Remark

This document contains chapter 8 from Lindeberg: *Scale-Space Theory in Computer Vision*. This material constitutes a revised presentation of "Scale-space behaviour of local extrema and blobs" first published in *Journal of Mathematical Imaging and Vision*, vol. 1, pp. 65–99, 1992.

8

Behaviour of image structures in scale-space: Deep structure

The treatment so far has been mainly concerned with the formal definition of the scale-space representation and the definition of image descriptors at any single scale. A complementary problem concerns how to relate structures at different scales. This subject has been termed *deep structure* by Koenderink (1984). When a pattern is subjected to scale-space smoothing, its shape changes and may be distorted. For example, features like local extrema, edges, blobs, etc. can be expected to drift when the underlying grey-level image is subject to blurring. More generally, transitions between objects of qualitatively different appearance may also take place. This gives rise to the notion of *dynamic shape,* which as argued by Koenderink and van Doorn (1986) is an essential component of any shape description of natural objects.

Aspects of these phenomena have been studied by several authors from different viewpoints. Canny (1986) discussed the general trade-off problem between detection and localization occurring in edge detection. Bergholm (1987) estimated the drift velocity of edges for a set of plausible configurations with the aim of estimating a step size for scale changes in the edge focusing algorithm. Berzins (1984) analyzed the localization error for zero-crossings of the Laplacian of the Gaussian.

Other kinds of phenomena affecting the topology may also occur. As developed by Koenderink and van Doorn (1986), blobs can disappear, merge, and split. Similar transitions apply to edges, zero-crossings of the Laplacian, corners, etc. Such events are usually called *bifurcations.*

In this chapter we shall study critical points, that is, local extrema and saddle points, and investigate in detail what happens to those features when an image undergoes scale-space smoothing. We shall

- develop how these feature points can be expected to behave when the scale parameter in scale-space changes,
- derive an expression for their drift velocity,
- classify their behaviour at bifurcation situations into a discrete set of generic situations, and

- give a coarse estimate to the global problem of how the number of local extrema in a signal can be expected to vary with scale.

The results that will be derived are not based on any specific models for the intensity variations in the image but are generally valid under rather weak *a priori* assumptions. Although the results will be expressed in a general form, the primary intention with the study is to provide a further theoretical basis of the scale-space primal sketch concept developed in chapter 7. In this context, the results to be presented will find their main application in

- the formal construction and definition of the primitives (scale-space blobs) in the scale-space primal sketch. The scale-space blobs are defined as families of grey-level blobs, which in turn are directly determined by pairs of critical points. This treatment allows for precise mathematical definitions of those concepts.

- providing a theoretical basis for the linking algorithm necessary when computing the representation.

- giving further motivations for the normalization process with respect to "expected scale-space behaviour," which is necessary when defining the significance measures of the scale-space blobs.

In other words, we shall try to explain what happens when scale changes in scale-space, especially with application to the scale-space primal sketch. Therefore, special attention will be given to the primitive objects of that representation, i.e. the grey-level blobs and scale-space blobs. The methodology of analysis is, however, general and can be applied to, for example, any features that can be expressed as zero-crossings of differential expressions (differential singularities).

Before starting, let us point out that some of the results to be presented are (at least partly) known or touched upon before, see e.g., Koenderink (1984, 1990), Koenderink and van Doorn (1986), and Clark (1988). Bifurcations in scale-space have also been studied by Johansen *et al.* (1986), who have shown that a band-limited one-dimensional signal up to a multiplicative constant is determined by its "top points," that is the points in scale-space where bifurcations take place. A work by Johansen (1993) extends the analysis in an earlier version of this presentation (Lindeberg 1991, 1992) with a differential geometric study of trajectories of critical points in scale-space.

The purpose of this treatment is to develop systematically and comprehensively what can be said about the behaviour in scale-space of critical points using elementary mathematical techniques and to convey an intuitive feeling for the qualitative behaviour in the different generic cases. Detailed calculations will also be given showing the behaviour of blobs in a set of "characteristic examples."

8.1. Trajectories of critical points in scale-space

In many situations it is of interest to estimate the drift velocity of critical points when the scale parameter varies. Such information is useful, for instance, when estimating the localization error of feature points due to scale-space smoothing, or when tracking local extrema or related entities across scales. In non-degenerate situations, that is when the second differential is a non-degenerate quadratic form, such an analysis can be based on the implicit function theorem.

DEFINITION 8.1. (CRITICAL POINT) *A point $x_0 \in \mathbb{R}^N$ is a critical point of a mapping $f : \mathbb{R}^N \to \mathbb{R}$ if the gradient at this point*

$$(\nabla f)(x_0) = \left(\begin{array}{c} \partial_{x_1} f \\ \vdots \\ \partial_{x_N} f \end{array} \right) \Bigg|_{x_0} \tag{8.1}$$

is zero. The critical point is said to be non-degenerate if the Hessian matrix in this point

$$(\mathcal{H}f)(x_0) = \left(\begin{array}{ccc} \partial_{x_1 x_1} f & \dots & \partial_{x_N x_1} f \\ \vdots & \ddots & \vdots \\ \partial_{x_1 x_N} f & \dots & \partial_{x_N x_N} f \end{array} \right) \Bigg|_{x_0} \tag{8.2}$$

is non-singular. Otherwise it is said to be degenerate.

LEMMA 8.2. (BEHAVIOUR OF CRITICAL POINTS IN SCALE-SPACE)
Let $L : \mathbb{R}^N \times \mathbb{R}_+ \to \mathbb{R}$ be the scale-space representation of an N-dimensional continuous signal given by the diffusion equation (2.27), and assume that at some scale level $t_0 > 0$ a point $x_0 \in \mathbb{R}$ is a non-degenerate critical point for the mapping $x \mapsto L(x;\ t_0)$.

Then, there exist an open set $S_{(x_0;\ t_0)} \subset \mathbb{R}^N \times \mathbb{R}_+$ and an open interval $I_{t_0} \subset \mathbb{R}_+$ with $(x_0;\ t_0) \in S_{(x_0;\ t_0)}$ and $t_0 \in I_{t_0}$ having the following property: To every $t_1 \in I_{t_0}$ there corresponds a unique $x_1 \in \mathbb{R}^N$ such that $(x_1;\ t_1) \in S_{(x_0;\ t_0)}$ and x_1 is a non-degenerate critical point for the mapping $x \mapsto L(x;\ t_1)$.

If this x_1 is defined to be $r(t_1)$, then r is a continuously differentiable mapping $I_{t_0} \to \mathbb{R}^N$ such that

- $r(t_0) = x_0$,
- *$r(t_1)$ is for every $t_1 \in I_{t_0}$ a non-degenerate critical point for the mapping $x \mapsto L(x;\ t_1)$.*
- *the derivative of r with respect to t in the point x_0 is given by*

$$\partial_t r(t_0) = -\tfrac{1}{2} (\mathcal{H}L)(x_0)^{-1} (\nabla^2 (\nabla L))(x_0). \tag{8.3}$$

Proof. The result follows easily by applying the implicit function theorem to the gradient function. The standard version of the implicit function theorem (Rudin 1976) basically gives that there exists a path $r : I_{t_0} \to \mathbb{R}^N$ of critical points, and that the derivative of r is

$$\partial_t r(t_0) = -(\mathcal{H}L)(x_0)^{-1}(\partial_t(\nabla L))(x_0). \tag{8.4}$$

Then, the fact that L satisfies the diffusion equation can be used for replacing derivatives of L with respect to t by derivatives of L with respect to the spatial coordinates in order to arrive at (8.3). Since the Hessian $(\mathcal{H}L)(r(t))$ along this path is a continuous function of t, it follows that $r(t)$ will remain non-degenerate provided that the initial point x_0 is non-degenerate, and I_{t_0} is selected as a sufficiently short interval. □

8.1.1. Interpretation: Drift velocity estimates

This lemma expresses how critical points in general can be expected to behave in scale-space. One of the most immediate interpretations is that (8.3) gives a straightforward estimate of the *drift velocity of critical points* under scale-space smoothing.

This estimate can also be extended to comprise edges. For simplicity, assume that the edge under study is sufficiently long and sufficiently close to a straight line such that a one-dimensional analysis is a valid approximation. Further, without loss of generality, assume that the coordinate system is oriented such that the edge is perpendicular to the x_1-axis. Then, use non-maximum suppression to define the location of the edge as those points where the gradient magnitude (here the x_1-derivative) assumes a maximum along the gradient direction (here the x_1-direction). In other words, edge points are defined as those points where the second derivative along the gradient direction is zero. Now, since under these conditions, critical points are given by zeros in the first derivative and edge points by zeros in the second derivative, the one-dimensional version of (8.3) can be applied just by replacing L by L_{x_1}. Hence, the drift velocity in the direction perpendicular to a *straight edge* is

$$\partial_t r(t_0) = -\frac{1}{2}\frac{L_{x_1^4}(x_0;\, t_0)}{L_{x_1^3}(x_0;\, t_0)}. \tag{8.5}$$

A similar idea, although based on an approximate derivation, has been expressed by Zhuang and Huang (1986).

This analysis is applicable also to edges given by zero-crossings of the Laplacian, provided that the second derivative along the edge direction (here the x_2-direction) is sufficiently small to be neglected. Trivially, an identical result holds for edges of one-dimensional signals. Observe that there are no specific assumptions about the shape of the intensity profile

perpendicular to the edge. Hence, the result is valid for any configuration that can be described by a one-dimensional analysis.

In particular, it means that the drift velocity may tend to infinity when two adjacent parallel edges are just about to merge into one. This result can, for example, be used for explaining an observation by Zhang and Bergholm (1991), who noted that configurations consisting of two adjacent edges (so-called "staircase edges"; see figure 8.1) can lead to a rapid edge drift when the scale parameter changes, which in turn violates the assumptions behind the estimate of the scale step used in the edge focusing algorithm (Bergholm 1987). In such situations the third derivative is in fact very close to zero. A more general analysis of curved edges is given in section 8.6.

Figure 8.1. (a) A "staircase edge" can lead to a rapid edge drift. This behaviour can be explained by noting that (b) after sufficient amount of blurring the configuration will tend to a "diffuse step edge" as well as by studying the derivatives of (c) the original signal, and (d) the signal after strong smoothing. (e) By considering the paths the zero-crossings of the Laplacian describe as scale changes it is easy to realize that when the edge points tend to each other the drift velocity will tend to infinity. See also section 8.5 for a more detailed description of the behaviour at bifurcation situations, in particular section 8.5.4 concerning this configuration.

Finally, regarding the drift velocity estimates for local extrema and edges, it should be pointed out that although the drift velocity *momentarily* may tend to infinity, the total drift (integrated over some scale interval of finite length) will always be finite. What the results mean, is that it is not possible to derive any *uniform* upper bound for the drift velocity of these features. Given any scale interval of length Δt and any distance $|\Delta x|$ it is always possible to find a signal such that the total drift of a feature during the time Δt exceeds $|\Delta x|$. This property emphasizes the need for algorithms based on *adaptive* sampling along the scale direction.

8.1.2. Interpretation: Extremum paths

Another consequence of lemma 8.2 is that a non-degenerate critical point existing at a certain level of scale can, in general, be traced to a similar critical point both at a slightly coarser and a slightly finer scale. By continuation, such local paths obtained from the implicit function theorem can be extended to curves as long as the Hessian remains non-zero. It can be easily shown that the type of critical point remains the same as long the Hessian matrix is non-singular.

> It is obvious that a local maximum (minimum) cannot be transformed into a saddle point or vice versa. If the Hessian would change sign, then it would first become zero (since it is a continuous function of the scale parameter). Then, however, the trajectory would by definition be cut off by a degenerate critical point into two separate segments.
>
> Moreover, a maximum point cannot be transformed into a minimum point or opposite, since then (at least) the partial derivative $L_{x_1 x_1}$ would need to change sign. Such a sign change implies that this derivative would first become zero (because of continuity), which in turn means that the quadratic form would become indefinite, i.e., the point would get transformed into a saddle point. This transition has to go through a degenerate critical point, which means that the trajectory would be cut off into at least two parts.

In other words, if $(x_0;\ t_0)$ is a local maximum (minimum/saddle), then there exists a curve through this point, such that every point on the curve is a local maximum (minimum/saddle) at that scale. The curve is delimited by two scale levels t_{min} and t_{max}, at which the Hessian matrix degenerates (except for the boundary cases $t_{min} = 0$ or $t_{max} = \infty$). At all interior points the extremum point is non-degenerate. Such a curve $r_0 \colon [t_{min}, t_{max}] \to \mathbb{R}^2$ is is called an *extremum path* (*saddle path*).

The situation in other dimensions is similar, although there are no stable saddle points in the one-dimensional case.

8.2. Scale-space blobs

The notion of extremum path in previous section allows for a formal definition of scale-space blob—the basic primitive in the scale-space primal sketch. In chapter 7, a grey-level blob was defined as a local extremum with extent and a scale-space blob in turn as a family of those. More precisely, a grey-level blob of a two-dimensional signal was given by a pair consisting of a local extremum and a saddle point, and in one dimension by a maximum and minimum point, implying a one-to-one correspondence between local extrema and grey-level blobs. The previous definition of scale-space blob was, however, intuitive: "similar blobs at adjacent levels of scale were linked into scale-space blobs." The linking process proceeded

until no such linking could be performed, i.e., until a bifurcation was encountered. The idea behind this construction was to identify and group similar features at different scales into higher order and unified objects.

8.2.1. Definition of scale-space blob

We can now express the linking criterion in a more formal way. Consider the two-dimensional case, and study a local extremum x_0 with associated grey-level blob $G_{blob}(x_0)$ in a non-degenerate (Morse) signal at some scale t_0 in scale-space. Then, there is a unique extremum path $r_0 : [t_{min}, t_{max}] \to \mathbb{R}^2$ associated with the extremum $x_0 = r_0(t_0)$ of the grey-level blob, and for each extremum $r_0(t)$ along this path there is a corresponding grey-level blob $G_{blob}(r_0(t))$.

For every scale level $t \in [t_{min}, t_{max}]$ where the scale-space representation is non-degenerate, there is a unique delimiting saddle point[1] $S_{delimit}(r(t))$ associated with the local extremum $r_0(t)$. All such saddle points associated with an extremum path need, however, not be on the same saddle path. At certain scales transitions between different saddle paths can be expected to take place. Generically, this occurs at a discrete set of scales, at which the local extremum point is non-degenerate, and the extent of the grey-level blob is delimited by two non-degenerate saddle points having the same grey-level. Such transitions will not be regarded as affecting the scale-space blobs.

On the other hand, if the delimiting saddle (or the extremum) is involved in a bifurcation, then the local topology will be changed—a *blob event* has occurred. It is therefore natural to proceed with the linking as long as the extrema and their delimiting saddle points are non-degenerate, and to stop it when either of the critical points degenerates. Hence, consider a (maximal) scale interval $[t'_{min}, t'_{max}] \subset [t_{min}, t_{max}]$ such that

- for all interior scales, the (possible multiple) delimiting saddle points $S_{delimit}(r_0(t))$ are non-degenerate, and

- at the end points either of $r_0(t'_{min})$ and $S_{delimit}(r_0(t'_{min}))$ and also either of $r_0(t'_{max})$ and $S_{delimit}(r_0(t'_{max}))$ are degenerate critical points.

Then, the *scale-space blob* associated with the segment $r'_0 : [t'_{min}, t'_{max}] \to \mathbb{R}^2$ of the extremum path is the (four-dimensional) set

$$S_{blob}(r'_0) = \overline{\{(x;\ z;\ t) \in \mathbb{R}^2 \times \mathbb{R} \times \mathbb{R}_+ :\ \overline{(t'_{min} < t < t'_{max}) \wedge ((x;\ z) \in G_{blob}(r'_0(t)))\}}}, \qquad (8.6)$$

[1] If an extremum point E and a saddle point S together define the extent of a grey-level blob, then S is said to be the delimiting saddle point of E (section 7.1).

and the *support region* of the scale-space blob is in turn the (three-dimensional) region

$$S_{support}(r'_0) = \{(x;\ t) \in \mathbb{R}^2 \times \mathbb{R}_+ : (x;\ z;\ t) \in S_{blob}(r'_0) \text{ for some } z\}.$$

In most figures with scale-space blobs (e.g., in chapter 7) it is the latter descriptor that has been illustrated.

8.2.2. Definition of scale-space blob volume

Strictly, in this coordinate system the scale-space blob volume is

$$S_{vol}(r'_0) = \int_{(x;\ z;\ t) \in S_{blob}(r'_0)} dx\, dz\, dt = \int_{t \in [t'_{min}, t'_{max}]} G_{volume}(r'_0(t))\, dt,$$

where $G_{vol}(r'_0(t))$ is the grey-level blob volume of the grey-level blob associated with the extremum point $r'_0(t)$. However, when the scale-space blob volume is to be used as a significance measure in the scale-space primal sketch, it turns out that some transformations of the coordinate axes must be performed (see section 7.7). One would like structures at different scales to be treated uniformly, such that the significance measure neither favours fine scales to coarse scales nor the opposite. Therefore, the *normalized scale-space blob volume* is defined by

$$S_{vol,norm}(r'_0) = \int_{t \in [t'_{min}, t'_{max}]} V_{trans}(G_{vol}(r'_0(t));\ t)\, d\tau_{eff}(t), \qquad (8.7)$$

where $\tau_{eff}: \mathbb{R}_+ \to \mathbb{R}$ is a transformation function mapping the ordinary scale parameter t to a transformed scale parameter τ called *effective scale* (see section 7.7.1), and $V_{trans}: \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ is a corresponding transformation function normalizing the variation of the grey-level blob volumes into a more uniform behaviour over scales (see section 7.7.2).

8.3. Bifurcation events for critical points: Classification

The implicit function theorem used in previous sections guarantees that linking of non-degenerate critical points is a well-defined operation. When the Hessian matrix becomes singular, the implicit function theorem is no longer applicable, and *bifurcations* may occur.

Useful tools for analysing the qualitative behaviour of functions around bifurcation points can be obtained from a branch of mathematics known as *singularity theory*; see Poston and Stewart (1978) or Gibson (1979) for application-oriented introductions, and Arnold (1981), Arnold *et al.* (1985, 1988), Golubitsky and Schaeffer (1985), or Lu (1976) for more rigorous treatments of the subject.

8.3.1. Thom's classification theorem

One of the fundamental results in singularity theory is that the typical qualitative behaviour of families given by a small number of parameters can be expressed completely by the qualitative behaviour of a finite set of families. A famous theorem by Thom classifies the generic behaviour of families of functions with the number of parameters $r \leq 4$ into seven elementary catastrophes. A summarizing result expressed by Poston and Stewart (1978) states that:

> THOM'S CLASSIFICATION THEOREM: *Typically an r-parameter family $\mathbb{R}^N \times \mathbb{R}^r \to \mathbb{R}$ of smooth functions $\mathbb{R}^N \times \mathbb{R}^r \to \mathbb{R}$, for any N and $r \leq 4$, is structurally stable and is in every point (locally) equivalent to one of the following forms:*
>
> - *non-critical:* x_1,
> - *non-degenerate critical, or Morse:*
> $x_1^2 + \cdots + x_i^2 - x_{i+1}^2 - \cdots - x_N^2$ $(0 \leq i \leq N)$,
> - *degenerate critical, catastrophe;*
> - *fold:*
> $x_1^3 + u_1 x_1 + (M)$,
> - *cusp:*
> $\pm(x_1^4 + u_2 x_1^2 + u_1 x_1) + (M)$,
> - *swallowtail:*
> $x_1^5 + u_3 x_1^3 + u_2 x_1^2 + u_1 x_1 + (M)$,
> - *butterfly:*
> $\pm(x_1^6 + u_4 x_1^4 + u_3 x_1^3 + u_2 x_1^2 + u_1 x_1) + (M)$,
> - *elliptic umbilic:*
> $x_1^2 x_2 - x_2^3 + u_3 x_1^2 + u_2 x_2 + u_1 x_1 + (N)$,
> - *hyperbolic umbilic:*
> $x_1^2 x_2 + x_2^3 + u_3 x_1^2 + u_2 x_2 + u_1 x_1 + (N)$,
> - *parabolic umbilic:*
> $\pm(x_1^2 x_2 + x_2^4 + u_4 x_2^2 + u_3 x_1^2 + u_2 x_2 + u_1 x_2) + (N)$,
>
> *where (M) and (N) indicate Morse functions on the forms*
>
> $$(M) = x_2^2 + \cdots + x_i^2 - x_{i+1}^2 - \cdots - x_N^2 \quad (2 \leq i \leq N),$$
>
> $$(N) = x_3^2 + \cdots + x_i^2 - x_{i+1}^2 - \cdots - x_N^2 \quad (2 < i \leq N),$$
>
> *which must be added on to the previously mentioned expressions in order to match up the dimensions.*

The intuitive explanation of "structurally stable" is that a sufficiently small perturbation does not change the qualitative behaviour at the singularity. The term "locally equivalent" essentially means that the function

can be (locally) transformed to the listed polynomial representative of the singularity by a change of variables (a diffeomorphism). For precise definitions of these concepts, the reader is referred to the sources cited above. A brief review can also be found in (Lindeberg 1992).

8.3.2. Generic singularities of one-parameter families

Applied to one-parameter families, like the scale-space representation of a signal, this result means that the natural type of singularity to expect is the *fold singularity*. It can be represented by the polynomial

$$G_N(x;\ u) = x_1^3 + 3x_1 u + \sum_{i=2}^{N} \pm x_i^2, \tag{8.8}$$

where x_i should be interpreted as offset coordinates around the bifurcation point, here translated to the origin, and u as the parameter. This singularity means that at the bifurcation point the first and second order directional derivatives are zero along a certain direction, while the third derivative in that direction is non-zero.

Hence, if one is interested in the behaviour of the critical points of a signal during the evolution of the diffusion equation, it should in principle be sufficient to study this situation. For simplicity, consider from now on the two-dimensional case, and replace the notation (x_1, x_2) for coordinates by (x, y). Then, the *singularity set* is given by the solutions of

$$\partial_x G_2(x, y;\ u) = 3x^2 + u = 0, \tag{8.9}$$

$$\partial_y G_2(x, y;\ u) = \pm 2y = 0, \tag{8.10}$$

and the *bifurcation set* by the solution of

$$\partial_x G_2(x, y;\ u) = 3x^2 + u = 0, \tag{8.11}$$

$$\partial_y G_2(x, y;\ u) = \pm 2y = 0, \tag{8.12}$$

$$\partial_{xx} G_2(x, y;\ u) = 6x = 0. \tag{8.13}$$

Thus, the singularity set is given by

$$-(x_1(u), y_1(u)) = (x_2(u), y_2(u)) = (\sqrt{-u/3}, 0) \quad (u \leq 0), \tag{8.14}$$

and the bifurcation occurs at an isolated point $(x, y;\ u) = (0, 0;\ 0)$. From the sign of the Hessian

$$\det(\mathcal{H} G_2)(x, y;\ u) = \pm 12x, \tag{8.15}$$

it follows that $(x_1(u), y_1(u))$ are saddle/maximum points and $(x_2(u), y_2(u))$ are minimum/saddle points for every $u < 0$. At $u = 0$ the points merge along a parabola and then disappear; see figure 8.2(a).

Similarly, for a one-dimensional signal the fold singularity is represented by the polynomial $G_1(x;\ u) = x^3 + xu$, and corresponds to a maximum and a minimum point merging with increasing u; see figure 8.2(b).

Figure 8.2. (a) The generic behaviour at a singularity of a one-parameter family of two-dimensional functions is described by the unfolding $G_2(x, y;\ u) = x^3 + ux \pm y^2$. The singularity set of this family, that is the set of critical points of the mapping $x \mapsto G_2(x, y;\ u)$, describes an extremum point and a saddle point that merge along a parabola and then disappear. (b) For a one-parameter family of one-dimensional functions the behaviour is instead given by $G_1(x;\ t) = x^3 + ux$. The singularity set in this case corresponds to a similar merge of a maximum point and a minimum point. (The notation S/M^+ means that the trajectory corresponds to either a saddle point or a maximum point etc.)

To summarize, the typical behaviour to be expected at singularities in a one-parameter family of continuous signals is:

- annihilations or creations of pairs of local extrema and saddle points in the two-dimensional case, and
- annihilations or creations of pairs of local maxima and local minima in the one-dimensional case.

8.3.3. Interpretations with respect to scale-space representation

By comparisons with earlier theoretical and experimental results we know that this describes the qualitative behaviour of critical points in scale-space. However, there is one apparent complication when to give a more detailed interpretation. Thom's classification theorem states that there exists a diffeomorphism such that the singularity set of a solution to the one-dimensional diffusion equation around a bifurcation point $(x_0;\ t_0)$ in scale-space can be transformed into the singularity set of G_2 around $(0;\ 0)$. However, there is obviously some directional information lost in the equivalence concept: In which direction should the u parameter be interpreted as running? If u and t are treated as increasing simultaneously, then the situation describes a minimum and a maximum merging with increasing t. On the other hand, if u runs in the reverse direction, then the interpretation would be that a pair with a minimum and a maximum

would be created when t increases. The latter phenomenon is, however, impossible, since the number of local extrema cannot increase under scale-space smoothing in the one-dimensional case (see chapter 3).

The diffusion equation apparently introduces a directional preference to its solutions (due to the causality requirements), which makes such creations impossible. How should this information be incorporated into the analysis of the singularities in scale-space?

One way of avoiding the previous blindness of the equivalence concept to the structural property of the diffusion equation could be by trying to develop results similar to Thom's classification theorem, which instead of being expressed in terms of the ordinary standard basis of polynomials could be expressed in terms of polynomials satisfying the diffusion equation. Natural modified representatives of the fold singularities in the one-dimensional and two-dimensional cases would then be

$$G_1(x;\ t) = x^3 + 3xt, \tag{8.16}$$

$$G_2(x, y;\ t) = x^3 + 3xt \pm (y^2 + t). \tag{8.17}$$

Another approach is to use the previous classification to state what configurations are possible in general one-parameter families of functions. Then, the special structure of the diffusion equation can be taken into account for judging which cases apply to the scale-space representation when the directional constraint of the diffusion equation has been added.

Yet a third approach will be considered in the next section, where it will be demonstrated that the results from the second approach agrees with what can be obtained from a differential geometric analysis of the paths that critical points form in scale-space.

8.3.4. Algebraic classification of singularities

The bifurcation events between critical points in scale-space can also be classified by using elementary techniques. Following Johansen (1993), consider critical paths defined by the solutions to

$$\begin{cases} L_x(x, y;\ t) = 0, \\ L_y(x, y;\ t) = 0. \end{cases} \tag{8.18}$$

Bifurcation points are points $r_0 = (x_0, y_0;\ t_0)^T$ where

$$\det(\mathcal{H}L)(r_0) = L_{xx}(r_0)L_{yy}(r_0) - L_{xy}^2(r_0) = 0. \tag{8.19}$$

Without loss of generality assume that the coordinate system is rotated such that

$$L_{xy}(r_0) = 0. \tag{8.20}$$

Then, either L_{xx} or L_{yy} must be zero at the bifurcation point. Without loss of generality assume that

$$L_{xx}(r_0) = 0, \tag{8.21}$$

and let a critical path r through this point be parameterized by an arc length parameter s such that $r(s) = (x(s), y(s);\ t(s))^T$ with $r_0 = r(s_0)$

$$|r'(s)|^2 = x'(s)^2 + y'(s)^2 + t'(s)^2 = 1. \tag{8.22}$$

Implicit differentiation of $L_x(x(s), y(s);\ t(s)) = 0$ and $L_y(x(s), y(s);\ t(s)) = 0$ with respect to s then gives

$$\begin{cases} L_{xx}(r(s))\, x'(s) + L_{xy}(r(s))\, y'(s) + L_{xt}(r(s))\, t'(s) = 0, \\ L_{xy}(r(s))\, x'(s) + L_{yy}(r(s))\, y'(s) + L_{yt}(r(s))\, t'(s) = 0. \end{cases} \tag{8.23}$$

Using $L_{xx}(r_0) = 0$ and $L_{xy}(r_0) = 0$ these relations reduce to

$$\begin{cases} L_{xt}(r_0)\, t'(s_0) = 0, \\ L_{yy}(r_0)\, y'(s_0) + L_{yt}(r_0)\, t'(s_0) = 0. \end{cases} \tag{8.24}$$

In the generic case we can assume that $L_{xt}(r_0) \neq 0$ and $L_{yy}(r_0) \neq 0$. Then, (8.24) gives $y'(s_0) = 0$ and $t'(s_0) = 0$. By selecting the positive root from the normalization condition $x'(s_0)^2 = 1$ obtained from (8.22), we get $r'(s_0) = (x'(s_0), y'(s_0);\ t'(s_0))^T = (1, 0;\ 0)^T$. This shows that the curve intersects the bifurcation point with a horizontal tangent.

Concerning the second order structure at the bifurcation point, implicit differentiation of (8.23) combined with the expression for $r'(s_0)$ and the assumptions $L_{xx}(r_0) = 0$ and $L_{xy}(r_0) = 0$ gives

$$\begin{cases} L_{xxx}(r_0) + L_{xt}(r_0)\, t''(s_0) = 0, \\ L_{xxy}(r_0) + L_{yy}(r_0)\, y''(s_0) + L_{yt}(r_0)\, t''(s_0) = 0, \end{cases} \tag{8.25}$$

from which explicit expressions can be obtained for the curvature components of the critical path

$$x''(s_0) = 0, \tag{8.26}$$

$$\begin{aligned} y''(s_0) &= \frac{L_{xxx}(r_0)\, L_{yt}(r_0) - L_{xxy}(r_0)\, L_{xt}(r_0)}{L_{yy}(r_0)\, L_{xt}(r_0)} \\ &= \frac{L_{xxx}(r_0)\, L_{yyy}(r_0) - L_{xxy}(r_0)\, L_{xyy}(r_0)}{L_{yy}(r_0)\, (L_{xxx}(r_0) + L_{xyy}(r_0))}, \end{aligned} \tag{8.27}$$

$$t''(s_0) = \frac{L_{xxx}(r_0)}{L_{xt}(r_0)} = -\frac{2L_{xxx}(r_0)}{L_{xxx}(r_0) + L_{xyy}(r_0)}. \tag{8.28}$$

The result $x''(s_0) = 0$ is a direct consequence of the fact that the derivative of a constant length vector is perpendicular to the original vector.

In the two other expressions alternative forms have also been given, in which derivatives with respect to t have been replaced by derivatives with respect to y and x using the fact that the Gaussian derivatives satisfy the diffusion equation. One example is $L_{xt} = \partial_t L_x = \frac{1}{2}(\partial_{xx} + \partial_{yy})L_x$.

By studying the sign of $t''(s_0)$ it can be seen that a pair of critical points is *annihilated* with increasing scale if

$$t''(s_0) = -\frac{2L_{xxx}(r_0)}{L_{xxx}(r_0) + L_{xyy}(r_0)} < 0, \tag{8.29}$$

while a pair of critical points is *created* with increasing scale if

$$t''(s_0) = -\frac{2L_{xxx}(r_0)}{L_{xxx}(r_0) + L_{xyy}(r_0)} > 0. \tag{8.30}$$

In order to analyse the types of critical points, consider a first order Taylor expansion of the coordinate functions

$$r(s) = (x(s), y(s);\ t(s))^T = (x'(s_0)\, s + \mathcal{O}(s^2), \mathcal{O}(s^2);\ \mathcal{O}(s^2))^T,$$

and approximate the partial derivatives of L linearly

$$L_{xx}(r) = L_{xx}(r_0) + (\nabla_{(x;\, t)} L_{xx})(r_0)^T r + \mathcal{O}(r^2), \tag{8.31}$$

$$L_{xy}(r) = L_{xy}(r_0) + (\nabla_{(x;\, t)} L_{xy})(r_0)^T r + \mathcal{O}(r^2), \tag{8.32}$$

$$L_{yy}(r) = L_{yy}(r_0) + (\nabla_{(x;\, t)} L_{yy})(r_0)^T r + \mathcal{O}(r^2), \tag{8.33}$$

where the symbol $\nabla_{(x;\, t)} = (\partial_x, \partial_y;\ \partial_t)^T$ means that differentiation should be performed with respect to x, y, and t. Then, the first order Taylor expansion of the Hessian can be written

$$\begin{aligned} \det(\mathcal{H}L)(r(s)) &= L_{xx}(r(s))L_{yy}(r(s)) - L_{xy}^2(r(s)) \\ &= L_{xxx}(r_0)\, L_{yy}(r_0)\, s + \mathcal{O}(s^2). \end{aligned} \tag{8.34}$$

Hence, provided that $L_{xxx}(r_0) \neq 0$ and $L_{yy}(r_0) \neq 0$, the critical points on one side of the bifurcation point are local extrema, and the critical points on the other side are saddle points.

To summarize, this analysis gives an alternative verification that the generic singularities for critical points are annihilations and creations of saddle-extremum pairs. Moreover, it states an explicit condition for when creations can occur. In a coordinate system selected such that $L_{pp}(r_0) = L_{pq}(r_0) = 0$ at the bifurcation point r_0, this condition can be expressed:

$$\begin{aligned} &((L_{ppp}(r_0) > 0) \wedge (L_{pqq}(r_0) < -L_{ppp}(r_0)) \vee \\ &((L_{ppp}(r_0) < 0) \wedge (L_{pqq}(r_0) > -L_{ppp}(r_0)). \end{aligned} \tag{8.35}$$

In one dimension, (8.28) reduces to $t'' = -2 < 0$, showing that critical points are always annihilated and never created in this case.

8.4. Bifurcation events for grey-level blobs and scale-space blobs

A natural question that arises in connection with the scale-space primal sketch concerns what types of blob events are possible in bifurcation situations. Since scale-space blobs are defined in terms of paths of critical points, the behaviour of a scale-space blob at a singularity is solely determined by the behaviour of the extremum/saddle paths during a short scale interval around the bifurcation moment.

Compared to the previous treatment, where critical points were analysed, there is one additional factor that must be taken into account when dealing with scale-space blobs, namely the fact that a saddle point delimiting the extent of a grey-level blob involved a bifurcation can be associated with other grey-level blobs as well. This leads to natural coupling between scale-space blobs sharing the same saddle path (of delimiting saddle points) in a neighbourhood of a bifurcation.

8.4.1. Shared and non-shared saddle path

In view of this observation let us define the following: A saddle path involved in a structurally stable bifurcation is said to be *non-shared* before (after) the bifurcation if there exists some scale interval before (after) the bifurcation during which every saddle point of the saddle path is not contained in more than one grey-level blob. Otherwise, the saddle path is said to be *shared*.

More precisely, a saddle path is said to be non-shared before (after) a bifurcation at t_{bifurc} if there exists some $\epsilon > 0$ such that for all scales in the interval $t \in]t_{bifurc} - \epsilon, t_{bifurc}[$ ($t \in]t_{bifurc}, t_{bifurc} + \epsilon[$) the saddle point of the saddle path at that scale does not belong to more than one grey-level blob, see figure 8.3. Another way to express this property is that a shared saddle point is the delimiting saddle point of two (or more) grey-level

Figure 8.3. The definition of grey-level blob for a two-dimensional signal. Every local extremum gives rise to a blob and the extent of the blob is given by a saddle point. A saddle point is said to be shared if it is contained in more than one grey-level blob, i.e., if it is a delimiting saddle point of two (or more) grey-level blobs of the same polarity.

blobs of the same polarity, while a non-shared saddle point either is the delimiting saddle point of one or no grey-level blob.

8.4.2. Generic bifurcation events for scale-space blobs

Hence, depending on whether a extremum-saddle pair is annihilated or created, and depending on whether the saddle path is shared or non-shared, four generic cases can be distinguished:

- A *non-shared* saddle path participating in the *annihilation* of an extremum-saddle pair with increasing scale describes an isolated blob that disappears. Such a blob event can aptly be called a *blob annihilation.*
- On the other hand, a *shared* saddle path involved in a similar *annihilation* describes a blob disappearing under the influence of a neighbour blob—a *blob merge.*
- Similarly, a *shared* saddle point taking part in an extremum-saddle pair that is *created* with increasing scale describes a *blob split.*
- Finally, a *non-shared* saddle path participating in an extremum-saddle creation describes an isolated blob which appears— a *blob creation.*

To summarize, (Below, the term annihilation (creation) of an extremum-saddle pair means that a pair consisting of an extremum path and a saddle path disappears (appears) when the scale parameter increases.)

PROPOSITION 8.3. (CLASSIFICATION OF BLOB EVENTS; 2D)
In the scale-space representation of two-dimensional continuous signal, the following blob events are possible at a structurally stable bifurcation:

- *blob annihilation—annihilation of an extremum-saddle pair, where the saddle path is non-shared before the bifurcation,*
- *blob merge—annihilation of an extremum-saddle pair, where the saddle path is shared before the bifurcation,*
- *blob split—creation of an extremum-saddle pair, where the saddle path is shared with another scale-space blob after the bifurcation,*
- *blob creation—creation of an extremum-saddle pair, where the saddle path is non-shared after the bifurcation.*

These four cases constitute the definitions of the terms annihilation, merge, split and creation with respect to grey-level blobs and scale-space blobs in the two-dimensional case.

Proof. From section 8.3 it follows that the typical behaviour at singularities are pairwise annihilations and creations of extremum-saddle pairs. Combined with the definition of shared saddle path this means that the class of possible blob events is restricted to the given four types, provided that only structurally stable bifurcations are considered.

Figure 8.4. (a) New local extrema can be created with increasing scale in the scale-space representation of a two-dimensional signal. Interpreted in terms of blobs the configuration describes a blob split. (b) By modifying the example slightly (by replacing the higher one of the two peaks with a double peak) one realizes that blob creations can occur as well. The base levels of the different grey-level blobs have been indicated.

What remains to verify is that all these four types can be instantiated, and that they are structurally stable. It is well-known that blob annihilations and blob merges can take place in scale-space (see also section 8.4 for illustrative examples). The fact that splits can occur is known as well (see the example given by Lifshitz and Pizer (1987) illustrated in figure 8.4(a)). The latter configuration can be modified to describe a blob creation as well, if the higher of the two peaks is replaced by a double peak (see figure 8.4(b)). Then, the extent of the two smaller blobs at the higher peak will be delimited by the grey-level in the valley between them, which means that when the narrow ridge has eroded and given rise to the creation of a saddle-extremum pair in which the saddle path is not be shared by any other blob. □

The assumption of structural stability is important in this context, since otherwise, there is an infinite variety of possible events. For instance, three or more blobs could merge into one blob at the same moment. Such events will however be unstable,[2] since a small perturbation of the input signal would perturb such a simultaneous merge of three blobs into a sequence of two successive pairwise merges.

Algorithmically, this means that an encountered actual situation with, say, three blobs at a fine scale seeming to belong all to the same blob at a coarser scale, can in general be decomposed into transitions of the four primitive types. This principle forms the idea behind the automatic scale refinement algorithm described in chapter 9, which essentially refines the scale sampling until all relations between scale-space blobs in scale-space can be decomposed into events of the previously listed types.

For one-dimensional signals the possible events[3] are blob annihilations and blob merges. Splits and creations are impossible due to the earlier mentioned property that new local extrema cannot be created in the one-dimensional case.

8.5. Behaviour near singularities: Examples

In previous sections a general methodology has been described for analysing the evolution properties of critical points. Moreover, the qualitative behaviour at bifurcation points has been classified. Here, a number of examples will be given demonstrating how the blob descriptors vary with scale in characteristic bifurcation situations.

8.5.1. Third order Taylor expansion in one dimension

A special property in one dimension is that it is possible to arrive at the generic representative of the fold unfolding (8.16) by a simple qualitative study. Consider a third order Taylor expansion of the scale-space embedding around a given point x_0 at some scale t_0

$$f_{t_0}(x) = \alpha + \beta(x - x_0) + \gamma(x - x_0)^2 + \epsilon(x - x_0)^3, \qquad (8.36)$$

where

$$\alpha = L(x_0;\ t_0), \quad \beta = L_x(x_0;\ t_0), \quad \gamma = \tfrac{1}{2}L_{xx}(x_0;\ t_0), \quad \epsilon = \tfrac{1}{6}L_{xxx}(x_0;\ t_0).$$

[2]Note in this context that for Morse functions no pair of critical points will have the same values. In other words, for generic functions all critical points will be distinct. Although, by definition, the grey-level function will not be Morse at a bifurcation, we can, in general, assume this latter property to hold at bifurcations. This means that situations with three or more blobs simultaneously merging into one can be expected not to occur.

[3]A formal statement of this result, including the relevant definitions, can be found in (Lindeberg 1992).

The scale-space representation of this signal (with $L(x;\ t_0) = f_{t_0}(x)$) is

$$\begin{aligned} L(x;\ t) = \alpha + \beta(x - x_0) + \gamma(x - x_0)^2 + \epsilon(x - x_0)^3 \\ + \delta_1(t - t_0) + \delta_2(x - x_0)(t - t_0), \end{aligned}$$

where $\delta_1 = \gamma$ and $\delta_2 = 3\epsilon$. For simplicity, introduce new (offset) variables $u = x - x_0$ and $v = t - t_0$. Then,

$$\tilde{L}(u;\ v) = L(u + x_0;\ v + t_0) = \alpha + \beta u + \gamma(u^2 + v) + \epsilon(u^3 + 3uv).$$

The critical points of the function $u \mapsto \tilde{L}(u;\ t)$ are given by

$$\partial_u \tilde{L}(u;\ v) = \beta + 2\gamma u + 3\epsilon(u^2 + v) = 0. \tag{8.37}$$

If $\epsilon = 0$ there is one single root $x = -\frac{b}{2\gamma}$, whose location is independent of t. Obviously, this case is not interesting, since it implies a stationary solution. Therefore, assume that $\epsilon \neq 0$. Then, there are two trajectories of critical points

$$u_{1,2}(v) = -\frac{\gamma}{3\epsilon} \pm \sqrt{\frac{\gamma^2}{9\epsilon^2} - (v + \frac{\beta}{3\epsilon})}, \tag{8.38}$$

which only exist when $v \leq \frac{\gamma^2}{9\epsilon^2} - \frac{\beta}{3\epsilon}$. They meet at the bifurcation point

$$(u_{bifurc};\ v_{bifurc}) = \left(-\frac{\gamma}{3\epsilon};\ \frac{\gamma^2}{9\epsilon^2} - \frac{\beta}{3\epsilon}\right). \tag{8.39}$$

From the second derivative $\tilde{L}_{uu}(u_{1,2};\ v) = 2\gamma + 6\epsilon u_{1,2} = \pm 6\epsilon\sqrt{\frac{\gamma^2}{9\epsilon^2} - (v + \frac{\beta}{3\epsilon})}$ it can be seen that the second derivative has different sign in the two critical points. Thus, the bifurcation consists of one maximum point and one minimum point that meet and annihilate (see figure 8.5).

Figure 8.5. Third order Taylor expansion of the scale-space embedding: Schematic view over the loci of the critical points as scale changes. The bifurcation consists of a maximum point and a minimum point that meet and annihilate.

Drift velocity estimates. Assume for a moment that $x = x_0$ is a critical point for the mapping $x \mapsto L(x;\ t_0)$, i.e., that at $u = 0$ is a critical point for the mapping $u \mapsto \tilde{L}(u;\ 0)$. Then, $\beta = 0$ and (8.39) gives an estimate of the time Δt_{bifurc} and the distance Δx_{bifurc} to the bifurcation

$$(\Delta x_{bifurc};\ \Delta t_{bifurc}) = (\frac{\gamma}{3\epsilon};\ (\frac{\gamma}{3\epsilon})^2) = (\frac{L_{xx}}{L_{xxx}};\ (\frac{L_{xx}}{L_{xxx}})^2). \qquad (8.40)$$

So far, no numerical experiments have been made testing the feasibility of using this estimate for scale step regulation in blob linking. Note, however, that despite the pessimistic upper bounds on the drift velocities discussed in section 8.1.1, the local extremum will hardly escape far outside the support region of the grey-level blob. This property turns out to be very useful in the blob linking algorithm to be described in section 9.2.

OBSERVATION 8.4. (COARSE BOUND ON THE DRIFT OF LOCAL EXTREMA) *Although the drift velocity of a local extremum point may momentarily be very large (tend to infinity near a bifurcation), when scale changes, the grey-level blob support region defines a natural spatial region to search for blobs in at the next level of scale.*

Reduction to the fold unfolding. To simplify further considerations, introduce again new coordinates by

$$\xi = u + \frac{\gamma}{3\epsilon}, \quad \tau = v + \frac{\beta}{3\gamma} - \frac{\gamma^2}{9\epsilon^2}, \quad \tilde{\tilde{L}}(\xi;\ \tau) = \tilde{L}(\xi - \frac{\gamma}{3\epsilon};\ \tau - \frac{\beta}{3\gamma} - \frac{\gamma^2}{9\epsilon^2}).$$

Then, the expression for the scale-space representation reduces to polynomial representative of the fold unfolding

$$\lambda(\xi;\ \tau) = \tilde{\tilde{L}}(\xi;\ \tau) - (\alpha - \frac{\gamma\beta}{3\epsilon} + \frac{2\gamma^3}{27\epsilon^2}) = \epsilon(\xi^3 + 3\xi\tau). \qquad (8.41)$$

All these coordinate shifts only mean that the coordinate axes have been translated such that the bifurcation occurs at $(\xi;\ \tau) = (0;\ 0)$, and a constant has been subtracted to achieve $\lambda(0;\ 0) = 0$. Hence, $\lambda\colon \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ still satisfies the diffusion equation.

8.5.2. The fold singularity in one dimension

Consider again the generic unfolding of the scale-space embedding in the neighbourhood of a bifurcation

$$L(x;\ t) = x^3 + 3xt, \qquad (8.42)$$

where x and t should be interpreted local coordinates in a coordinate system centered at the bifurcation point. As mentioned above, the critical points of this function are given by

$$\partial_x L(x;\ t) = 3(x^2 + t) = 0, \qquad (8.43)$$

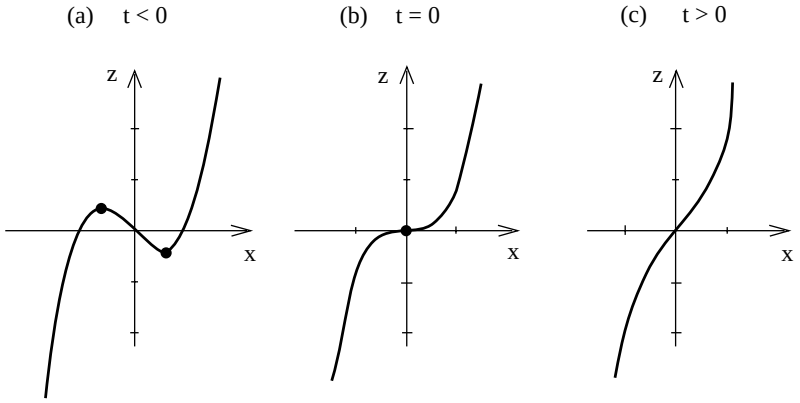

Figure 8.6. Fold unfolding in the one-dimensional case: Schematic view of the smoothed signal; (a) before the bifurcation, (b) at the bifurcation, and (c) after the bifurcation.

that is by

$$\xi_1(t) = -\sqrt{-t}, \quad \xi_2(t) = +\sqrt{-t}. \tag{8.44}$$

Moreover, the critical values are

$$\lambda_1(t) = -\lambda_2(t) = L(\xi_1(t);\, t) = -L(\xi_2(t);\, t) = +2(-t)^{\frac{3}{2}}. \tag{8.45}$$

In terms of grey-level blobs, this bifurcation describes the simultaneous annihilation of a bright and a dark blob. Obviously, the contrasts of the two one-dimensional blobs have equal magnitude

$$C_1(t) = C_2(t) = \mid \lambda_2(t) - \lambda_1(t) \mid = 4(-t)^{\frac{3}{2}}, \tag{8.46}$$

and the boundaries ρ_1 and ρ_2 of the support regions are determined by the equations $\lambda(\rho_1, t) = \lambda_2(t)$ and $\lambda(\rho_2, t) = \lambda_1(t)$, with solutions $\rho_1(t) = +2\sqrt{-t}$ and $\rho_2(t) = -2\sqrt{-t}$. Hence, the variation of the area of the support region follows

$$A_1(t) = |\xi_2(t) - \rho_2(t)| = A_2(t) = |\rho_1(t) - \xi_1(t)| = 3\sqrt{-t}, \tag{8.47}$$

Figure 8.7. The situation before the bifurcation occurs. Illustration of the definitions of ξ_1, ξ_2, λ_1, λ_2, ρ_1 and ρ_2.

and the grey-level blob volume is

$$V_1(t) = V_2(t) = \int_{x=\rho_2(t)}^{\xi_2(t)} |\lambda(x;\, t) - \lambda_2(t)|\, dx = \frac{27(-t)^2}{4}. \quad (8.48)$$

Assuming that the scale-space blob is delimited by some minimum scale $t_{min} < 0$, its scale-space blob volume can be computed by

$$S_1 = \int_{t_{min}}^{0} V_1(t)\, dt = \frac{9(-t_{min})^3}{4}. \quad (8.49)$$

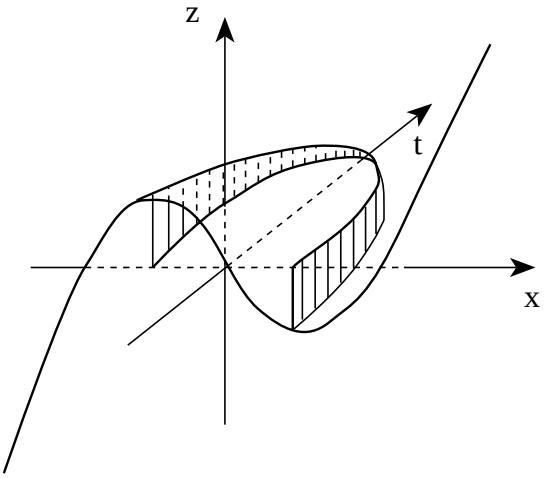

Figure 8.8. The fold singularity in the one-dimensional case.

8.5.3. The fold singularity in two dimensions

In the two-dimensional case the generic fold unfolding is

$$x^3 + ux \pm y^2. \quad (8.50)$$

By replacing each polynomial by a corresponding polynomial satisfying the diffusion equation (i.e., x^3 by $x^3 + 3xt$, (x by x), and y^2 by $y^2 + t$), we get

$$L(x, y;\, t) = x^3 + (u + 3t)x \pm (y^2 + t). \quad (8.51)$$

Obviously, u corresponds to an unessential translation of the scale parameter and can be omitted. Moreover, assume that the sign of the $\pm(y^2+t)$ term is positive. Then, the scale-space family to be studied is

$$L(x, y;\, t) = x^3 + 3xt + y^2 + t, \quad (8.52)$$

where x, y and t should again be interpreted as offset coordinates. The critical points of this mapping are given by

$$\begin{cases} \partial_x L = 3(x^2 + t) = 0, \\ \partial_y L = 2y = 0, \end{cases} \tag{8.53}$$

and their type by the sign of

$$\det(\mathcal{H}L) = L_{xx} L_{yy} - L_{xy}^2 = 12x. \tag{8.54}$$

If $t < 0$ there are two solutions:

$$r_1(t) = (x_1(t), y_1(t)) = (-\sqrt{-t}, 0), \quad r_2(t) = (x_2(t), y_2(t)) = (+\sqrt{-t}, 0),$$

where r_1 describes a saddle path, and r_2 a minimum path. The values at the critical points are

$$L_1(t) = L(r_1(t);\, t) = t - 2t\sqrt{-t}, \quad L_2(t) = L(r_2(t);\, t) = t + 2t\sqrt{-t}.$$

In terms of grey-level blobs, this singularity describes the annihilation of a dark grey-level blob. The boundary of the support region is obtained by solving the equation $L(x, y;\, t) = L_1(t)$, which can be reduced to

$$x^3 + 3tx + y^2 - 2(-t)^{\frac{3}{2}} = 0 \tag{8.55}$$

(see figure 8.9). Let y^- and y^+ denote the results of solving for y as a function of x and t in (8.55). Closed-form expressions for the blob descriptors can then be calculated as

$$C(t) = L_1(t) - L_2(t) = 4(-t)^{\frac{3}{2}}, \tag{8.56}$$

$$A(t) = \int_{x=-\sqrt{-t}}^{2\sqrt{-t}} \int_{y=y^-(x;\, t)}^{y=y^+(x;\, t)} dy\, dx = \frac{24\sqrt{3}(-t)^{\frac{5}{4}}}{5}, \tag{8.57}$$

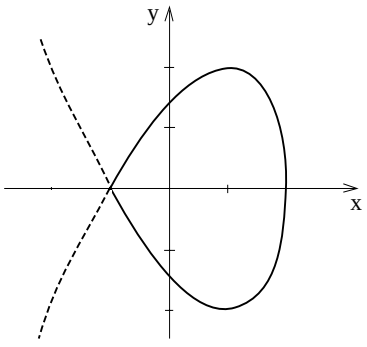

Figure 8.9. The support region of the grey-level blob. (The dashed line indicates the continuation of the level curve corresponding to the base level of the blob).

$$V(t) = \int_{x=-\sqrt{-t}}^{2\sqrt{-t}} \int_{y=y^-(x;\,t)}^{y^+(x;\,t)} (2(-t)^{\frac{3}{2}} - 3tx - x^3 - y^2)\, dy\, dx = \frac{3456\sqrt{3}(-t)^{\frac{11}{4}}}{385}.$$

If the sign of the $(y^2 + t)$ term in (8.51) instead is selected negative, the trajectories of critical points will be similar. The only difference is that the minimum point is replaced by a saddle point and the saddle point by a maximum point. In terms of blobs, that bifurcation corresponds to the annihilation of a bright blob.

Figure 8.10. The fold unfolding in two dimensions $L(x, y;\, t) = x^3 + 3xt \pm (y^2 + t)$ describes (a) a saddle point and a minimum (maximum) point merging with increasing scale, i.e., (b) the annihilation of a dark (bright) grey-level blob.

8.5.4. The cusp singularity in two dimensions

Consider next the generic unfolding of the cusp singularity

$$x^4 + ux^2 + vx \pm y^2. \tag{8.58}$$

To make this function satisfy the diffusion equation, replace x^4 by $x^4 + 6tx^2 + 3t^3$, x^2 by $x^2 + t$, and y^2 by $y^2 + t$. This gives

$$L(x, y;\, t) = x^4 + (6t + u)x^2 + vx + ut + 3t^2 \pm (y^2 + t). \tag{8.59}$$

Here, the parameter u corresponds to a non-essential translation of the scale parameter and can be disregarded. Moreover, assume that the sign of the $\pm(y^2 + t)$ term is positive. Then, we get the unfolding

$$L(x, y;\, t) = x^4 + 6x^2t + vx + 3t^2 + y^2 + t, \tag{8.60}$$

where v is a free parameter. The critical points satisfy

$$\begin{cases} \partial_x L = 4x^3 + 12tx + v = h(x) = 0, \\ \partial_y L = 2y = 0, \end{cases} \tag{8.61}$$

and their type is determined by the sign of the Hessian

$$(\mathcal{H}L)(x, y;\, t) = 24(x^2 + t). \tag{8.62}$$

After some calculations it can be shown that the roots to $h(x) = 4x^3 + 12tx + v = 0$ obey the following qualitative behaviour:

- If $t > -(\frac{v}{8})^{\frac{2}{3}}$ then the equation $h(x) = 0$ has only one real root, and the unique stationary point is a local minimum.
- If $t < -(\frac{v}{8})^{\frac{2}{3}}$ then $h(x) = 0$ has three distinct roots x_i, satisfying $x_1(t) < -\sqrt{-t} < x_2(t) < +\sqrt{-t} < x_3(t)$. From (8.62) it follows that $x_1(t)$ and $x_3(t)$ are local minima and that $x_2(t)$ is a saddle.
- If $t = -(\frac{v}{8})^{\frac{2}{3}}$ then $h(x)$ has either one root of multiplicity three or two roots of multiplicity one and two. The bifurcation occurs at $x = (\frac{v}{8})^{\frac{1}{3}}$, and corresponds to the root with multiplicity greater than one. The behaviour around this point depends on the value of v, see figure 8.11.

Hence, the singularity describes a minimum-saddle pair annihilating under the influence of another minimum. In terms of blobs, this corresponds to two grey-level blobs merging into one.

Note that variation of the parameter v affect the bifurcation diagram of the critical points (figure 8.11), while the bifurcation diagram for the grey-level blobs remains the same (figure 8.12). This demonstrates that *bifurcation relations between grey-level blobs are more stable than bifurcation relations between critical points.*

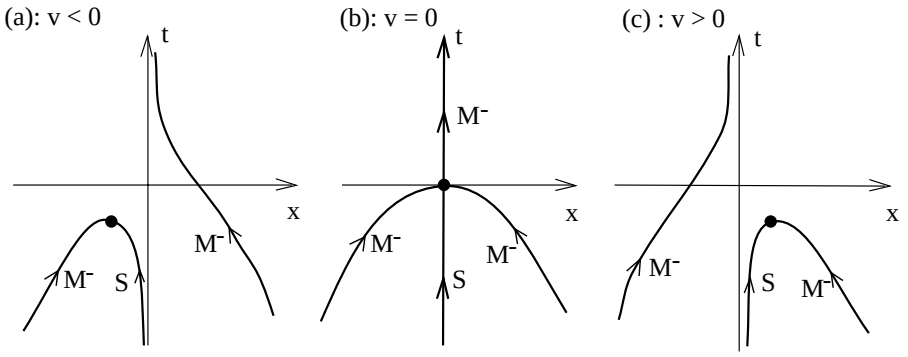

Figure 8.11. The cusp unfolding in two dimensions $L(x, y;\ t) = x^4 + 6x^2t + vx + 3t^2 + (y^2 + t)$ describes a minimum and a saddle merging under the influence of another minimum. Different events may occur depending on the value of v.

Figure 8.12. In terms of grey-level blobs, the three situations in figure 8.11 correspond to two dark grey-level blobs merging into one (independent of the value of v).

If the sign of the $\pm(y^2+t)$ is instead selected negative, then x_1 and x_2 will be saddle points, and x_2 a maximum, while if the sign of the entire unfolding is changed, all maxima are replace by minima and vice versa.

Zero-Crossings of the Laplacian. Introduce a parameter α such that

$$L_\alpha(x,y;\,t) = x^4 + 6x^2t + vx + 3t^2\alpha(y^2+t), \tag{8.63}$$

and analyse the zero-crossings of the Laplacian, which are given by

$$\frac{\partial^2 L_\alpha}{\partial x^2} + \frac{\partial^2 L_\alpha}{\partial y^2} = 12x^2 + 12t + 2\alpha = 0. \tag{8.64}$$

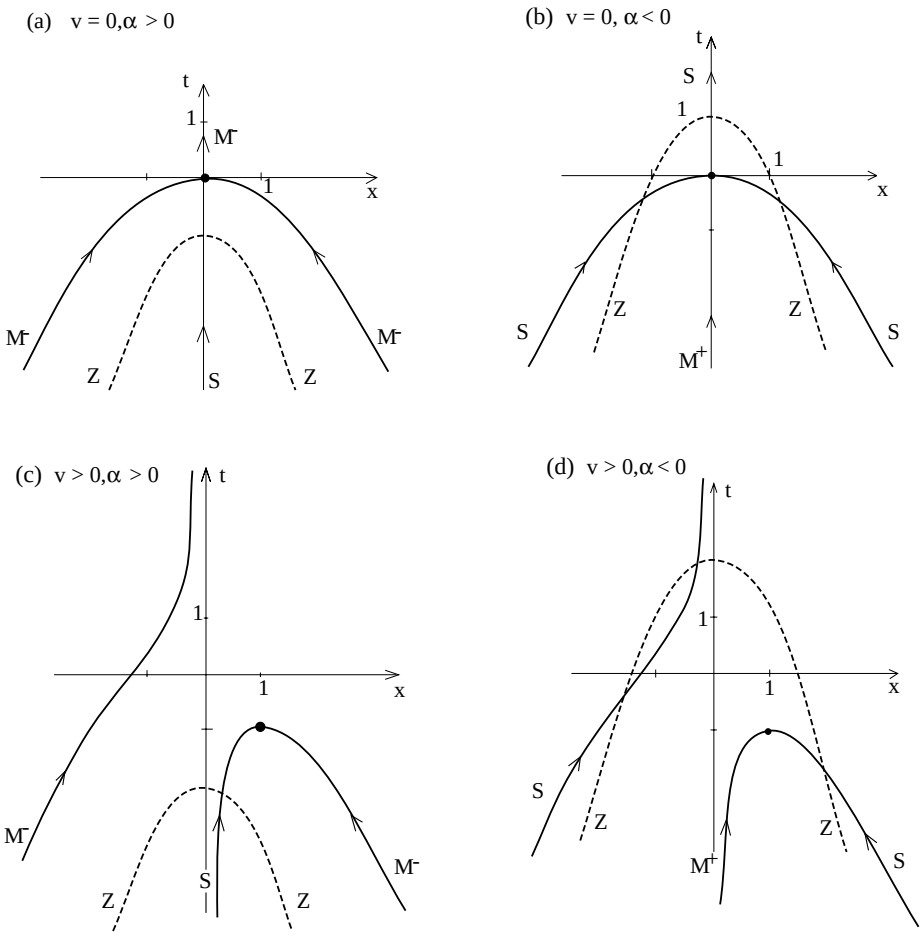

Figure 8.13. Locations of the zero-crossings of the Laplacian (marked by Z) and critical points (marked by M^+, M^-, and S) for the cusp unfolding (8.63) in the two-dimensional case. Note that during a certain scale interval the zero-crossings of the Laplacian fail to enclose isolated local extrema.

Obviously, there are two solutions $x = \pm\sqrt{-t - \alpha/6}$ if $t \leq -\alpha/6$, and we can observe that these zero-crossing curves do not give a correct subdivision around the local extrema for all t (see figure 8.13).

This example demonstrates that *in two (and higher) dimensions there is no absolute relationship between the locations of the Laplacian zero-crossing curves and the local extrema of a signal.* A Laplacian zero-crossing curve may enclose either no extremum, one extremum, or more than one local extremum. Only in the one-dimensional case it holds that there is exactly one local extremum point between two zero-crossings of the second derivative.

Drift velocity analysis. To analyse the drift velocity of the local extremum point *not* involved in the bifurcation, consider the one-dimensional version of (8.60) (delete the $(y^2 + t)$ term), and differentiate with respect to t:

$$\partial_t x = -\frac{x}{x^2 + t}. \tag{8.65}$$

To find the scale where the drift velocity is maximal, differentiate again and set the derivative to zero:

$$\partial_{tt} x = \frac{2xt}{(x^2 + t)^3} = 0. \tag{8.66}$$

Here, we are not interested in the case $x = 0$, since the behaviour at the bifurcation has already been analysed. Thus, as expected, the maximum drift velocity occurs when $t = 0$. Then, $x = (-\frac{v}{4})^{1/3}$ and

$$|\partial_t x|_{max} = -\frac{1}{x} = (\frac{4}{v})^{\frac{1}{3}}, \tag{8.67}$$

which shows that the maximum drift velocity tends to infinity as v tends to zero. This example demonstrates a further consequence of the results in section 8.1.1, namely that *even for critical points not directly involved in bifurcations there is no absolute upper bound one their drift velocity*, a conclusion which valid both in one and two dimensions.

Application to edge tracking. This analysis gives a further explanation to some of the problems that occur when edge focusing is applied to "staircase edges" (see figure 8.1 and the brief discussion in section 8.1.1). From experiments (Bergholm 1989) it is known that, in general, only one of the two edges in such configurations will be found by the focusing algorithm, and that in certain cases even that edge might get lost.

The fact that only one of the edges will be found is obvious from the bifurcation diagram in figure 8.11 provided that the focusing procedure is initiated from a sufficiently coarse scale and the bifurcation takes places

sufficiently far away from the edge subject to tracking. The bifurcation diagram and the previous analysis for local extrema also indicate that the drift velocity of an edge point may increase rapidly even though the edge is not directly involved in any bifurcation, and hence exceed the finite drift velocity estimate used by the edge focusing algorithm.

8.6. Relating differential singularities across scales

Although the analysis so far has been concerned with critical points, the ideas behind it are general, and can be extended to other aspects of image structures. Differential singularities (features which can be defined as zero-crossings of (possibly non-linear) differential expressions) are conceptually easy to relate across scales.

For simplicity, let us restrict the analysis to the two-dimensional case, and consider features that at any scale t can be defined by

$$h(x, y;\ t) = 0 \tag{8.68}$$

for some function $h\colon \mathbb{R}^2 \times \mathbb{R}_+ \to \mathbb{R}^N$, where N is either 1 or 2. Using the implicit function theorem it is easy to analyze the dependence of (x, y) on t in the solution to (8.68). The results to be derived give estimates of the drift velocity of different features due to scale-space smoothing, and provides a theoretical basis for linking and identifying corresponding features at adjacent scales.

8.6.1. Zero-dimensional entities (points)

Assume first that $N = 2$, and write $h(x, y;\ t) = (h_1(x, y;\ t), h_2(x, y;\ t))$ for some functions $h_1, h_2\colon \mathbb{R}^2 \times \mathbb{R}_+ \to \mathbb{R}$. The derivative of the mapping h at a point $P_0 = (x_0, y_0;\ t_0)$ is

$$h'|_{P_0} = \left(\begin{array}{ccc} \partial_x h_1 & \partial_y h_1 & \partial_t h_1 \\ \partial_x h_2 & \partial_y h_2 & \partial_t h_2 \end{array} \right)\Big|_{P_0} = \left(\begin{array}{cc} \frac{\partial(h_1, h_2)}{\partial(x, y)} & \frac{\partial(h_1, h_2)}{\partial(t)} \end{array} \right)\Big|_{P_0}. \tag{8.69}$$

If the matrix $\partial(h_1, h_2)/\partial(x, y)$ is non-singular at P_0, then the solution (x, y) to $h(x, y;\ t_0) = 0$ will be an isolated point. Moreover, the implicit function theorem guarantees that there exists some local neighbourhood around P_0 where (x, y) can be expressed as a function of t. The derivative of that mapping $t \mapsto (x, y)$ is:

$$\left(\begin{array}{c} \partial_t x \\ \partial_t y \end{array} \right)\Big|_{P_0} = - \left(\begin{array}{cc} \partial_x h_1 & \partial_y h_1 \\ \partial_x h_2 & \partial_y h_2 \end{array} \right)\Big|^{-1}_{P_0} \left(\begin{array}{c} \partial_t h_1 \\ \partial_t h_2 \end{array} \right)\Big|_{P_0}. \tag{8.70}$$

If h is a function of the spatial derivatives of L only, which is the case, for example, for the feature extractors treated in section 6.1.4, then the fact that spatial derivatives of L satisfy the diffusion equation

$$\partial_t L_{x^i y^j} = \tfrac{1}{2}(\partial_{xx} + \partial_{yy}) L_{x^i y^j}, \tag{8.71}$$

can be used for replacing derivatives with respect to t by derivatives with respect to x and y. In this way, closed form expression can be obtained containing only partial derivatives of L with respect to x and y.

For example, the *junction candidates* given by (6.14) satisfy $(\tilde{\kappa}_{\bar{u}}, \tilde{\kappa}_{\bar{v}}) = (0, 0)$. In terms of directional derivatives, (8.70) can then be written

$$\left(\begin{array}{c} \partial_t u \\ \partial_t v \end{array} \right) \Big|_{P_0} = - \left(\begin{array}{cc} \tilde{\kappa}_{\bar{u}\bar{u}} & \tilde{\kappa}_{\bar{u}\bar{v}} \\ \tilde{\kappa}_{\bar{u}\bar{v}} & \tilde{\kappa}_{\bar{v}\bar{v}} \end{array} \right) \Big|_{P_0}^{-1} \left(\begin{array}{c} \partial_t \tilde{\kappa}_{\bar{u}} \\ \partial_t \tilde{\kappa}_{\bar{v}} \end{array} \right) \Big|_{P_0} . \tag{8.72}$$

By differentiating the expressions for $\tilde{\kappa}_{\bar{u}}$ and $\tilde{\kappa}_{\bar{v}}$ with respect to t, by using the fact that the spatial derivatives satisfy the diffusion equation, and by expressing the result in terms of directional derivatives along the preferred u- and v-directions (see section 6.1.1), the following expressions can be obtained (the calculations have been done using Mathematica[4])

$$\begin{aligned}
\tilde{\kappa}_{\bar{u}} &= L_{\bar{v}}^2 L_{\bar{u}\bar{u}\bar{u}}, \\
\tilde{\kappa}_{\bar{v}} &= L_{\bar{v}}^2 L_{\bar{u}\bar{u}\bar{v}} + 2L_{\bar{v}}(L_{\bar{u}\bar{u}} L_{\bar{v}\bar{v}} - L_{\bar{u}\bar{v}}^2), \\
\tilde{\kappa}_{\bar{u}\bar{u}} &= L_{\bar{v}}^2 L_{\bar{u}\bar{u}\bar{u}\bar{u}} + 2L_{\bar{u}\bar{u}}(L_{\bar{u}\bar{u}} L_{\bar{v}\bar{v}} - L_{\bar{u}\bar{v}}^2) + 2L_{\bar{v}}(L_{\bar{u}\bar{v}} L_{\bar{u}\bar{u}\bar{u}} - L_{\bar{u}\bar{u}} L_{\bar{u}\bar{u}\bar{v}}), \\
\tilde{\kappa}_{\bar{u}\bar{v}} &= L_{\bar{v}}^2 L_{\bar{u}\bar{u}\bar{u}\bar{v}} + 2L_{\bar{u}\bar{v}}(L_{\bar{u}\bar{u}} L_{\bar{v}\bar{v}} - L_{\bar{u}\bar{v}}^2) + 2L_{\bar{v}}(L_{\bar{v}\bar{v}} L_{\bar{u}\bar{u}\bar{u}} - L_{\bar{u}\bar{v}} L_{\bar{u}\bar{u}\bar{v}}), \\
\tilde{\kappa}_{\bar{v}\bar{v}} &= L_{\bar{v}}^2 L_{\bar{u}\bar{u}\bar{v}\bar{v}} + 2L_{\bar{v}\bar{v}}(L_{\bar{u}\bar{u}} L_{\bar{v}\bar{v}} - L_{\bar{u}\bar{v}}^2) \\
&\quad + 2L_{\bar{v}}(L_{\bar{u}\bar{u}} L_{\bar{v}\bar{v}\bar{v}} + 2L_{\bar{v}\bar{v}} L_{\bar{u}\bar{u}\bar{v}} - 3L_{\bar{u}\bar{v}} L_{\bar{u}\bar{v}\bar{v}}), \\
\partial_t \tilde{\kappa}_{\bar{u}} &= L_{\bar{v}}^2 (L_{\bar{u}\bar{u}\bar{u}\bar{u}\bar{u}} + L_{\bar{u}\bar{u}\bar{u}\bar{v}\bar{v}})/2 \\
&\quad + (L_{\bar{u}\bar{u}} L_{\bar{v}\bar{v}} - L_{\bar{u}\bar{v}}^2)(L_{\bar{u}\bar{u}\bar{u}} + L_{\bar{u}\bar{v}\bar{v}}) + L_{\bar{v}}(L_{\bar{u}\bar{u}\bar{u}} L_{\bar{v}\bar{v}\bar{v}} - L_{\bar{u}\bar{u}\bar{v}} L_{\bar{u}\bar{v}\bar{v}}), \\
\partial_t \tilde{\kappa}_{\bar{v}} &= L_{\bar{v}}^2 (L_{\bar{u}\bar{u}\bar{u}\bar{u}\bar{v}} + L_{\bar{u}\bar{u}\bar{v}\bar{v}\bar{v}})/2 + (L_{\bar{u}\bar{u}} L_{\bar{v}\bar{v}} - L_{\bar{u}\bar{v}}^2)(L_{\bar{u}\bar{u}\bar{v}} + L_{\bar{v}\bar{v}\bar{v}}) \\
&\quad + L_{\bar{v}}(L_{\bar{v}\bar{v}}(L_{\bar{u}\bar{u}\bar{u}\bar{u}} + L_{\bar{u}\bar{u}\bar{v}\bar{v}}) + L_{\bar{u}\bar{u}}(L_{\bar{v}\bar{v}\bar{v}\bar{v}} + L_{\bar{u}\bar{u}\bar{v}\bar{v}}) \\
&\quad - 2L_{\bar{u}\bar{v}}(L_{\bar{u}\bar{u}\bar{u}\bar{v}} + L_{\bar{u}\bar{v}\bar{v}\bar{v}})) \\
&\quad + L_{\bar{v}}(L_{\bar{u}\bar{u}\bar{v}}(L_{\bar{u}\bar{u}\bar{v}} + L_{\bar{v}\bar{v}\bar{v}}) - L_{\bar{u}\bar{v}\bar{v}}(L_{\bar{u}\bar{u}\bar{u}} + L_{\bar{u}\bar{v}\bar{v}})).
\end{aligned}$$

(These expressions simplify somewhat if we make use of $L_{\bar{u}\bar{u}\bar{u}}|_{P_0} = 0$, which follows from $\tilde{\kappa}_{\bar{u}} = 0$.) Note that as long as the Hessian matrix of $\tilde{\kappa}$ is non-degenerate, the sign of the $\tilde{\kappa}_{\mathcal{H}}$ and $\tilde{\kappa}_{\bar{u}\bar{u}}$ will be constant. This means that the type of extremum remains the same. For *local extrema* of the grey-level landscape, given by $(L_x, L_y) = (0, 0)$, the expression for the drift velocity, of course, reduces to previously derived expression (8.3).

8.6.2. One-dimensional entities (curves)

If $N = 1$, then there will no longer be any unique correspondence between points at adjacent scales. An ambiguity arises, very similar to what is called the aperture problem in motion analysis. Nevertheless, we can determine the drift velocity in the normal direction of the curve.

[4]Mathematic is a registered trademark of Wolfram Research Research Inc., U.S.A.

Given a function $h\colon \mathbb{R}^2 \times \mathbb{R}_+ \to \mathbb{R}$ consider the solution to $h(x, y;\ t) = 0$. Assume that $P_0 = (x_0, y_0;\ t_0)$ is a solution to this equation and that the gradient of the mapping $(x, y) \mapsto h(x, y;\ t_0)$ is non-zero. Then, in some neighbourhood around (x_0, y_0) the solution (x, y) to $h(x, y;\ t_0) = 0$ defines a curve. Its normal at (x_0, y_0) is given by $(\cos\phi, \sin\phi) = (h_x, h_y)/(h_x^2 + h_y^2)^{1/2}$ at P_0. Consider next the function $\tilde{h}\colon \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ defined by $\tilde{h}(s;\ t) = h(x_0 + s\cos\phi, y_0 + s\sin\phi;\ t)$. It has the derivative

$$\tilde{h}_s(0;\ t_0) = h_x(x_0, y_0;\ t_0)\cos\phi + h_y(x_0, y_0;\ t_0)\sin\phi = \sqrt{h_x^2 + h_y^2}\Big|_{P_0}.$$

Since this derivative is non-zero, we can apply the implicit function theorem. It follows that there exists some neighbourhood around P_0 where $\tilde{h}(s;\ t) = 0$ defines s as a function of t. The derivative of this mapping is

$$\partial_t s|_{P_0} = -\tilde{h}_s\Big|_{P_0}^{-1} \tilde{h}_t\Big|_{P_0} = -\frac{h_t}{\sqrt{h_x^2 + h_y^2}}\Big|_{P_0}. \tag{8.73}$$

As an example of this, consider an *edge* given by non-maximum suppression

$$h = \alpha = L_x^2 L_{xx} + 2L_x L_y L_{xy} + L_y^2 L_{yy} = 0. \tag{8.74}$$

By differentiating (8.74), by using the fact that the derivatives of L satisfy the diffusion equation, and by expressing the result in terms of the directional derivatives we get

$$\alpha = L_{\bar{v}}^2 L_{\bar{v}\bar{v}} = 0, \tag{8.75}$$

$$\alpha_{\bar{u}} = L_{\bar{v}}^2 L_{\bar{u}\bar{v}\bar{v}} + 2L_{\bar{v}} L_{\bar{u}\bar{v}} L_{\bar{u}\bar{u}}, \tag{8.76}$$

$$\alpha_{\bar{v}} = L_{\bar{v}}^2 L_{\bar{v}\bar{v}\bar{v}} + 2L_{\bar{v}} L_{\bar{u}\bar{v}}^2, \tag{8.77}$$

$$\alpha_t = L_{\bar{v}}^2 (L_{\bar{u}\bar{u}\bar{v}\bar{v}} + L_{\bar{v}\bar{v}\bar{v}\bar{v}})/2 + L_{\bar{v}} L_{\bar{u}\bar{v}} (L_{\bar{u}\bar{u}\bar{u}} + L_{\bar{u}\bar{v}\bar{v}}). \tag{8.78}$$

To summarize, the drift velocity in the normal direction of a *curved edge* in scale-space is (with $\alpha_{\bar{u}}$ and $\alpha_{\bar{u}}$ according to (8.76)–(8.77))

$$(\partial_t u, \partial_t v) = -\frac{L_{\bar{v}}(L_{\bar{u}\bar{u}\bar{v}\bar{v}} + L_{\bar{v}\bar{v}\bar{v}\bar{v}}) + 2L_{\bar{u}\bar{v}}(L_{\bar{u}\bar{u}\bar{u}} + L_{\bar{u}\bar{v}\bar{v}})}{2((L_{\bar{v}} L_{\bar{u}\bar{v}\bar{v}} + 2L_{\bar{u}\bar{v}} L_{\bar{u}\bar{u}})^2 + (L_{\bar{v}} L_{\bar{v}\bar{v}\bar{v}} + 2L_{\bar{u}\bar{v}}^2)^2)} \Big(\frac{\alpha_{\bar{u}}}{L_{\bar{v}}}, \frac{\alpha_{\bar{v}}}{L_{\bar{v}}}\Big). \tag{8.79}$$

Unfortunately, this expression cannot be further simplified unless additional constraints are posed on L. For a *straight edge*, however, where all partial derivatives with respect to u are zero, it reduces to

$$(\partial_t u, \partial_t v) = -\frac{1}{2}\frac{L_{\bar{v}\bar{v}\bar{v}\bar{v}}}{L_{\bar{v}\bar{v}\bar{v}}}(0, 1), \tag{8.80}$$

which agrees with the result in (8.5). For a curve given by the *zero-crossings of the Laplacian* we have

$$(\partial_t u, \partial_t v) = -\frac{\nabla^2(\nabla^2 L)}{2((\nabla^2 L_{\bar{u}})^2 + (\nabla^2 L_{\bar{v}})^2)}(\nabla^2 L_{\bar{u}}, \nabla^2 L_{\bar{v}}), \tag{8.81}$$

which also simplifies to (8.80) if all directional derivatives in the u-direction are set to zero. Similarly, for a *parabolic curve*, given by $\det(\mathcal{H}L) = L_{xx}L_{yy} - L_{xy}^2 = 0$, the drift velocity in the normal direction is

$$\begin{aligned}(\partial_t p, \partial_t q) = -\frac{L_{qq}L_{pppp} + (L_{pp} + L_{qq})L_{ppqq} + L_{pp}L_{qqqq}}{2((L_{pp}L_{pqq} + L_{qq}L_{ppp})^2 + (L_{pp}L_{qqq} + L_{qq}L_{ppq})^2)} \\ \times (L_{pp}L_{pqq} + L_{qq}L_{ppp}, L_{pp}L_{qqq} + L_{qq}L_{ppq}).\end{aligned} \tag{8.82}$$

Here, the result has been expressed in a pq-coordinate system, with the p- and q-axes aligned to the principal axes of curvature such that the mixed second-order directional derivative $L_{\bar{p}\bar{q}}$ is zero.

8.7. Density of local extrema as function of scale

In some applications it is of interest to know how the density of local extrema can be expected to vary with scale. One example is the derivation of effective scale, a transformed scale parameter intended to capture the concept of "scale-space lifetime" in a proper manner (see section 7.7). Of course, this question seems to be very difficult or even impossible to answer to generally, since such a quantity can be expected to vary substantially from one image to another. How should one then be able to talk about "expected behaviour"? Should one consider all possible (realistic) images, study how this measure evolves with scale and then form some kind of average?

In this section a simple study will be performed. We will consider random noise data with normal distribution. Under these assumptions it turns out to be possible to derive a compact closed form expression for this quantity. The analysis will be based on a treatment by Rice (1945) about the expected density of zero-crossings and local maxima of stationary normal processes; see also (Papoulis 1972; Cramer and Leadbetter 1967).

8.7.1. Continuous analysis

The density of local maxima μ for a stationary normal process is given by the second and fourth derivatives of the autocorrelation function R

$$\mu = \frac{1}{2\pi}\sqrt{-\frac{R^{(4)}(0)}{R''(0)}}. \tag{8.83}$$

This expression can also be written as (Rice 1945; Papoulis 1972)

$$\mu = \frac{1}{2\pi}\sqrt{\frac{\int_{-\infty}^{\infty}\omega^4 S(\omega)d\omega}{\int_{-\infty}^{\infty}\omega^2 S(\omega)d\omega}}, \tag{8.84}$$

where S is the spectral density

$$S(\omega) = \int_{-\infty}^{\infty} e^{-i\omega\tau} R(\tau)\, d\tau. \tag{8.85}$$

Since the scale-space representation L is generated from the input signal f by a linear transformation, the spectral density of L, denoted S_L, is given by

$$S_L(\omega) = |H(\omega)|^2 S_f(\omega), \tag{8.86}$$

where S_f is the spectral density of f, and $H(\omega)$ is the Fourier transform of the impulse response h

$$H(\omega) = \int_{-\infty}^{\infty} h(t) e^{-i\omega t}\, dt \tag{8.87}$$

In our scale-space case, h is of course the Gaussian kernel

$$g(\xi;\, t) = \frac{1}{\sqrt{2\pi t}} e^{-\xi^2/2t}, \tag{8.88}$$

which has the Fourier transform

$$G(\omega;\, t) = e^{-\omega^2 t/2}. \tag{8.89}$$

Assuming that f is generated by white noise with $S_f(w) = 1$ this gives

$$S_L(\omega) = e^{-\omega^2 t}. \tag{8.90}$$

Using the formula (Spiegel 1968: 15.77)

$$\int_0^{\infty} x^m e^{-ax^2} dx = \frac{\Gamma((m+1)/2)}{2a^{(m+1)/2}}, \tag{8.91}$$

a closed form expression can be calculated for the density of local maxima of a continuous signal, $p_c(t)$:

$$p_c(t) = \frac{1}{2\pi}\sqrt{\frac{\int_{-\infty}^{\infty}\omega^4\, e^{-\omega^2 t} d\omega}{\int_{-\infty}^{\infty}\omega^2\, e^{-\omega^2 t} d\omega}} = \frac{1}{2\pi}\sqrt{\frac{2\frac{\Gamma(5/2)}{2t^{5/2}}}{2\frac{\Gamma(3/2)}{2t^{3/2}}}} = \frac{1}{2\pi}\sqrt{\frac{3}{2}}\frac{1}{\sqrt{t}}. \tag{8.92}$$

Of course, an identical result applies[5] to local minima. To summarize,

[5]Observe that the same type of qualitative behaviour ($p_c(t) \sim t^{-\frac{1}{2}}$) applies also to the local extrema in the *spatial derivatives* of the scale-space representation (just replace $H = G$ by $H = (i\omega)^n G$ in the previous analysis).

PROPOSITION 8.5. (DENSITY OF LOCAL EXTREMA; WHITE NOISE; 1D) *In the scale-space representation of a one-dimensional continuous signal generated by a white noise stationary normal process, the expected density of local maxima (minima) in a smoothed signal at a certain scale decreases with scale as* $t^{-1/2}$.

This scale dependence implies that a graph showing the density of local maxima (minima) as function of scale can be expected[6] to be a *straight line* in a log-log diagram

$$\log(p_c(t)) = \tfrac{1}{2}\log(\tfrac{3}{2}) - \log(2\pi) - \tfrac{1}{2}\log(t) = \text{constant} - \tfrac{1}{2}\log(t).$$

In section 7.7.1 it was shown that a natural way to convert the ordinary scale parameter t into a transformed scale parameter, effective scale τ, is by $\tau(t) = A + B\log(p(t))$, where $p(t)$ again denotes the expected density of local extrema at a certain scale t and A and B are arbitrary constants. This result gives:

COROLLARY 8.6. (EFFECTIVE SCALE FOR CONTINUOUS SIGNALS; 1D) *For continuous one-dimensional signals the effective scale parameter* τ_c *as function of the ordinary scale parameter* t *is (up to an arbitrary affine transformation, i.e., for some arbitrary constants* A' *and* $B' > 0$*) given by a logarithmic transformation*

$$\tau_c(t) = A' + B'\log(t). \tag{8.93}$$

An interesting question concerns what will happen if the uncorrelated white noise model for the input signal is changed. A self-similar spectral density that has been applied to fractals (Barnsley *et al.* 1988; Gårding 1988) is given by

$$S_f(w) = w^{-\beta}. \tag{8.94}$$

For one-dimensional signals, reasonable values of β are obtained between 1 and 3. Of course, such a distribution is somewhat non-physical, since $S_f(w)$ will tend to infinity as t tends to zero and neither one of the spectral moments is convergent. However, when multiplied by a Gaussian function the second and fourth order moments in (8.84) will converge provided that $\beta < 3$. We obtain,

$$p_{c,\beta}(t) = \frac{1}{2\pi}\sqrt{\frac{\int_{-\infty}^{\infty}\omega^4\, e^{-\omega^2 t}\omega^{-\beta}d\omega}{\int_{-\infty}^{\infty}\omega^2\, e^{-\omega^2 t}\omega^{-\beta}d\omega}} = \frac{1}{2\pi}\sqrt{\frac{3-\beta}{2}}\frac{1}{\sqrt{t}} \quad (\beta < 3). \tag{8.95}$$

[6] Of course, we cannot expect that a graph showing this curve for a particular signal to be a straight line, since this would require some type of ergodicity assumption that in general will not be satisfied. However, the average behaviour over many different types of imagery can be expected to be close to this situation.

PROPOSITION 8.7. (DENSITY OF LOCAL EXTREMA; FRACTAL NOISE)
In the scale-space representation of a one-dimensional continuous signal generated by a stationary normal process with spectral density $\omega^{-\beta}$ ($\omega \in [0, 3[$), *the expected density of local maxima (minima) in a smoothed signal at a certain scale decreases with scale as* $t^{-1/2}$.

Note that also this graph will be a straight line in a log-log diagram.

8.7.2. Discrete analysis

From the previous continuous analysis it is apparent that the density of local extrema may tend to infinity when the scale parameter tends to zero. As earlier indicated, this result is not applicable to discrete signals, since in this case the density of local extrema will have an upper bound because of the finite sampling. Hence, to capture what happens in the discrete case, a genuinely discrete treatment is necessary. The treatment will be based on the discrete scale-space concept developed in chapter 3. Given a discrete signal $f : \mathbb{Z} \to \mathbb{R}$ the scale-space representation $L : \mathbb{Z} \times \mathbb{R}_+ \to \mathbb{R}$ is defined by

$$L(x;\, t) = \sum_{n=-\infty}^{\infty} T(n;\, t)\, f(x - n), \tag{8.96}$$

where $T(n;\, t) = e^{-t} I_n(t)$ is the discrete analogue of the Gaussian kernel, and I_n are the modified Bessel functions of integer order (Abramowitz and Stegun 1964). Equivalently, this scale-space family can be defined in terms of a semi-discretized version of the diffusion equation

Consider the scale-space representation of a signal generated by a random noise signal. The probability that a point at a certain scale is say a local maximum point is equal to the probability that its value is greater than (or possibly equal to)[7] the values of its nearest neighbours:

$$\begin{aligned} &P(x_i \text{ is a local maximum at scale } t) \\ &\qquad = P((L(x_i;\, t) \geq L(x_{i-1};\, t)) \wedge (L(x_i;\, t) \geq L(x_{i+1};\, t))). \end{aligned} \tag{8.97}$$

If we assume that the input signal f is generated by a stationary normal process then also L will be a stationary normal process and the distribution of any triple $(L_{i-1}, L_i, L_{i+1})^T$, from now on denoted by $\xi = (\xi_1, \xi_2, \xi_3)^T$, will be jointly normal, which means that its statistics will be completely determined by the mean vector and the autocovariance matrix. Trivially, the mean of ξ is zero provided that the mean of f is zero.

[7] Although there are several ways to define a local extremum of a discrete signal using different combinations of "$>$" and "$\geq$", these definitions will yield the same result with respect to this application.

Since the transformation from f to L is linear, the autocovariance C_L for the smoothed signal L will be given by

$$C_L(\cdot;\ t) = T(\cdot;\ t) * T(\cdot;\ t) * C_f(\cdot) = T(\cdot;\ 2t) * C_f(\cdot), \tag{8.98}$$

where C_f denotes the autocovariance of f. In the last equality we have made use of the semigroup property $T(\cdot;\ s) * T(\cdot;\ t) = T(\cdot;\ s+t)$ for the family of convolution kernels. If the input signal consists of white noise, then C_f will be the discrete delta function and $C_L(\cdot;\ t) = T(\cdot;\ 2t)$. Taking the symmetry property $T(-n;\ t) = T(n;\ t)$ into account, the distribution of ξ will be jointly normal with mean vector m_{3D} and covariance matrix C_{3D} given by:

$$m_{3D} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad C_{3D} = \begin{pmatrix} T(0;\ 2t) & T(1;\ 2t) & T(2;\ 2t) \\ T(1;\ 2t) & T(0;\ 2t) & T(1;\ 2t) \\ T(2;\ 2t) & T(1;\ 2t) & T(0;\ 2t) \end{pmatrix}. \tag{8.99}$$

By introducing new variables $\eta_1 = \xi_2 - \xi_1$ and $\eta_2 = \xi_2 - \xi_3$, it follows that $\eta = (\eta_1, \eta_2)^T$ will be jointly normal and its statistics completely determined by

$$m_{2D} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad C_{2D} = \begin{pmatrix} a_0(t) & a_1(t) \\ a_1(t) & a_0(t) \end{pmatrix}. \tag{8.100}$$

From well-known rules for the covariance $C(\cdot,\cdot)$ of a linear combination it follows that

$$\begin{aligned} a_0(t) &= C(\eta_1, \eta_1) = C(\eta_2, \eta_2) \\ &= 2(T(0;\ 2t) - T(1;\ 2t)), \end{aligned} \tag{8.101}$$

$$\begin{aligned} a_1(t) &= C(\eta_1, \eta_2) = C(\eta_2, \eta_1) \\ &= T(0;\ 2t) - 2T(1;\ 2t) + T(2;\ 2t). \end{aligned} \tag{8.102}$$

From $a_0(t) - a_1(t) = T(0;\ t) - T(2;\ t)$ and the unimodality property of T ($T(i;\ t) \geq T(j;\ t)$ if $|i| > |j|$) it follows that $a_0(t) > a_1(t)$ and trivially $a_0(t) > 0$ for all t. Now $p_d(t)$ can be expressed in terms of a two-dimensional integral

$$p_d(t) = \int\int_{\{\eta=(\eta_1,\eta_2):\ (\eta_1 \geq 0) \wedge (\eta_2 \geq 0)\}} \frac{1}{\sqrt{(2\pi)^2 |C_{2D}|}} e^{-\eta^T C_{2D}^{-1} \eta/2}\, d\eta_1\, d\eta_2. \tag{8.103}$$

After some calculations (Lindeberg 1991: appendix A.5.4) it follows that

$$p_d(t) = \frac{1}{4} + \frac{1}{2\pi} \arctan\left(\frac{a_1(t)}{\sqrt{a_0^2(t) - a_1^2(t)}} \right). \tag{8.104}$$

Observe that for any $a_0(t)$ and $a_1(t)$ this value is guaranteed to never be outside the interval $[0, \frac{1}{2}]$. With the expressions for $a_0(t)$ and $a_1(t)$,

given by smoothing with the discrete analogue of the Gaussian kernel, the maximum value over variations in t is obtained for $t = 0$:

$$p_d(0) = \frac{1}{3}. \tag{8.105}$$

PROPOSITION 8.8. (DENSITY OF LOCAL EXTREMA;1D)
In the scale-space representation (8.96) of a one-dimensional discrete signal generated by a white noise stationary normal process, the expected density of local maxima (minima) in a smoothed signal at a certain scale t is given by (8.104) *with $a_0(t)$ and $a_1(t)$ according to* (8.101) *and* (8.102).

It is interesting to compare the discrete expression (8.104) with the earlier continuous result (8.92). The scale value where the continuous estimate gives a density equal to the discrete density at $t = 0$ is given by the equation $p_c(t) = p_d(0)$, that is by

$$\frac{1}{2\pi}\sqrt{\frac{3}{2}}\frac{1}{\sqrt{t}} = \frac{1}{3} \tag{8.106}$$

which has the solution

$$t_{c-d} = \frac{27}{8\pi^2} \approx 0.3420 \tag{8.107}$$

This corresponds to a σ-value of about 0.5848. Below this scale value the continuous analysis is, from that point of view, definitely not a valid approximation of what will happen to discrete signals. By combining proposition 8.8 with the effective scale concept we get:

COROLLARY 8.9. (EFFECTIVE SCALE FOR DISCRETE SIGNALS; 1D)
For discrete one-dimensional signals the effective scale parameter τ_d as function of the ordinary scale parameter t is given by

$$\tau_d(t) = A'' + B'' \log \left(\frac{4\pi}{3\pi + 6 \arctan\left(\frac{a_1(t)}{\sqrt{a_0^2(t) - a_1^2(t)}} \right)} \right) \tag{8.108}$$

for some arbitrary constants A'' and $B'' > 0$ with $a_0(t)$ and $a_1(t)$ given by (8.101) *and* (8.102).

When defining the effective scale τ_d for discrete signals it is natural to let $t = 0$ correspond to $\tau_d = 0$. In that case A'' will be zero. Without loss of generality, we will from now on set $A'' = 0$ and $B'' = 1$.

8.7.3. Asymptotic behaviour at fine and coarse scales

A second order MacLaurin expansion of $p_d(t)$ (Lindeberg 1991: appendix A.5.5) yields

$$p_d(t) = \frac{1}{3} - \frac{1}{2\sqrt{3}\pi}t + \frac{1}{6\sqrt{3}\pi}t^2 + O(t^3). \tag{8.109}$$

This means that the effective scale $\tau_d(t)$ can be MacLaurin expanded (Lindeberg 1991: appendix A.5.5)

$$\tau_d(t) = \log\left(\frac{p_d(0)}{p_d(t)}\right) = \frac{\sqrt{3}}{2\pi}t + \left(\frac{1}{2\sqrt{3}\pi} + \frac{3}{8\pi^2}\right)t^2 + O(t^3), \tag{8.110}$$

showing that *at fine scales the effective scale τ for one-dimensional discrete signals is approximately an affine function of the ordinary scale parameter t*. On the other hand, a Taylor expansion of $p_d(t)$ at coarse scales gives

$$p_d(t) = \frac{1}{2\pi}\sqrt{\frac{3}{2}}\frac{1}{\sqrt{t}}\left(1 + \frac{1}{8t} + O(\frac{1}{t^2})\right) \tag{8.111}$$

which asymptotically agrees with the continuous result in (8.92). By inserting this expression into the expression for effective scale and by using $p_d(0) = \frac{1}{3}$ it follows that

$$\tau_d(t) = \log\left(\frac{p_d(0)}{p_d(t)}\right) = \log\left(\frac{2\pi}{3}\sqrt{\frac{2}{3}}\right) + \frac{1}{2}\log(t) + \log\left(1 - \frac{1}{8t} + O(\frac{1}{t^2})\right).$$

Hence, *at coarse scales the effective scale τ for one-dimensional discrete signals is approximately (up to an arbitrary affine transformation) a logarithmic function of the ordinary scale parameter t.*

The term $\log(1 - \frac{1}{8t} + O(\frac{1}{t^2}))$ expresses how much the effective scale derived for discrete signals differs from the effective scale derived for continuous signals, provided that the same values of the (arbitrary) constants A and B are selected in both cases.

8.7.4. Comparisons between the continuous and discrete results

As an illustration of the difference between the density of local maxima in the scale-space representation of a continuous and a discrete signal, we show the graphs of p_c and p_d in figure 8.14 (linear scale) and figure 8.15 (log-log scale). As expected, the curves differ significantly for small t and approach each other as t increases.

Numerical values quantifying this difference for a few values of t are given in table 8.1. It shows the ratio

$$\tau_{diff}(t) = \frac{\tau_d(t) - \tau_c(t)}{\tau_c(2t) - \tau_c(t)} = \frac{\tau_d(t) - \tau_c(t)}{\log(2)/2}, \tag{8.112}$$

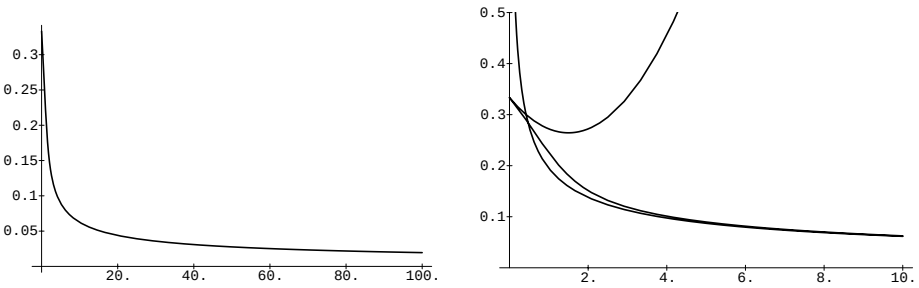

Figure 8.14. The density of local maxima of a discrete signal as function of the ordinary scale parameter t in linear scale. (a) Graph for $t \in [0, 100]$. (b) Enlargement of the interval $t \in [0, 10]$. For comparison the graphs showing the density of local extrema for a continuous signal $p_c(t)$ and the second order Taylor expansion of $p_d(t)$ around $t = 0$ have also been drawn. As expected, the continuous and discrete results differ significantly for small values of t but approach each other as t increases. The MacLaurin expansion is a valid approximation only in a very short interval around $t = 0$.

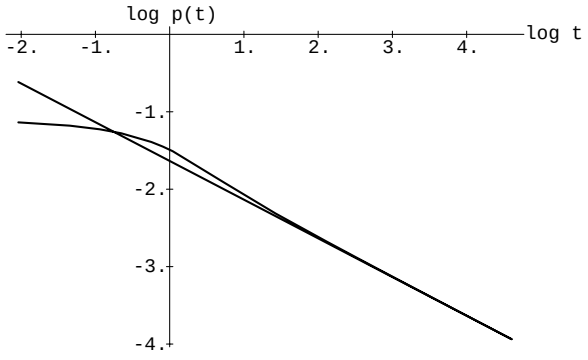

Figure 8.15. The density of local maxima of a continuous and a discrete signal as function of the ordinary scale parameter t in log-log scale ($t \in [0, 100]$). The straight line shows $p_c(t)$ and the other curve $p_d(t)$. It can be observed that p_c and p_d approach each other as the scale parameter increases. When t tends to zero, $p_c(t)$ tends to infinity while $p_d(t)$ tends to a constant ($\frac{1}{3}$).

t	$\tau_{diff}(t)$
0	∞
0.0625	250.30 %
0.25	67.46 %
1.0	-41.82 %
4.0	-10.47 %
16.0	-2.32 %
64.0	-0.56 %
256.0	-0.14 %
∞	0

Table 8.1. Indications about how the effective scale obtained from a discrete analysis differs from the effective scale given by the continuous scale-space theory. The quantity $\tau_{diff}(t)$ expresses the difference between $\tau_d(t)$ and $\tau_c(t)$ normalized such that one unit (100 %) in $\tau_{diff}(t)$ corresponds to the increase in τ_c induced by an increase in t with a factor of two.

which is a natural measure for how much the effective scale obtained from a continuous analysis differs from a discretely determined effective scale. The quantity is normalized so that one unit in τ_{diff} corresponds to the increase in τ_c induced by an increase in t with a factor of two.

8.7.5. Extension to two dimensions

The same type of analysis can, in principle, be carried out also for two-dimensional discrete signals. The probability that a specific point at a certain scale is a local maximum point is again equal to the probability that its value is greater than the values of its neighbours. Depending on the connectivity concept (four-connectivity or eight-connectivity on a square grid) we then obtain either a four-dimensional or an eight-dimensional integral to solve. However, because of the dimensionality of the integrals, no attempts have been made to calculate explicit expressions for the variation of the density as function of scale. Instead, for the purpose of implementation, the behaviour over scale has been simulated for various uncorrelated random noise signals (see section 7.7.1). From those experiments it has been empirically demonstrated that the $t^{-\alpha}$ dependence (with $\alpha \approx 1.0$) constitutes a reasonable approximation at coarse levels of scale.

> The reason the exponent α changes from 0.5 to 1.0 when going from one to two dimensions can intuitively be understood by a dimensional analysis: Assume (as in appendix 7.7.1) that the standard deviation of the Gaussian kernel, $\sigma = \sqrt{t}$, can be linearly related to a characteristic length, x, in the scale-space representation of an N-dimensional signal at scale t. Moreover, assume that a characteristic distance d between the local extrema in that signal is linearly

related to x. Then, the density of local extrema will be proportional to $d^{-N} \sim x^{-N} \sim \sigma^{-N}$, that is to $t^{-N/2}$.

8.8. Summary

We have analysed the behaviour of critical points in scale-space and shown that non-degenerate critical points will in general form regular curves across scales. Along those we have provided generally valid estimates of the drift velocity. At degenerate critical points the behaviour is more complicated and bifurcations may take place. For one-dimensional signals, the only bifurcation events possible when the scale parameter increases, are annihilations of pairs of local maxima and minima, while for two-dimensional signals both annihilations and creations of pairs of local extrema and saddle points can occur. Applied to grey-level and scale-space blobs only annihilations and merges will take place in the one-dimensional case, while the list of possibilities in two-dimensions comprises four types: annihilations, merges, splits and creations.

Finally, it should be pointed out that this analysis has been mainly concerned with the scale-space concept for continuous signals. When implementing this theory computationally it is obvious that one has to consider sampled (that is, discrete) data. At coarse scales, when a characteristic length of features in the image can be regarded as large compared to the distance between adjacent grid points, it seems plausible that the continuous results should constitute a reasonable approximation to what will happen in the scale-space representation of a discrete signal and vice versa. However, as indicated in section 8.7 this similarity will not necessarily hold[8] at fine scales. In those cases, a genuinely discrete theory might be needed. A thorough understanding of what happens to continuous signals under scale-space smoothing constitutes a first step towards this goal.

[8]Some conceptual complications arise in this context (for instance, what should be meant by drift velocity for discrete signals). It seems difficult to estimate such a quantity accurately, especially at fine scales, since in the discrete case local extrema will not move continuously, but rather in steps from one pixel to the next. Thus, one cannot talk about velocity, but rather about how long it takes until an extremum point moves one pixel. An alternative approach to this problem would be to analyse the feature points with sub-pixel accuracy (although this idea has not been carried out). Other conceptual problems concern what should be meant by singularities or degenerate and nondegenerate critical points in the discrete case. One possibility is to define these in terms of adjacent pixels having equal values, or by transitions (say e.g., blob bifurcations). But, will the classification of possible blob events still be valid in the discrete case?

Bibliography

[1] M. Abramowitz and I. A. Stegun, eds., *Handbook of Mathematical Functions*. Applied Mathematics Series, National Bureau of Standards, 55 ed., 1964.

[2] V. I. Arnold, *Singularity Theory, Selected papers*, vol. 53 of *London Mathematical Society Lecture Note Series*. Cambridge: Cambridge University Press, 1981.

[3] V. I. Arnold, S. M. Gusein-Zade, and A. N. Varchenko, *Singularities of Smooth Maps, Volume I*. Boston: Birkhäuser, 1985.

[4] V. I. Arnold, S. M. Gusein-Zade, and A. N. Varchenko, *Singularities of Smooth Maps, Volume II*. Boston: Birkhäuser, 1988.

[5] M. F. Barnsley, R. L. Devaney, B. B. Mandelbrot, H.-O. Peitgen, D. Saupe, and R. F. Voss, *The Science of Fractals*. New York: Springer-Verlag, 1988.

[6] F. Bergholm, "Edge focusing", *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 9, pp. 726–741, Nov. 1987.

[7] F. Bergholm, *On the Content of Information in Edges and Optical Flow*. Dissertation, Dept. of Numerical Analysis and Computing Science, Royal Institute of Technology, May. 1989. ISRN KTH/NA/P–8904–SE.

[8] V. Berzins, "Accuracy of Laplacian edge detectors", *Computer Vision, Graphics, and Image Processing*, vol. 27, pp. 195–210, 1984.

[9] J. Canny, "A computational approach to edge detection", *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 8, no. 6, pp. 679–698, 1986.

[10] J. J. Clark, "Singularity theory and phantom edges in scale-space", *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 10, no. 5, pp. 720–727, 1988.

[11] H. Cramer and M. R. Leadbetter, *Stationary and Related Stochastic Processes*. New York: John Wiley and Sons, 1967.

[12] J. Gårding, "Properties of fractal intensity surfaces", *Pattern Recognition Letters*, vol. 8, pp. 319–324, Dec. 1988.

[13] C. G. Gibson, *Singular Points of Smooth Mappings*. Research Notes in Mathematics, London: Pitman Publishing, 1979.

[14] M. Golubitsky and D. G. Schaeffer, *Singularities and Groups in Bifurcation Theory I*, vol. 51 of *Applied Mathematical Sciences*. New York: Springer–Verlag, 1985.

[15] P. Johansen, "On the classification of toppoints in scale space", *J. of Mathematical Imaging and Vision*, vol. 4, pp. 57–67, 1994.

[16] P. Johansen, S. Skelboe, K. Grue, and J. D. Andersen, "Representing signals by their top points in scale-space", in *Proc. 8:th Int. Conf. on Pattern Recognition*, (Paris, France), pp. 215–217, Oct. 1986.

[17] J. J. Koenderink, "The structure of images", *Biological Cybernetics*, vol. 50, pp. 363–370, 1984.

[18] J. J. Koenderink, *Solid Shape*. Cambridge, Massachusetts: MIT Press, 1990.

[19] J. J. Koenderink and A. J. van Doorn, "Dynamic shape", *Biological Cybernetics*, vol. 53, pp. 383–396, 1986.

[20] L. M. Lifshitz and S. M. Pizer, "A multiresolution hierarchical approach to image segmentation based on intensity extrema", tech. rep., Departments of Computer Science and Radiology, University of North Carolina, Chapel Hill, N.C., U.S.A, 1987.

[21] T. Lindeberg, *Discrete scale space theory and the scale space primal sketch*. PhD thesis, Dept. of Numerical Analysis and Computing Science, Royal Institute of Technology, Stockholm, May. 1991. A revised and extended version published as book *Scale-Space Theory in Computer Vision* in Kluwer Int. Series in Engineering and Computer Science, 1994.

[22] T. Lindeberg, "Scale-space behaviour of local extrema and blobs", *J. of Mathematical Imaging and Vision*, vol. 1, pp. 65–99, Mar. 1992.

[23] Y.-C. Lu, *Singularity Theory and an Introduction to Catastrophe Theory*. New York: Springer-Verlag, 1976.

[24] A. Papoulis, *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, 1972.

[25] T. Poston and I. Stewart, *Catastrophe Theory and its Applications*. London: Pitman, 1978.

[26] S. O. Rice, "Mathematical analysis of random noise", *The Bell System Technical J.*, vol. XXIV, no. 1, pp. 46–156, 1945.

[27] W. Rudin, *Principles of Mathematical Analysis*. McGraw-Hill, 1976.

[28] M. R. Spiegel, *Mathematical Handbook of Formulas and Tables*. Schaum's Outline Series in Mathematics, McGraw-Hill, 1968.

[29] W. Zhang and F. Bergholm, "An extension of Marr's "signature" based edge classification", in *Proc. 7th Scandinavian Conference on Image Analysis*, (Aalborg, Denmark), pp. 435–443, Aug. 1991.

[30] X. Zhuang and T. S. Huang, "Multi-scale edge detection with Gaussian filters", in *Int. Conf. on Acoustics, Speech and Signal Processing*, (Tokyo, Japan), pp. 2047–2050, 1986.