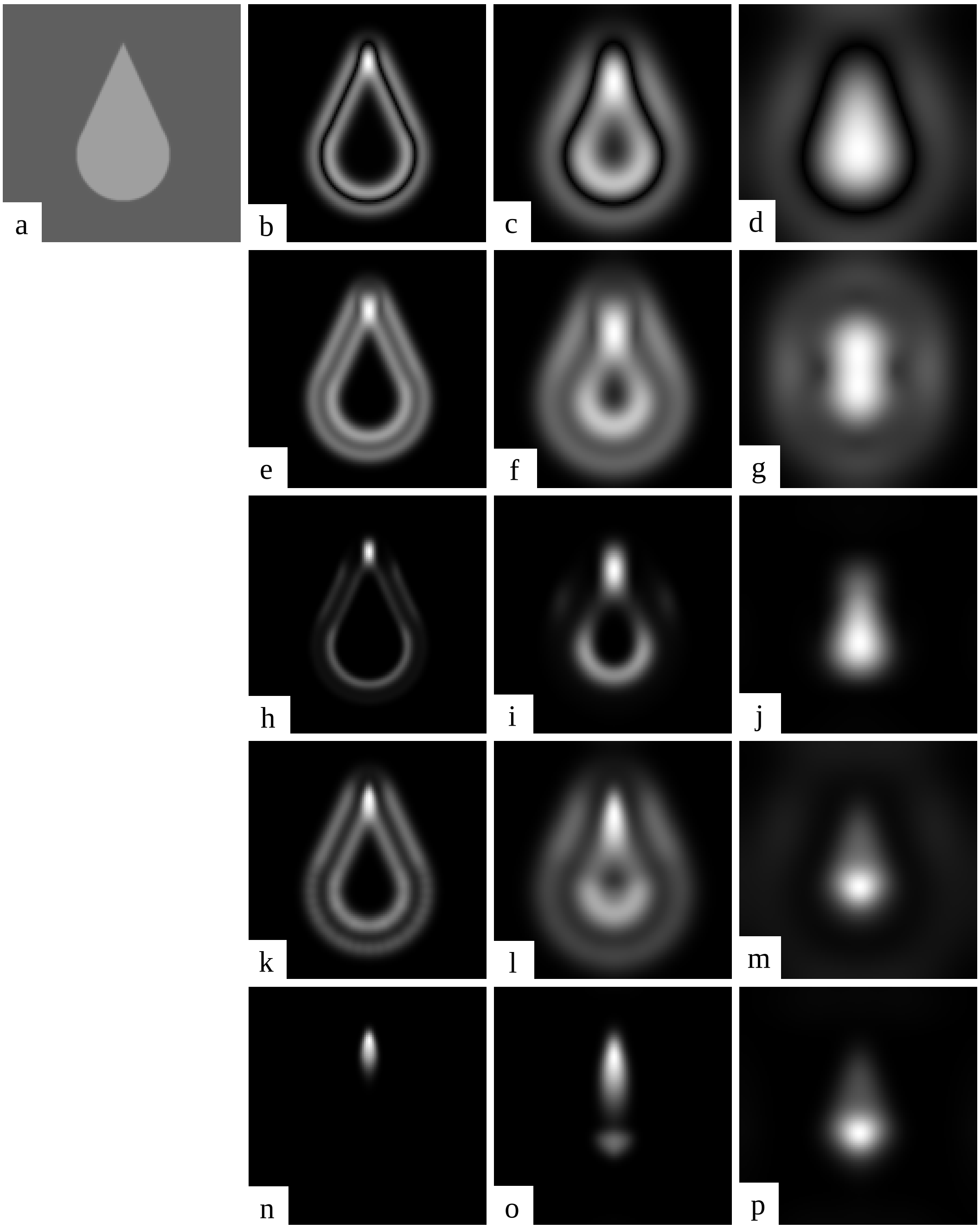
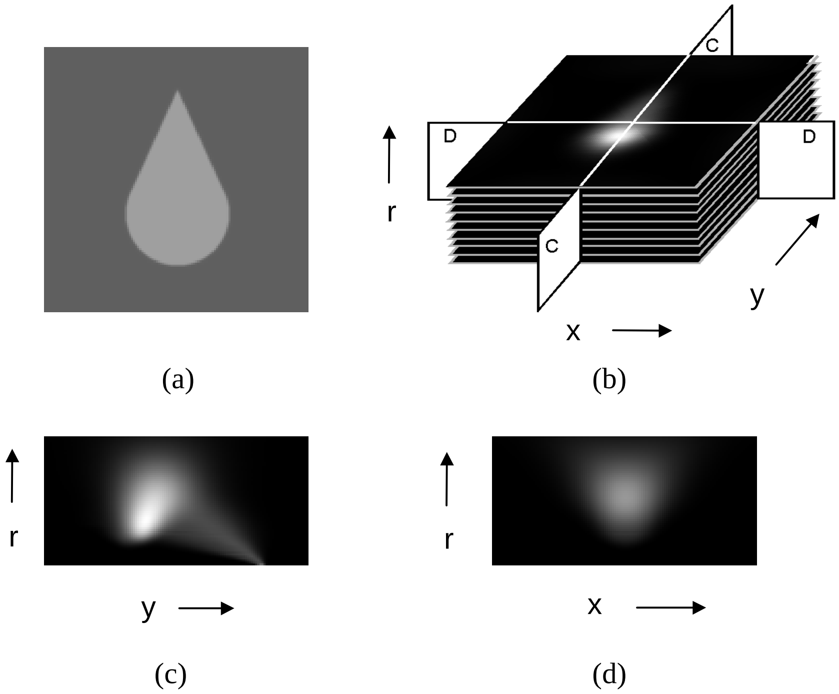


Fig. 4.21. The medialness response at selected scales for a teardrop shape: (a) Original image; (b)-(d), LoG; (e)-(g), HMAT; (h)-(j), credit attribution algorithm after 5 iterations; (k)-(m), HMAT-2; (n)-(p), CMAT. The columns from left to right are at operator radii of 6, 12 and 24 pixels, respectively. The radius of the circular arc at the bottom of the shape is 24 pixels.



(a) (b) (c) (d)

Fig. 4.22. Sections of the CMAT medialness in scale-space for the teardrop shape: (a) original image, (b) diagrammatic representation of scale-space, (c) section on plane C and (d) section on plane D.

Fig. 4.23 shows the medialness response, for a longitudinal MR image of a pair of legs (an image from the Visible Human Project), at three selected scales. This image was chosen as an example of natural image, in which the amplitude of the grey-level boundary varies, the shapes are relatively complex and multiple “objects” are in close proximity. The medialness responses for the LoG (first row) and HMAT (second row) can be seen to be a mixture of medial and boundary responses. The credit attribution algorithm greatly refines the result of the HMAT, but the contour of both legs is still visible (third row). In the results of the CMAT (fifth row), only those responses for symmetry are retained. The CMAT medialness reflects the bones and the fat layers (bright regions in the inner side of legs) at a small radius, at a medium operator radius the knees are emphasised, and at a large operator radius only the major structure of the limbs is maintained. Note that the CMAT response between the legs correctly reflects the medial axis for a triangular shape.

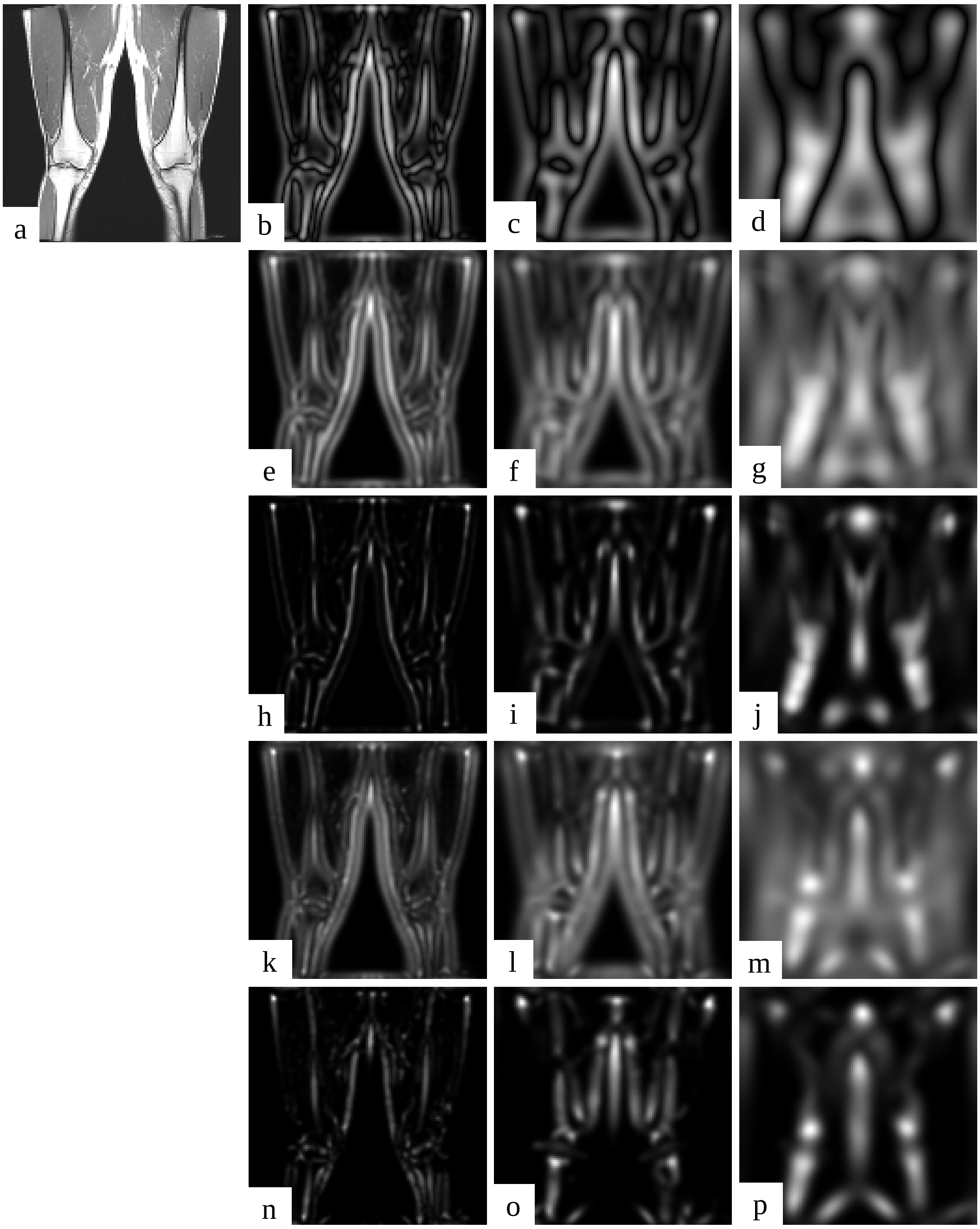


Fig. 4.23. The medialness responses at selected scales for an MR leg image: (a) Original image; (b)-(d), LoG; (e)-(g), HMAT; (h)-(j), credit attribution algorithm after 5 iterations; (k)-(m), HMAT-2; (n)-(p), CMAT. The columns from left to right are at operator radii of 3, 6 and 12 pixels, respectively.

4.2.2 Computational Cost of the CMAT

For each medialness operator, the computational cost at a single scale increases linearly with operator radius, *r*, and the image size, N^2. Therefore the computational complexity at a single scale is $O(r, N^2)$. The times taken to compute the medialness response for a 128×128 pixel image at an operator radius of 20 pixels on a SUN Ultra-2 workstation are shown in Table 4.1. Five iterations were used in the credit attribution (CA) algorithm. We used LUTs to compute the Gaussian and its derivatives in each algorithm, and the weighting function in the HMAT and credit attribution algorithms.

Algorithms	LoG	HMAT	CA	HMAT-2	CMAT
Time (seconds)	10	40	365	11	13

Table 4.1. Time taken to compute medialness for several medialness algorithms at a selected operator radius.

To compute a medialness scale-space at operator radii from 1 to S_H pixels and in steps of 1 pixel, the computation cost is in proportion to:

$$N^2 \sum_{r=1}^{S_H} r = N^2 \frac{S_H (S_H + 1)}{2}$$

where $S_H >> 1$. Therefore the corresponding computational complexity is $O\left(S_H^2, N^2\right)$.

4.2.3 Alternative Definitions of Contribution Confidence

In Section 3.2.3, we presented alternative definitions of contribution confidence for the CMAT algorithm. Compared to the definition of contribution confidence $f(x) = 1 - x$ (Section 3.2.1), used in the preceding experiments, these definitions may emphasise (or de-emphasise) end-point medialness with respect to the medialness of parallel boundaries. Fig. 4.24 shows the CMAT medialness responses of a bar shape at a selected scale, using different

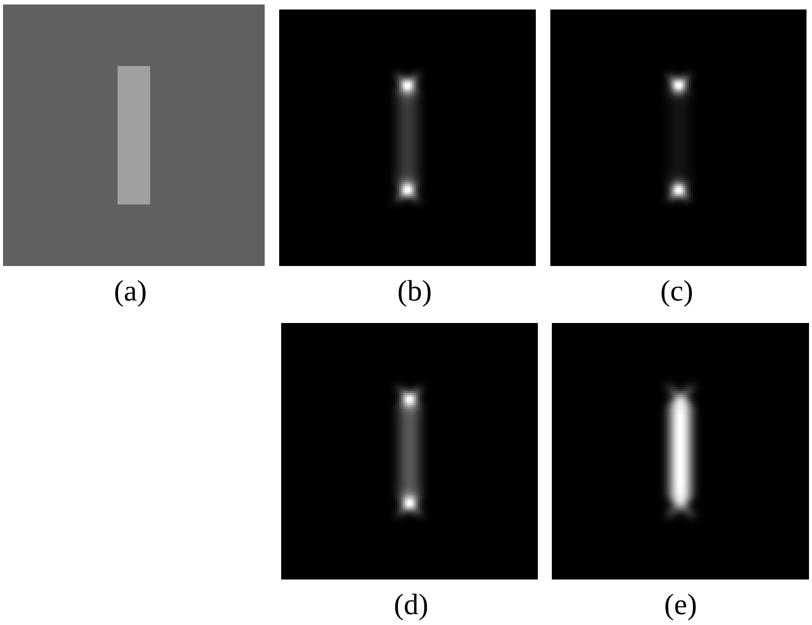


Fig. 4.24. Different definitions of contribution confidence: (a) the original image of a bar shape, and its CMAT medialness responses, at a operator radius equal to object half-width, using curve functions: (b) I, (c) II, (d) III, and (e) IV.

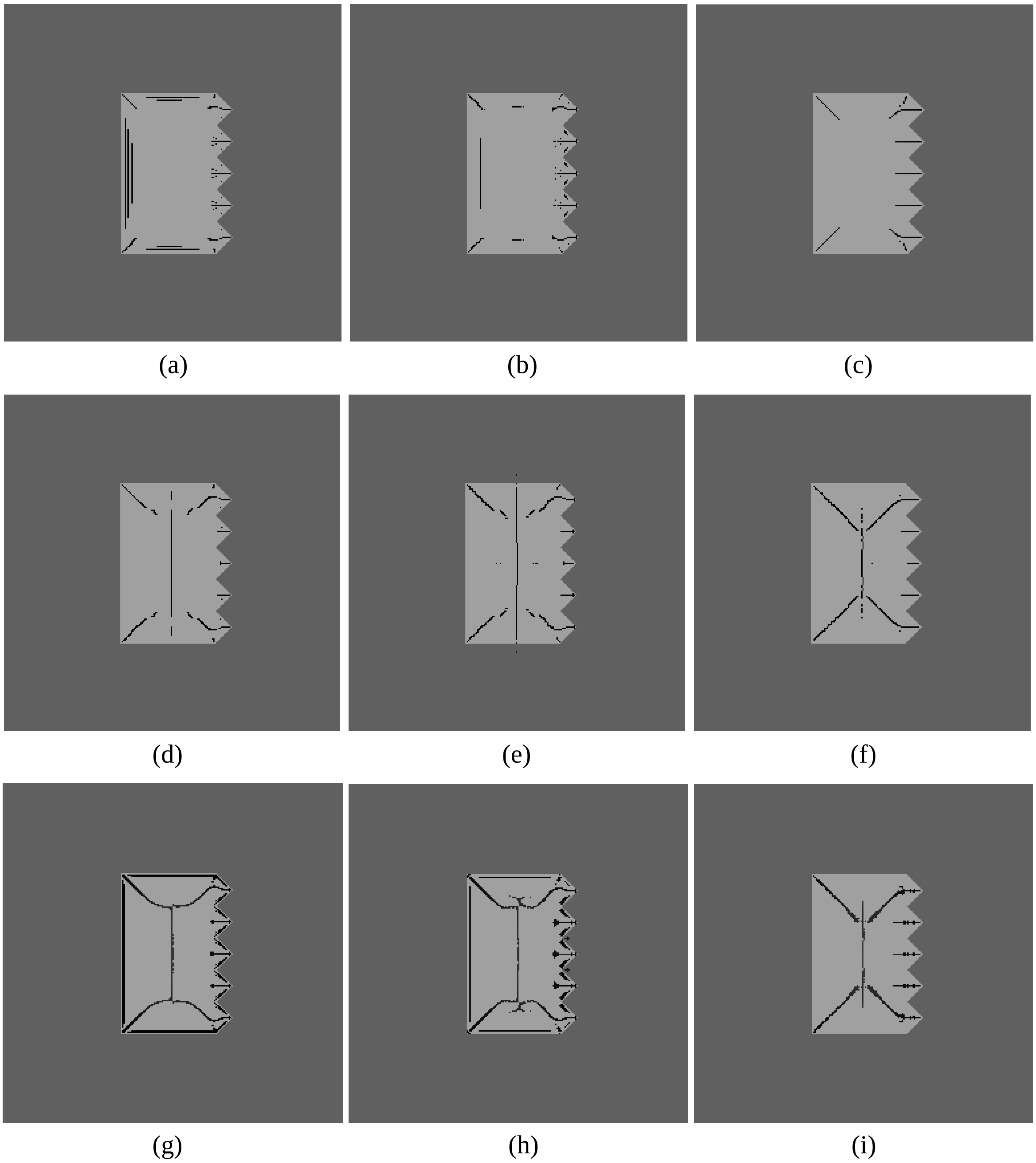
definitions for contribution confidence. Function I is the definition of contribution confidence $f(x)=1-x$ used in the preceding experiments and the CMAT medialness response is shown in Fig. 4.24(b). Function II is $f(x)=(1-x)^2$ and suppresses the relative CMAT medialness for parallel boundaries as shown in Fig. 4.24(c). Function III is $f(x)=(1-x)^{1/2}$, which enhances the relative CMAT medialness for parallel boundaries as shown in Fig. 4.24(d). Function IV is triangular in shape, as defined in Eq. 3.48, producing CMAT medialness evenly distributed along the middle line, as shown in Fig. 4.24(e).

4.2.4 Ridges of CMAT Medialness Using Existing Algorithms

Fig. 4.25 shows the results of the ridge extraction on the medialness responses, computed using the LoG, HMAT, and CMAT, for the rectangle with a sawtooth edge. These ridges were obtained using the optimal scale ridge definition at operator radii of 1-15 pixels (top row) and 1-60 pixels (middle row), and the height ridge definition at operator radii of 1-60 pixels (bottom row). The medialness response, for a large-sized feature (at large scales), at positions away from the medial axis may dominate the response, for a small-sized feature (at small scales), at positions on the medial axis as discussed in Section 3.3.2.2. In addition, the optimal scale ridge definition only considers the global medialness maximum over scale at each pixel. To avoid the medial axes for small-sized features (here including the spurious MMA arising from edge responses) being hidden or shortened by large-sized features (compare the top and middle rows), the operator radius was first limited to 1-15 pixels (first row).

The ridge extraction results shown here are after a thresholding operation, in which the ridge points that have medialness responses less than 15% of the globally maximum through scale-space are ignored. For the height ridge definition, an additional thresholding operation, based on the absolute value of the second greatest eigenvalue (i.e. the second derivative in the direction $\mathbf{e}_2$ of Fig. 2.10), was used. This value was required to be at least 10% of the greatest eigenvalue (the second derivative in the direction $\mathbf{e}_1$ of Fig. 2.10). This removed some of spurious ridges caused by the fluctuation of edge responses for the linear LoG and HMAT algorithms. In the results of the LoG (Figs. 4.25(a) and (g)) and HMAT (Figs. 4.25(b) and (h)), the remaining spurious ridges exist in parallel with boundaries. On the other hand, the results of the CMAT (Figs. 4.25(c) and (i)) contain no spurious ridge points, only the symmetric points.

The results of the ridge extraction for an MR image of a pair of legs are shown in Fig. 4.26. These ridges were obtained using both the optimal scale ridge and height ridge definitions. The operator radius was limited to 1-10 pixels. In the results of the LoG (Figs. 4.26(b) and (e)) and HMAT (Figs. 4.26(c) and (f)), spurious MMA branches appear in parallel with bones and fat layers. The CMAT results (Figs. 4.26(d) and (g)) give a more appropriate representation.

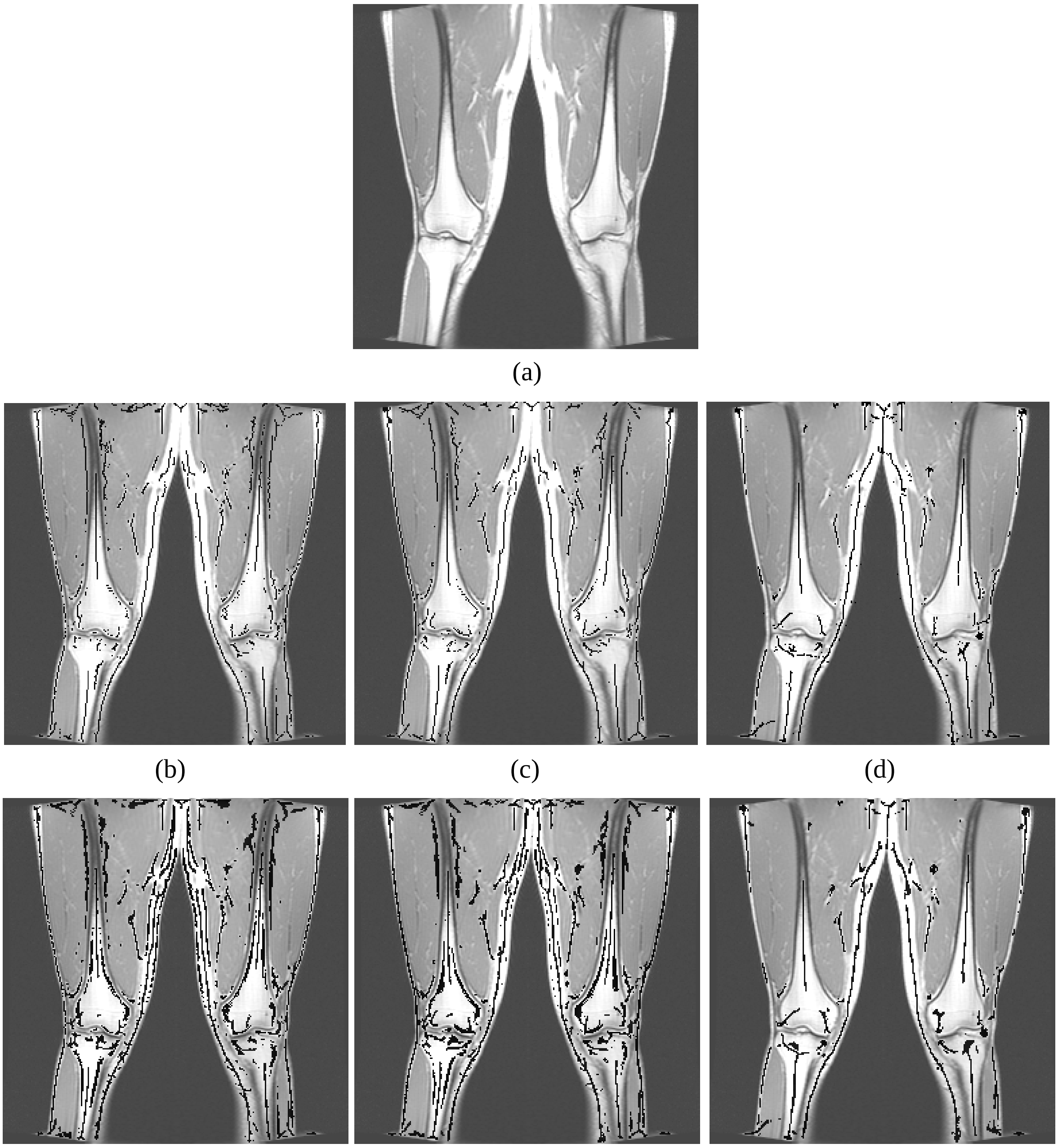


(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Fig. 4.25. The medialness ridges for a rectangle with a sawtooth edge: optimal scale ridges at operator radii of 1-15 pixels (top row) and 1-60 pixels (middle row), and height ridges at operator radii of 1-60 pixels (bottom row). The medialness was computed using the LoG (left column), HMAT (middle column), and CMAT(right column). The ridges are overlaid on the original image.



(a)

(b) (c) (d)

(e) (f) (g)

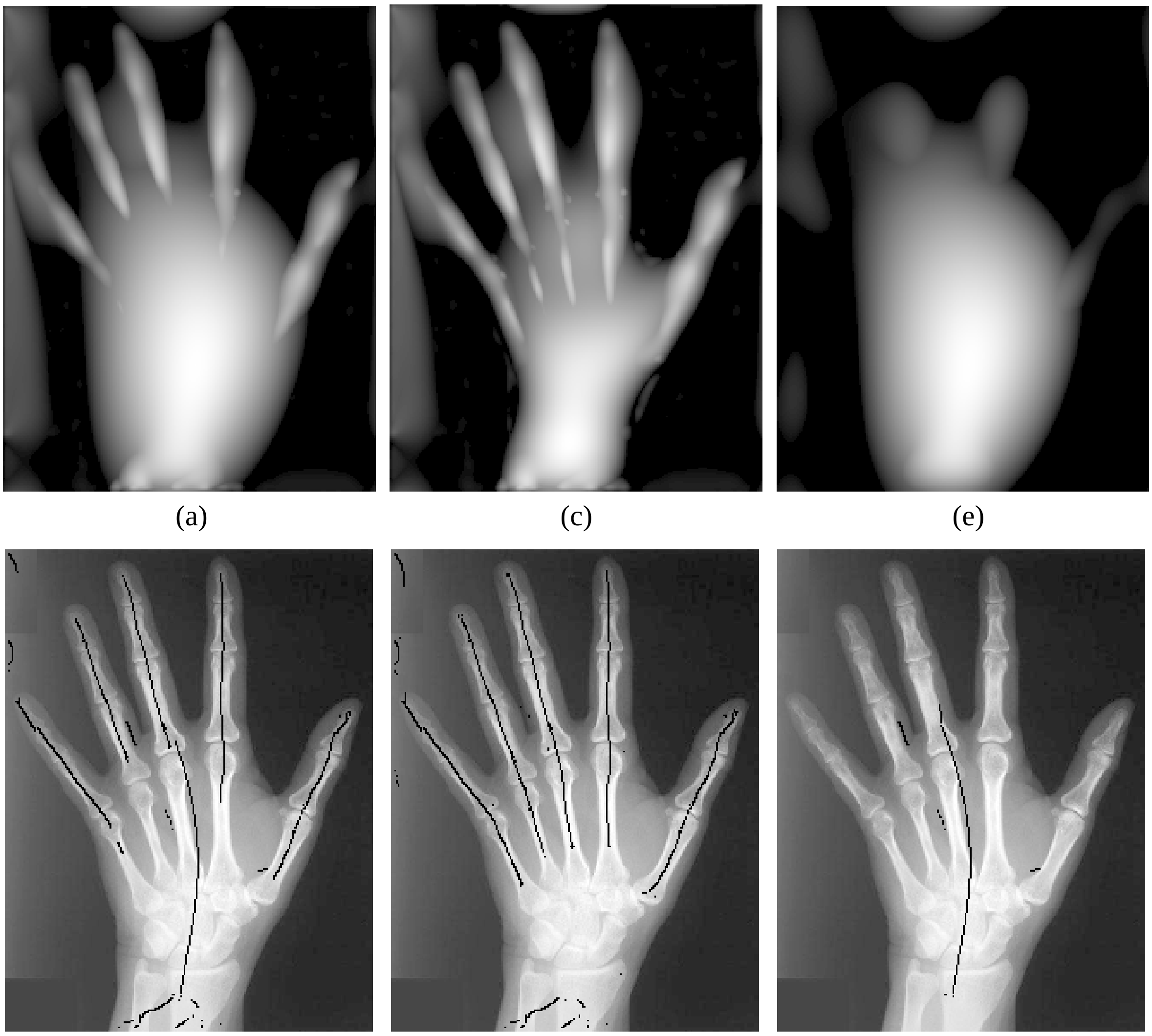
Fig. 4.26. The medialness ridges for an MR leg image: (top row) the original image, (middle row) optimal scale ridges and (bottom row) height ridges at operator radii of 1-10 pixels. The medialness was computed using the LoG (left column), HMAT (middle column), and CMAT(right column). The ridges are overlaid on the original image.

4.3 Sliding Window Ridges

4.3.1 Ridges of LoG Medialness Using Sliding Windows

To assess the significance of the sliding window algorithm for detecting locally optimal scale ridges, we have applied this algorithm to both synthetic and natural images. The results were also compared with those produced using other scale-space ridge definitions, such as the globally optimal scale ridge and height ridge. The LoG medialness operator, rather than the CMAT algorithm, was used initially in the demonstration of the formation and effect of locally optimal scale ridges (for the hand image in Figs. 4.27-4.30). This distinguishes the improvement in performance of the CMAT medialness algorithm from that due to the sliding window algorithm for ridge detection. The CMAT medialness algorithm was used throughout the following experiments (Figs. 4.31-4.33) to demonstrate of the improvement in performance due to the combination of both the algorithms. Each image is 256×256 pixels, except the hand image which is 216×282 pixels. For each image, only the MMA branches for bright objects on darker backgrounds are presented to simplify the visual interpretation of the results.

Fig. 4.27 shows the globally optimal scale responses (top) and ridges (bottom), computed for different ranges of optimal radius, for an X-ray hand image. The left column shows the results for a full radius range of 4-72 pixels. Note that the axis for the palm is complete, but the axes for the fingers are shortened (Fig. 4.27(b)). From the corresponding optimal scale response (Fig. 4.27(a)), it can be seen that the “off-axis” response at large scales, due to the palm, may be stronger than the “on-axis” response at small scales due to the fingers. Consequently, the medialness response of large dimension components dominates that of smaller components and the ridges for small objects adjacent to or embedded in a large object are



(a) (c) (e)

(b) (d) (f)

Fig. 4.27. The globally optimal scale responses (top) and ridges (bottom), under different ranges of operator radius, for an X-ray image of the hand: 4-72 pixels (left), 4-30 pixels (middle), and 31-72 pixels (right). On the bottom row the extracted ridges are shown overlaid on the original image.

shortened. This observation in a 2-D image is consistent with the behavior of the 1-D embedded pulse in scale-space presented in Section 3.3.2.2.

The interference between globally optimal scale responses at large and small scales can be avoided by separately computing responses at small and large scales. This is illustrated in the middle and right columns of Fig. 4.27, which are the globally optimal scale responses and ridges computed at small (4-30 pixels) and large (30-72 pixels) operator radii, respectively. At

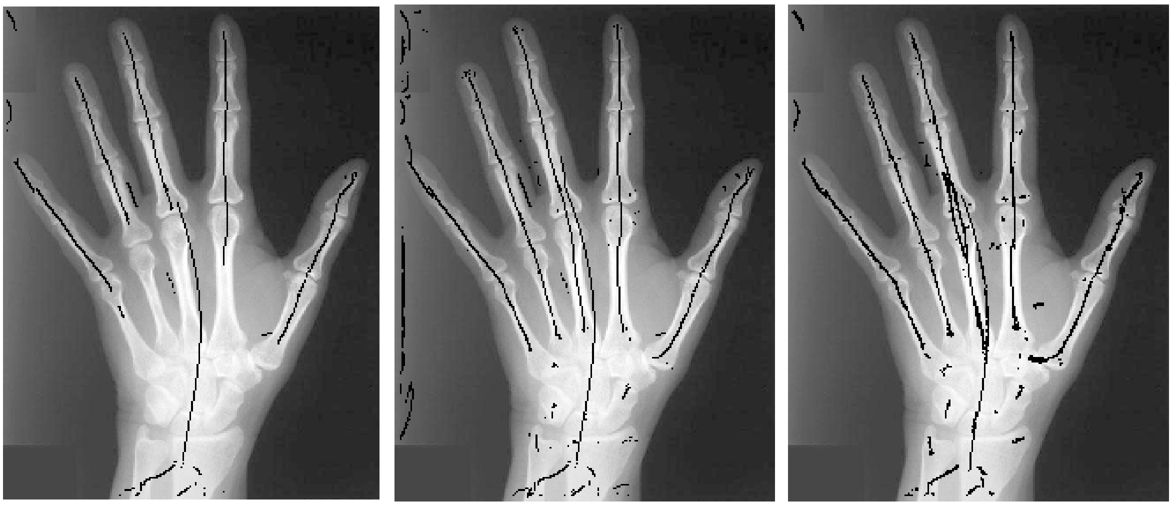

(a) (b) (c)

Fig. 4.28. Ridge points for the hand image, extracted using different definitions: (a) the globally optimal scale, (b) the locally optimal scale, and (c) height ridge definitions, at operator radii of of 4-72 pixels.

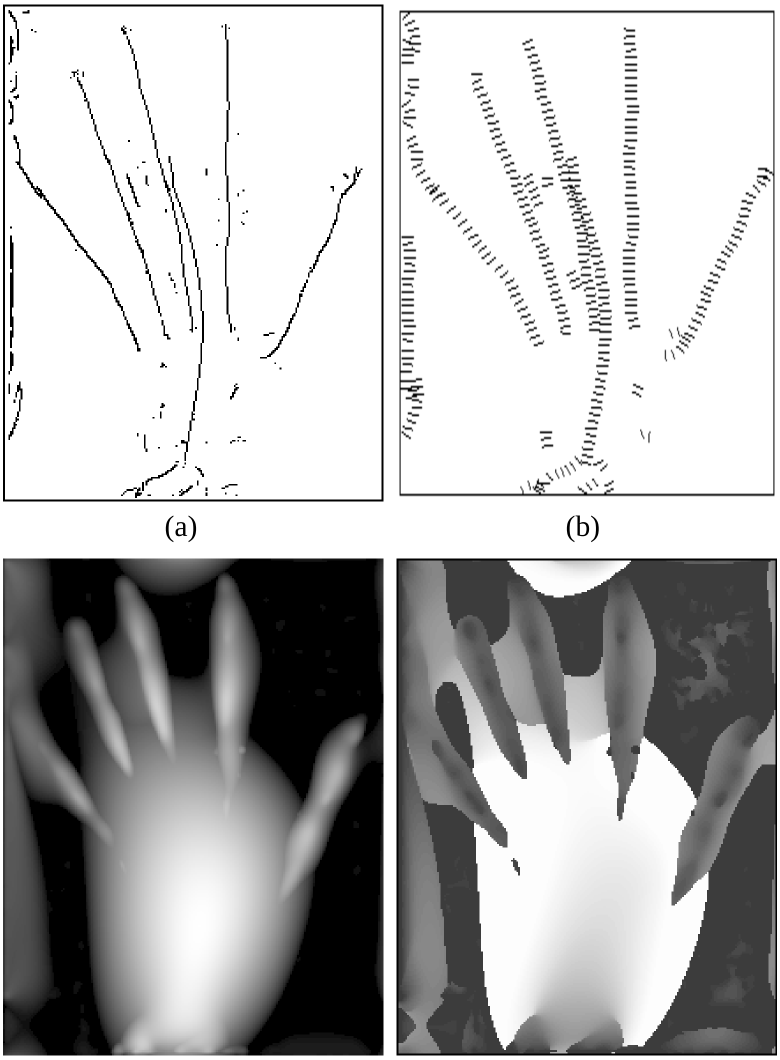
the operator radii of 4-30 pixels, Fig. 4.27(c), the optimal scale response for the fingers is stronger than that for the palm, and the axes for the fingers are fully extended along the fingers, see Fig. 4.27(d). There exist some ridges in the optimal scale response due to the palm (see the elongated highlight in the lower part of Fig. 4.27(c)). However, their corresponding “optimal scales” are at the largest scale (operator radius) of the scale range, 30 pixels, and they cannot be determined as maxima with respect to scale at this scale range. At the operator radii of 30-72 pixels, the optimal scale response for the fingers almost disappears (Fig. 4.27(e)), and only the axis for the palm is presented (Fig. 4.27(f)). Thus, the extraction of ridges for small and large object components is separated and there is little interference. The union of the two set of ridges, extracted at two segmented scale ranges, constitutes a complete MMA representation for the hand image.

The selection of the scale used above was, in this case, based on a judgement of image structure. However, such prior knowledge is not always available. In this case the sliding window algorithm is beneficial for complete ridge extraction.

The MMA representation of the hand image, obtained using the sliding window algorithm for locally optimal scale ridges, is shown in Fig. 4.28(b). It is almost the same as the union of the MMA branches extracted at two segmented scale ranges using the globally optimal scale ridge procedure, except that some short MMA segments emerge in the palm, see Fig. 4.27(d) and (f). Compared with the globally optimal scale ridges computed over a scale (operator radius) range of 4-72 pixels, Fig. 4.28(a), use of a sliding window results in complete axes for the fingers and a long axis for the palm. This is similar to the result achieved using the height ridge definition, shown in Fig. 4.28(c), even with respect to the short MMA segments that emerge in the palm. The use of too coarse a segmentation of the scale dimension, see Figs. 4.27(d) and (f), can cause MMA branches to be lost, as concluded in the analysis of scale distinguishibility, in Section 3.3.2.5. In Fig. 4.28, it is interesting to note that each finger and its bone does not produce separate MMA branches. This is because the widths of each finger and its bone are too close for scale-space analysis to distinguish them. The width ratio of the finger bones (embedded objects) to the fingers (outer objects) is much greater than the critical ratio of 1:4 identified in Section 3.3.2.2.

In Figs. 4.27 and 4.28, the ridge points are overlaid on the original image to show the spatial location of the MMA. This visualisation does not reflect the richness of the MMA representations that is shown in Fig. 4.29. Figs. 4.29(a)-(d) show the locations, normal orientations, medialness responses, and associated scales of the ridge points, respectively. The associated (optimal) scales are represented in gray levels, with a brighter region corresponding to a larger scale and object width. Using these attributes, the ridge points can be linked into MMA branches.

The linked MMA branches, longer than 10 pixels, for the hand image are shown overlaid on the original image in Fig. 4.30(a). A 3-D visualisation, in scale-space, of these axes is shown in Fig. 4.30(b). The MMA branch for the palm is at a higher level of scale than those of the fingers, reflecting the fact that the palm is wider than the fingers. In computing the medialness response, the operator kernel can extend beyond the limits of the image. To deal with the effects of a limited field of view, each image in the experiments is treated as periodic. This can cause false wraparound symmetries between actual boundaries that lie on opposite sides of the original image. For example, the left border in Fig. 4.30(a) is brighter than the right border, which can be interpreted as "ridges" and the MMA under the wraparound of the image.



(a) (b)

(c) (d)

Fig. 4.29. The richness of ridge point attributes: (a) the ridge points, (b) the normals to the ridges, (c) the medialness response at the ridges (optimal scale response), and (d) the scales (object widths) associated with the ridges (the optimal scales). In (d) a brighter region corresponds to a larger scale and object width.

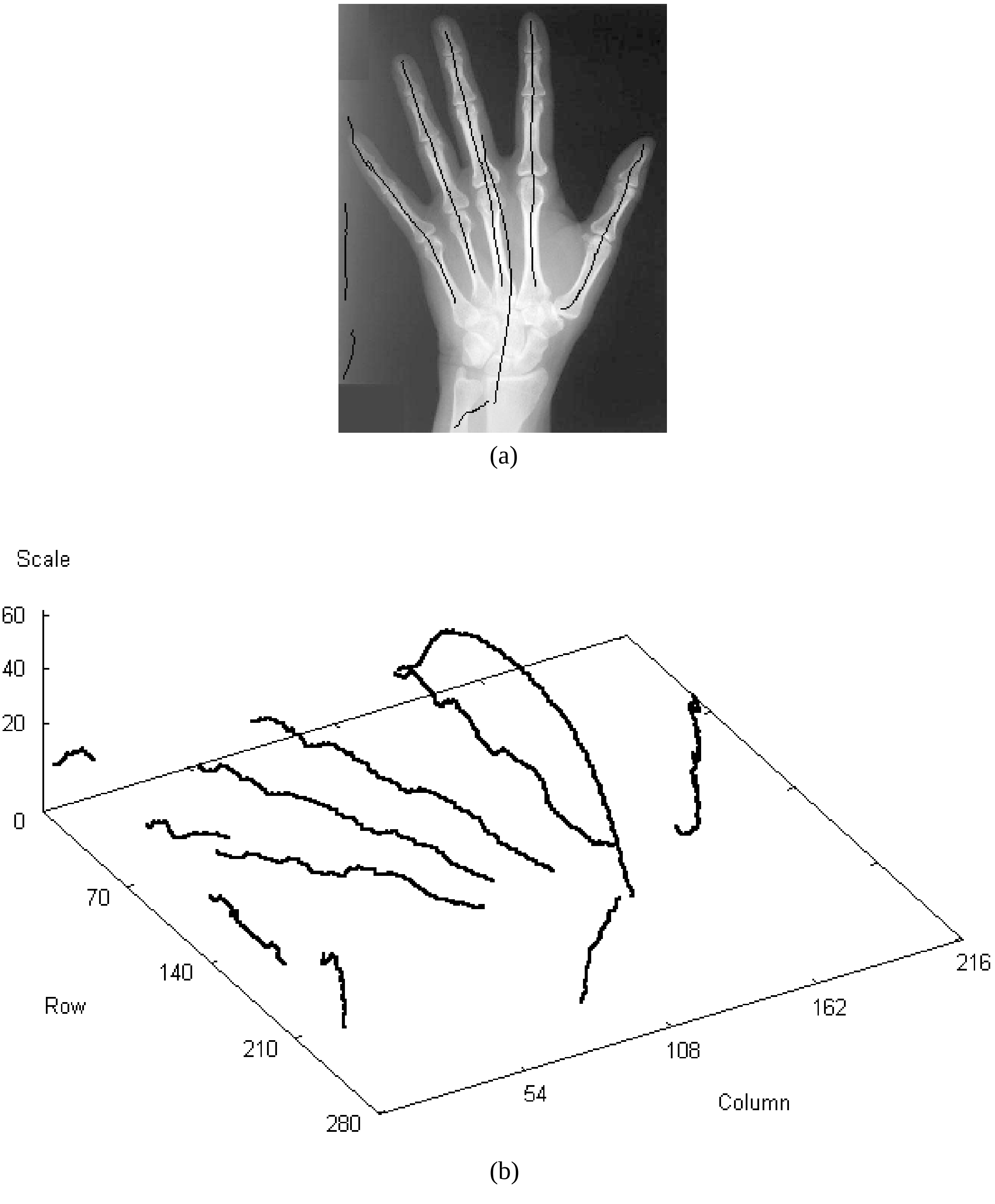


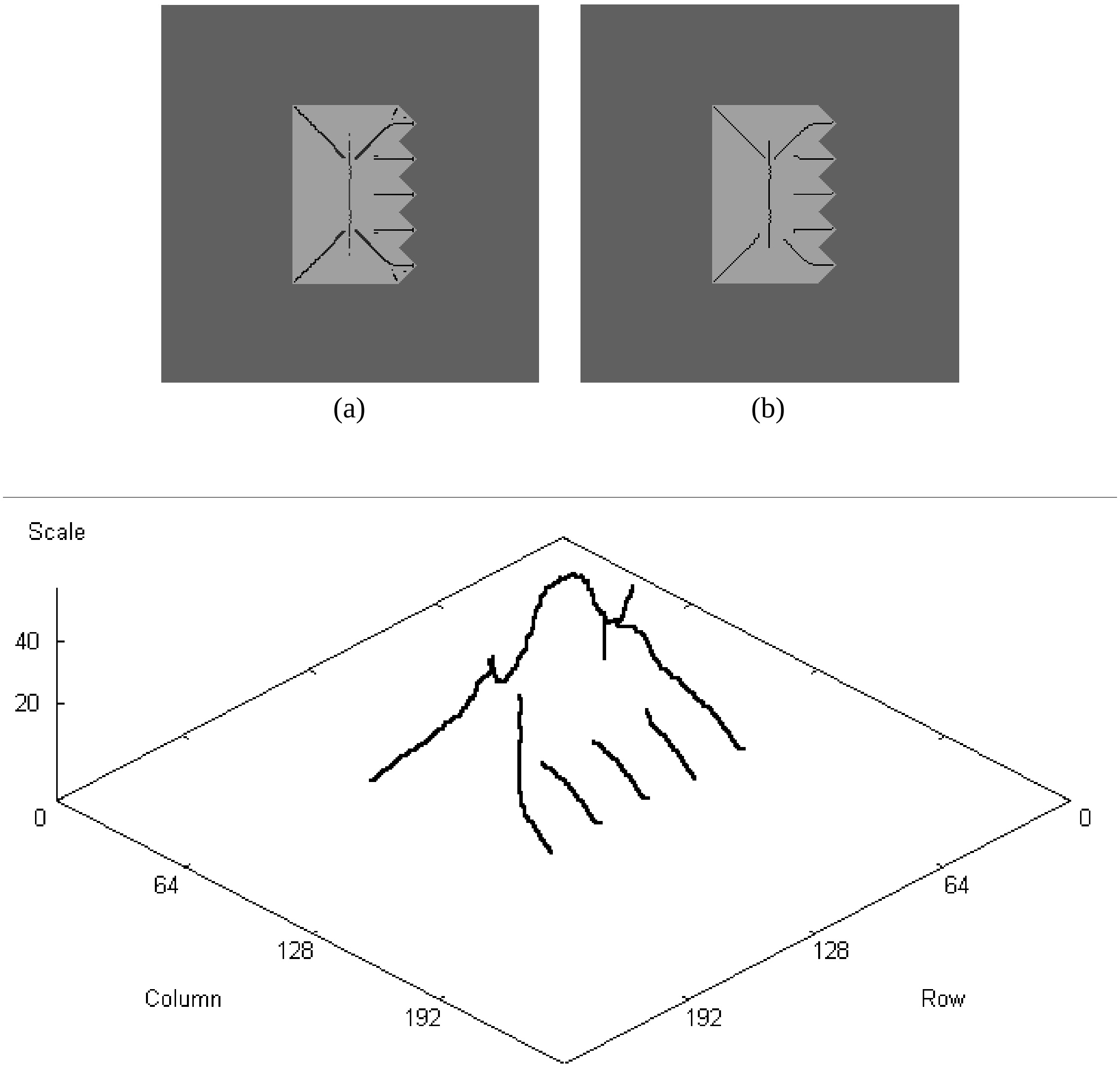
Fig. 4.30. Medial axes (longer than 10 pixels), obtained using the LoG and sliding window algorithm, for an X-ray hand image: (a) axes superimposed on original image, (b) 3-D visualization.

4.3.2 Ridges of CMAT Medialness Using Sliding Windows

In Fig. 4.31(b) the linked MMA branches, longer than 5 pixels, for a rectangle with one “sawtooth” edge are shown overlaid on the original image. At small scales, there are five short, disjoint MMA branches representing the “sawteeth” of the object. The rectangular property is also represented by the central branch and the four diagonal branches bisecting the corners. The representation of the coarse scale properties, such as the rectangle, is little influenced by those of fine details such as the “sawteeth”. A 3-D visualisation, in scale-space, of these axes is shown in Fig. 4.31(c).

In Fig. 4.32(b) the linked MMA branches, longer than 5 pixels, for a teardrop shape are shown overlaid on the original image. A 3-D visualisation, in scale-space, of these axes is shown in Fig. 4.32(c). Note that the scale of the axis appears to increase linearly along the bisector of the shape, except near the large radius end of the teardrop.

In Fig. 4.33(b) the linked MMA branches, longer than 25 pixels, for the MR leg image are shown overlaid on the original image. A 3-D view, in scale-space, of these axes are shown in Fig. 4.33(c). Note that each MMA branch for a bone structure increases in scale near the joints between the upper and lower legs, reflecting the widening of local structures. The MMA branches change in scale as the width of the fat layers changes.

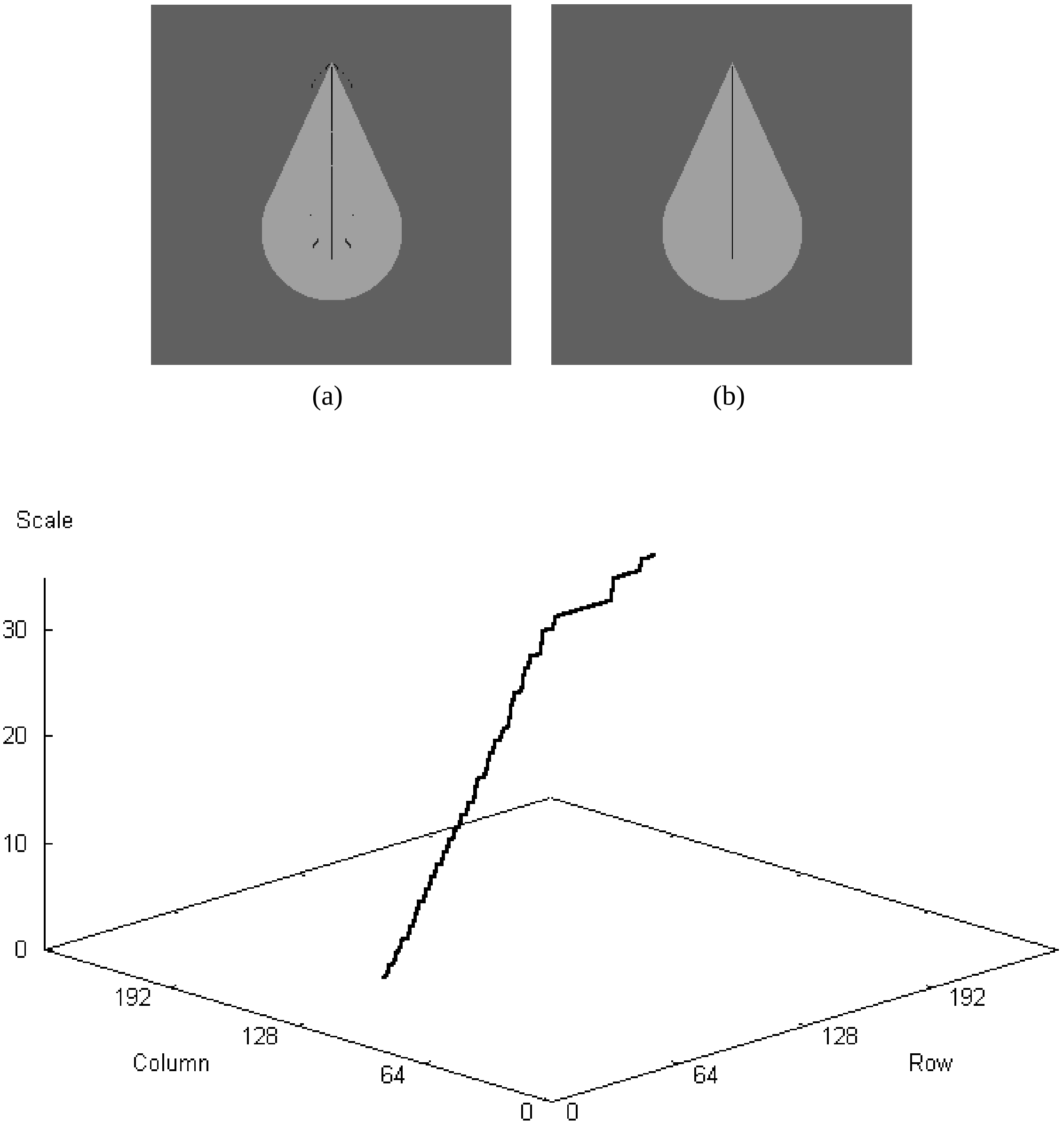


(a)

(b)

(c)

Fig. 4.31. Medial axes, obtained using the CMAT and sliding window algorithm, for a rectangle with a sawtooth edge: (a) the original image overlaid with axes before linking and (b) after linking (longer than 5 pixels), and (c) 3-D visualization.

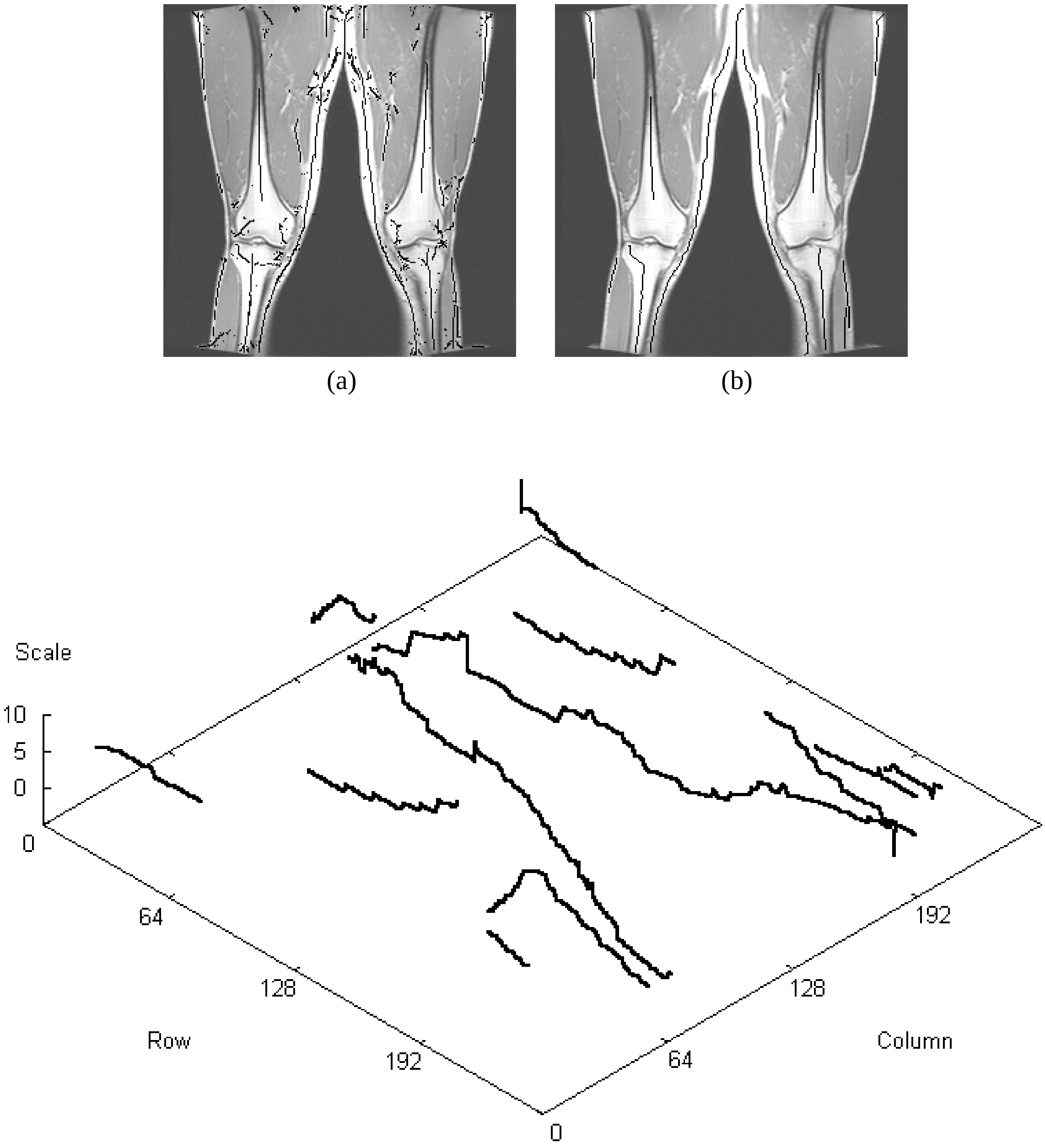


(a)

(b)

(c)

Fig. 4.32. Medial axes, obtained using the CMAT and sliding window algorithm, for a teardrop shape: (a) the original image overlaid with axes before linking and (b) after linking (longer than 5 pixels), and (c) 3-D visualization.


(a) (b)

(c)

Fig. 4.33. Medial axes, obtained using the CMAT and sliding window algorithm, for an MR image of a pair of legs: (a) the original image overlaid with axes before linking and (b) after linking (longer than 25 pixels), and (c) 3-D visualization.

4.3.3 Computational Cost of Sliding Windows

Suppose that O is a functional expression of the order of the complexity, and that S_L and S_H are the smallest and largest scales being considered. The computational cost for computing globally optimal scale ridges is $O(N^2)$; that for computing locally optimal scale ridges and height ridges is $O(S_H - S_L, N^2)$. We have tested the three algorithms for multi-scale ridge extraction on a 128×128 image, over a scale range of 1-30 pixels, using a SUN Ultra-5/10 workstation. The computation of the LoG medialness scale-space takes 115 seconds. The time to extract ridges of the medialness response and the MMAs, for each scale-space ridge algorithm, is provided in Table 4.2.

Algorithms	Globally Optimal Scale	Locally Optimal Scale	Height
Ridge Extraction (seconds)	0.2	5.0	10.0
MMA Total (seconds)	115.2	120.0	125

Table 4.2. Time taken to compute ridges and the MMAs for several scale-space ridge algorithms.

Therefore, the sliding window algorithm is still efficient, compared with the large portion of time used in the computation of medialness responses.

Chapter 5 Discussion

5.1 The CMAT with Other Edge-Free Medialness Algorithms

5.1.1 Credit Attribution

The credit attribution (CA) algorithm is an iterative process to enhance medialness responses for symmetry and suppress edge responses. It assumes that each boundary point makes a constant contribution to the medialness response on both sides of the boundary, as shown in Fig. 5.1. For each boundary point, the contribution is weighted by the ratio of the medialness, in the previous iteration, to the total medialness on both sides. When a boundary contribution interacts with those from other boundaries it is enhanced, and the contribution to medialness will iteratively grows on one (interacted) side of the boundary with respect to the other side, as shown in Fig. 5.1.

The CMAT is similar to the credit attribution in that both use an additional voting process, in which boundary contribution is weighted according to the result of the previous voting phase. However, the credit attribution is iterative and therefore computationally expensive. In the credit attribution, the suppression of the edge responses on one side of a boundary depends on interaction with other boundary contributions. For an isolated edge (or a symmetric object examined at a small scale, compared to the object width), there is little or no interaction between the boundary contributions. Therefore, the medialness response due to an isolated edge is strong, see Fig. 5.1(a). In addition, the credit attribution cannot be used when only one polarity of boundary transition is considered, such as looking only for a bright object on darker background as in for X-ray images. In this case all the contributions of a boundary point are cast on one side of the boundary. Therefore there is no enhancement of the responses for

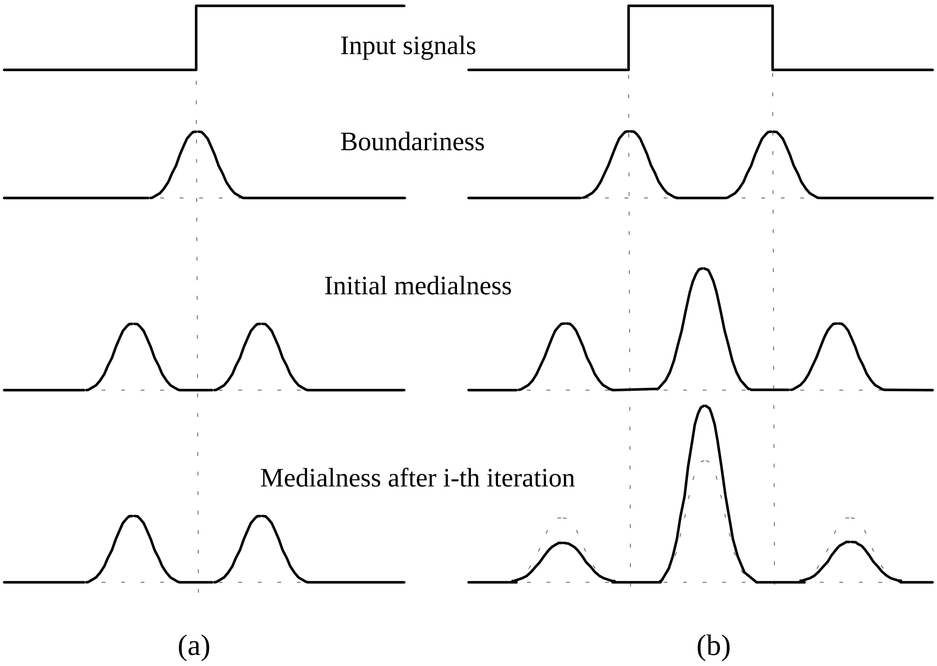


Fig. 5.1. The competition for boundariness contributions between the two sides of each edge, in the presence of (a) an isolated edge and (b) a pair of symmetric edges.

symmetry and no suppression of edge responses.

5.1.2 Global Ridge Strength Measure

The global ridge strength measure (GRSM) can differentiate the HMAT medialness responses for symmetric and isolated edges. It is defined as the ratio of the total medialness at a point to the maximum boundariness that contributes to it. The CMAT is similar to the GRSM in using the ratio between boundariness and medialness responses. However, the contribution confidence of the CMAT is attributed to each boundary point while the GRSM is defined for each medial point. Also contribution confidence is combined into the final medialness response while the GRSM is treated as an independent factor of the HMAT medialness response. These differences lead to the distinctive behaviour of the CMAT. First, the GRSM is an enhancing process that differentiates between structural symmetries and edges, while the CMAT inhibits the medialness response of single edges. Secondly, the GRSM is intensity invariant but the CMAT is not. The invariance to linear transforms of intensity can be argued to be important but ignores the difference between a high-contrast object with strong boundariness and a low-contrast object (possibly noise or computational errors) with weak boundariness. Intensity information is important in identifying medial axes, and it is a mistake to discard such information. The aim of intensity invariance can be partly reached by an adaptive MMA branch selection, e.g. selecting “the ten most salient axes” as in [LIN98] or “the axes whose saliency measures are above 20% of the greatest one”.

5.1.3 Koller Line Detector

The Koller line detector (KLD) belongs to another approach to the suppression of edge responses in medialness [WAN82] [SUB93][KOL95]. For each potential medial point, a pair of boundariness contributions (shifted edge responses), R_L and R_R, are computed in a direction that maximises the second derivative of the blurred intensity. To suppress the medialness response to a single edge, Koller noted that the final response must be large, if both R_L and R_R are large and zero, if either R_L or R_R is zero. Therefore the minimum operation, $min\{R_L, R_R\}$ is used and the geometric mean, $\sqrt{R_L R_R}$, is also considered [KOL95].

The CMAT is similar to the KLD in that both prevent edge responses from contaminating medialness. In the case of 1-D pulses or 2-D line structures, the CMAT is also computed as the non-linear combination of a pair of boundariness contributions using $2R_L R_R/(R_L + R_R)$. However in the KLD, the assumption that a pair of boundary contributions is steered to one direction implies that this method is designed for the detection of line structures, rather than for general objects which may be a wedge, bent and branched (multiple boundary points involved). In addition, even for a 1-D pulse or 2-D line structures, the KLD is found to be easily biased by disturbances of the symmetry, unlike the CMAT, LoG, and HMAT.

5.2 Sliding Window Algorithm with Other Ridge Definitions

5.2.1 Globally Optimal Scale Ridges

The sliding window algorithm is an efficient way to implement the original definition of the optimal scale ridge [PIZ92][FRI92]. The closest related method is Fritsch's simplified strategy, which uses globally optimal scale ridges to approximate optimal scale ridges. At each pixel, this simplified strategy projects the global medialness maximum over the full scale range onto the image plane and the ridge search is conducted over this single image. The sliding window algorithm is similar to the simplified strategy in that it extracts globally optimal scale ridges within each sliding window on scale dimension. If the length of the sliding window, L, is set to S_H-S_L (where S_H and S_L are the largest and smallest scales being considered), then the sliding window algorithm becomes Fritsch's simplified strategy. The sliding window algorithm considers a small scale segment at each iteration, rather than the whole scale range as in Fritsch's method. The analyses in Sections 3.3.2.5 and 3.3.3.1 showed that such a small scale segment has an improved ability to distinguish embedded objects in scale-space by separating the ridge detection for small and large objects. Therefore, the sliding window algorithm can be used to extract locally optimal scale ridges and allows each point to belong to MMA segments at different scales, as required for an embedded object.

5.2.2 Optimal Position Ridges

A direct extension of the optimal position ridge definition from 1-D signals to 2-D images has been proposed and implemented by Lindeberg [LIN98]. This method extracts spatial ridges of 2-D medialness response at each scale. The scale-space ridges are those 2-D spatial ridge points that are local maxima with respect to scale. The sliding window algorithm is similar to Lindeberg's optimal position ridge definition in that it also extracts spatial ridge points from 2-D medialness responses and selects those that are local maxima over scale. If we set the length of each sliding window, *L*, to 1, then the sliding window algorithm becomes Lindeberg's optimal position ridge definition. However, the medialness response used in the sliding window algorithm is the projection of the global maxima over each sliding window onto the image plane, rather than the medialness response at a single scale. This difference of the sliding window algorithm is advantageous in avoiding spurious MMA branches caused by the medialness response for isolated edges. The medialness response for an isolated edge, resulting from a traditional operator, is a inclined fuzzy sheet in scale-space (Fig. 2.8(a)). Each point on the core of this fuzzy sheet is a ridge point in the medialness at a single scale and tends to be a local maximum over scale. Therefore the medialness responses for isolated edges satisfy the optimal position ridge definition. On the other hand, due to the size invariance of the medialness response for a step edge (See Section 3.1.2), the projection of the maxima over scale onto the image plane, over a scale range, tends to be flat and does not readily produce spurious ridges.

5.2.3 Height Ridges

In the height ridge method a ridge point is defined as a local maximum of the underlying function in the direction that (negatively) maximises the second directional derivative [HAR83]. This definition has been extended to 3-D space [EBE94a] and 2+1 dimensional scale-space [MOR94]. The height ridge definition is also used in the sliding window algorithm to extract spatial ridges of the 2-D optimal scale response with Euclidean geometry. However, this differs from direct ridge extraction in a 2+1 dimensional scale-space using a Riemanian geometry, which is required by the incommensurability between spatial and scale dimensions.

Chapter 6 Conclusions and Future Work

6.1 Conclusions

The conclusions of this thesis are:

(1) CMAT medialness computation

- An edge-free medialness function (CMAT) can be computed by adding a second stage of boundariness accumulation, in which the weight for each boundary point is a function of the ratio between the boundariness and total medialness at the first stage.
- The CMAT medialness obtained as above increases with the increasing sum and decreasing sample variance of the contributing boundariness responses. Therefore, it prohibits isolated boundaries from generating "medialness" responses.
- The CMAT medialness depends more on the sum of contributing boundariness, while its rate of increase is determined by the sample variance of the contributing boundary responses.
- Alternative definitions for boundariness weights emphasizes or suppresses the CMAT medialness at end points with respect to that along middle of parallel boundaries.

(2) The sliding window algorithm for scale-space ridge extraction

- The sliding window algorithm can efficiently extract locally optimal scale ridges and thus detect a complete set of MMA branches for assemblies of objects, e.g. embedded or adjacent objects.

(3) Comparison of medialness functions

- The localization accuracy for the position and width of an object, as well as the robustness under noisy conditions, in the CMAT is as good as that for other medialness functions.

- Compared with linear medialness operators, the CMAT response is more selective to the position and width of an object, which facilitates maximum (or ridge) detection.
- For an asymmetric object, the trace of the CMAT medialness maxima (ridges) in scale-space has no structural change and is shifted less from the object centre than for other medialness functions. Thus, the CMAT algorithm has better spatial localization at any single scale.

(4) The behaviour of embedded objects in scale-space

- The ability of scale-space analysis to discern objects in an embedded grouping is limited. It can only identify embedded objects with at most 1/4 the width of the outer object, and the outer objects with at least 1/3 the height contrast of the embedded object.
- The position and width of the embedded object can be correctly identified. For the outer object, its position is biased to the embedded object and its width is often underestimated.

6.2 Future Work

(1) Detection of ridges in the scale direction

There are occasionally some places in medialness scale-space where ridges running almost in the direction of scale axis, e.g. at branching points and centres of round “blobs”. The medialness response along such a ridge is smooth over scale, but appears as a local spatial maximum (rather than a ridge) in the medialness response at a single scale or even the optimal scale response over multiple scales. The scale-space ridge definitions, which consider the spatial and scale dimensions separately, such as the optimal scale and Lindeberg’s ridge definitions, may lose MMA at these places. A potential solution to this problem is to consider these points as ridge points. The attention of focus would be on those points whose medialness responses are local maxima at each of successive scales and the maximum values are smooth over scale. This approach is, in spirit, similar to Crowley's scheme of linking peaks [CRO84].

(2) Extensions of the sliding window algorithm to 3-D images

Because the sliding window algorithm can readily separate the scale dimension from the search space, it simplifies ridge detection in higher dimensional spaces. For a 3-D image, the medialness scale-space is 4 dimensional. We can first compute the 3-D maximal response over each sliding window in the scale dimension. The points, which are both 2-D spatial ridges (cores of fuzzy sheets) in the 3-D maximal response and local maxima over scale, are determined as on 2-D medial surfaces. The points, which are both 1-D spatial ridges in the 3-D maximal response and local maxima over scale, are determined as on 1-D medial axes. The 2-D medial surface can be used to identify the skulls in 3-D CT images. The medial axes can be used to identify blood vessels in 3-D MR angiogram images.

(3) Connectors in sliding window algorithm

Connector curves [DAM99] can fill in the gaps between ridges (and the MMA). The combination of the MMA with connector curves provides a more robust representation for image structures. Because extracting ridges (the MMA) from an image directly applies to the extraction of connectors, connector curves are an inexpensive and useful addition to the set of ridges. However, the current definition of connector curves is based on the eigenvalue analysis of the Hessian of second derivatives (i.e. maximal convexity ridges), it cannot be directly applied to the optimal scale ridge definition including sliding window algorithm. Considering the simplicity of the sliding window algorithm, the extension of connector curves into the sliding window algorithm has promising applications.

(4) Sparse scale sampling in line detection

In Section 4.1.4.4, it was found that, in 1-D, the optimal position maxima of the medialness response, computed using the KLD geometric mean operator (s=2σ), give the best estimation of the pulse centre at a single scale. For 2-D images, this means that the ridges of the medialness at a single scale approximate the medial axes. This has a useful application in decreasing the scale sampling rate. If the width of the lines of interest has a limited range (e.g. roads in remote-sensing images or blood vessels in medical images), it would be possible to locate the line centres, with negligible error, by computing the KLD medialness response at only one scale. This is beneficial in practice, because computing medialness scale-space at a full scale range is time-consuming.

References

[AHU78] N. Ahuja, L. S. Davis, D. L. Milgram, and A. Rosenfeld, “Piecewise approximation of pictures using maximal neighborhoods”, *IEEE Trans. on Computers*, 27:375-379, 1978.

[ARC78] C. Arcelli and G. Sanniti di Baja, “On the sequential approach to medial line transformation”, *IEEE Trans. on System., Man, and Cybernetics.*, 8:139-144, 1978.

[ARC89] C. Arcelli and G. Sanniti di Baja, “A one-pass two-operations process to detect the skeletal pixels on the 4-distance transform”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 11:411-414, 1989.

[ARC95] C. Arcelli and G. Ramella, “Finding grey-skeletons by iterated pixel removal”, *Image and Vision Computing*, 13(3):159-167, 1995.

[BAB86] J. Babaud, A. P. Witkin, M. Baudin and R. O. Duda, “Uniqueness of the Gaussian kernel for scale-space filtering”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8(1):26-33, 1986.

[BER87] F. Bergholm, “Edge focusing”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 9(6):726-741, 1987.

[BHA99] A. A. Bharath and C. J. Huberson, “Obtaining medial responses from steerable filters”, *IEE Proc. Vision, Image and Signal Processing*, 146(5):286-292, 1999.

[BLU67] H. Blum, “A transformation for extracting new descriptors of shape”, in *Models for the perception of speech and vision forms*, W. Wathen-Dunn (Ed.), MIT Press, pp. 362-380, 1967.

[BLU73] H. Blum, “Biological shape and visual science (Part I)”, *J. Theor. Biol.*, 38:205-287, 1973.

[BLU78] H. Blum and R. N. Nagel, “Shape description using weighted symmetric axis features”, *Pattern Recognition*, 10:167-180, 1978.

[BOO79] F. L. Bookstein, “The line-skeleton”, *Computer Vision, Graphics, and Image Processing*, 11:123-137, 1979.

[BOR98] G. Borgefors, G. Ramella, and G. Sanniti di Baja, “Coarse-to-fine skeletons from grey-level pyramids", in *Proc. Int. Conf. on Pattern Recognition.*, pp. 400-402, 1998.

[BRA84] M. Brady and H. Asada, "Smoothed local symmetries and their implementation", *Int. J. Robotics Research,* 3:36-61, 1984.

[BRA92] J. W. Brandit and V. R. Algazi, "Continuous skeleton computation by Voronoi diagram", *CVGIP: Image Understanding*, 55:329-338, 1992.

[CAN86] J. Canny, "A computational approach to edge detection", *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8:679-697, 1986.

[CHI87] R. T. Chin, H. K. Wan, D. L. Stover, and R. D. Iverson, "A one-pass thinning algorithm and its parallel implementation", *Computer Vision, Graphics, and Image Processing,* 40: 30-40, 1987.

[CRO84] J. L. Crowley and A. C. Parker, "A representation for shape based on peaks and ridges in the difference of low-pass transform", *IEEE Trans. Pattern Analysis and Machine Intelligence*, 6(2):156-169, 1984.

[DAM99] J. Damon, "Properties of ridges and cores for two-dimensional images", *J. Mathematical Imaging and Vision*, 10:163-174, 1999.

[DAN80] P. E. Danielsson, "Euclidean distance mapping", *Computer Vision, Graphics and Image Processing*, 14:227-248, 1980.

[DIL87] A. R. Dill, M. D. Levine and P. B. Noble, "Multiple resolution skeletons", *IEEE Trans. Pattern Analysis and Machine Intelligence*, 9:495-504, 1987.

[EBE94a] D. Eberly, R. B. Gardner, B. S. Morse, S. M. Pizer, and C. Scharlach, "Ridges for image analysis*", J. Mathematical Imaging and Vision.*, 4(4):351-371, 1994.

[EBE94b] D. Eberly, "A differential geometric approach to anisotropic diffusion", *in Geometry-Driven Diffusion in Computer Vision*, B. ter Haar Romeny (ed.), Kluwer, Amsterdam, pp. 371-391,1994.

[FEL97] J. Feldmar, G. Malandain, and N. Ayache, "Matching 3D MR angiography data and 2D X-ray angiograms", *Lecture Notes in Computer Science*, 1205:129-138, 1997.

[FID96] M. Fidrich, "Iso-surface extraction in 4D with applications related to scale space", Research Report 2833, INRIA, 1996.

[FLO92] L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink and M. A. Viergever, “Scale and the differential structure of images”, *Image and Vision Computing,* 10(6):376-388, 1992.

[FRE91] W. T. Freeman and E. H. Adelson, “The design and use of steerable filters”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13(9):891-906, 1991.

[FRI92] D. S. Fritsch, “Medial description of gray-scale image structure by gradient-limited diffusion”, *Visualization in Biomedical Computing, SPIE,* 1808:105-117, 1992.

[FRI93] D. S. Fritsch, “Registration of radiotherapy images using multiscale medial descriptions of image structure”, *Ph. D. Dissertation*, Dept. of Biomedical Eng., Univ. of North Carolina at Chapel Hill, 1993.

[FRI95] D. Fritsch, D. Eberly, and S. M. Pizer, “Stimulated cores and medical applications*”, Proc. Information Processing in Medical Imaging (IPMI’95),* pp. 365-368, 1995.

[FUR96a] J. D. Furst, R. S. Keller, J. E. Miller, and S. M. Pizer, “Image loci are ridges in geometric spaces”, Technical Report TR96-046, University of North Carolina Computer Science Department, 1996.

[FUR96b] J. D. Furst, S. M. Pizer, and D. Eberly, “Marching cores: a method for extracting cores from 3D medical images”, Technical Report TR96-003, Dept. Computer Science, Univ. North Carolina, 1996.

[FUR98] J. D. Furst, S. M.Pizer, “Marching optimal-parameter ridges: an algorithm to extract shape loci in 3D images”, *Lecture Notes in Computer Science*, 1496:780-787, 1998.

[GAG86] M. Gage and R. S. Hamilton, “The heat equation shrinking convex plane curve*”, J. Differential Geometry*, 23:69-96, 1986.

[GAU89] J. M. Gauch, “A multiresolution intensity axis of symmetry and its application to image segmentation”, Technical Report TR89-047, Department of Computer Science, University of North Carolina at Chapel Hill, 1989.

[GAU93] J. M. Gauch and S. M. Pizer, “Multiresolution analysis of ridges and valleys in grey-scale images”, *IEEE Trans. Pattern Analysis and Machine Intelligence,* 15(6):635-646, 1993.

[GEL66] I. M. Gel'fand, M. I. Graev, and N. Y. Vilenkin, *Generalized Functions*, Vol. 5, Academic Press, New York, 1966.

[GRI91] L. D. Griffin, A. C. F. Colchester, and G. P. Robinson, “Scale and segmentation of gray-level images using maximum gradient paths”, *Lecture Notes in Computer Science*, 511: 256-272, 1991.

[GON92] R. C. Gonzalez and P. Wintz, *Digital Image Processing*, Addison-Wesley, 1992.

[GUO89] Z. Guo and R. W. Hall, “Parallel thinning with two-subiteration algorithms”, *Communications of ACM*, 32:359-373, 1089.

[HAR83] R. Haralick, “Ridges and valleys in digital images”, *Computer Vision, Graphics, Image Processing*, 22:28-38, 1983.

[HAR92] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*, Volume I, Addison-Wesley, Reading, MA, 1992.

[HIL69] C. J. Hilditch, “Linear skeletons from square cupboards”, *Machine Intelligence*, 4:403-420, 1969.

[HO86] S. Ho and C. R. Dyer, “Shape smoothing using medial axis properties”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8:512-520, 1986.

[HUE86] A. Huertas and G. Medioni, “Detection of intensity changes with subpixel accuracy using Laplacian-Gaussian masks”, *IEEE Trans. Pattern Analysis and Machine Intelligence,* 8(5): 651-664, 1986.

[KIM95] B. Kimia, A. R. Tannenbaum, and S. W. Zucker, “Shapes, shocks, and deformations I: The components of two-dimensional shape and the reaction-diffusion space”, *Int. J. Computer Vision*, 15:189-224, 1995.

[KLE87] F. Klein and O. Kubler, “Euclidean distance transformation and model guided image interpretation”, *Pattern Recognition letters*, 5:19-30, 1987.

[KOE84] J. J. Koenderink, “The structure of images”, *Biological Cybernetics*, 50:363-370, 1984.

[KOE86] J. J. Koenderink and A. J. Doorn, “Dynamic shape”, *Biological Cybernetics*, 53:383-396, 1986.

[KOE91] J. J. Koenderink, *Solid Shape*, MIT Press, Cambridge, MA, 1991.

[KOE94] J. J. Koenderink and A. J. Doorn, “Two-plus-one-dimensional differential geometry”, *Pattern Recognition Letters*, 15:439-443, 1994.

[KOL95] T. M. Koller, G. Gerig, G. Szekely, and D. Dettwiler, “Multiscale detection of curvilinear structures in 2-D and 3-D image data”, in *Proc. Int. Conf. on Computer Vision*, pp. 864-869, 1995.

[KRI98] K. Krissian, G. Malandain, and N. Ayache, “Model based multiscale detection and reconstruction of 3D vessels”, Research Report 3442, INRIA Sophia Antipolis, 1998.

[LEE82] D. T. Lee, “Medial axis transformation of a planar shape”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 4:363-369, 1982.

[LEY87] M. Leyton, “Symmetry-curvature duality”, *Computer Vision, Graphics, and Image Processing*, 38:327-341, 1987.

[LIN92] T. Lindeberg, “Scale-space behaviour of local extrema and blobs”, *J. Mathematical Imaging and Vision*, 1(1):65-99, 1992.

[LIN94] T. Lindeberg, *Scale-space Theory in Computer Vision*, Kluwer Academic Publisher, Dordrecht, Netherlands, 1994.

[LIN98] T. Lindeberg, “Edge detection and ridge detection with automatic scale selection”, *Int. J. Computer Vision*, 30(2), 1998.

[LIU94] A. Liu, S. M. Pizer, D. Eberly, B. S. Morse, J. Rosenman and V. Carrsaco, “Volume registration using the 3D core”, submitted to *Visualization in Biomedical Computing’94*, 1994.

[LOR87] W. E. Lrenson and H. E. Cline, “Marching cubes: a high resolution 3D surface construction algorithm”, *Computer Graphics*, 21(4):163-169, 1987.

[MAI96] J. B. A. Maintz, P. A. van den Elsen, and M. A. Viergever, “Evaluation of ridge seeking operators for multimodality medical image matching”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18(4):353-364, 1996.

[MAR80] D. C. Marr and E. C. Hildreth, “Theory of edge detection”, *Proc. Royal Society London B*, 207:187-217, 1980.

[MOK86] F. Mokhtarian and A. Mackworth, “Scale-based description of planar curves and two-dimensional shapes”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8:34-43, 1986.

[MON68] U. Montanari, “A method for obtaining skeletons using a quasi-Euclidean distance”, *J. of ACM*, 15:600-624, 1968.

[MON69] U. Montanari, “Continuous skeletons from digitized images”, *J. of ACM*, 16:534-549, 1969.

[MON95] O. Monga, N. Armande, and P. Montesinos, “Thin nets and crest lines: application to satellite data and medical images”, Research Report 2480, INRIA, 1995.

[MOR91] B. S. Morse, S. M. Pizer and C. A. Burbeck, “A Hough-like medial axis response function”, Technical Report *TR91-044*, Dept. of Comp. Sci., Univ. of North Carolina at Chapel Hill, 1991.

[MOR93] B. S. Morse, S. M. Pizer and A. Liu, “Multiscale medial analysis of medical images”, *Lecture Notes in Computer Science*, 687:112-131, 1993.

[MOR94] B. S. Morse, S. M. Pizer and A. Liu, “Multiscale medial analysis of medical images”, *Image and Vision Computing*, 12(6):327-338, 1994.

[MOR95] B. S. Morse, “Computation of object cores from greyscale images”, *Ph.D. Dissertation*, Dept. of Comp. Sci., Univ. of North Carolina at Chapel Hill, 1994.

[MOR98] B. S. Morse, S. M. Pizer, D. T. Puff, and C. Gu, “Zoom-invariant vision of figure shape: effects on cores of image disturbances”, *Computer Vision and Image Understanding*, 69(1):72-86, 1998.

[OGN94] R. L. Ogniewicz, "Skeleton-space: a multiscale shape description combining region and boundary information", *in Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 746-751, 1994.

[OGN95] R. L. Ogniewicz and O. Kubler, "Hierarchic Voronoi skeletons", *Pattern Recognition*, 28: 343-359, 1995.

[ONE66] B. O’Neill, *Elementary Differential Geometry*, Academic Press, London, 1966.

[OPE83] A. V. Oppenheim and A. S. Willsky, *Signals and Systems*, Prentice-Hall, 1983.

[PAL92] S. K. Pal and L. Wang, “Fuzzy medial axis transformation (FMAT): practical feasibility”, *Fuzzy Sets and Systems*, 50:15-34, 1992.

[PEL81] S. Peleg and A. Rosenfeld, “A min-max medial axis transformation”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 3:208-210, 1981.

[PER90] P. Perona and J. Malik, “Scale-space and edge detection using anisotropic diffusion”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(7):629-639, 1990.

[PIZ87] S. M. Pizer, W. R. Oliver, and S. H. Bloomberg, “Hierarchical shape detection via the multiresolution symmetric axis transform”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 9(4):505-511, 1987.

[PIZ94] S. M. Pizer, C. A. Burbeck, J. M. Coggins, D. S. Fritsch and B. S. Morse, “Object shape before boundary shape: scale-space medial axes”, *J. Mathematical Imaging and Vision*, 4:303-313, 1994.

[PIZ98] S. M. Pizer, B. S. Morse, D. Eberly, and D. S. Fritsch, “Zoom-invariant vision of figural shape: The mathematics of cores”, *Computer Vision and Image Understanding*, 69(1):55-71, 1998.

[PUF95] D. Puff, “Human vs. model performance for medical image estimation tasks”, *Ph. D. dissertation*, Dept. of Biomedical Engineering, Univ. of North Carolina, 1995.

[ROM93] H. Rom and G. Medioni, “Hierarchical decomposition and axial shape description", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 15:973-981, 1993.

[ROS66] A. Rosenfeld and J. L. Pfaltz, “Sequential operations in digital picture processing”, *J. of ACM*, 13:471-494, 1966.

[ROS82] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, Academic Press, 1982.

[SAT98] Y. Sato, T. Araki, M. Hanayama, H. Naito, and S. Tamura, “A viewpoint determination system for stenosis diagnosis and quantification in coronary angiographic image acquisition”, *IEEE Trans. Medical Imaging*, 17(1):121-137, 1998.

[SET85] J. A. Sethian, “An analysis of flame propagation”, *Ph. D. dissertation*, University of California, Berekely, 1985.

[SHA81] B. Shapiro, J. Pisa and J. Sklansky, “Skeleton generation from x, y boundary sequences”, *Computer Vision, Graphics, and Image Processing*, 15:136-153, 1981.

[SHA98] D. Shaked and A. M. Bruckstein, “Pruning medial axis”, *Computer Vision and Image Understanding*, 69(2):156-169, 1998.

[STE98] C. Steger, “An unbiased detector of curvilinear structures”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20(2):113-125, 1998.

[TAR97] S. Tari, J. Shah and H. Pien, “Extraction of shape skeletons from greyscale images”, *Computer Vision and Image Understanding*, 66(2):133-146, 1997.

[THI92] J. P. Thirion and A. Gourdon, “The 3D marching line algorithm and its application to crest line extraction”, Research Report 1672, INRIA, 1992.

[ter91] B. M. ter Haar Romeny, L. M. J. Florack, J. J. Koenderink, and M. A. Viergever, “Scale-space: its natural operators and differential invariants”, *Lecture Notes in Computer Science*, 511:239-255, 1991.

[STE98] C. Steger, “An unbiased detector of curvilinear structures”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20(2):113-125, 1998.

[SUB93] J. B. Subirana-Vilanova and K. K. Sung, “Ridge detection for the perceptual organization without edges, in *Proc. Int. Conf. on Computer Vision*, pp. 57-64, 1993.

[WAN82] S. Wang, A. Rosenfeld, and A. Y. Wu, “A medial axis transformation for greyscale pictures”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 4(4):419-421, 1982.

[WHI93] R. T. Whitaker and S. M. Pizer, “A multiscale approach to nonuniform diffusion”, *Com-*

puter Vision, Graphics, and Image Processing: Image Understanding, 57(1):99-110, 1993.

[WID75] D. V. Widder, *The Heat Equation*, Academic Press, New York, 1975.

[WIT83] A. Witkin, “Scale space filtering”, in *Proc. Int. Joint Conf. on Artificial Intelligence*, pp. 1019-1023, 1983.

[WRI93] M. W. Wright and F. Fallside, “Skeletionisation as model-based feature detection”, *IEE Proc. Vision, Image and Signal Processing*, 140(1):7-11, 1993.

[XU98] M. Xu and D. Pycock, “Estimating true symmetry in scale space”, in *Proc. IEEE Int. Conf. on Systems, Man ,and Cybernetics*, pp. 4620-4625, 1998.

[XU99] M. Xu and D. Pycock, “A scale-space medialness transform based on boundary concordance voting”, *J. Mathematical Imaging and Vision*, 11(3):277-299, 1999.

[YUI86] A. L. Yuille and T. A. Poggio, “Scaling theorems for zero-crossing”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8:15-25, 1986.

Appendix

A.1 Publication List

1) M. Xu and D. Pycock, “A scale-space medialness transform based on boundary concordance voting”, *J. Mathematical Imaging and Vision*, 11(3): 277-299, 1999.
2) M. Xu and D. Pycock, “Estimating true symmetry in scale space”, in *Proc. IEEE Int. Conf. on Systems, Man ,and Cybernetics*, Vol. 5, pp. 4620-4625, San Diego, 1998.
3) M. Xu and D. Pycock, “Multiscale medial response of grey-level images”, *Proc. 9th British Machine Vision Conference*, Vol. 1, pp. 224-234, Southampton, 1998.
4) M. Xu and D. Pycock, “Multiscale medial axis through a complete set of optimal scale ridges”, *Proc. IEEE International Conference on Image Processing*, Vol. 4, pp. 390-394, Kobe, Japan, 1999.
5) R. Jackson, D. Pycock, M. Xu, Knowles M. and Harman S., “Event detection and period extraction using multi-scale symmetry and entropy-based methods”, submitted to *IEEE Transactions on Signal Processing*, 1999.

A.2 The JMIV Paper