ELSEVIER

Image and Vision Computing 15 (1997) 369–398

image AND vision COMPUTING

Scale-imprecision space

Lewis D. Griffin*

Department of Vision Sciences, University of Aston, Aston Triangle, Birmingham B4 7ET, UK

Received 28 February 1996; accepted 5 October 1996)

Abstract

It is argued that image measurements should satisfy two requirements of physical plausibility: the measurements are of non-zero scale and non-zero imprecision; and two required invariances, nothing is lost by expanding the image and nothing is lost by increasing the contrast of the image. A model of measurements satisfying these constraints, based on blurring the graph of the incident luminance, is described. Within this framework, several types of filtering can be expressed: mean filtering (ordinary scale space); median filtering; and mode filtering. It is found that of these possibilities, a system based on mode filtering produces interesting results. In particular, edges–defined as discontinuities–naturally appear, and their behaviour over scale and imprecision is presented.

Keywords: Measurement theory; Edge detection; Visual invariances; Median filtering

1. A conceptual justification

There is no doubting the extraordinary effectiveness of mammalian visual systems. While empirical studies progressively reveal how this performance is achieved with neural hardware, it is also possible to progress more theoretically. One such a strategy is to assume *as premise* that the functional design is as near to perfect as is possible. No doubt one day this assumption will be shown to be false, but in the interim this position provides a sturdy foothold from which to attack. To use this approach one starts with a model that is of ideal performance, and then in incremental steps one modifies it in the direction of physical plausibility. The recognition that the earliest measurements of the retinal irradiance must be achieved with apertures of *non-zero area*, which leads to the idea of scale space [1], is an example of this approach in action. In this paper I will explore the consequences of the fact that these same initial measurements cannot be of *unlimited precision*.

The need for measurements to be of non-zero scale and limited precision

The first task of the visual system is to measure the retinal irradiance. If one ignores the quantized nature of light, then the measurements of an ideal visual system would produce an everywhere-defined real-valued function over a continuous two-dimensional domain. However, complete determination of this function using physically plausible measurements is not possible for several reasons: neither arbitrarily fine spatial detail nor arbitrarily small differences in scalar magnitude can be resolved; the image can only be sampled at a finite number of locations, etc. Taking account of these limitations is not easy, particularly while maintaining a tractable mathematical model, and an opportunistic strategy is reasonable. The impossibility of resolving arbitrarily fine spatial detail has been dealt with by the notion of scale space. In this paper, I describe how this constraint may be addressed simultaneously with the impossibility of resolving arbitrarily small differences in scalar magnitude. Rather than modify the scale space theory, I will build the new theory from scratch and then show how it includes scale space.

An idealized device for measuring the value of a scalar field at a point would be something like a sharp *sensor* connected to a readout device consisting of a slender *pointer* that roves along a copy of the real line. The sensor is placed into the scalar field and the pointer registers the value measured by the sensor by taking up a location on the real line. In the idealized device, *the sensor has an infinitely sharp tip* and *the pointer a vanishing width*. To

* E-mail: l.d.griffin@aston.ac.uk

PII S0262-8856(96)1139-0

modify this picture in the direction of physical plausibility, the sensor tip must be admitted to be of non-zero area (i.e. blunt) and the pointer of non-zero width. Note that these two modifications are related to, but distinct from, the image processing terms spatial resolution and grey-level quantization. Resolution and quantization are to do with the density of sampling of, respectively, space and value (i.e. pixels and grey-levels); and while this is crucial for physical plausibility, these issues do not concern me here[1].

I shall start with the modification of the sensor tip from zero to non-zero area. The challenge is to modify the concept of a zero area sensor to produce the concept of a non-zero area sensor while retaining its nature as a 'point' operator. One can be optimistic about the possibility of success, since Huntingdon [2] has shown how a consistent and unique geometry can be constructed out of the spheres whose radii are not less than a certain value–'a perfectly rigorous geometry in which the 'points', like the schoolmaster's chalk-marks on the blackboard, are of definite, finite size, and the 'lines' and 'planes' of definite finite thickness'. At first blush one might imagine that changing to a sensor with an aperture of non-zero area would not necessitate modifying the read-out device (i.e. the real line and pointer); the intuition is that if we are to be able to treat the non-zero area sensor like a non-zero area point, just a single number should be measured as it was for a zero-sized sensor. This is too hasty. Euclid, according to Heath's translation [3], characterized a point as 'that which has no part'; and this is all that is required for the characterization of point within the subject matter of geometry. If this is accepted, then when the concept of a point is replaced with that of a point operator (as Koenderink [4] recommends) it is only obligatory to retain this property; so all that is required is that *it is not possible to discern any structure within the point by examination of the output of the operator*. Thus, there is nothing wrong with considering the blunt sensor as being composed of an infinite collection of infinitely sharp sensors, as long as the 'wires' from the sharp sensors are not labelled as to their location within the blunt tip. So, when the change is made to a blunt sensor, an infinite collection of pointers taking up positions on the real line read-out is required, one for each zero area point in the sensor tip! This is easier to imagine if the metaphor is changed midstream by replacing the pointers of the read-out device with pale shadows, so that one may more easily imagine them superimposing and adding. Then it becomes clear that the infinite collection of shadow pointers associated with a blunt sensor depicts the *histogram of the luminance values within the sensor aperture*; this is what remains when all spatial information is discarded.

Now I consider modifying the pointer, without modifying the sensor. This is more straightforward: one just uses a pointer with width, no further modifications are entailed. This is a convenient point to distinguish imprecision from error. *I will be using imprecision to indicate the pointer width*, i.e. the imprecision is small when the pointer is narrow and large when the pointer is wide. To clarify this, consider a simple measurement method that measures, say, the density of a liquid. It is well known that even if instrumental, personal and systematic errors are eliminated a random error will always persist [5]. What this means in practice is that if the method is repeatedly applied to the same sample of liquid, different results will be obtained. To calibrate the technique one can take a large (preferably infinite) number of measurements of the density of a liquid of already known density. If these results are plotted, it will be found that they form a distribution centred on the true value (the Central Limit Theorem hints at why this is often well modelled as a normal distribution). Some measure of the width of this distribution is defined to be the imprecision of the method. One can now imagine applying the method to measure the density of an unknown liquid. If it is applied repeatedly, different readings will be obtained, which again one can plot. The familiar shape of the already determined distribution will gradually re-appear, but since there is only a finite number of samples there will be some uncertainty as to the location of the centre of the distribution. This uncertainty is lessened with more samples but only eliminated with an infinite number of readings, which would allow the full distribution to be recovered. The uncertainty in position of the centre of the distribution is the error of the reading. The measurements that I am discussing have no concept of error, only of imprecision (so there is still a long way to go for true physical plausibility); they are, as it were, like the final report from an infinite set of measurements. Thus, the modification from a vanishingly thin pointer to a pointer with width is unimportant as long as the sensor still has zero area–one simple looks at the centre of the pointer ignoring its width–however, as I will explain, it does become important once the sensor is modified to have non-zero area. From this point on *I will use the standard term scale to refer to the aperture area.*

If both modifications are put together, blunt sensor and wide pointers, the readout device becomes a superimposition of an infinite collection of wide pale shadow pointers. As with zero imprecision, this can be interpreted as a histogram of luminance values within the sensor area; but, crucially, this time it is a *blurred* version of the true histogram. Since the shadows are unlabelled (or uncoloured, say) they cannot be disentangled once

[1] It is essential to distinguish scale and resolution. One reason for the common confusion of the two is that in many imaging systems (e.g. CCD cameras) the scale and the sampling density are intimately linked because apertures do not overlap. This does not need to be the case, though. For example, in mammalian visual systems, although receptors do not overlap, receptive fields (groups of receptors) do.

they start to overlap. So one cannot use the trick of noting the position of the centres of the shadows. This shows how non-zero imprecision becomes important if one must 'calculate' with such 'numbers' before a chance is had to note their midpoints.

Now I return to ideal measurements, and ask how they look within this new framework that accommodates non-zero scale and non-zero imprecision measurements. They look the same as before: the real line readout is marked with a single, vanishingly thin and infinitely dark shadow at some location. However, whereas before it seemed that this was the proper state of affairs and blurred histograms a curiosity, from the new position such readouts may be seen as an exceptional form of the more general blurred histogram output; exceptional in the sense that the histogram is zero-valued everywhere apart from at a single value (i.e. the histogram is a delta function).

Having characterized the nature of single measurements, I now consider what the output of an infinite collection of such measurements, one at each location in the field, looks like. This is easiest to do for a 1-D image (see Fig. 1). This figure depicts the (scalar) value axis of each measurement as extending vertically and the single spatial axis of the field extending horizontally. This gives a function on a 2-D position × value space, which I shall call a measurement function. It becomes clear what this is when the output of ideal measurements is considered: *the zero-scale and zero-imprecision measurement function is the graph of the luminance function.* An ideal measurement function and four examples of measurement functions of non-zero scale and non-zero imprecision are shown in Fig. 1. The differences between the measurement functions may be explained in terms of changing aperture area and/or pointer width, but can also be understood more directly in terms of the measurement functions themselves: increasing scale has the effect of horizontally blurring the measurement function; increasing imprecision has the effect of vertically blurring; and increasing both simultaneously blurs both horizontally and vertically. In the appendix I describe how measurement functions may be calculated in practice.

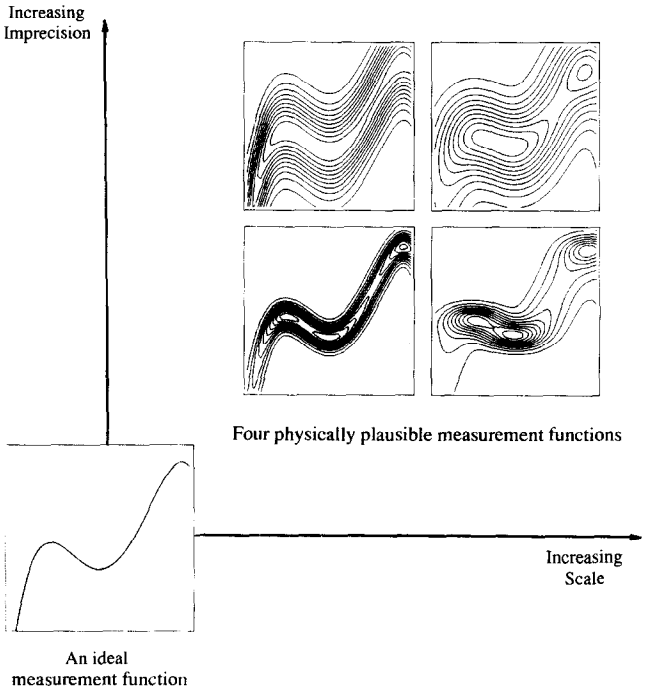


Fig. 1. An ideal measurement function and four physically plausible measurement functions.

The need for measurements to be performed at multiple scales and imprecisions

Preferably, what a visual system could determine about the world would be affected as little as possible by changes in viewpoint or the overall lighting level. Viewpoint invariance is much more difficult to satisfy than lighting level invariance. In principle, with measurements of unlimited precision, homogeneous linear lighting changes could be dealt with without any loss of information; but the collapse from three dimensions to two, which is a core feature of the visual process, means that complete invariance to changes in viewpoint cannot be achieved however idealized the measurements of retinal irradiance. Leaving open the question of what the limits of viewpoint invariance are, the physical necessity of measurements being of non-zero scale and limited precision makes the attainment of either type of invariance distinctly more difficult. These invariances are discussed in turn below.

The case of lighting level invariance is fairly simple. If the measurements are of unlimited precision, luminance differences of 0.00001 units can be distinguished just as readily as differences of one unit; but with limited precision measurements, re-scaling the luminance by reducing the lighting level to (say) 1% of what it was will cause some previously possible discriminations to drop below the level of distinguishability. A related phenomenon in image processing is when a reduction in the number of grey-levels used for display can cause image structures to disappear.

The modifications of the retinal irradiance pattern due to changes in viewpoint are challengingly complex. However, a sequence of approximations to these modifications can be given,. The first level of approximation is that of translations and rotations of the retinal irradiance pattern: caused by swivelling of the eyeball without change in location. In this work I am assuming that such changes can be tolerated without loss of information. To reach the next level of approximation, one adds isotropic scaling of the retinal irradiance pattern. This approximates the effect of moving towards or away from the scene; the approximation being most valid in the case of distant scenes with little variation in depth. The third level is achieved by allowing affine transformations which occur

when a visible surface alters in its slant with respect to the viewer [6]. The fourth level occurs when it is admitted that all the previous transformations should actually be perspective transformations. The fifth level, when it is acknowledged that these transformations vary across the scene due to differences in depth. And the sixth level when occlusion is acknowledged. In this paper, I consider only as far as the second level: translations, rotations and isotropic re-scalings. Isotropic re-scaling does not present a problem when using measurements made with zero scale apertures–a separation of 0.00001 retinal units is resolved as readily as a separation of one unit–but is a problem when using measurements of non-zero scale; shrinking the image by a linear factor of (say) 100, as is caused by moving to a vantage point 100 times as far away, will cause some spatial details to drop below the scale of the measurement apertures. Again, this is similar to the situation in image processing where a reduction in the spatial resolution (i.e. sampling density) can cause a disappearance of some image structures.

Now I will consider what sort of invariance *is* possible with measurements of non-zero scale and limited precision. I loosely formalize the measurement process as $MF(s,i,l)$, i.e. the measurement function produced by operators of scale s and imprecision i when applied to the incident luminance l. Consider a particular set of operators of scale S and imprecision I. Now ask what happens to the measurement function produced using these devices if the incident luminance pattern is spatially expanded by a linear factor of two (i.e. what is the difference between $MF(S,I,l)$ and $MF(S,I,expand[l]))$? Clearly, the measurement function will change since spatial features are now twice the size (i.e. a feature that was one unit long will now be two units long). Even if this difference is removed by spatially shrinking the measurement function by a factor of two, there will still be a difference; the shrunk measurement function will have more spatial detail than the original (i.e. $shrink[MF(S,I,expand[l])] \neq MF(S,I,l)$). This makes perfectly good sense, as fixing the measurement scale and doubling the size of the image allows more detail to be discerned. However, if larger apertures are used then the same measurement function will result, i.e. $shrink[MF(2S,I,expand[l])] = MF(S,I,l)$. A similar relationship holds when the incident luminance is shrunk, i.e. $expand[MF(\frac{1}{2}S,I,shrink[l])] = MF(S,I,l)$. The imprecision equivalent of expanding the incident luminance pattern is brightening it by multiplying the luminance values by a factor greater than unity; the same measurement function can be obtained by using less precise operators and re-scaling the luminance axis. Similarly, darkening the image can be combatted by using more precise operators, etc.

If a visual system produced only one measurement function at a particular scale and imprecision it would have no invariance at all. Expanding, shrinking, brightening or darkening the image all would change the measurement function. On the other hand, if all the (non-ideal) measurement functions were produced it would have complete invariance. Unfortunately, a physical system cannot produce all the measurement functions: physical plausibility commits us, not only to measurements being of non-zero scale and non-zero imprecision, but also to there being a non-zero lower bound on the scale and imprecision of the measurements in a system. Thus, the most that is possible is to have all the measurement functions coarser and more imprecise than some 'best' set of measurements. A visual system equipped with such a set of measurement devices will lose information when the scene is shrunk or darkened, but will not lose information when it is expanded or brightened[2].

Aperture and needle shapes

Finally, the shapes of the sensor aperture and pointer profile must be considered. Obvious choices would be a circular aperture and a bar for the pointer profile, but experience with scale space suggests otherwise. In fact, the scale space arguments apply equally well to scale-imprecision space, and so tell us that the aperture should not have a hard cutoff and instead should have the tapering form of a Gaussian. Of course, one has to accept the physical implausibility of an infinite aperture as an outstanding, but hopefully temporary, problem. Considering the profile of the pointer, it is fairly easy to hazard a guess as to its correct form: a Gaussian. This is perhaps easier to swallow than it is for the aperture function due to familiarity with the Gaussian as a manifestation of 'pure uncertainty'.

Several authors [7, 8] have addressed the problem of proving that the Gaussian is uniquely suitable to play the role of a point operator of non-zero area. There are two main lines of argument. The first is to demand that the operation of the apertures should satisfy an 'image in, image out' property [9]; which amounts to the requirement that an observation of an observation should be equivalent to a single observation or, even more formally, that the convolution product of two apertures should be another aperture. This seems to be contestable as it is at least conceivable (though maybe not practical) that a visual system, could make all of its measurements once and for all directly on the luminance and so have no need of this property; and also, as Pauwels et al. [10] have pointed out, this requirement must be supplemented with others such as smoothness and positivity in order to uniquely specify the Gaussian. The other line of argument, pioneered by Koenderink [1], is via the 'causality' property which I will briefly describe.

[2] This statement applies to an idealized system in which one ignores issues of spatial sampling, quantization, boundedness of the field of view and saturation.

The causality property applies to a continuous family of continuous functions, say $f_t : \mathbb{R}^n \to \mathbb{R}, t\epsilon\mathbb{R}$ (for rhetorical reasons I will refer to the indexing parameter t as if it is time). There are several equivalent ways to characterize it. The most provocative is that if the causality property holds, then values of the function may always be continuously traced back to earlier times. To be explicit, consider the $n+1$ dimensional space formed as the cross product of the function domain and the indexing parameter. This space is foliated (like an onion) by n-dimensional iso-surfaces of constant function value. The causality property is that if one picks a point of any iso-surface, it is possible to trace a continuous path on the surface that: starts at the point, always progresses from later to earlier times and can be extended to arbitrarily early times. To determine whether a given family of functions has this property or not it is simplest to check the condition $\nabla f = \vec{0} \Rightarrow \nabla^2 f \cdot \partial f/\partial t \geq 0$ which in words means that the extrema of the function are attenuated as time passes. This extrema attenuation property holds if and only if the family is causal, and so may be used instead of the iso-surface condition.

Koenderink [1] has shown that out of all the possible shapes one might consider for a blurring kernel (i.e. the aperture weighting function or the pointer cross-section), only with the Gaussian will one obtain a causal family. Whether the causal property (or its equivalent extremum attenuation) is an essential feature for an entity claiming to play the role of a non-zero area point is a difficult question to answer. However, what is certain is that the Gaussian plays the role very well. In conclusion, I assume that both the aperture and the pointer profile should have the form of Gaussians; with width increasing with scale and imprecision, respectively.

Previous work by other authors

The idea of looking at the behaviour of a histogram at multiple Gaussian scales has been described by Carlotto [11] and Minnotte and Scott [12]. The idea of working with the position × value cross product space is described by Noest and Koenderink [13] and Noest [14]. The primary concern of this work is with neural mechanisms that can cope with "...'transparency', that is overlaid or finely interleaved data from multiple objects". To this end, Noest proposes the use of 'blurred relations' defined as functions on position × value space (i.e. what I call measurement functions). The 'value' he is concerned with is velocity and, to a lesser extent, orientation. He proposes a 1-D family of such measurement functions produced by blurring them with an isotropic Gaussian. Thus his space of measurement functions corresponds to a sloping line through the origin in my 2-D scale-imprecision space. Similar proposals are made by Simoncelli [15] and Snowden et al. [16].

A formalism for scale-imprecision space

In this section I will introduce notation to describe scale-imprecision space in precise terms. The formalism will be set-up for the general case of an image of integer dimensionality ($n\epsilon\mathbb{N}$). I define $\mathcal{L} : \mathbb{R}^n \to \mathbb{R}$ to be the incident luminance across this field. I do not assume $\mathcal{L}$ to be continuous but I shall assume that there exists a polynomial $p : \mathbb{R}^n \to \mathbb{R}$, that bounds its magnitude; thus $\forall \vec{x} \in \mathbb{R}^n |\mathcal{L}(\vec{x})| < p(\vec{x})$. Note that I use $\vec{x}$ as a variable location within the n-dimensional domain (i.e. it varies across *position*); this usage will be consistent in what follows, so for brevity I will drop the $\forall \vec{x} \in \mathbb{R}^n$.

Next, I define isotropic Gaussian kernels $\forall d \in \mathbb{N}\ \forall w \in \mathbb{R}^+ G_w^d : \mathbb{R}^d \to \mathbb{R}^+$ of dimensionality (d) and width (w) by $G_w^d(\vec{r}) = (4\pi w)^{-d/2} \cdot e^{-\vec{r}\cdot\vec{r}/4w}$.

I define a 1-D family of aperture operators $\forall s \in \mathbb{R}^+ A_s : \mathbb{R}^n \to \mathbb{R}^+$ and a 1-D family of pointer operators $\forall i \in \mathbb{R}^+ P_i : \mathbb{R}^n \to \mathbb{R}^+$ to be Gaussian kernels on position and value, respectively, i.e. $A_s(\vec{x}) = G_s^n(\vec{x})$ and $P_i(v) = G_i^1(v)$. The spreads of these kernels, which are positive real numbers, will be referred to as *scale* (s) and *imprecision* (i), respectively. Note that they have dimensions **LENGTH**^{2} and **VALUE**^{2}, respectively. It should be noted that the choice of **LENGTH**^{2} rather than **LENGTH** is nothing to do with the dimensionality of the domain, but is merely that the formulae are simpler when using a variance rather than a standard deviation. This usage of s and i will be consistent in what follows, so for readability, I will drop the $\forall s, i \in \mathbb{R}^+$.

Next I define a 2-D family of measurement functions $L_{s,i} : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^+$ by

$$L_{s,i}(\vec{x}; v) = \int_{\vec{a}\in\mathbb{R}^n} A_s[\vec{a} - \vec{x}] \cdot P_i[\mathcal{L}(\vec{a}) - v]$$

Measurement functions are defined across position × value space. I refer to the entire two-dimensional family of measurement functions as a scale-imprecision space. Scale-imprecision is the cross-product of two separate spaces produced by Gaussian blurring. So it inherits the properties of such spaces:

(1) Infinite differentiability over the $3+n$-dimensional space of position, value, non-zero scale and non-zero imprecision: $L_{s,i}(\vec{x}; v)|_{s,i>0} \in C^{\infty}_{3+n}$.
(2) Satisfaction of the diffusion equation[3]:

$$\frac{\partial L}{\partial s} = \nabla^2_{\vec{x}} L, \quad \frac{\partial L}{\partial i} = \frac{\partial^2 L}{\partial v^2}.$$

[3] This is the most attractive feature of using Gaussian kernels for blurring: it is equivalent to diffusing (like heat or a gas) whatever is being blurred. Diffusion processes can be described by this very simple differential equation making for tractability.

(3) Causality[4]:

$$\frac{\partial L}{\partial \vec{x}} = \vec{0} \Rightarrow \frac{\partial L}{\partial s} \cdot \nabla^2_{\vec{x}} L \geq 0, \quad \frac{\partial L}{\partial v} = 0 \Rightarrow \frac{\partial L}{\partial t} \cdot \frac{\partial^2 L}{\partial v^2} \geq 0$$

Summarizing images

At this point I introduce a new idea: instead of working with the full measurement histogram for each position, pick out one (or more) representative values from each histogram so that a real-(but possible multi-)valued function is defined over the domain. It is difficult to find a compelling justification for this step. One possible line of reasoning is that the measurement functions have an undeserved number of degrees of freedom, and hence contain redundancy. The question then is how to pick out an irredundant core? A clue is offered by the fact that the measurement histograms always have unit integral. So they may be treated as probability density functions, and definitions of *central location* may be used to pick out key values (these will always exist because of the polynomial bound on the luminance). This will be my program, and I now review such measures. I will eventually pick out one (the stable mode) as particularly relevant for visual processing, but the others undoubtedly have their uses. In the appendix I describe an algorithm for computing these summarizing images for discrete data.

Definitions of central location

An example histogram is shown in Fig. 2. Several possible central locations have been marked. I will review them one by one with reference to a probability distribution function p.

The mean is the most commonly used of all the measures of central location. I will denote it by $\mu[p]$. It is normally defined as

$$\mu[p] = \int_{x \in \mathbb{R}} x \cdot p(x)$$

but may also be defined as the value that minimizes the squared distance to the rest of the distribution, i.e. $x^2 \otimes p$ attains its minimum at $\mu[p]$, and it can be shown that this is the only minimum. A simple situation in which the mean can be shown to be superior to the median or modes is when dealing with samples from a normally distributed population. If the task is to estimate the population centre (mean, median and mode all being equal for the population distribution); and the question is, what is best, the sample mean, the sample median or the sample mode? It can be shown that, assuming the population is normal, the sample mean is best. For example, one would expect as reliable an answer from a sample of 64, when using the sample mean, as from a sample of 100 when using the sample median[5]. However, this is a use of the mean as an estimator of the centre of a population, where, due to symmetry, the centre is identical to the mean, median and mode, and so it fails to elucidate its meaning distinctly. A less ambiguous use of the mean is in calculating the expected, or average, value. This is an appropriate concept only for interval data, meaning that it is believed that the structure of the measured magnitudes is isomorphic with the real line all the way up to its metrical properties. It gives the value which if repeatedly picked, rather than a random sample from the population, would give the same sum of values in the long run. Note that, when using the mean in this way, one completely disregards questions of whether the populations is uni- or multi-modal, symmetric or asymmetric: it is (or should be) the gambler's measure, only the long term sum is of concern. This should not be taken as a condemnation of the mean, since if one has interval data and calculates only the median or modes, then the structure of the data is not being fully exploited. This is reflected in the relative weakness of non-parametric statistical tests as compared to parametric.

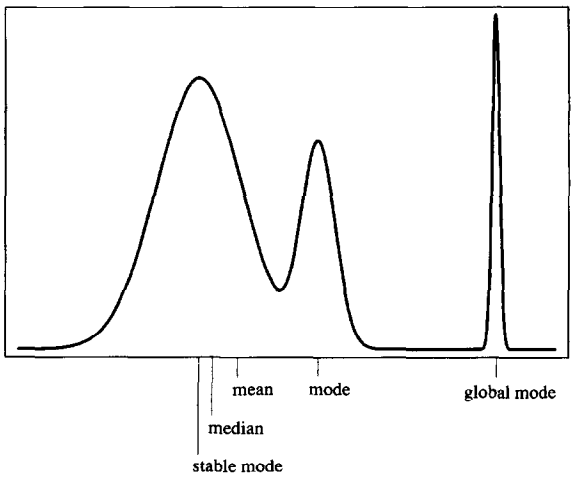


Fig. 2. An example probability distribution function.

The *median* ($m[p]$) is defined as the value that splits the distribution into two equal halves, i.e.

$$\int_{-\infty}^{m[p]} p(x)dx = \int_{m[p]}^{\infty} p(x)dx = \frac{1}{2}$$

It may also be defined as the value that minimizes the absolute distance to the rest of the distribution (i.e.

[4] This is the property discussed in the section on pointer and aperture shapes. In words it means that maxima decrease with blurring and minima increase; thinking of the blurring process as diffusion this is obvious. It is a useful property for deciding (without resorting to algebra) what types of changes can happen during blurring, i.e. events in which extrema are enhanced never occur.

[5] The asymptotic ratio is $2/\pi \approx 0.64$.

$|x| \otimes p$ is minimized at $m[p]$). Like the mean, the median has two uses. Firstly, the sample median, like the sample mean, can be used to estimate the centre of a symmetrical population distribution. Although as explained above, this is not a good idea when dealing with a pure normal population it is worthwhile if the population is contaminated. For example, if the population is of the form $p(x) = \frac{9}{10}N\ (x|\mu = 0, \sigma^2 = 1) + \frac{1}{10}N\ (x|\mu = 0, \sigma^2 = 9)$, then the sample median is a more robust estimator of the population centre (in this case 0) than the sample mean. This is what is meant by 'robustness to outliers'. With reference to Fig. 2, observe that (speaking informally) the mean has been dragged to the right by the thin spike whereas the median has not. The second use of the median is as a definition of centre for ordinal data. In such a case the mean cannot be used since ordinal, or rank, data is isomorphic with the real line only up to its topology. The fact that the median is only dependent on the topology of the values gives it powerful properties. It means that if the data is re-ranked, in a way that preserves the topology, then the median of the re-ranked data is at the same point in the distribution. This is obvious if you think of its definition as the point which divides the distribution into equal halves. Another curious property of the median is that the weight of the distribution may be shifted around, and as long as none is transferred across the median point, the median will not move. This is not true of the mean. This is one reason why proposals for a minimum wage are phrased in terms of the median earnings rather than the mean.

The *modes* (denoted $h[p]$) are defined as the maxima of the distribution. Like the median, the modes are invariant to monotonic re-ranking of the data: and in fact they are invariant under a re-distribution of the weight of the distribution as long as no maxima are created or destroyed. Unlike the mean and the median, there may be multiple modes. A characterization of a distribution as multi-modal will often be followed by a hypothesis that the distribution can be explained as the superposition of several distinct sub-types.

Generally one of the modes can be distinguished as being higher than the others, this is the *global mode* which I denote $h^+[p]$. The global mode is invariant under any one-to-one re-mapping of the ranks, and so its most important use is for domains with no sense of topology: what statisticians call nominal data. For such data the mean, median and modes cannot be defined so the global mode is the only option. Even when the data is of a real-valued variable, it is troublesome to use for empirically samples distributions since, in general, there will be no duplications at all. The way around this is to put the sample data into coarsely defined bins and choose the centre of the most popular bin as the global mode. This is a far from satisfactory procedure since the global mode may change discontinuously as the bin width is varied. Consider the distribution in Fig. 2. If one had a very large set of samples from this distribution and one used narrow bins, then the right-hand mode would (probably) be identified as the global mode; but with larger bins (as might be required with fewer samples) the left-hand mode would be selected.

The observation above concerning bin widths and the fact that the left-hand mode actually looks like a better 'typical' value than the global mode creates the suspicion that there might be some other way to select one of the modes as best. And in fact, I believe there is. I call it the *stable mode* (denoted $H[p]$) for reasons that will become clear. If a 1-D histogram is Gaussian blurred (c.f. increasing bin-width), then the modes and anti-modes (histogram minima) will alter in position and will annihilate in pairs via fold catastrophes, until eventually only a single mode which never disappears is all that remains. That remaining mode may be traced back to one of the modes in the original histogram; I call that mode the stable mode. As shown in Fig. 2, the stable mode need not be the same as the global mode, but it 'often' is. The attractive property of this definition for my purposes is that as a measurement histogram is blurred (i.e. as imprecision increases), the stable mode will move continuously. This property is not shared by the global mode; in the example it will start as the right-hand mode and after some blurring will jump (discontinuously) to the left-hand mode.

The above describes the pattern of modes during blurring for a generic histogram; but for scale-imprecision space it is necessary to consider versal histograms[6]. For one-dimensional luminance patterns it is necessary to consider histograms with one or two degrees of versality and extend the definition of the stable mode to cover them. This can be done naturally. The imprecision family generated by a versal-I histogram will contain a cusp catastrophe. There are two variants of this. Firstly, a cusp catastrophe where two anti-modes and a mode merge to produce an anti-mode; this is not relevant to the stable mode, since clearly neither of the two modes which are destroyed will be the final mode that survives blurring. Secondly, a cusp catastrophe where two modes and an anti-mode merge to produce a mode; this event is significant if the surviving mode goes on to be the mode that survives blurring (i.e. it is the stable mode). For such a histogram, it makes sense to regard the stable mode as initially bi-valued but becoming single valued after sufficient blurring. For a versal-II family there are two possibilities. Firstly, that two modes and two anti-

[6] The generic/versal distinction can be illustrated by the example of picking, at random, three real numbers. The generic case is that all three numbers are distinct; for a single triple of numbers this will be the case with probability infinitesimally close to 1. If a one dimensional family of triples are generated, it is to be expected that there will be triples where two numbers are equal; this would be a versal-I state of affairs. If a two dimensional family is generated there can be cases where all three numbers are equal; this is versal-II.

modes annihilate together (via a swallowtail catastrophe) to produce nothing. This is irrelevant to the stable mode as it cannot involve the mode which goes on to be the one that survives blurring. The other possibility is that two cusp catastrophes should occur. Of the various ways in which this might happen, only one effects the definition of the stable mode. Namely, that two modes and an anti-mode annihilate into a mode and that this mode then goes on to annihilate with another mode and an anti-mode to produce a mode that goes on to be the one that survives the blurring process. When this occurs, the stable mode starts triple-valued and becomes bi-valued and then single-valued with blurring. All these possibilities will be clarified by diagrams later in the paper.

The stable mode certainly looks to be a representative value whenever it is calculated for actual histograms, and because of its obvious link with imprecision it has very nice behaviour in scale imprecision space. However, it would be nice to have more solid grounds for believing its representativeness. In the appendix I present a conjecture about the stable mode which links it to more standard definitions of central location. If true, the conjecture shows that the stable mode is indeed a particularly representative value of a histogram.

Definitions and properties of the summarizing images

The mean images $\mu_{s,i}$: $\mathbb{R}^n \rightarrow \mathbb{R}$ are created from measurement functions by taking the mean of the value histogram for each position $\mu_{s,i}(\vec{x}) = \mu[L_{s,i}(\vec{x}; \cdot)]$. As the definition stands, they are a 2-D family. But an anti-climax is in store because in fact imprecision is irrelevant to the mean image, they only vary with scale, i.e. $\mu_{s,i} = A_s \oslash \mathcal{L}$. This is because blurring a histogram with a symmetric kernel such as the Gaussian does not alter its mean. So the mean images are a 1-D family and form an ordinary scale space, as is shown by the relation $\partial\mu/\partial_s = \nabla^2_{\vec{x}}\mu$. This is one way in which scale-imprecision space contains scale space, but I will show others. Henceforth, I shall refer to the mean images as indexed by a single scale parameter (i.e. $\mu_s \equiv A_s \oslash \mathcal{L}$). All the properties of scale space described in previous works (e.g. [17,18]) apply to the 1-D family of mean images. In particular, the sequence is causal since $\nabla^2_{\vec{x}}\mu \cdot \partial\mu/\partial s \geq 0$ holds, making $\partial\mu/\partial\vec{x} = \vec{0} \Rightarrow \nabla^2_{\vec{x}}\mu \cdot \partial\mu/\partial s \geq 0$ trivially true (so extrema are attenuated as scale increases).

The median images $m_{s,i} : \mathbb{R}^n \rightarrow \mathbb{R}$ are created from measurement functions by taking the median of the value histogram for each position $m_{s,i}(\vec{x}) = m[L_{s,i}(\vec{x}; \cdot)]$. Although the median image is a non-linear function of the luminance it, like the mean image, evolves causally with respect to scale. To see this, consider the scale derivative

$$\frac{\partial m}{\partial s} = \nabla^2_{\vec{x}} m + \frac{\partial m}{\partial \vec{x}} \cdot \left(2 \left[\frac{\partial L}{\partial \vec{x}} \right]_{v=m} + \frac{\partial m}{\partial \vec{x}} \left[\frac{1}{L} \frac{\partial L}{L \partial v} \right]_{v=m} \right).$$

This shows the evolution with scale to be more than simple diffusion, since there is a gradient-based term as well as a Laplacean term. This does not prevent the evolution being causal, since the condition for causality is that extrema are attenuated, and the gradient term makes no contribution at an extremum; so $\partial m/\partial\vec{x} = \vec{0} \Rightarrow \nabla^2_{\vec{x}} m \cdot \partial m/\partial s \geq 0$. In the appendix it is proved that $\lim_{i\rightarrow\infty} m_{s,i} = \mu_s$ (i.e. as imprecision increases, the median images tend to the mean images); thus the median images contain ordinary scale space.

The multi-modal images $h_{s,i} : \mathbb{R}^n \rightarrow \mathcal{P}(\mathbb{R}) - \oslash$ are created from measurement functions by taking the modes of the value histogram for each position $h_{s,i}(\vec{x}) = h[L_{s,i}(\vec{x}; \cdot)]$. They are different from the mean and median images in that they are multi-valued functions, i.e. at each point of the domain, the function attains, not a single value, but a non-empty set of values. It is natural to ask whether the multi-modal image is causal with respect to scale. This is slightly troublesome to answer due to its multi-valued nature. But if attention is restricted to some open region of the position × value cross-product space where the modal image is single-valued and continuous, then the answer is yes. Taking liberties with notation, the scale derivative in such a region is

$$\frac{\partial h}{\partial s} = \nabla^2_{\vec{x}} h + \frac{\partial h}{\partial \vec{x}} \left[\frac{2 \dfrac{\partial^3 L}{\partial \vec{x} \partial v^2} + \dfrac{\partial h}{\partial \vec{x}} \dfrac{\partial^3 L}{\partial v^3}}{\dfrac{\partial^2 L}{\partial v^2}} \right]_{v=h}$$

As with the median image there is a gradient-based term in addition to the Laplacean term; again this does not prevent the evolution being causal, so $\partial h/\partial\vec{x} = \vec{0} \Rightarrow \nabla^2_{\vec{x}} h \cdot \partial h/\partial s \geq 0$ is true. In the appendix it is proved that $\lim_{i\rightarrow\infty} h_{s,i} = \{\mu_s\}$ (i.e. as imprecision increases, the multi-modal images tend to the mean images); thus the multi-modal images, like the median, contain ordinary scale space.

The stable mode images $H_{s,i} : \mathbb{R}^n \rightarrow \mathcal{P}(\mathbb{R}) - \oslash$ are created from measurement functions by taking the stable mode of the value histogram for each position $H_{s,i}(\vec{x}) = H[L_{s,i}(\vec{x}; \cdot)]$. Like the multi-modal image, the stable mode image is defined as set-valued. However, whereas the multi-modal image will be multi-valued for intervals of position (even for the entire domain when scale is high and imprecision low), the stable mode will be single-valued almost everywhere. Where it is not single-valued, it may be up to $n + 2$-valued. As will be seen in later sections, these points mark discontinuities in the stable mode image. Since $H_{s,i} \subseteq h_{s,i}$, the evolution with

scale (disregarding the discontinuities) is causal and the stable mode images contain an ordinary scale space at infinite imprecision.

Comparison of the summarizing images

Fig. 3 shows measurement functions and summarizing images for an example magnetic resonance image of a human brain (showing part of a transaxial section through the lateral ventricles). Each row consists of four sub-panels showing the mean, median and stable mode images and part of the measurement function. The portion of the measurement function shown corresponds to the horizontal yellow line; which is at the same height in all panels. Note that the measurement functions were calculated for the two dimensional image and so are 2 + 1-dimensional, although only a 1 + 1-dimensional slice through them is shown. On the measurement functions

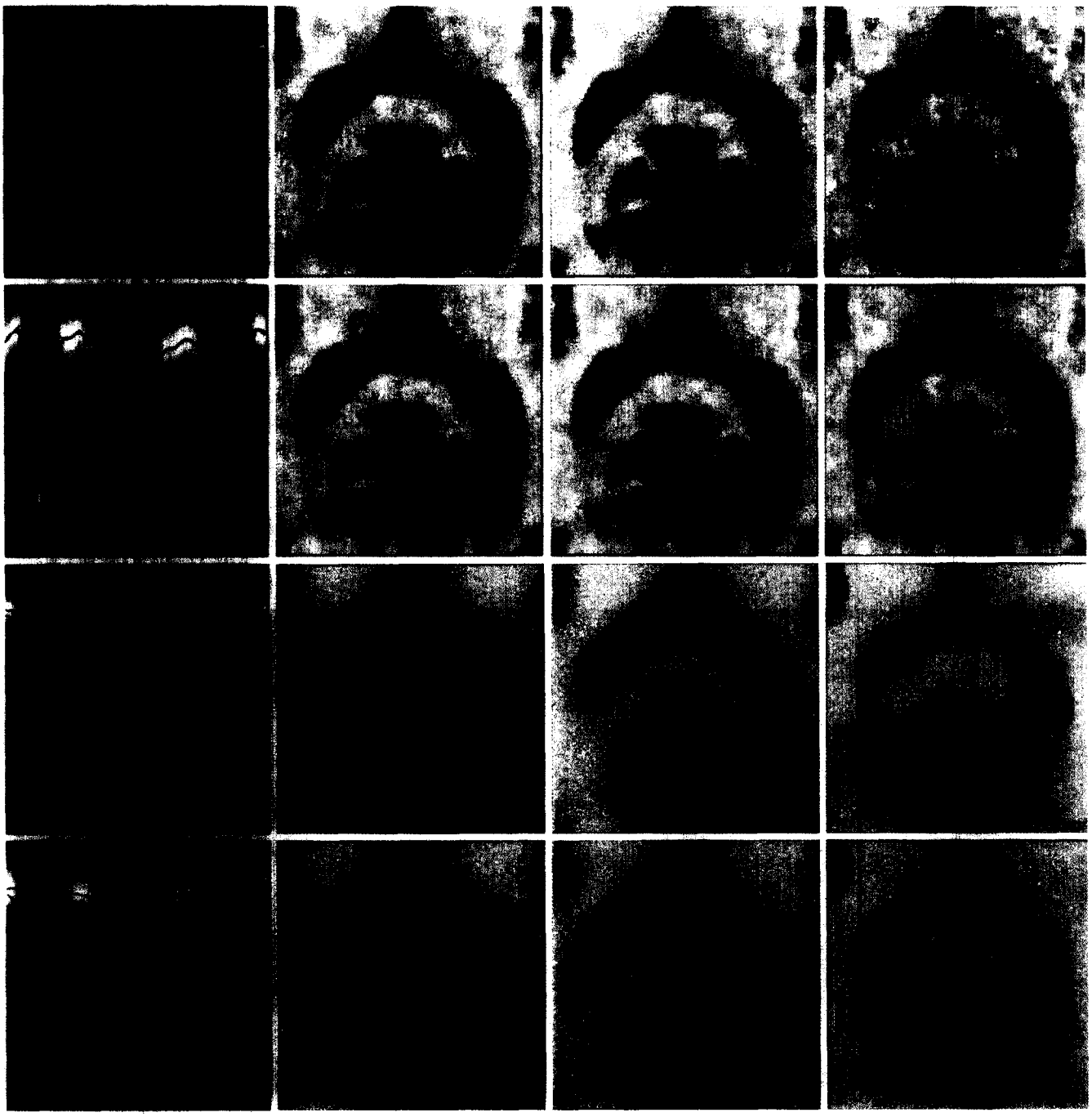


Fig. 3. Each row relates to measurements at a particular scale and imprecision. The left-hand column shows a cross-section (corresponding in position to the yellow line in the other panels) through the measurement function. Overlaid are the mean image (light blue), median image (green) and the stable mode image (red). The other three columns, from left to right, show the mean, median and stable mode images.

are overlaid the mean, median and stable mode images shown in blue, green and red respectively. The scales and imprecisions of the four rows are different. The second row is at the same scale but a higher imprecision than the first; the third row is at the same imprecision as the first but at a coarser scale; and the final row is at the same imprecision as the second and the same scale as the first.

The figure illustrates several points. (1) The relative smoothness of the mean image compared to the median, and of the median compared to the stable mode; this can be seen both in the grey-level images and in the measurement functions; (2) that the mean image is unchanged by changes in imprecision without a change in scale, shown by the equality of the mean images in rows one and two and rows three and four; (3) that the median and stable mode image become smoother as imprecision is increased while scale is held constant (row two vs. row one and row four vs. row three); this is slightly difficult to see because of difficulties with reproduction; (4) that the stable mode image is discontinuous while the mean and median images are continuous; (5) that stable mode discontinuities disappear, but do not appear, with increasing imprecision.

Superficial and deep structure of the multi-modal image

The motivation for examining the mean, median and modal images in the previous section was that a return to something more image-like was desirable given the obvious redundancy of measurement functions. One finding of the previous section was that each of the 2D families of summarizing images contains the 1D family of mean images as a subset; so the mean image always comes for free. Whilst the median image is certainly very interesting, particularly for image processing (e.g. [19, 20]), it is the two modal images which really catch one's eye in relation to the vision problem.

The multi-modal image is appealing as it offers a way to produce multi-valued images which may be necessary for the modelling of transparency. Also, regions can be defined by looking at the spatial extent of a continuous stretch of mode; these regions are overlapping, since there may be multiple modes (and thus regions) at a point. Neither multi-valuedness nor naturally defined regions are properties displayed by the mean or median images. The multi-modal image seems an expensive solution to the fairly rare phenomenon of transparency in luminance images, but may well be justified when dealing with attributes other than luminance, for instance, velocity. On the other hand, the stable mode images share the property of breaking the image into objects; but rather than being overlapping, they have well-defined discontinuities at the interfaces between them (Fig. 4, right). The occurrence of discontinuities in a clearly defined model of the measured visual image is extremely appealing given the difficulties in defining edges as some kind of locus of points of high gradient. The question of whether edges defined as discontinuities in the stable mode image are a good model for perceptual edges is finally one for physiology and psychophysics. Nevertheless, in the remainder of this paper, I will present evidence which is consistent with the hypothesis, to wit, that they behave in a 'nice' and 'edge-like' manner with changes in scale and imprecision; and that crude computations of their

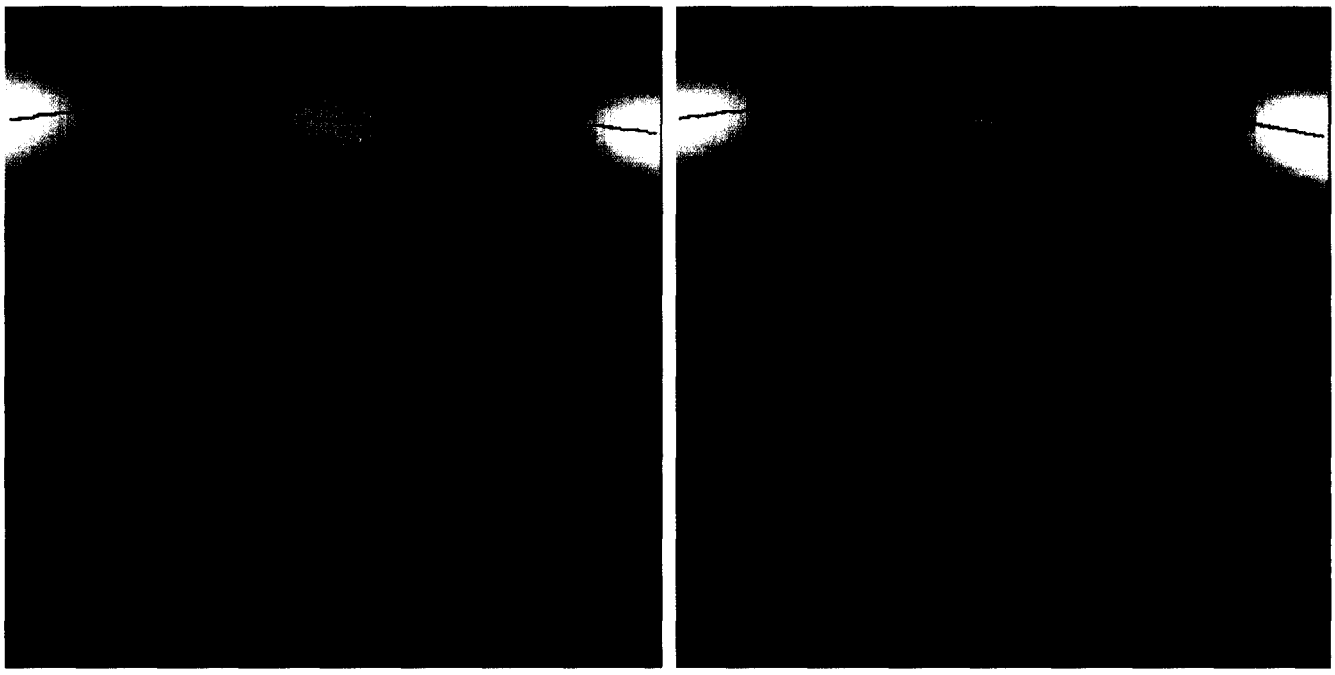


Fig. 4. A measurement function is shown overlaid with the augmented modal image (left) and the stable mode image (right).

location on actual images produces reasonable results. Towards this goal, in the next section I describe the scale and imprecision behaviour of the stable mode image for 1D images. To make this easier, and because it is interesting in its own right, in this section I describe the scale and imprecision behaviour of the multi-modal image for 1D images. In the final section, I present the scale-imprecision behaviour of some standard functions (1D) and empirical data (2D). The results of all three sections are an application of Catastrophe Theory [21, 22]. The standard theory needs some modifications for the present problem since the functions being examined are not of the normal form, i.e. infinitely differentiable single-valued functions from $\mathbb{R}^p$ to $\mathbb{R}^q$. However, the changes are not significant enough to require a re-working of the formal theory; a simple understanding of the basic catastrophes together with an appreciation of the effect of the causality constraint of Gaussian blurring has been sufficient to produce the results shown.

My first step is to define the *augmented modal image* which contains the multi-modal image but also contains anti-modes (i.e. histogram minima). The presentation is simplified by defining $D = \partial L/\partial v$ and $D_x = \partial^2 L/\partial x \partial v$, etc. The augmented modal image $h^*_{s,i} : \mathbb{R}^n \rightarrow P(\mathbb{R}) - \varnothing$ may then be simply defined as $h^*_{s,i}(\vec{x}) = \{r : D(\vec{x}; r) = 0\}$, i.e. the zero-crossing of a linear differential of the measurement function; this is not possible for the unaugmented modal image, and so its behaviour over scale and imprecision is more difficult to analyse. This definition is valid whatever the dimensionality of the domain; however, for the majority of the remainder of the paper I am only concerned with a 1-D domain. Fig. 4 (left) shows a measurement function with the augmented multi-modal image overlaid. Modes are shown in red, anti-modes in blue, tie points (explained below) in purple and extrema in green. Whereas the modes mark local maxima in measurement histograms, the anti-modes mark local minima.

Since the augmented modal image is defined as the zero-crossing of a linear differential of the measurement function, it generally consists of closed loops and curves extending indefinitely to the left and right. Some thought shows that the tie points (in purple) occur where the multi-modal image has a dangling endpoint; these occur whenever the tangent to the D zero-crossings is vertical; at these points a section of anti-mode tunnels back, under or over, the section of mode. Other simple lines of reasoning lead to the following conclusions: every column has at least one mode; in any column modes and anti-modes alternate; and, in any column, the outermost points of the augmented modal image are always modes. I have also marked the extrema (in green) as these are important features whose behaviour over scale and imprecision can be analysed, as it is for extrema in scale space. Note that I have marked extrema on both the modes and anti-modes, as although anti-mode extrema are unlikely to be of importance, the presentation is clearer if the modes and anti-modes are treated equally.

Fig. 5 shows the different types of points that can occur in the augmented modal image. The apex of the large pyramid shows the most common type of point: *a slope point* of mode or anti-mode. The layout of this box is as follows. The top-left square shows a canonical appearance of such a point; this is a diagram in position × value space. The four smaller squares in the lower half of the box show all possible appearances of this type of point, i.e. mode or anti-mode, positive or negative slope. The pyramid in the top-right of the box is a Pascal's triangle showing which derivatives of D must be zero for such a point to occur; in this case just the zeroth order, i.e. $D = 0$.

The second row of Fig. 5 shows the *extremal* and *tie points*. Their boxes are linked to the slope point above as they are specializations of it, i.e. they share the condition $D = 0$ but also have the additional conditions $D_x = 0$ (extremum) and $D_v = 0$ (tie-point), as is shown by the extra zero in their Pascal's triangles. The difference between the rows is to be understood as follows. If you pick a scale and imprecision at random, and then pick a position at random out of the corresponding modal image then it will (almost always) be a slope point (i.e. the first row); whereas if you examine all the positions of the randomly chosen augmented multi-modal image, it will typically include point types of the second row. Such points exist over a region of scale-imprecision space (like an extremum exists over a range of scales in scale space). The third row are points that occur in each of the measurement functions corresponding to a curve in scale-imprecision space; I call these versal-I events. As such, they are vanishingly likely to occur in a single randomly selected measurement function but can be expected to be encountered when considering a 1-D family of modal images. The fourth row of Fig. 5 shows points that exist at isolated points of scale-imprecision space, thus they can be expected to be encountered when considering all the measurement functions in a region of scale-imprecision space, but not when considering either a single randomly chosen measurement function or a random 1-D family of measurement functions; I call these versal-II events (their structure will be explained following this paragraph). Together, these points and the inter-relationships between them induced by the catastrophic events in which some occur while others are created and destroyed capture the qualitative structure of scale-imprecision space. This is analogous to the way that critical point events, etc. in scale space capture and organize the image structure, except here the events occur in a space which is two-dimensional. The listings of types of point in Fig. 5 is exhaustive. In particular, there cannot be another row of the large

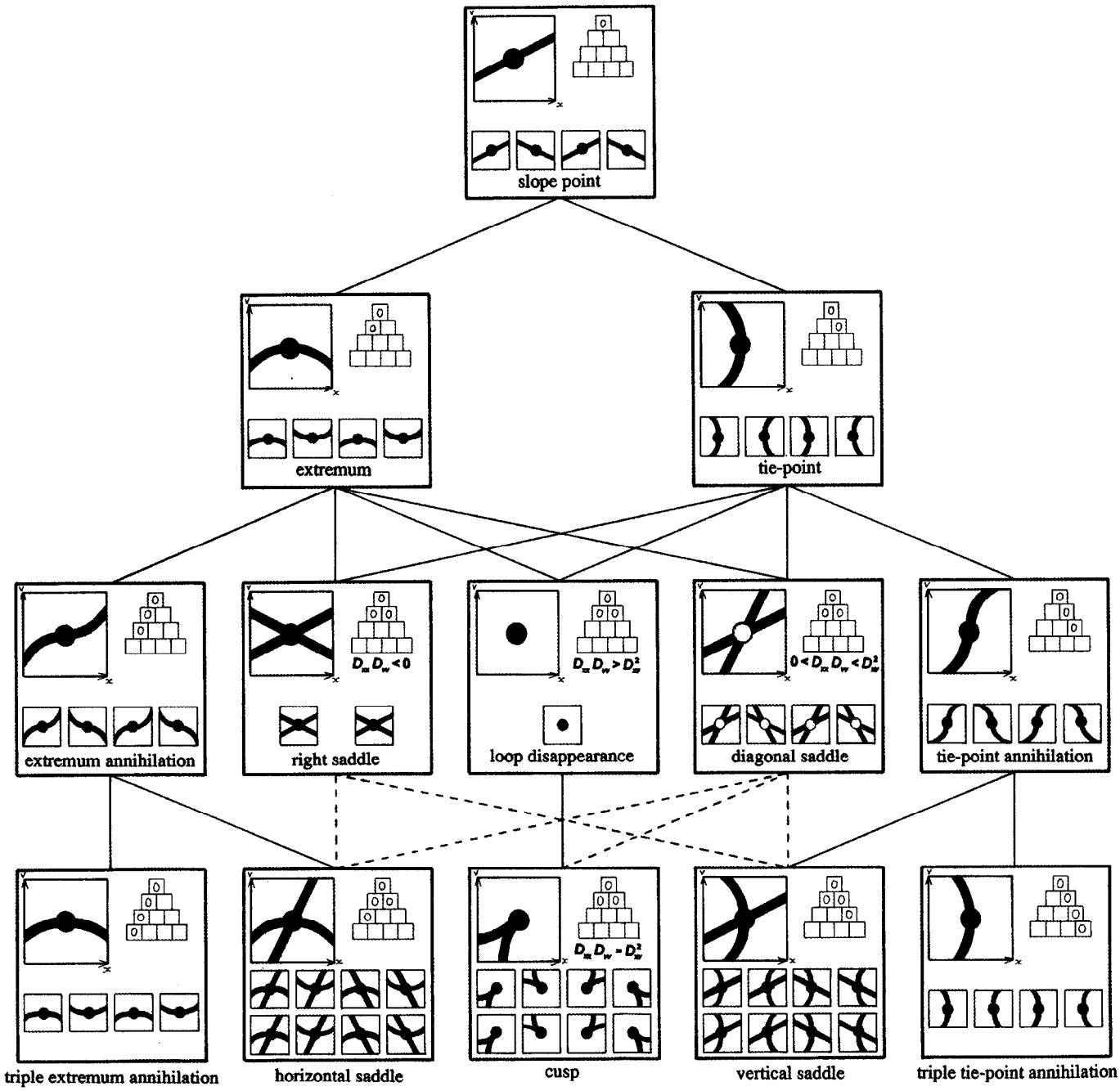


Fig. 5. The different point types that may be displayed by the augmented modal image in scale-imprecision space.

pyramid (i.e. versal-III events) as such points would be non-generic. This means that should such a point exist it could be removed from the scale-imprecision space by infinitesimal perturbation of the incident luminance. Such a process would not work for a versal-II event, as it would merely be shifted to a slightly different scale, imprecision, position and/or value but not removed from the space.

Fig. 6 shows the embeddings of the versal-I events. Firstly, on the left are *extremum annihilation* and *tie-point annihilation*, via, respectively, horizontal and vertical inflections. Referring back to Fig. 5, it can be seen that in terms of derivatives, these two events are specializations of extremum and tie-points, respectively, and so as can be seen, involve instances of these points. The large panels in the boxes of Fig. 6 show the structure of the event in scale-imprecision space. The small inset panels within this large panel are snapshots of position × value space showing the qualitative structure at this point in scale-imprecision space. Whilst the extremum annihilation event corresponds to not being able to resolve a nearby maximum and minimum as the aperture becomes too large, the tie-point annihilation event corresponds to not being able to detect a spatial region

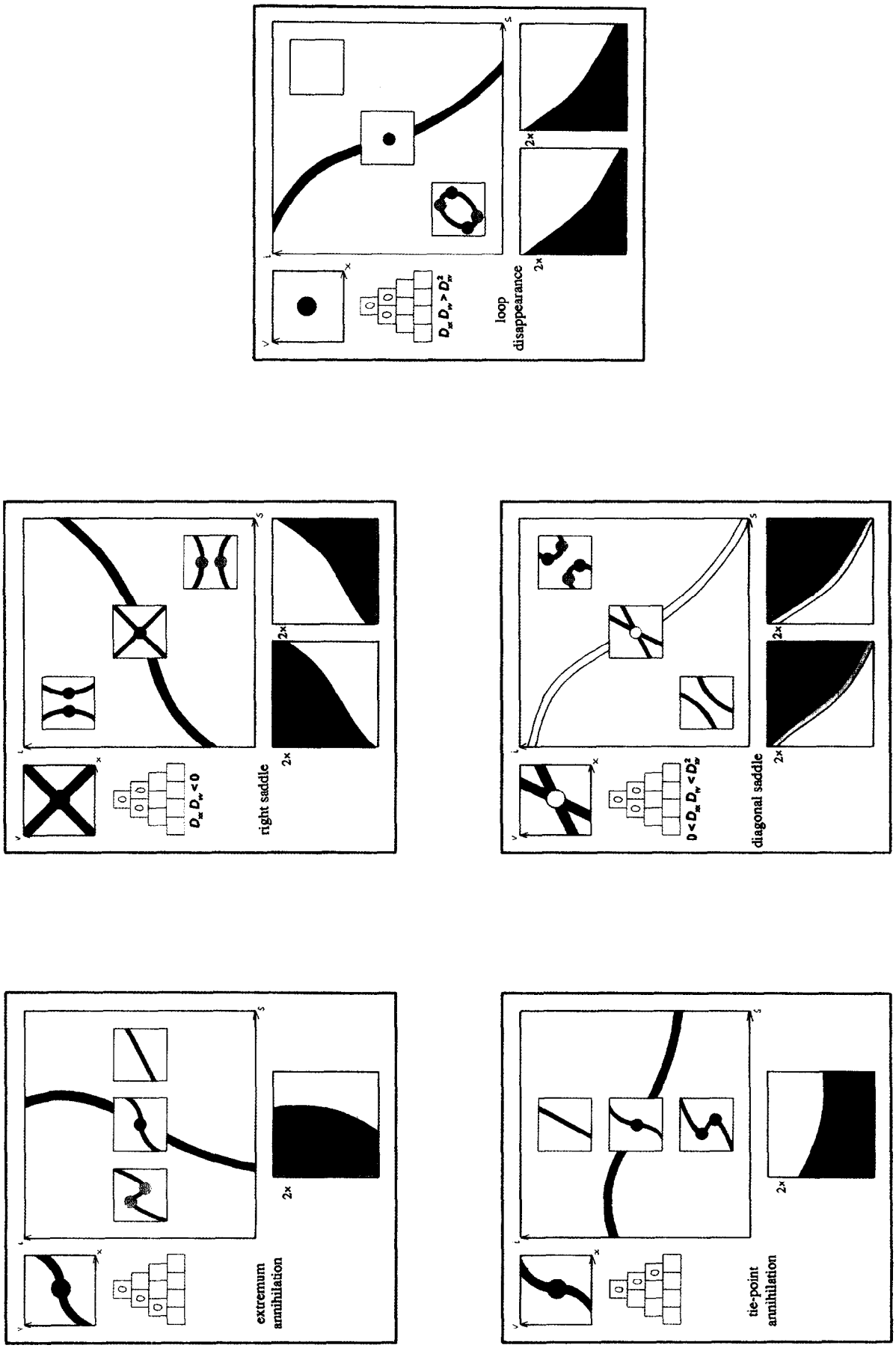


Fig. 6. The scale-imprecision embeddings of the versal-I point types of the augmented modal image.

of bi-valuedness as measurements become less precise. The annihilations occur at every point along a curve in scale-imprecision space; this is why they are (generically) missed by single-modal images, but may be encountered by a 1-D family. But note the spatial location and value of the annihilation point will vary along this curve. Focusing on the extremum annihilations, the event curve in scale-imprecision space has variable tangent but is never horizontal, and the annihilation will always occur from left to right (i.e. with increasing scale), never *vice versa*; if it were otherwise, the evolution would not be causal with respect to increasing scale. The bottom panel of the extremum annihilation box shows the areas of scale-imprecision space over which the extrema involved in the event exist. The tie-point annihilation box is very similar to the extremum annihilation; in this case, the event curve can never be vertical, and the annihilation only ever occurs with increasing not decreasing imprecision.

The other three boxes of Fig. 6 show the other three versal-I events; these correspond to the central three boxes of the third row of Fig. 5. As can be seen from this figure, these three versal-I events are specializations of both extremum and tie-points, and thus involve both types of point. Firstly, there are the two cross-shaped point types (Fig. 6, middle column); these occur at a zero-valued saddle point of D. In both events, two non-overlapping pieces of modal surface approach each other and join; or read the other way around, a single modal sheet splits into two pieces which move apart. The breakdown into two types of saddle depends on the relationship of the arms of the cross to the horizontal: I shall call them a *right saddle* and a *diagonal saddle*. The reason for this distinction is that the two points have different embeddings, as shown by the diagrams in the large panels. The curves along which these events happen can vary in tangent but not so much to attain the limits of, let alone escape from, the quadrant in which they are shown. Finally, the box on the right of Fig. 6 shows the last of the versal-I events: a *loop disappearance event*. It occurs at a zero-valued extremum of D and corresponds to a loop, formed of a stretch of mode and a stretch of anti-mode, collapsing to a point. Again, this versal-I curve must have its tangent within the quadrant shown.

Figs. 7 and 8 shows the five versal-II events. These occur at isolated points of scale-imprecision space where two or four versal-I event curves meet. The two boxes on the left of Fig. 7 show events which are specializations of extremum and tie-point annihilations (c.f. links between boxes in Fig. 5); I call them *triple extremum annihilation* and *triple tie-point annihilation*. For instance, the versal-II extremum event is a collapse of three extremum (two of one sign and one of the other), via a quartic hump, into a single extremum. As shown by the large panel in the top left box of Fig. 7 this event occurs where two extremum annihilation event curves meet in a cusp, with the tangent of each curve approaching horizontal. Along one arm of the cusp a maximum annihilates with a minimum; along the other arm, a different maximum annihilates with the same minimum. The two medium sized panels show the domains of existence of the extremum involved. The upper panel represents the central minimum which only exists to the left of the event curves. The lower panel represents the extremum that exists to the right of the event curves. The left-hand side is coloured more intensely as the extremum has a double-existence in this region; it can be traced continuously to either of the maxima to the left of the event curves–which maximum is reached depends on which event curve is crossed. The triple tie-point annihilation event is similar to the triple extremum annihilation event, as is apparent from the figure.

The third box of Fig. 7 shows an event that occurs when D has the form of a zero-valued shoe-surface (i.e. zero gradient and Hessian of rank one); I call this a *cusp event*. The small panels of the central box of the bottom row of Fig. 5 show the possible forms of this point. This is the first event that has been described that has two distinct ways in which it may be embedded in scale-imprecision space. The two embeddings are shown by the two medium sized panels in the box of Fig. 7. Both embeddings involve a diagonal saddle event curve and a loop disappearance event curve; they meet, with equal limiting tangent, in a cusp as shown. The angle of the cusp where the two event curves meet can be different from the example, but must be within the same quadrant. Referring to the figure, as one moves along the diagonal saddle event curve, towards the cusp event, the small loop shrinks to a point; whereas as one moves in a similar direction along the loop disappearance event curve, the point of the loop disappearance, the extremum and the tie point all approach each other. The domains of existence of the extremum and tie-points involved are shown in the medium sized panels. Like the triple annihilation events, there is a doubling up where an extremum or tie-point has a single existence in one part of scale-imprecision space and a double existence in another part. Note that, although the diagram for a cusp event involving a right rather than a diagonal saddle can be drawn, the tangent of the cusp of the embedding of the event in scale-imprecision space would be horizontal or vertical, and such a point would be a versal-III event and so does not occur generically.

Finally, Fig. 8 shows the last pair of versal-II events. Both events are specializations of the right and diagonal saddles, and can be understood as transitions between the two. One of them (Fig. 8 top) makes this transition by one pair of the arms of the saddle drifting across vertical; I shall call this a *vertical saddle*. The other by a pair of arms drifting across horizontal-a *horizontal saddle*. Note that the horizontal and vertical pairs of arms are shown as curves as their second order structure is significant once the tangent is fixed; this accounts for the eight

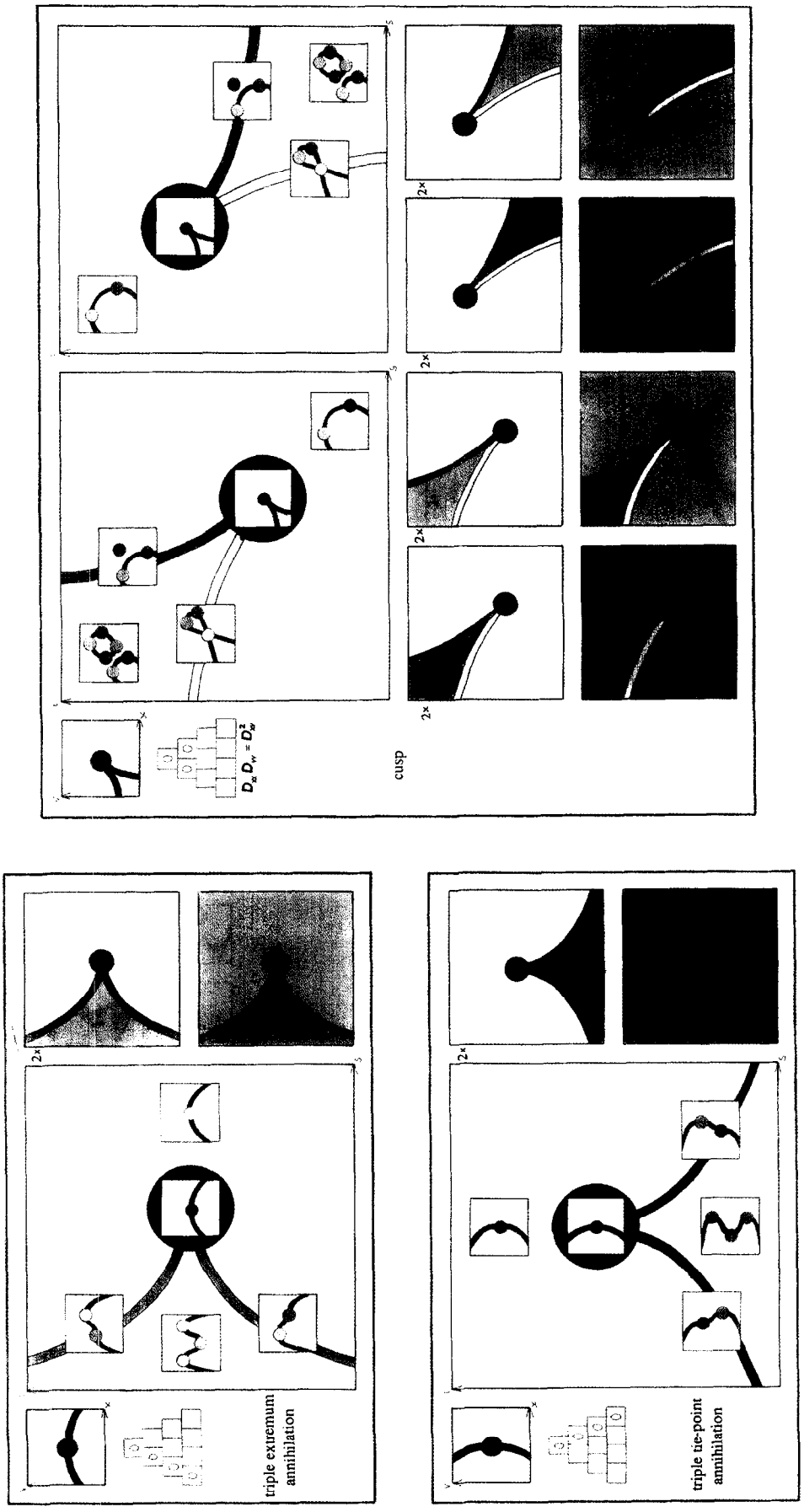
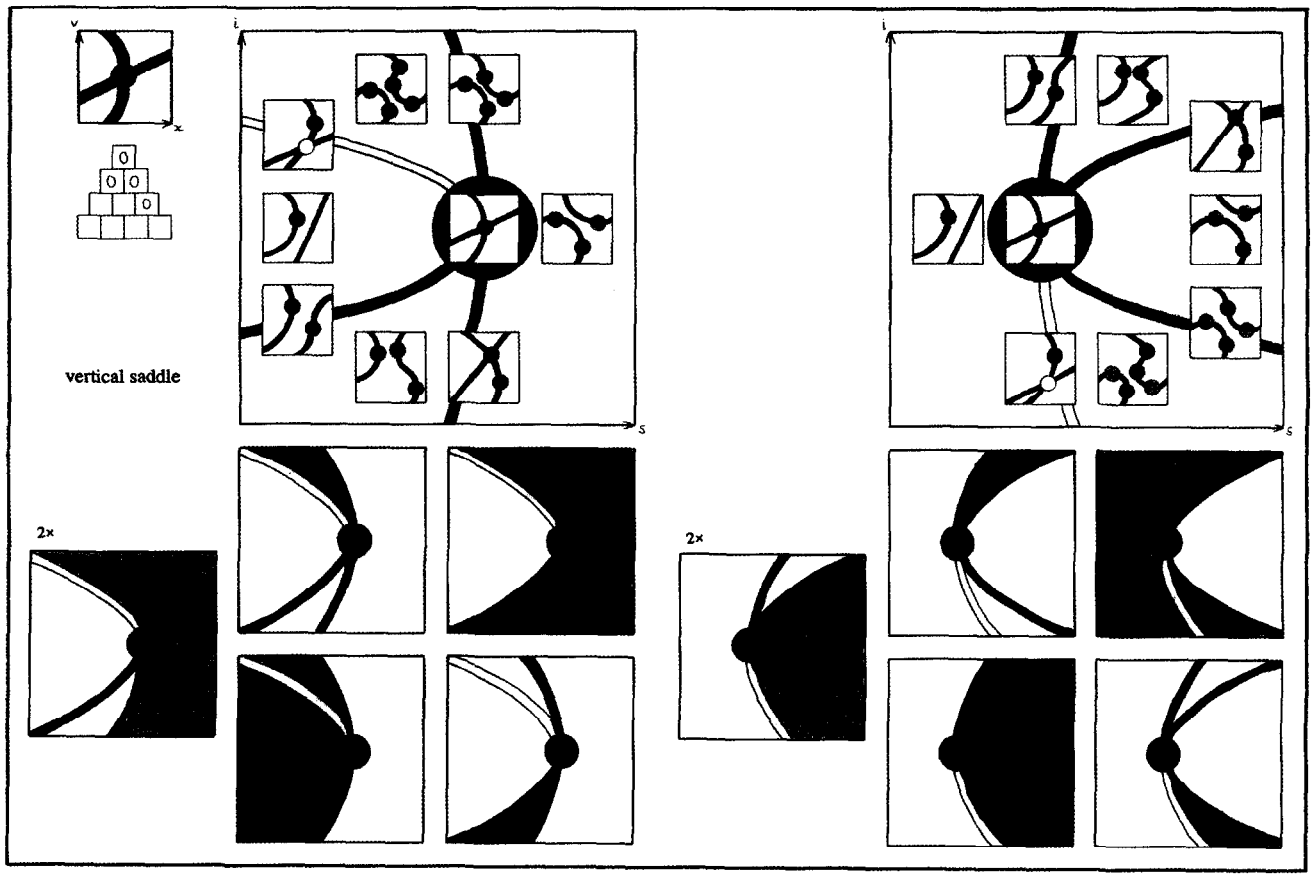
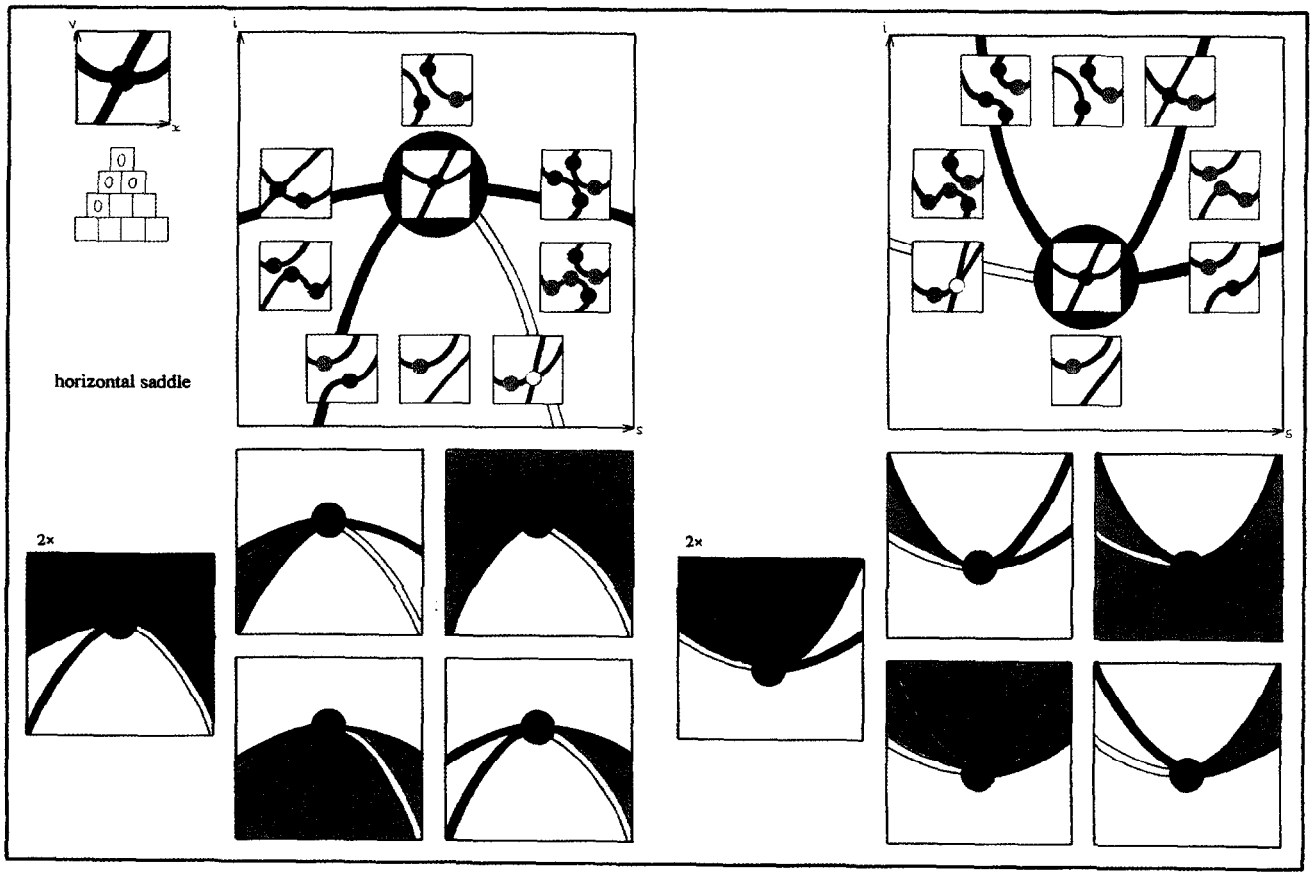


Fig. 7. The scale-imprecision embeddings of some of the versal-II point types of the augmented modal image.


vertical saddle
2×
2×


horizontal saddle
2×
2×

variations shown in the small panels of Fig. 5. Also shown in Fig. 5 is that, as well as being specializations of both types of saddle on the second row, the horizontal saddle is also a specialization of the extremum annihilation event and the vertical saddle a specialization of the tie-point annihilation event. Looking at the large panels of Fig. 8, one sees that each event occurs at the meeting of a right saddle event curve, a diagonal saddle event curve and two curves of the annihilation event of which it is a specialization. The curves all meet with an equal limiting tangent, which is vertical for the vertical saddle and horizontal for the horizontal. Both types of saddle have two possible embeddings in scale-imprecision space. These events are quite difficult to grasp in a single mental picture, but it is helpful to consider as a main backbone the transition along the saddle versal-I curves. Along this curve an extremum or tie-point drifts across the saddle, causing the transition between right and diagonal saddle.

Superficial and deep structure of the stable mode image

Fig. 4 (right) shows an example of the stable mode image. Compared to the augmented multi-modal image to its left, one can see that (i) there are no stretches of anti-mode, and (ii) there is a new point type, the discontinuity. Fig. 9 shows this new feature, and all the other types of point that may occur. Like Fig. 5, the organization by rows shows the different levels of versality of the points. Many, but not all, of the point types of the augmented modal image survive into the stable mode image. Only point types which occur at a maximum of the measurement histogram at that position can be part of the stable mode. For instance, when a tie point occurs in the augmented modal image, the measurement histogram has the form of a horizontal inflection, which clearly is not a mode let alone the stable mode, and so tie-points do not occur in the stable mode image. In terms of the Pascal's triangles shown in Fig. 5, the rule is quite simple: there must be an odd number of zeroes in the right-hand edge of the pyramid. What cannot be seen from the diagrams is that the first non-zero value in the right-hand row of the pyramid must be negative (i.e. at a maximum in the v-direction), and thus only the 'red' versions of points occur.

The layout of each of the boxes in Fig. 9 is similar to that of Fig. 5, but with two differences. Firstly, the small depictions of local forms have been reduced in number by eliminating forms which are equivalent under horizontal and/or vertical reflection. Secondly, the Pascal's triangle has been replaced with a small figure showing the topology (not value) over imprecision of the stable mode(s) at that position. In these diagrams the vertical axis is value, like the box to the left, but the horizontal axis is imprecision (increasing to the right). The dotted vertical line shows the imprecision at which the point type occurs. So considering the *slope point* (top row), this diagram shows that the stable mode is single-valued at this position, imprecision and scale and will continue to be as imprecision is increased while position and scale are held constant. The diagram may give the imprecision that the value remains constant, but this is misleading and the trace of the stable mode will vary in value with increasing imprecision, i.e. the line will wobble up and down.

Moving down to the second row. On the left is the *extremum* which also occurs for the multi-modal image. The right-hand box of the panel shows that this position will remain single-valued as imprecision is increased while scale stays constant. But note that, since the extremum will in general move with changes in imprecision, the position will not remain an extremum as imprecision is increased. In diagrammatic terms, this would mean that the horizontal line of the right-hand box of the extremum panel could be coloured red, except for the odd isolated point such as the green one shown.

The other panel of the second row shows the *discontinuity* feature. It is a point where the stable mode is bi-valued; it is drawn as two red slope points connected by a vertical dark blue line. The use of the two red points reflects the fact that the gradient of the stable mode function tends to a non-zero value (i.e. like a slope point) as the discontinuity is approached from the left or right. The limiting gradients will not in general be the same, and may be of different signs as shown in the small boxes. The vertical blue line is a graphic device, and the reader should not follow Fourier [23] and treat it as part of the function. The right-hand box shows that at this position, scale and imprecision, the stable mode is bi-valued but will necessarily become single-valued when imprecision is sufficiently increased and scale is unchanged (it will effect this transition via a tie-point annihilation, as will be seen later). What this also shows is that as imprecision is increased (without changing scale), discontinuities never move or appear: they just disappear one-by-one. This is in contrast to extrema, which wander with changes of scale or imprecision.

The third row of the pyramid of Fig. 9 shows the versal-I events of the stable mode image. With the exception of the extremum annihilation event (far left), the embeddings for these events are shown in Fig. 10. The embedding for *extremum annihilation* has been omitted as it is the same as for the multi-modal image (Fig. 9). On the right-hand end of the third row of Fig. 9 is the *tie-point annihilation* event. This is a point type that

Fig. 8. The scale-imprecision embeddings of the remainder of the versal-II point types of the augmented modal image.

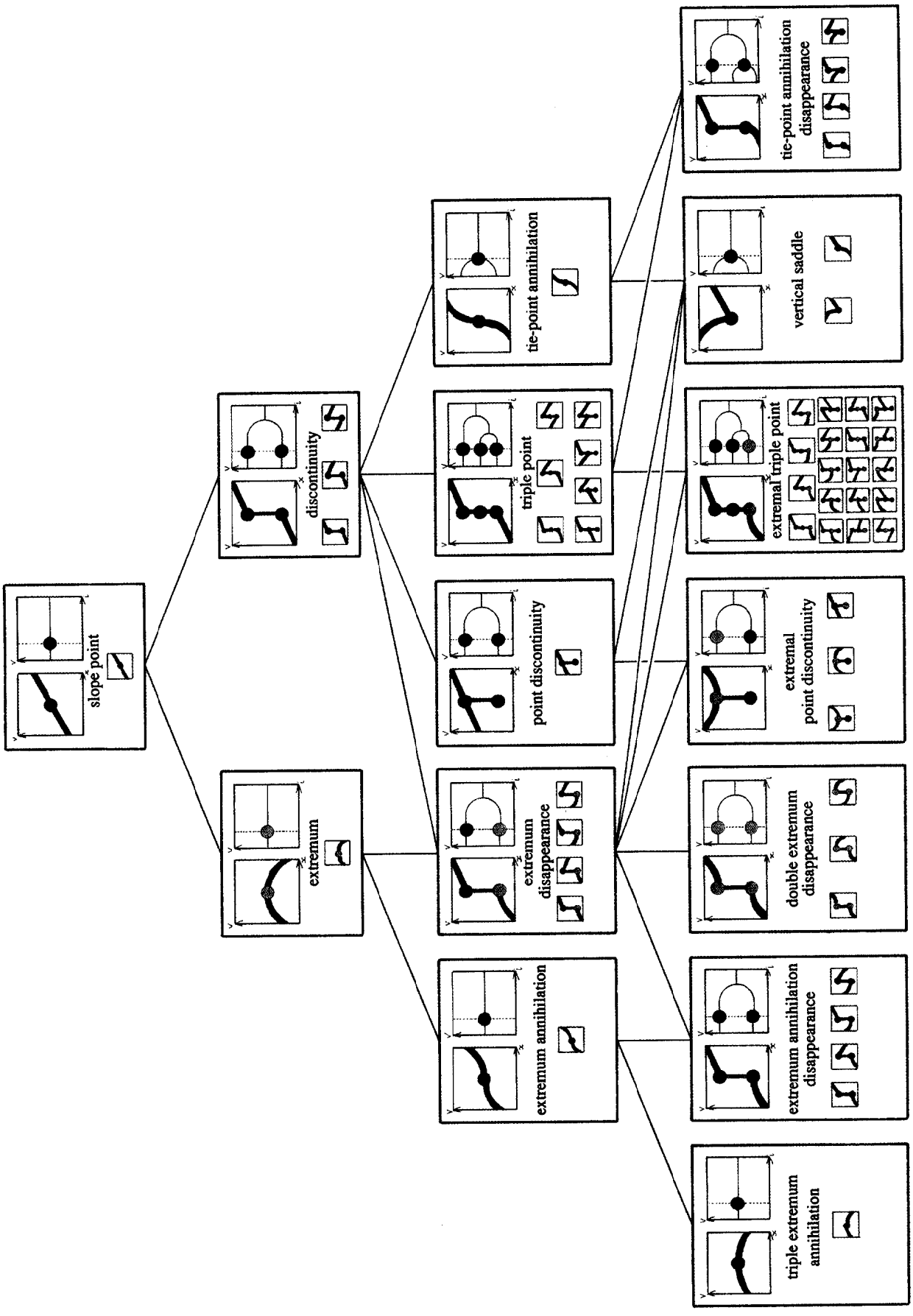


Fig. 9. The different point types that may be displayed by the stable mode image in scale-imprecision space.

occurs in the stable mode image even though tie-points themselves do not. The explanation for this can be seen by inspecting its embedding in Fig. 10 and comparing it with the embedding in Fig. 6. The comparison shows that the embeddings are almost identical, except that to the lower imprecision side of the curve there is in one case two tie-points and in the other a single discontinuity. The bottom box of the tie-point annihilation panel in Fig. 10 shows the domain of existence of the discontinuity.

The other three point types in the third row of Fig. 9 are new for the stable mode image. On the left is an extremum disappearance. This occurs as an extremum slips beneath a discontinuity. In the middle of the third row is a point discontinuity which is a specialization of discontinuity (the point is bi-valued, but the image has equal left and right limits); this will be seen to occur at the annihilation of two discontinuities. Finally, there is the triple point. This also will be seen to be the result of the collision of two discontinuities. The right-hand box of the triple point panel shows that, as imprecision is increased, the point will at some point change from tri-valued to bi-valued (a discontinuity), and eventually to single valued. Turn again now to Fig. 10 to see the embeddings for these three events.

Looking first at the panel for *extremum disappearance*, one can see the reason for its name. This versal curve, unlike any of the others described so far, is completely unrestricted in its embedding in scale and imprecision, and so an extremum is able to disappear under an edge with increasing or decreasing scale or imprecision. Note that, as well as marking the disappearance of an extremum, it also marks a change between sub-types of discontinuity (the small diagrams in the discontinuity panel of Fig. 9). This is significant, because looking across to the tie-point annihilation you will see that the event can occur with only one of the three discontinuity sub-types, i.e. the one whose left and right limits of the gradient are of the same sign. The final two versal-I curves are highly restricted in their embedding in scale and imprecision: they are both always vertical. The *point discontinuity* event is also constrained in the scale direction (i.e. the event shown cannot play out for decreasing scale as it does for increasing). The event corresponds to the disappearance, with increasing scale, of a small dark object on a lighter background, or *vice versa*. Note that the gradient of the disappearing object can be different in sign from the gradient of the background, in contrast to what is shown. The *triple point* event, unlike the point discontinuity event, is not constrained in the scale direction, and so has two possible embeddings, as shown. In the left embedding an object sandwiched between two objects disappears; in the right an object appears between two objects. Note that, although in the left embedding the disappearing object corresponds to the central value of the triple point and in the right to the top value, this distinction can in fact happen with either direction of embedding.

Now I will consider the final row of Fig. 9, i.e. versal-II events that occur at isolated points of scale-imprecision space. I will deal with them from left to right. The first event is the *triple extremum annihilation*, already encountered for the multi-modal image; its embedding was given in Fig. 7 and so has not been repeated. Next is the *extremum annihilation disappearance* whose embedding is shown in Fig. 12. It occurs when an extremum annihilation disappears by slipping out of sight beneath a discontinuity. Three versal-I curves are involved in the embedding: two extremum disappearances and an extremum annihilation. The two extremum disappearance curves meet in a cusp, and this can make any angle with the extremum annihilation curve, e.g. it can be an obtuse angle rather than the acute angle shown. The embedding is easiest to grasp by treating the extremum annihilation curve as a backbone; seeing it slip under the discontinuity as the versal-II event is approached; and appreciating how the two extremum disappearance curves are dragged in as an inevitable consequence.

Third is the *double extremum disappearance* event whose embedding is shown in Fig. 11. It occurs at the crossing of two extremum disappearance curves; one affecting the top point of the discontinuity and the other the bottom. The two event curves can cross at any angle.

The central panel of the third row of Fig. 9 shows the *extremal point discontinuity*; the embeddings are shown in Fig. 11. It has two distinct types of local form; they have identical embeddings in terms of the versal-I curves, but they differ in the domain of existence of the extremum that is involved. Note that the cusp formed by the two extremum disappearance curves, although constrained to be to the left of the point discontinuity curve (or rather, vertical line), can make any angle with the vertical.

Next is the *extremal triple point* event, which occurs when an extremum drifts across a triple point. There are four distinct embeddings, as shown in Fig. 11, which fall into two classes. One class (the left pair) occurs when an extremum drifts across one of the two points of the triple point which are left or right limits of the stable mode image. This is a fairly insignificant type of event, involving a vertical triple point curve and an extremum disappearance curve which cuts across it at any angle. The other type of embedding is more interesting. It occurs when the disappearing object of the triple point event has an extremum within it. In this case the extremum disappearance curves form a cusp. Each of the two types of embedding has two versions, reflecting the fact that a triple point event can be embedded either way around with respect to scale.

The next versal-II event is the stable mode image manifestation of the *vertical saddle* already encountered for the multi-modal image. If you compare Figs. 12 and 8, you will see that it has different local forms in the two

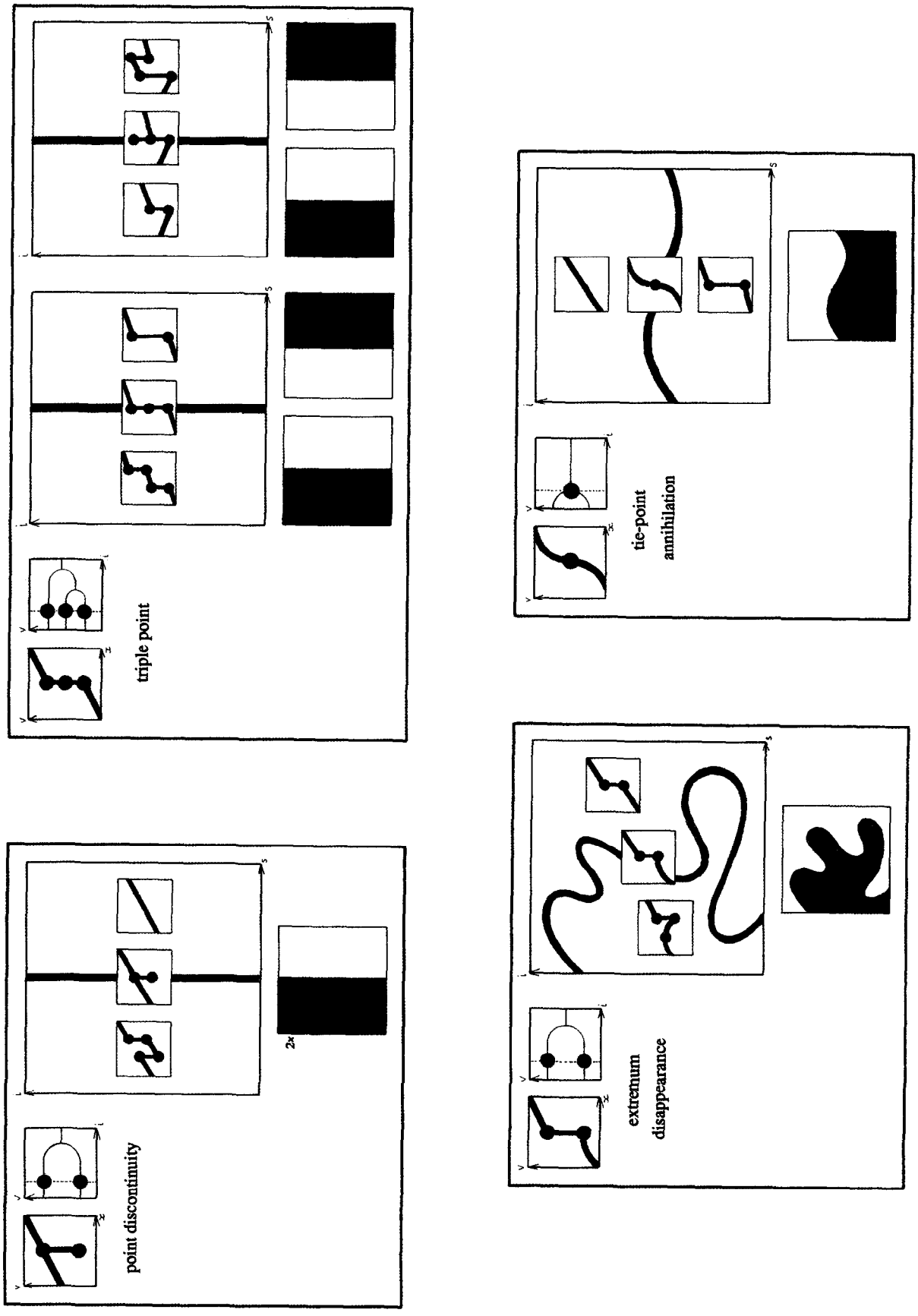


Fig. 10. The scale-imprecision embeddings of some of the versal-I point types of the stable mode image.

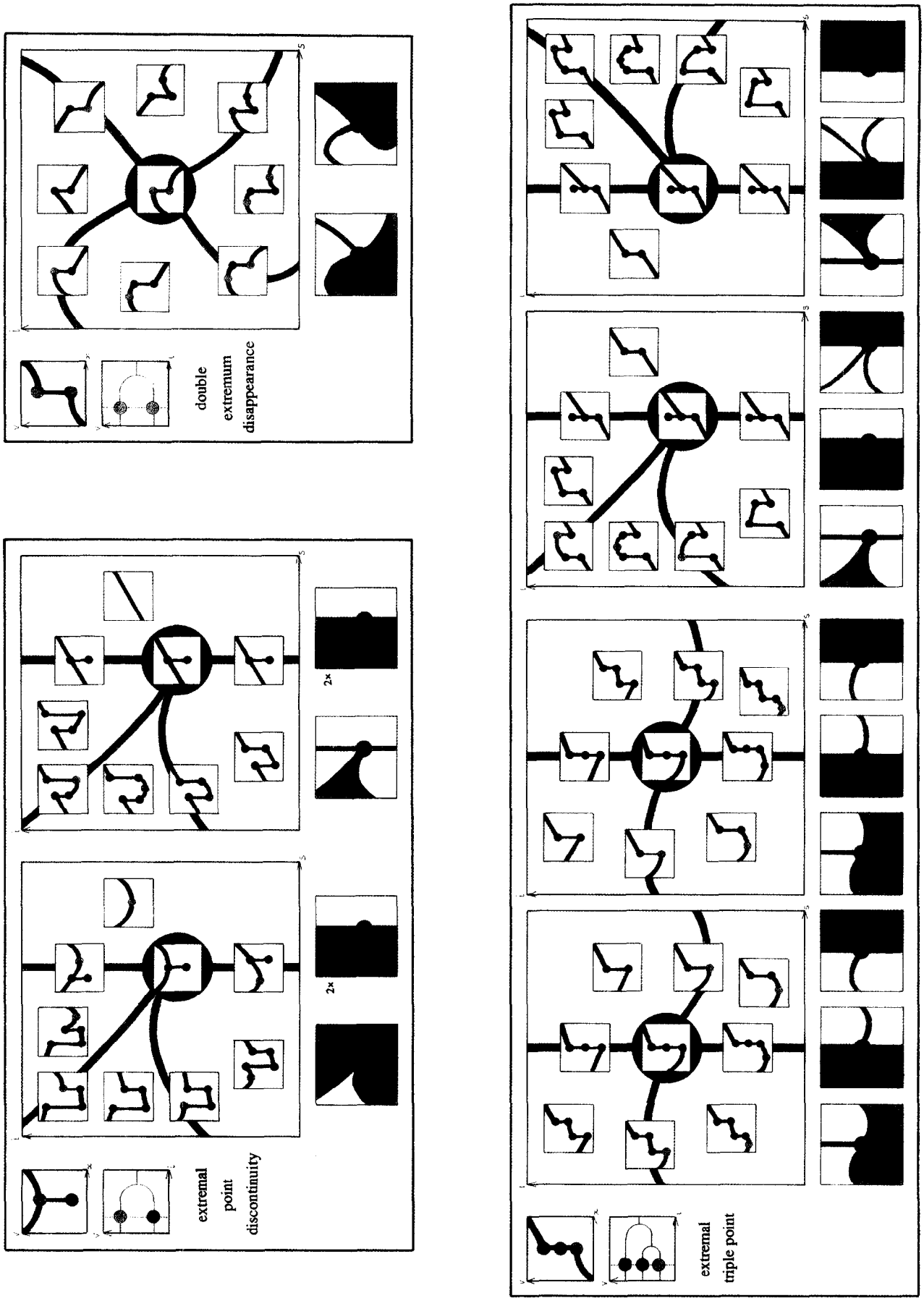


Fig. 11. The scale-imprecision embeddings of some of the versal-II point types of the stable mode image.

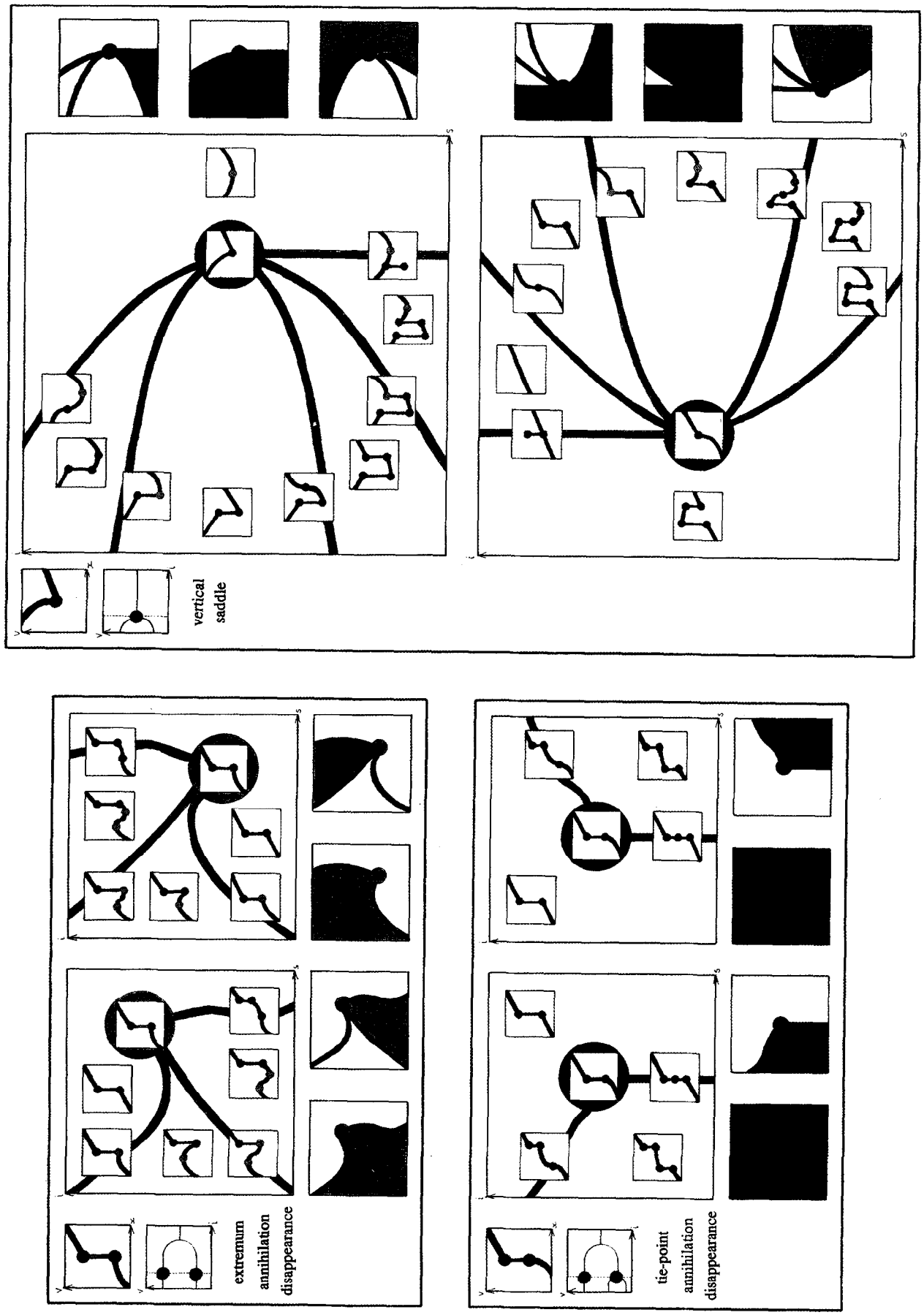


Fig. 12. The scale-imprecision embeddings of some of the versal-II point types of the stable mode image.

images. In fact in the stable mode image it has two distinct local forms, corresponding to which of the two branches of the vertical tangent part of the saddle is the stable mode (the other being an ordinary mode), and these different local forms have different embeddings as shown. This is not a case where either local form can occur with either embedding. These embeddings are quite difficult to grasp as a gestalt, but a couple of pointers may help. The upper embedding is most easily read as two sequences, one reading clockwise and the other counter-clockwise. Both sequences start at the region in the bottom left between a tie-point annihilation curve and an extremum disappearance curve, and both end in the large region to the right. The lower embedding can be read in a similar manner, but the sequence starts in the large region and ends in the region between the point discontinuity curve (vertical) and a tie-point annihilation curve. A final note of caution. When comparing the embeddings with those in Fig. 8, do not make the mistake of thinking that the extremum disappearance curves are at the same position as the saddle event curves (which remember cannot be part of the stable mode image). Careful examination will show where they do in fact lie.

The final versal-II event is the *tie-point annihilation disappearance*, the two possible embeddings of which are shown in Fig. 12. It is another incidence of a versal-II event being caused by a versal-I event slipping beneath a discontinuity. The reason that there are two embeddings is because of the reversibility (with respect to scale) of the triple point event. Note that the curve of the tie-point annihilation can make any angle with the vertical triple point curve; but that the versal-II event must top the triple point event curve, it cannot be at the lower end of it (cf. the embeddings of the extremum annihilation disappearance). Also note the lower boxes showing the domains of existence of the two discontinuities involved. The left-hand box of each pair appears to be superfluous as the discontinuity exists all around the versal-II event, but although it can be tracked continuously across the tie-point annihilation curve, it is not continuous across the triple point curve.

Examples

In this section I present modal images for some standard functions and on a medical image. The first function examined is a step function. The modal images are shown in the top half of Fig. 13. As can be seen, the only event is a tie-point annihilation. In the stable mode image this corresponds to not being able to detect the discontinuity when the imprecision is too high. Note that the annihilation curve is horizontal in scale-imprecision space, and so the imprecision at which the discontinuity cannot be detected does not depend upon scale; this is not surprising given that the original step function is unchanged by being spatially shrunk or expanded. In the bottom half of Fig. 13 the behaviour of a ramp edge function is shown. This also (perhaps surprisingly) exhibits a tie-point

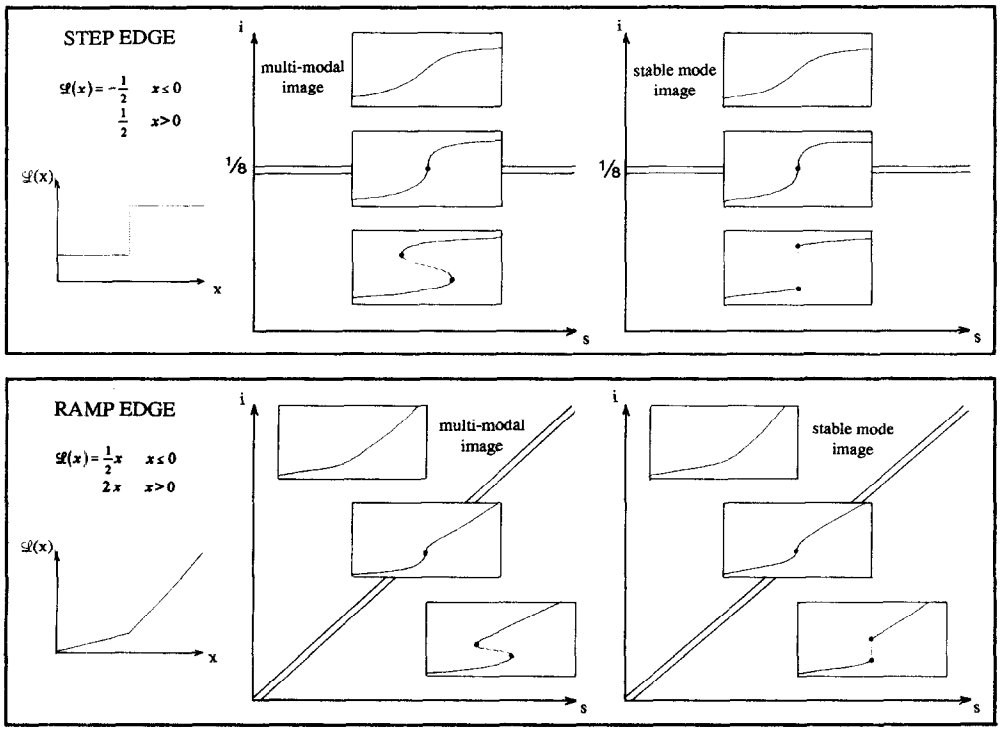


Fig. 13. The scale-imprecision behaviour of the modal images of the step and ramp edge.

annihilation. This might be taken as meaning that discontinuities can occur in scale-imprecision space, not only when there is a discontinuity in the original functive, but even when there is a discontinuity in the first derivative while continuity in the function itself. However, things are not quite so simple, as no discontinuity occurs when the change in the first derivative is such as to cross zero (e.g. $|\mathbf{x}|$).

The next example is a Gaussian function, shown in Fig. 14. Several different catastrophes occur with this function; most of them are actually non-generic due to the symmetry of the Gaussian, but their nature is quite clear. The left-most catastrophe curve is non-generic because two tie-point annihilations occur simultaneously. The other catastrophe curves are non-generic because, throughout their length, there is also an extremum at that position. Actually they are only non-generic or linear, such events occur as versal-II events at isolated points. They were omitted, for brevity, in the main treatment. Finally, the versal-II catastrophe (enclosed in an ellipse) is non-generic, as it is a cusp catastrophe involving a right saddle. The example of the Gaussian shows that discontinuities may occur even when the original function is continuous. Given that any discontinuous function may be approximated arbitrarily closely by a continuous function, it would be an unattractive property of scale-imprecision space if this were otherwise. The figure also shows that the catastrophe structure of the stable mode is much simpler than that of the augmented modal image, as catastrophes effecting non-stable modes are irrelevant. Focusing on the stable mode image, the figure shows that a Gaussian is always measured as continuous except in a certain bounded portion of scale-imprecision space. In this region a pair of discontinuities isolating the central hump of the Gaussian will be detected. It is important to realize that there are infinite stretches of imprecision (above) and scale (to the right) where the Gaussian is also measured as continuous. In fact, if one looks at the (approximate) figures marked on the axis, it is clear that the scale and, particularly, the imprecision must be small compared to the Gaussian in order to detect these discontinuities, i.e. it must be big and bright.

The final synthetic example is for a sinusoid (see Fig. 15). In this case, all of the catastrophes are non-generic, since the symmetry and the periodicity of the sinusoid guarantees that any catastrophe will occur at multiple spatial locations. Focusing first on the augmented modal image, the surprising part of the diagram is the small region directly below the versal-II catastrophe. This can perhaps be explained by realizing that the separation of the flat parts and the sloping parts of the sinusoid is very similar to the discontinuity that occurred for the ramp edge. Looking now at the stable mode, there are three regions: a region of continuity; a region where a discontinuous, positive then negative then positive, function is measured; and a small region where ascenders and descenders are also detected. Again, notice the (approximate) figures on the axis show that these details are only detected at scales and imprecisions that are small compared to that of the sinusoid.

Fig. 16 shows the computed discontinuities of the stable mode image at different scales and imprecisions, scale increasing to the right and imprecision upwards, overlaid over the original image (see the appendix for algorithmic details). This is the same original image, with scales and imprecisions as shown in Fig. 3; thus the discontinuities shown in red can be compared to the discontinuities discernible in the final column of that figure. Calculating these discontinuities is difficult, because an algorithm must operate in the discrete domain whereas the definition of the discontinuities is phrased within the continuous domain. The primary difficulty is not calculating the stable mode, which can be done quite easily and reliably, but properly linking together the modes at different positions to form continuous sheets. From visual inspection of the results I estimate that 10% of the edgels are incorrectly present due to incorrect linkage and a similar number missing. The figure illustrates that increasing scale or imprecision leads to fewer edges. With an increase in imprecision (top row vs. bottom) the edges that remain have not moved in position. With an increase in scale there is the expected movement of edges (e.g. rounding of sharp features); the scale of the images may be assessed by looking at the mean images of Fig. 3 (second column). It is also possible to calculate edges for the global mode rather than the stable mode. I have not presented the results, but they are of noticeably poorer quality. The points of disagreement occur when the stable mode and the global mode of the measurement histogram at a point are not equal. In all such cases examined, the stable mode appears to be a more representative value than the global mode, in the sense of being supported by a greater bulk of the histogram.

Summary and concluding remarks

I will summarize the course of the paper. It was observed that scale-space led from two considerations: (i) image measurements are determined by a non-zero area of the visual field; (ii) it is desirable to have the property–what can be seen in a shrunk version of the image can be seen in an expanded version. Next, I analogically, claimed that (a) image measurements are of non-zero imprecision; (b) it is desirable that what can be seen in a lower contrast version of an image can be seen in a high contrast version. Together these two pairs of demands required a 2-D scale-imprecision space for their satisfaction. I argued that the way to achieve this was to consider blurring the graph of the incident luminance; with horizontal blurring being equivalent to an

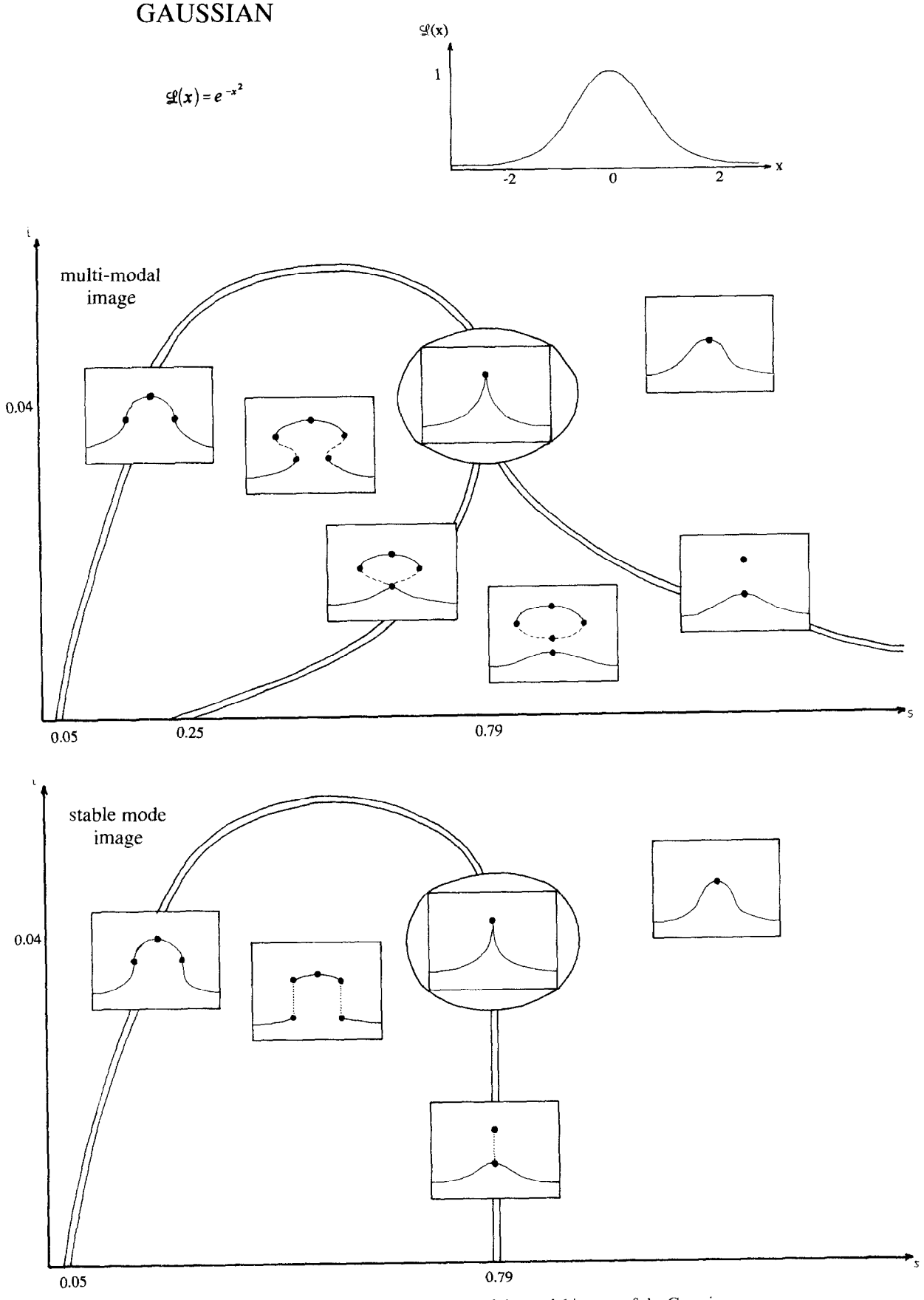


Fig. 14. The scale-imprecision behaviour of the modal images of the Gaussian.

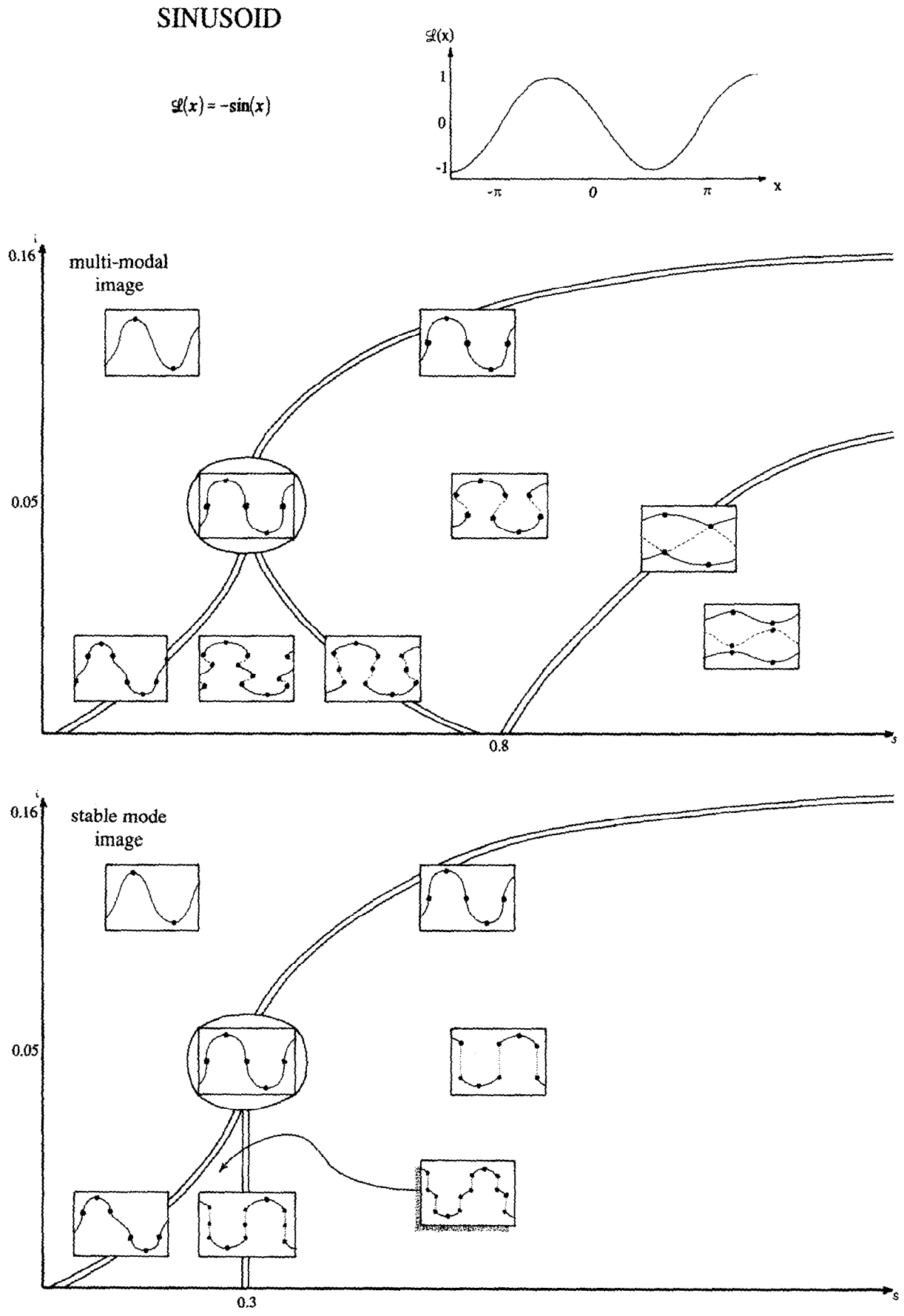


Fig. 15. The scale-imprecision behaviour of the modal images of the Sinusoid.

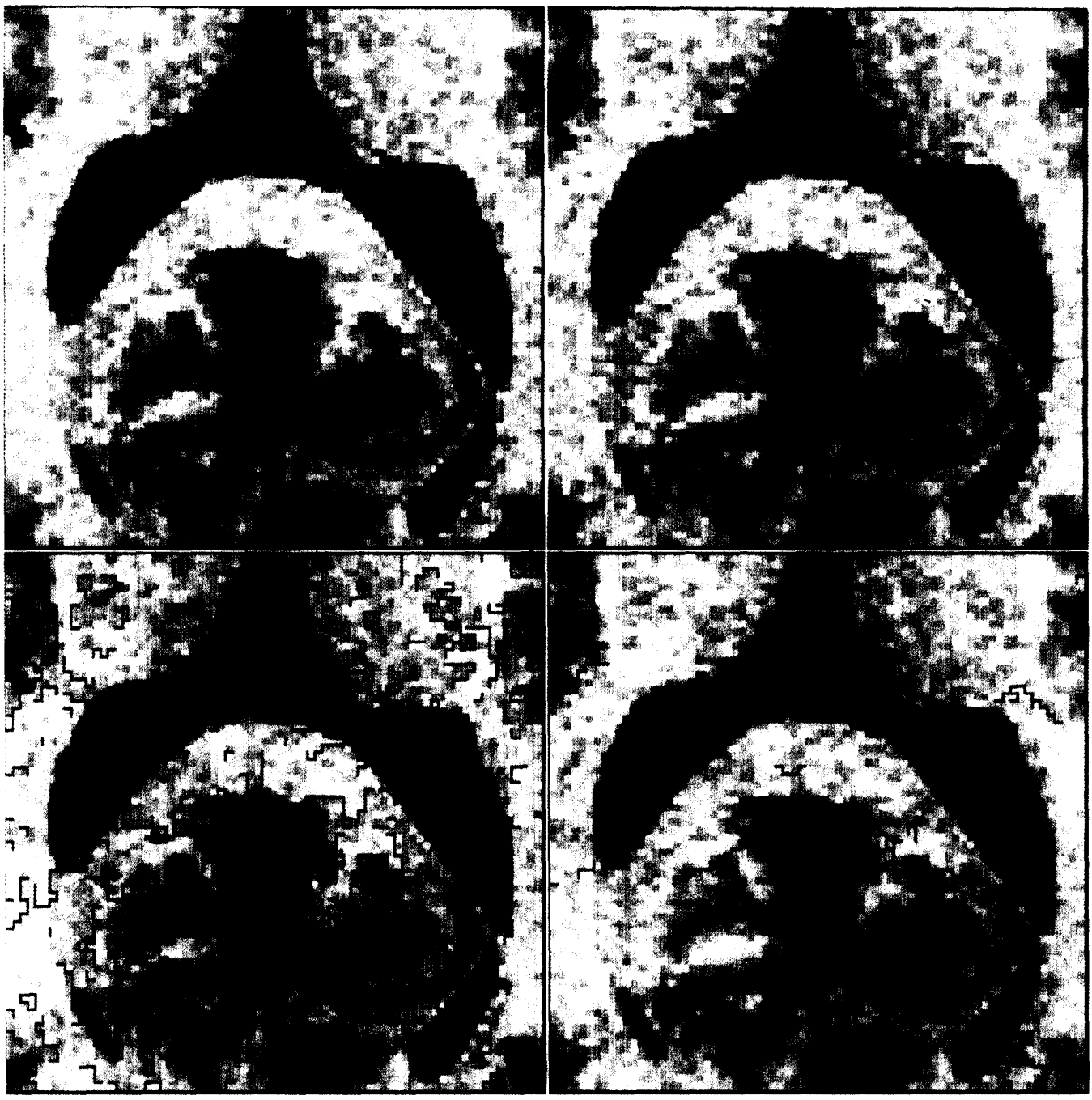


Fig. 16. Algorithmically computed discontinuities of the stable mode calculated for a transaxial MR image of the brain.

increase in scale, and vertical blurring to an increase in imprecision. The next step was to define summarizing images (i.e. a finite number of values, rather than a histogram, for each position) from these blurred graphs. I identified several possibilities. Firstly, the mean image, where the mean of each column of the blurred graph was taken as representative: it was shown that for this image the imprecision was irrelevant and a normal scale space was generated. Secondly, the median image, where the median value of each column is taken as representative: it was seen that for this image the imprecision was relevant and, compared to the mean image, destruction of detail with increasing scale was less rapid. Also shown was that, although the median image is non-linear with respect to the incident luminance, it was still continuous and causal; and that at infinite imprecision the median image is equal to the mean image. Lastly, images based on the modes were considered. The multi-modal image was defined as taking the values of all the modes of the histogram at a position. Again, this image was shown to

evolve causally with scale and to contain the mean images at infinite imprecision. However, the multi-modal image is generally multi-valued, and so it was suggested that a single mode should be selected. Two alternatives, the global and stable modes, were defined and considered. The stable mode was chosen over the global mode for two reasons: (i) it is often, subjectively, a better representative of a histogram than the global mode; (ii) it has very nice behaviour with respect to changing imprecision; (iii) it leads to better looking edges than does the global mode. In addition, in the appendix a conjecture is presented, which if true, leads to another attractive property of the stable mode. The rest of the paper dealt with the properties of the stable mode image (for 1D images), but to develop these fully it was necessary to explore the behaviour of the multi-modal image.

The most attractive property of the stable mode image is that it has discontinuities which, through consideration of how they form and through examination of examples on actual images, are candidates for perceptual edges. The reader may wonder why discontinuities occur in the scale-imprecision framework but not in the scale framework. My answer is to quote the classic Cauchy [24] ϵ-δ definition of continuity of a function:

f is continuous at x_0 if and only if
$\forall \epsilon > 0\ \exists \delta > 0\ \forall x |x - x_0| < \delta \Rightarrow |f(x) - f(x_0)| < \epsilon$

Loosely, ϵ and δ may be identified as, respectively, imprecision and scale parameters; both are needed for a definition of continuity.

It was shown that, generically, a 1-D stable mode image consists of infinitely differentiable stretches separated by discontinuities. As imprecision is increased, while holding scale fixed, these discontinuities do not change their location. Each discontinuity will disappear when the imprecision is high enough but none ever appear with increasing imprecision. In contrast, with changing scale, the discontinuities change their position and may appear or disappear. For a 2-D image the discontinuities will, generically be smooth arcs. These arcs may be closed or open. If they are open, they will end at endpoints or T-junctions. The full details of the catastrophe structure of the 2-D image has not been worked out and although it may contain some quite complicated events it will still be the case that: discontinuities will not change location with imprecision, and the discontinuities at a high imprecision will always be a subset of the discontinuities at a lower imprecision.

Acknowledgements

This research was funded at the Department of Neurology, Guys Hospital, London by the EPSRC (then SERC) and by bridging funding from the Wellcome Trust; at the Epidaure Group of INRIA-Sophia, France by an EC Human Capital Mobility Fellowship.

Appendix

Theorem: *As imprecision tends to infinity, the median image tends to the mean image:*

$$\lim_{i\to\infty} m_{s,i} = \mu_s$$

Proof

It is sufficient to consider the Gaussian blurring of a single histogram p, where p has the properties that the histograms $L_{s,i}|\vec{x};\cdot]$ have as a consequence of the incidence luminance $\mathcal{L}$ being bounded by a polynomial, i.e. $|\int_{v\in\mathbb{R}} v^r \cdot p(v)| < \infty$. So it must be shown that $\lim_{w\to\infty} m[G_w \otimes p] = \mu[p]$.

Assume, without loss of generality, that $\mu[p] = 0$; so $\int_{v\in\mathbb{R}} v \cdot p(v) = 0$. Define

$$s_w(v) = \sqrt{\pi w}\left[\int_{-\infty}^{v} (G_w \otimes p)(r)dr - \int_{v}^{\infty} (G_w \otimes p)(r)dr\right]$$

so that $s_w(v) = 0 \Leftrightarrow v = m[G_w \otimes p]$. If it can be shown that $[\lim_{w\to\infty} S_w(v) = 0] \Leftrightarrow v = 0$, then the theorem is proved. So

$$\begin{aligned}
S_w(v) &= \sqrt{\pi w}\int_{r\in\mathbb{R}} p(r)\left[\int_{-\infty}^{v} G_w(c-r)dc - \int_{c}^{\infty} G_w(c-r)dc\right] \\
&= \sqrt{\pi w}\int_{r\in\mathbb{R}} p(r)\int_{r-v}^{v-r} G_w(t)dt \\
&= \int_{r\in\mathbb{R}} p(r)\left[(v-r) - \frac{(v-r)^3}{12w} + \frac{(v-r)^5}{160w^2} - \cdots\right] \\
&= v - \frac{1}{12w}\int_{r\in\mathbb{R}} (v-r)^3 p(r) + \frac{1}{160w^2}\int_{r\in\mathbb{R}} (v-r)^5 p(r) - \cdots
\end{aligned}$$

Thus $\lim_{w\to\infty} S_w(v) = v$. **Q.E.D.**

Theorem: *As imprecision tends to infinity, the modal images tend to the mean image:*

$$\lim_{i\to\infty} H_{s,i} = \lim_{i\to\infty} h_{s,i} = \{\mu_s\}$$

Proof

It is not necessary to consider h and H separately, since for sufficiently large i they will both be single-valued and equal. It is sufficient to consider the Gaussian blurring of

a single histogram p, where p has the properties that the histograms $L_{s,i}[\vec{x}; \cdot]$ have as a consequence of the incident luminance $\mathcal{L}$ being bounded by a polynomial, i.e. $|\int_{v\in\mathbb{R}} v^r \cdot p(v)| < \infty$. So it must be shown that $\lim_{w\to\infty} h[G_w \otimes p] = \mu[p]$. Assume, without loss of generality, that $\mu[p] = 0$; so $\int_{v\in\mathbb{R}} v \cdot p(v) = 0$. Define

$$P_w : \mathbb{R} \to : \mathbb{R}, \quad P_w = 4\sqrt{\pi w^3} \cdot \frac{\partial}{\partial v}(G_w \otimes p)$$

so that $P_w(v) = 0 \Leftrightarrow v = h[G_w \otimes p]$. If it can be shown that $[\lim_{w\to\infty} P_w(v) = 0] \Leftrightarrow v = 0$, then the theorem is proved. So

$$\begin{aligned} P_w(v) &= 4\sqrt{\pi w^3} \int_{r\in\mathbb{R}} G'_w(v-r) \cdot p(r) \\ &= \int_{r\in\mathbb{R}} p(r) \\ &\quad \cdot (r-v) \cdot \left[1 - \frac{(v-r)^2}{4w} + \frac{(v-r)^4}{32w^2} - \cdots\right] \\ &= -v + \frac{1}{4w}\int_{r\in\mathbb{R}} (v-r)^3 \cdot p(r) \\ &\quad - \frac{1}{32w^2}\int_{r\in\mathbb{R}} (v-r)^5 \cdot p(r) + \cdots \end{aligned}$$

So $\lim_{w\to\infty} P_w(v) = -v$. **Q.E.D.**

A conjecture with implications for the stable mode

The first point to realize is that the mean, median and modes of a probability distribution function p can be expressed within a single framework by considering the local minima of $|x|^r \otimes p$ [25, 26]; I will call these r-modes. There are two subtleties with this definition that must be addressed. Firstly, the definition only works for $r > 0$, but may be extended to $r = 0$ by considering the limit of the positions of the minima as $r \downarrow 0$. The second problem occurs for $r < 1$. For $r \geq 1$ there is only one minimum, but for $0 \leq r < 1$ there may be several. The problem is that for $0 \leq r \leqslant 1$ these minima may be interval-like; but this possibility may be removed by assuming that p is a Gaussian blurred version of some other distribution. This is clearly legitimate for my purposes, as I am only interested in non-zero imprecision. The relationship of this definition to the standard definitions of central location is that: for $r = 2$ one obtains the mean, for $r = 1$ the median and for $r = 0$ the modes.

The second point to realize is that the family of functions $|x|^r \otimes p$ is continuous, and so the r-modes evolve continuously with r, but they may be involved in catastrophes. Assuming p is generic, the only catastrophes that can occur are the mutual annihilation or creation of an r-mode and an r-anti-mode (i.e. a local maxima of $|x|^r \otimes p$). From (i) there will be as many 0-modes as p has modes, and (ii) there can only be one r-mode for $r \geq 1$, it follows that r-modes can disappear as r is increased.

I conjecture that *r-modes are never created with increasing r*. If this is true, two things follow: (1) a unique mode always exists that is singled out by the fact that it never disappears with increasing r, and instead evolves continuously for all r; becoming the median at $r = 1$ and the mean at $r = 2$. Let this mode be called the *continuous mode*; (2) *the continuous mode and the stable mode are identical*. Hence, if the conjecture is true the stable mode is definable purely by reference to the classic definitions of central location, and with no mention being made of Gaussian blurring.

If p is versal of degree n the stable mode may be up to $n + 1$-valued. It also follows from the conjecture that the continuous mode will be identically multi-valued.

Computing measurement functions and summarizing images

The theory as presented is continuous in the sense of disregarding spatial sampling and grey-level quantization. However, to produce some of the results I needed a discretized implementation which is described below. There is as yet no theory on how to do this 'properly' so what is presented is *ad hoc* and unoptimized.

To calculate a measurement function for a 2D image I use a three-dimensional array of floating point numbers, initialized to zero, to represent position × value space. Two of the array's dimensions are the same as the image dimensions, the third dimension (a column over each pixel) is (say) 128. I first create the ideal measurement function by setting one cell in each column to 1.0. To choose which cell, I map the grey levels of the original image into the central 64 cells of that dimension (the surrounding 32 cells at either end being for padding). The volume is then Gaussian blurred by a three-dimensional kernel. This results in a measurement function of non-zero scale and non-zero imprecision. To calculate the summarizing images, each column (which is now a histogram) is examined in turn. The mean and median of a histogram are calculated in the obvious manner, the modes are detected by looking for maxima in the histogram (i.e. a cell with more weight than its two neighbours in the column). This gives the mean, median and multi-modal image. The stable mode image is more complicated.

To calculate the stable mode of a histogram I first blur the histogram repeatedly until it is uni-modal (the result of each step of blurring being saved in memory). I then attempt to link the mode at the most blurred level to a mode in the level below. To do this linking, I look at the modes and anti-modes of the level below. If the mode/anti-mode in the lower level that is closest to the mode in the upper level is in fact a mode then the link is satisfactory, if not I produce another level of blurring intermediate between the two blurring levels and repeat the linking process to this new level (similar to ref. [27]). Once a link

is established, I then proceed to link to the next level (possibly with further creation of new blurring levels). Eventually, the original histogram is reached and one of its modes is identified as the stable mode by the chain of linkages reaching up to the stable mode.

The above allows computation of the stable mode image, but does not locate discontinuities. To locate these it is necessary to consider each pair of neighbouring pixels (i.e. each pixel crack) and their associated histograms. The values of the stable modes in a pair of histograms defines an interval of values. I regard the stable mode as being continuous between the two pixels if neither histogram contains an anti-mode within that interval. If there is an anti-mode, then the pixel crack is marked as a discontinuity and a red line along the pixel crack is displayed to reflect the fact.

References

[1] J.J. Koenderink, 'The structure of images', Biol. Cybern., 50, 363–370 (1984).

[2] E.V. Huntingdon, 'A set of postulates for abstract geometry, expressed in terms of the simple relation of inclusion', Math. Annal., 73, 522–559 (1913).

[3] Euclid, The Thirteen Books of The Elements. Vol. 1 (Books I and II). (trans. & intro. Heath, T.L.) Written 300 BC, New York, Dover, 1956.

[4] J.J. Koenderink, 'Embodiments of geometry', in Brain Theory: Spatio-Temporal aspects of Brain Function, A. Aertsen (ed), Elsevier, Amsterdam, 1993, pp. 3–28.

[5] H. Poincaré, Science and Hypothesis, Dover, New York, 1905.

[6] L.D. Griffin, 'Critical point events in affine scale space', in Gaussian Scale-Space Theory, J. Sporring, M. Nielsen, L. Florack and P. Johansen, (eds), Kluwer (in press).

[7] T. Lindeberg, 'Scale-space for discrete signals', IEEE Trans. Patt. Anal. Mach. Intell, 12(3), 234–254 (1990).

[8] L. Alvarez, F. Guichard, P.L. Lions and J.-M. Morel, 'Axioms and fundamental equations of image processing', Arch. Rational Mech. & Anal., 123(3), 199–257 (1993).

[9] L.M.J. Florack, 'Grey-scale images', ERCIM Technical Report 09/95-R039, 1995.

[10] E.J. Pauwels, L.J. Van Gool, P. Fiddelaers and T. Moons, 'An extended class of scale-invariant and recursive scale space filters', IEEE Trans. Patt. Anal. Mach. Intell., 17(7), 691–701 (1995).

[11] M.J. Carlotto, 'Histogram analysis using a scale space approach', IEEE Trans. Patt. Anal. Mach. Intell., 9(1), 121–129 (1987).

[12] M.C. Minnotte and D.W. Scott, 'The mode tree: a tool for visualization of nonparametric density features', J. Comput. Graph. Stat., 2(1), 51–68 (1993).

[13] A.J. Noest and J.J. Koenderink, 'Visual coherence despite transparency or partial occlusion', Perception, 19, 384 (1990).

[14] A.J. Noest, 'Neural processing of overlapping shapes', in Shape in Picture NATO ASI Series F vol. 126, Y.-L. O, A. Toet, D. Foster, H.J.A.M. Heijmans and P. Meer (eds), 1994, pp. 383–392.

[15] E.P. Simoncelli, 'Distributed analysis and representation of visual motion', PhD thesis, MIT, 1993.

[16] R.J. Snowden, S. Treue and R.A. Andersen, 'The response of area MT and V1 neurons to transparent motion', J. Neurosci., 11, 2768–2785 (1991).

[17] L.D. Griffin and A.C.F. Colchester, 'Superficial and deep structure in linear diffusion scale space: isophotes, critical points and separatrices', Image & Vision Computing, 13(7), 543–557 (1995).

[18] L.D. Griffin, 'Descriptions of image structure', PhD thesis, University of London, 1995.

[19] P.L. Torroba, N.L. Cap, H.J. Rabal and W.D. Furlan, 'Fractional order mean in image processing', Optical Engineering, 33(2), 528–533 (1994).

[20] E.R. Davies, 'On the noise suppression and image enhancement characteristics of the median, truncated median and mode filters', Patt. Recog. Lett., 7(2), 87–97 (1988).

[21] A. Majthay, Foundations of Catastrophe Theory, Pitman, London, 1985.

[22] R. Gilmore, Catastrophe Theory for Scientists and Engineers, Dover, Mineola, NY, 1981.

[23] I. Lakatos, Proofs and Refutations. The Logic of Mathematical Discovery, Cambridge University Press, Cambridge, 1976.

[24] A.L. Cauchy, Cours d'Analyse de l'École Royale Polytechnique, de Bure, Paris, 1821.

[25] J.B. Souto, 'El modo y otras medias: casos particulares de una misma expresión matemática', Cuadernos de Trabajo, No. 3, Universidad Nacional de Buenos Aires, 1938.

[26] M. Fréchet, 'Les éléments aléatoires de nature quelconque dans un espace distancié', Annales de l'Institut Henri Poincaré, X, 215–308 (1948).

[27] A. Goshtasby, 'On edge focusing', Image & Vision Computing, 12(4), 247–256 (1994).