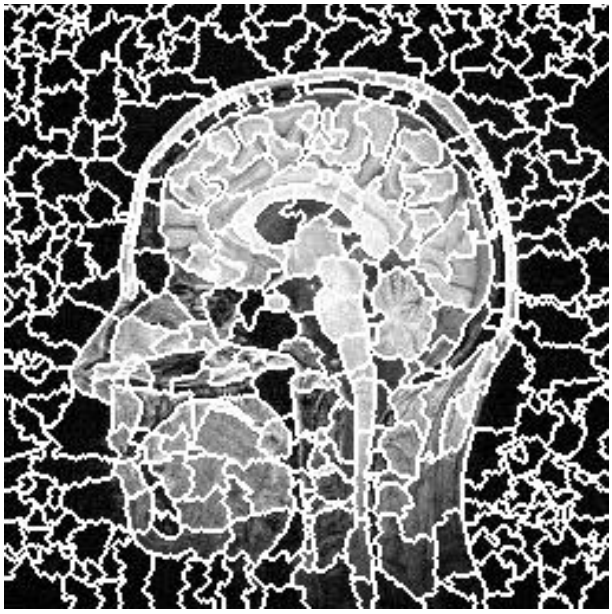
Multi–Scale Segmentation of Grey–Scale Images



Ole Fogh Olsen
Department of Computer Science
University of Copenhagen

September 11, 1996

Abstract

The topic of this thesis is segmentation of grey–scale images by accessing and relating the structures of the image for all reasonable scales. Grey–scale images are regarded as physical measurements obtained by an ensemble of detectors with finite aperture and finite readout precision. Based on this point of view, the notion of image is established using linear Gaussian scale space and Morse theory.

The problem of segmentation is addressed within the framework of linear Gaussian scale space. The image at a given scale is partitioned according to a dissimilarity measure. Each minimum in the measure corresponds to exactly one segment. The segment is defined as the catchment basin belonging to the minimum. The gradient squared is used as the measure of dissimilarity.

An essential part of the scale–space paradigm is to study the family of images indexed by scale as a family and not as a distinct set of images. The structure of the segments across scale are analysed exploiting the established duality between the minima in the dissimilarity measure and the segments. Catastrophe theory is used to analyse the structure of the minima across scale. The generic catastrophe events for the minima of the gradient squared are proved to be the fold and the cusp catastrophes.

The generic events of annihilation, creation, split and merge for the minima are used to analyse possible linking schemes for the segments. The implemented linking scheme produces a hierarchical structure. Two segments are merged when the saddle connecting the two corresponding minima is annihilated. A segment is divided when a border is created within the segment. The creation of a border corresponds to the creation of a saddle connecting the resulting two minima.

The duality between minima and catchment basins is exploited to make a robust implementation of the linking scheme. The actual detection and tracking of segments is region based. A semi–automatic segmentation tool based on this uncommitted segmentation scheme has been implemented. Experimental results obtained using the segmentation tools on artificial, camera and medical data are presented. The experimental data are two and three dimensional.

The singularity analysis and the implementation is performed for two and three dimensional images. The theory can be generalised to arbitrary finite dimension.

This thesis also includes an introduction to scale–space theory, differential invariants, differential description of local forms for two and three dimensional functions, single scale and multi–scale segmentation methods, Morse theory and Catastrophe theory

Contents

Preface

This is my Master thesis. The work was carried out in the period from October 1995 until submission August 1996 under the supervision of Professor Peter Johansen and Post Doc. Mads Nielsen at the Department of Computer Science, University of Copenhagen. An article based on this work will be published as a chapter in the book "Gaussian Scale–Space" [51].

The main bulk of the work was made in 1996. The final inspiration came with the presentation of preliminary work and participation in the PhD-school "Classical Scale–Space Theory". Discussions before, during and after this school were very inspiring. I thank the people who have helped me to an understanding of some of the concepts in Scale-Space Theory, Differential Geometry, Descriptions of superficial structure and Catastrophe theory.

Ole Fogh Olsen
fogh@diku.dk
Department of Computer Science
University of Copenhagen
Universitetsparken 1
DK-2100 Copenhagen East
Denmark

1 Introduction

This thesis is on multi-scale segmentation of grey–scale images. The offset will be taken from the title introducing the keywords in reverse order of appearance.

Grey–scale images

Grey–scale images are physical measurements acquired by a spatially ordered ensemble of detectors. An essential characteristic of physical measurements is finite resolution. The size of the detector, the aperture, has finite size and the precision of the measurement itself, the scalar, is finite as well. This property is essential in the understanding of structure and by that the task of segmentation.

The finite size of the detector has the effect that localization of details is only possible as long as their size is greater than the aperture of the detector. Details simply get too small compared to the aperture and the precision to be detected. This effect is illustrated in Figure 1. The images in (a) to (d) are recorded with the same "camera" but from increasing distance[1].

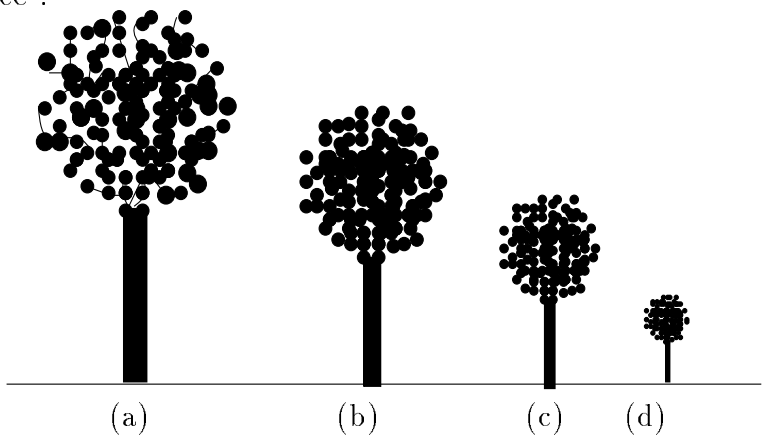


Figure 1: The same tree seen from different distances

Close to the tree (a) in Figure 1 all the leaves are connected in a network by branches. If we back off the width of the branches will be below the width of our aperture, this causes the funny effect of leaves floating freely in the air (b). Moving even further away will change the topology even more ending up with a stereotype of a tree: a connected blob on top of a trunk as in Figure 1 (d). An important point here is that objects only exist over a limited range of scales. There is no point in looking for leaves in (d) and in (a) the tree is not present as it is in (d). One can not see the forest for all the trees so to speak. The scale under which (a) is viewed has to be increased for the blob on a trunk to appear. If we can imitate the process taking place from (a) to (d) there is hope to identify an instance of (a) and (d) as a tree even though the topology of the two is very distinct in the recorded images.

[1] Approaching or withdrawing from a scene only results in a unified rescaling of the objects if the objects are located in approximately the same depth. A scenario where this is not the case could be the moon behind a tree.

Scale–space theory [56, 33, 19, 17, 40] has been developed over the last decade or so to handle the property of finite aperture size in a firm manner. In linear scale–space theory the signal is embedded in a family of functions by blurring the image with a Gaussian filter. The width of the filter is interpreted as the width of the aperture and in this sense the width indicates the scale on which we study the image structure. By this embedding we get the scale–space representation of the image which is indexed by a scale parameter and by usual spatial coordinates. It has been argued [17] that such a scale–space representation form a possible front–end for a general imaging system because it explicitly includes the essential scale aspect and is uncommitted with respect to position, direction and scale.

The property of finite precision for the intensity means that the scalar value representing the measurement is closer to represent an interval (or distribution over an interval) than a point in the set of possible values. This effect calls for a model that captures the notion of typicality. That is, the image measured with finite precision does not pinpoint one particular state (represented by a function) of the world but determines a class of possible states (a set of functions) of the world. An appropriate model is a typical member of the class.

Segmentation

The idea of a visual front–end in the form of a scale–space representation is fairly well established and recognised as useful in the field of computer vision. But it is still not common to build systems on top of this front–end. One of the goals in this thesis is to outline how a segmentation tool can be build based on the visual front–end, linear Gaussian scale space. This work is not restricted to solve a specific task [2] and therefore we will be without task specific knowledge. The resulting description, the segmentation, should therefore be general enough to capture essential image structures which are interesting to a range of tasks.

The analysis and development of the segmentation tools are limited to the study of the interesting cases of two and three dimensional images. The temporal aspect of imaging is not considered, hence only static images are considered. In two dimensional imaging the typical measured entity is a flux of photons in the visible spectrum. In three dimensions the measurements reflect a density. A rapidly growing production of three dimensional images takes place in medical imaging. An example of this could be measuring magnetic resonance (MR) often used for measuring the density of hydrogen (which is closely related to the type of tissue).

On the path from acquisition to a useful description of an image one of the biggest hurdles is the segmentation. That is, partitioning the image into basic structures corresponding to meaningful regions in image. Describing an image in terms of significant regions provides a powerful basis for further processing. Further processing could be: matching the shape against a database, visualization, quantitative measurements of the size and changes of this, manipulation (for instance in simulated surgery), etc. The usefulness of segmentations does of course not imply that they are difficult to obtain but experience in the computer vision community has proven it to be a non-trivial problem.

[2] With a task is meant a specific job like counting knots in planks.

The common strategy for segmentation algorithms is to detect discontinuities and/or smoothly varying regions in the image. This line will also be followed in this work. Here we will give a short rationale for why it is a good way of finding interesting regions.

Let us look at a simplified model of the world consisting of volumetric objects with smooth surfaces with constant surface reflectance. In such a world there are only three kinds of rays: the bundle of rays coming from infinity (background points); rays coming from a point on the surface (surface points) and finally rays gracing the boundaries of the objects (called rim points [34]). The pattern formed by these rays has the following form. Background points and surface points always occur in patches never in curves or points. The rim points occur in curves never patches or points; these curves can have terminations and junctions. It is not possible to draw a path from a surface point of one object to a surface point of another object(or to a background point) without crossing a rim. So surface and background patches are separated by rim curves albeit that not all rim curves separate distinct regions and one object may consist of several regions. In this scenario the luminance may change across the rims or across surface regions. In surface regions the luminance change is caused by a smooth change in depth and surface normal, consequently the luminance usually changes smoothly across surfaces. Across a rim the change is caused by a discontinuity in depth and in addition the surface normal usual changes as well, hence the luminance typically is discontinuous across a rim. Relaxing the assumptions on the world introduces lots of phenomena (shadows, texture, reflections,...) where discontinuities arise without corresponding to the rim of an object. It is by no means trivial to determine whether or not a discontinuity actually correspond to the rim of an object. What can be concluded is that the set of discontinuities in an image is a good basis for a segmentation since they form a superset for the rims.

In 3D medical imaging, the used imaging techniques capture to a high degree the structures of interest. In other words the probed densities are about the same in the same kind of tissue and change relatively much when a tissue type changes. Consequently the 3D scalar function representing the measurement of the density has discontinuities when the tissue type changes and that is why the hunt for discontinuities is also of interest in 3D medical images.

Multi–scale

An important consequence of the lack of task specific knowledge is that no scale for the image is preferred over another. The segmentation shall assess the image structure at all scales and even more important must probe the structure across scale. The importance of this was seen from Figure 1.

The blurring with a Gaussian filter causes the image structures to move, hence the localization of the boundaries is poorly determined for the blurred structure unless the localization of the equivalent unblurred version of the boundary can be found. To determine the connection between unblurred and more and more blurred versions of the image a description of the connection of structure across scale is needed.

In other words, in order to figure out that all the leaves (disconnected blobs) are in fact part of the same tree, then the structure must be probed at a sufficiently coarse scale (sufficiently blurred) for the leaves to become one blob. But the blurring process

causes the structure to move, therefore we loose the precise localization. To reestablish the localization we need to know how the structure moved when the scale was increased.

The structure across scale is often referred to as the deep structure. Already from the beginning of scale–space theory it was emphasised by Koenderink [33] that the family of blurred functions should be studied as a family not as a collection of images on different scales.

Feature detectors are often defined in terms of singularities of image differentials. The deep structure of such image features in scale–space can be analysed using Catastrophe theory (also known as singularity theory). We shall use Catastrophe theory to analyse the structure of our proposed segmentation. The second main goal in this thesis is to show the usefulness of Catastrophe theory for analysing the deep structure of features in scale space.

The structure of this thesis

Three essential questions must be addressed in image segmentation: the notion of an image, the notion of a segmentation and the method for producing the latter from the former. At a coarse scale these three points are the skeleton for the main body of this thesis.

The notion of image is established in section 2 and 3 with the introduction of concepts from scale–space theory and Morse theory. The notion of segmentation is founded in section 4 by a discussion on superficial structure and commonly used approaches to describe the structure. Based on this, a segmentation is chosen, rigorously defined and analysed in section 5. The behaviour of the chosen segmentation method in a multi–scale framework is analysed in section 8 using catastrophe theory. Catastrophe theory is introduced in section 6. Before embedding the chosen single scale segmentation in a multi–scale framework based on the singularity analysis a presentation of related multi–scale segmentation methods is given in section 7. In section 9 the results from the catastrophe analysis is used to form the possible linking schemes for segments at different scale. With this, the ideas and results can be used as basis for an implementation of interactive segmentation tools applicable to two and three dimensional grey scale images. These tools and the results obtained with these tools are presented in section 10. The thesis ends with general conclusions and comments on theoretical and practical results. An appendix with the most frequently used notation is attached at the end.

2 Scale–space theory

Multi–scale descriptions of images have been regarded as important from the early days of image analysis. Rosenfeld and Thurston [1] observed the advantage of using operators of different size for edge detection. Klinger [32] introduced the quadtree representation for image representation. Later the pyramid representation was developed by Burt [6] and Crowley [11]. This representation has been used for feature detection and image coding [7]. Marr [43] emphasised the importance of an *explicit* internal representation where information is directly available for decision processes. A clean formalism for image representation with explicit representation of scale was established with the introduction of scale–Space by Witkin [56] and its further development by Koenderink [33].

The main idea of multi–scale descriptions is to generate a series of images where the fine–scale details are successively more and more suppressed. The main difference between a scale–space representation and a pyramid representation of the signal is that the latter has a very coarse sampling in the scale parameters. This coarse sampling creates a combinatorially difficult problem when matching structures across scale, which is an undesirable property since the study of structure across scale is essential. In the scale–space representation the scale parameter is continuous and the image is not downsampled when scale is increased. This allows a formulation of the matching problem using differential entities. The left column of Figure 18 on page 75 in section 10.2.1 shows slices from a scale–space representation.

The Gaussian filter as the generator of a multi–scale representation can be derived in several ways. The basic assumptions for the derivation are based on the following ideas. A desirable representation of the image does not have a preferred direction, position or scale, and coarse scale structure must be caused by fine scale structure. The latter requirement states that spurious details should not be generated out of the blue. The reason is straight forward, causeless structure cannot be distinguished from structure reflecting the measurements, consequently the task is not solved based on the measurements but on artifacts in the representation. The first assumptions are based on non commitment. If no a priori knowledge of data is possessed the best way to proceed is to stay uncommitted and consequently process all data in the same way.

Koenderink [33] emphasised from the very beginning that the problem of scale must be handled in every imaging situation. This arises from the fact that relevant details in an image only exist over a finite range of scale, consequently it is a must to study image structure on an appropriate scale.

This section gives a brief introduction to scale–space theory and the concepts of the theory used in this work.

2.1 Inner scale and outer scale

An image posses a natural outer scale given by the limited extent of the image. The finite area of the individual measurements sets a lower bound for the resolution, the inner scale. The inner and outer scale of the image are intrinsic properties arising from the acquisition process, e.g. for a Charged Coupled Device (better know as a CCD) array the size and the spacing of photosensitive elements give the inner scale for the image, and the number

of elements gives the outer scale. For the visual system the discrete structure and the finite size of the receptive fields give the inner scale and the size of the retina gives the outer scale.

It is also appropriate to speak of the inner and outer scale of objects, and hence image features! The inner scale of an object is the coarsest scale where substructures of the object appear. The outer scale of an object corresponds to something like the minimum size of a window completely covering the object. So the scale range of an object is the range where an object has an existence of its own. It is not a collection of substructures and it is not smaller than the aperture of measurement device. As an example one can think of a treetop. The top is an object existing somewhere between the scale of leaves and the scale of the forest.

The inner and outer scale of the acquisition system bounds the range over which subjects can be studied, since the scale range for the subject has to overlap with the scale range for the image. A simple example is reading the letters in this text. If the text is too far away it is impossible to read because the letters are blurred together. The inner scale of the acquisition system is above the outer scale of the object. Being too close to the text has the opposite effect: it is impossible to form a letter because the outer scale of our visual system is below the inner scale of a letter[3].

The problem of finding the outer and inner scale of objects is definitely not trivial. The inner scale is the best level to localize the object and the outer scale is appropriate for detection.

If no preference for scale is given a priori the system should be able to study the structure over all scales. The real challenge is to exploit the structure over scale (the deep structure) and not just assess the structure as a collection of the images at different scales. Koenderink [33] states it in the following way: "Study the family as a family, i.e. define deep structure, the relations between structural features of different derived images."

The relation between structure at different scales is made accessible by adding a scale parameter to the original image.

2.2 The scale–space representation

Before surveying the derivation of the scale–space the resulting scale–space representation is presented: A scale–space representation of an image is constructed by convolving the image I with Gaussian kernels G:

$$G(x;\sigma) = \frac{1}{(2\pi\sigma^2)^{d/2}} \exp(\frac{-x^2}{2\sigma^2}) \tag{1}$$

The point x is a point in the space $\mathbb{R}^d$. Hence, Equation 1 is stated for arbitrary, finite number of dimensions d. The standard deviation of the Gaussian is denoted by σ and is often referred to as the width.

We denote the scale–space image by the symbol L defined in the following way:

$$L(x;\sigma) = G(x;\sigma) * I = \int_{a=-\infty}^{\infty} G(x-a;\sigma)I(a)da \tag{2}$$

[3]Other effects besides scale come into play in this example such as focusing and our inability to zoom.

The convolution operator denoted $*$ is defined in the standard way written in the right hand side of Equation 2. The measurements, the orignal image is denoted I. The scale–space image is a scalar function $L : \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}$ indexed by spatial variables and one scale parameter for the linear scale–space.

2.3 Derivation of the Gaussian

Historically, the concept of scale–space [56] arose from a wish to study the structure of signals (mainly one dimensional signals) on different resolutions. Witkin [56] noted that embedding a one dimensional signal in a one-parameter family by convolution with the Gaussian yields a family of signals where extrema could not be created, only annihilated. Witkin chose the Gaussian filter as generating filter because of its well–behavedness.

Drawing the paths of the maxima and minima in the scale–space plane for a one dimensional signal yields a result similar to Figure 2.

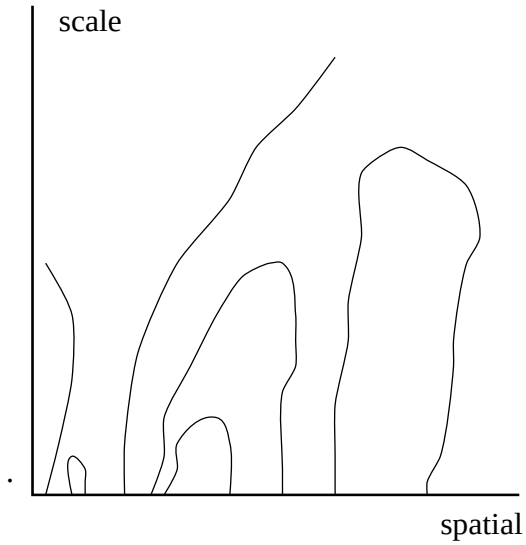


Figure 2: Free hand outline of possible paths for maxima and minima in a scale–space representation of a one dimensional signal.

The point where a maxima and a minima meet and annihilate is called a toppoint in scale space. It has been proven by Johansen *et al.* [29] that a band–limited one dimensional signal is uniquely determined up to a multiplicative and a additive constant by these toppoints. So given the top points the original signal can be reconstructed.

The Gaussian kernel from causality

The non–creation of extrema does not hold in arbitrary dimensions. Koenderink [33] introduced the concept of causality and on this basis generalized the scale–space to arbitrary dimensions. Koenderink [33] deduced from basic requirements of causality of structure and isotropy and homogeneity of the scale–space that the family of images should satisfy the diffusion Equation. Green's function for the diffusion Equation is the Gaussian, thus

the solution to the diffusion Equation can be found by convolving the initial conditions (the raw image) with a Gaussian. And the other way around the result of convolving a function with a Gaussian is a solution to the diffusion Equation. The two views are equivalent and can be interchanged when convenient.

The principle of causality states that all coarse scale structure must have a cause at a finer scale. The requirements of isotropy and homogeneity follows from the idea of non–commitment: without a priori knowledge of the signal all directions and positions should be treated equally.

Formally the criteria of causality is that new level surfaces in the scale space must not be created as the scale parameter increases. A level surface in scale–space is defined by:

$$\{(x;t) \in \mathbb{R}^d \times \mathbb{R} \mid L(x;t) = K, K \in \mathbb{R}\} \tag{3}$$

The causality requirement demands that the level surfaces have their concave opening towards finer scales. Figure 3 (a) shows a special presentation of a scale–space image. Figure 3 (a) consists of two parts, namely an isophote and the bottom slice of the scale–space image. The isophote has be been visualised in the case of 3D dimensional scale–space image. The viewing angle is from above the coarse scales. The concave openings of the isophote is directed away from viewer. The finest scale slice of the scale–space image is visible at the bottom of the isophotes. As reference, Figure 3 (b) shows the same slice as in (a). Hence in (b) is visualised a 2D slice of the 3D scale–space image sliced orthogonal to the scale direction. In (a) the same slice is tilted with respect to the viewer in an imaging 3D space and on top of this slice rises the isophote for the value 147.

The isophotes clearly illustrate how structures merge when the image is blurred. Bridges are formed between previously parted structures. A bridge is for instance visible at the neck of the man. The requirement of concavity towards fine scale implies that the Laplacian $I_{xx} + I_{yy}$ should equal the first order derivative in the scale direction I_s in extrema in the scale–space. Since the extrema can be located anywhere in scale–space the requirement should hold anywhere. So the causality principle leads to the diffusion Equation:

$$\triangle I = (\alpha(x;s))^2 I_s \tag{4}$$

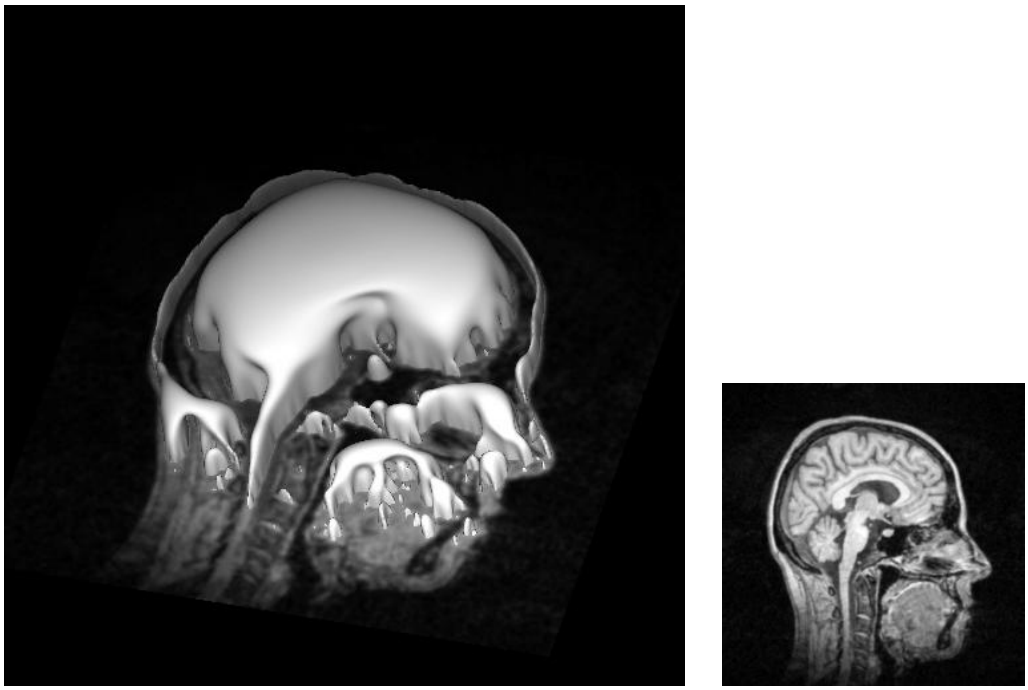
where $\alpha(x;s)$ is a nowhere vanishing real function. The additional requirements of isotropy and homogeneity limit the class of feasible functions α to functions depending solely on s. A monotonic re-parametrisation $t = \phi(s)$ puts the diffusion Equation 4 in this simple form:

$$\triangle I = I_t \tag{5}$$

This is the well–known physical Equation describing the evolution over time of a heat distribution given an initial distribution $I(.;0) = I_0$. The evolution takes place in a homogeneous medium with a constant conductance coefficient. The diffusion Equation is also referred to as the heat Equation!

The creation of new extrema

A typical misunderstanding of the causality principle is that new extrema cannot be created. This is only true in one spatial dimension. For two dimensions and higher



(a) (b)

Figure 3: (a) An image with the isophote for intensity 147 in scale space rising from the image. Scale increases out of the paper perpendicular to the image plane.(b) The same image as reference for (a)

creation of extrema is possible and happens generically [13]. The typical example of creation of new extrema is in a landscape scene: two hills connected by a bridge. The example is due to Lifshitz and Pizer [38], who also noted the creation empirically.

Example 1

Imagine a two–dimensional image function consisting of two hills, one of them somewhat higher than the other. Assume that they are smooth, wide, rather bell-shaped surfaces situated some distance apart clearly separated by a deep valley running between them. Connect the two tops by a narrow sloping ridge without any local extrema, so that the top point of the lower hill no longer is a local maximum. Let this configuration be the input image. When the operator corresponding to the diffusion Equation is applied to the geometry, the ridge will erode much faster than the hills. After a while it has eroded so much that the lower hill appears as a local maximum again.

Since the ridge between the tops is a fine–scale phenomenon compared to the coarse–scale hills it is quite reasonable that it should disappear before the hills[4]*. Thus, a new local extremum has been created.*

[4]As noted previously this effect can be noticed by looking at trees. The leaves are a coarser scale phenomenon than the thin branches supporting them, so at a suitable distance from the tree the outer branches will be too small to detect by our finite resolution measurement device but the leaves are still detectable. Therefore the leaves will appear to hang next to the tree but unconnected to it. Try it!

The Gaussian from group theory

Another approach to find the appropriate family of scale–space kernels was taken by Florack *et al.* [19]. Florack *et al.* proposed the following set of symmetry requirements for the scale–space representation:

- Semi–group structure (intermediate steps in the transformation makes no difference).
- Linearity (allowing superposition of input stimuli).
- Spatial shift invariance (absence of preferred position)
- Isotropy (absence of preferred direction)
- Scale invariance (absence of preferred scale)

The semi–group structure for the family of filter kernels states that the result of convolving two kernels yields a member of the family:

$$h(.;t_1) * h(.;t_2) = h(.;t_1 + t_2) \tag{6}$$

This implies that there should be no difference between transforming directly from a fine to a coarse scale and doing the transformation via intermediate steps. An element in the semi–group of rescaling operators do not in general have an inverse this implies that rescaling is irreversible. In other words, the rescaling gives rise to irreversible catastrophes in the topological structure in the image. Since the rescaling is causal by construction for increasing scale there is only causal bifurcations. For decreasing scale there is no guaranty for causality, consequently the blurring process cannot be reversed. (The concepts of catastrophes and bifurcations will be further described in section 6.)

On top of the assumptions Florack *et al.* [19] further assume separability of the kernel in a Cartesian coordinate system. The derivation [19] from all the above assumptions leads to the unique scale–space kernel in the Fourier domain:

$$\mathcal{G}(\omega;\sigma) = exp(-\frac{1}{2}\sigma^2\omega^2) \tag{7}$$

where ω is the frequence in the Fourier (frequence) domain. Equation 7 is in the spatial domain given by Equation 1. If the constraint of separability is withdrawn the more general result for the kernels are given by:

$$\hat{h}(\omega;\sigma) = exp(-\alpha \mid \omega\sigma \mid^p), \text{ where } \alpha > 0, p > 0 \tag{8}$$

The major steps in the derivation will now be outlined. The reader can refer to the literature [19, 52] for more details.

Any linear and shift–invariant operator h can be expressed as a convolution operator which in the Fourier domain transforms to a product:

$$\hat{L}(\omega;\sigma) = \hat{h}(\omega;\sigma)\hat{I}(\omega) \tag{9}$$

where $\hat{L}, \hat{h}$ and $\hat{I}$ denote the Fourier transforms of L, h and I. The Pi–theorem is then used to limit the possible form of operator $\hat{h}$. This fundamental theorem comes from physics and states that "Any physical law must be independent of the choice of fundamental parameters". This is also known as dimensional analysis. Dimensional analysis boils in practice down to the useful "lemma": "The unit on the left hand side of the equality sign has to equal the right hand". The Pi–theorem implies that "a function relating physical observations must be independent of the choice of dimensional units" [20]. Since $\hat{I}$ and $\hat{L}$ have the unit of illuminance then $\hat{h}$ can only depend on the dimensionless variable $\omega\sigma$. The assumption of isotropy implies that the operator $\hat{h}$ actually can only depend on the magnitude of the variable $\omega\sigma$:

$$\frac{\hat{L}(\omega;\sigma)}{\hat{I}(\omega)} = \hat{h}(\omega;\sigma) = \tilde{h}(\mid \omega\sigma \mid) \tag{10}$$

where $\tilde{h} : \mathbb{R} \rightarrow \mathbb{R}$ is a real function. The Equation states that the function $\hat{h}$ is a composed function and describe the first part of the composition. It was assumed that the parameter σ holds the unit length and is directly associated with each layer of the scale–space. The parameter t in the semi–group requirement depends monotonically on this parameter σ. The dependency is monotonic to preserve the ordering of scale values. We call the monotonic transformation ϕ: $t = \phi(\sigma)$. Using the semi–group constraint it can be shown that ϕ has this form:

$$t = \phi(\sigma) = C\sigma^p = \sigma^p \tag{11}$$

where the constant C is set to one, since it correspond to an indifferent scaling. The semi–group constraint can now be written in the form using the property 6 and the Equations 10 and 11:

$$\tilde{h}(\mid \omega\sigma_1 \mid)\tilde{h}(\mid \omega\sigma_2 \mid) = \tilde{h}(\mid \omega\phi^{-1}(\phi(\sigma_1) + \phi(\sigma_2)) \mid) \tag{12}$$

Where ϕ^{-1} denotes the inverse function of ϕ. We now define a function $\tilde{H}(x^p) = \tilde{h}(x) : \mathbb{R} \rightarrow \mathbb{R}$. Equation 12 can now be rewritten using Equation 11:

$$\tilde{H}(\mid \omega\sigma_1 \mid^p)\tilde{H}(\mid \omega\sigma_2 \mid^p) = \tilde{H}(\mid \omega\sigma_1 \mid^p + \mid \omega\sigma_2 \mid^p) \tag{13}$$

From Equation 13 we can deduce that the function $\hat{H}$ must be the exponential function. So finally we have as promised:

$$\hat{h}(\omega;\sigma) = \tilde{h}(\mid \omega\sigma \mid) = \tilde{H}(\mid \omega\sigma \mid^p) = exp(-\alpha \mid \omega\sigma \mid^p) \tag{14}$$

It is natural to require that measuring with an infinite scale always yields zero, formally stated:

$$\lim_{\sigma\rightarrow\infty} \hat{h}(\omega;\sigma) = 0 \tag{15}$$

This implies that $\alpha > 0$. The value for p remains to be determined.

Florack *et al.* imposed the separability constraint of the kernel to uniquely determine the value for p to 2. The constant α was fixed to the convenient value of $\frac{1}{2}$. This gives

the result presented in Equation 7. Demanding separability of the kernel together with rotational symmetry fix the function to exponential right away, so it might be advocated that the separability constraint is too strong [41], in the sence of being highly redundant to the other assumptions.

An alternative way is to consider differentiability with respect to the scale parameter. This constraint implies according to Lindeberg *et al.* [41], referring to Pauwel, that p must be even. Among the even values for p only the value 2 corresponds to non-negative convolution kernel. The extra requirement of non-negative kernel has been assumed in recent work by Florack [18], and with this requirement the Gaussian is the unique solution. The rationale of the requirement is that the aperture function can be viewed as a distribution and consequently have to be non–negative. The Gaussian $p = 2$ is the only kernel which fulfils the principle of causality.

It should be noted that it is the scale parameter $t = \sigma^2/2$ which has the additive property in the semi–group constraint.

In this section we have seen several ways to determine the Gaussians as the appropriate filter family. The basic assumptions have been non–commitment with respect to position, direction and scale.

2.4 Natural scale

The scale-space formalism gives a continuous scale representation of the image, but it has to be represented by a discrete set of sampled scales. In this section we will establish a natural dimensionless scale parameter τ for which an equi–distant sampling is natural. A dimensionless scale parameter $\tilde{\sigma}$ is established by using the intrinsic inner scale inherited from the image, the pixel width σ_0:

$$\tilde{\sigma} = \sigma/\sigma_0 \tag{16}$$

It is natural to work with order of magnitudes instead of absolute scale steps. This implies that the differential [5] of the natural scale parameter τ should equal the differential of $\tilde{\sigma}$ relative to $\tilde{\sigma}$:

$$d\tau = d\tilde{\sigma}/\tilde{\sigma} \tag{17}$$

Integration of Equation 17 reveals with the integration constant equal to zero:

$$\tau = \ln(\sigma/\sigma_0) \quad or \quad \sigma = \sigma_0 \exp(\tau), \quad \tau \in (-\infty; \infty) \tag{18}$$

Note that the role of σ_0 is to fix the origin in this case to zero for the scale corresponding to the inner scale of image. Values of $\tau < 0$ correspond to the sub pixel range which are not represented in the image. Hence in practice only positive values of τ are used. An equidistant step in τ is a natural step size in the scale parameter. An equidistant step size in σ would violate the required scale–invariance.

[5] The differential of the scale parameter is our yard stick in the scale direction

2.5 Properties of the scale–space image

Definition 1 (well–posedness)
According to Hadamard, a problem is well–posed if:

- *A solution exists*
- *The solution is unique*
- *The solution depends continuously on the input data*

Several problems in computer vision are ill–posed. Differentiation is one of them. Derivatives are not intrinsic properties of digital data. A data model has to be enforced to make them well–posed. The Linear Gaussian scale–space provides such a model. In the scale–space representation spatial derivative operators can be defined in the following way:

$$L_{i_1,\ldots,i_n}(x;\sigma) = \partial_{i_1,\ldots,i_n} L(x;\sigma) = \int_{a=-\infty}^{\infty} (\partial_{i_1,\ldots,i_n} G)(x-a;\sigma) I(a) da \tag{19}$$

Result 1
The Cartesian partial derivatives of order n of the rescaled image $L(x;\sigma)$ can be calculated by convolving the original image $I(x)$ with the corresponding partial derivatives of order n of the zeroth-order Gaussian kernel $G(x;\sigma)$

We follow the normal convention and write scale–space image derivatives as in Equation 19. For instance will L_{xx} mean the second order spatial derivative in the x-direction. It is implicit that differentiation is on a particular scale.

This way of taking scale–space derivatives with integral operators has a strong regularization effect. How strong can be formally stated: if I is absolute bounded by some polynomial the integral in equation 19 is guaranteed to converge. This guarantees that the Gaussian derivatives are well defined even for a discontinuous, non–differentiable original I. We have the important result:

Result 2
The scale–space image $L(x;\sigma) = G(x;\sigma) * I$ is infinitely differentiable

So introducing the scale–space representation of the image makes the notion of scale explicit and imposes a regularization that makes differentiation well–posed. These properties allow the liberty to apply the powerful tools of differential geometry, tensor calculus and invariant theory in image analysis.

2.6 Non-linearity

In the previous subsections it was established how linear differentials could be calculated in a well–posed manner from the image by explicitly imposing the notion of scale. Of course linear derivatives are not sufficient for extracting information from images. Some kind of a non–linear element must be introduced. There are two ways to study non linear phenomena in the scale frame work. Either by constructing non–linear scale spaces with a non linear diffusion scheme or by studying non–linear differentials (for instance

the gradient magnitude). We shall in the following subsections briefly touch non–linear diffusion and elaborate somewhat more on the non–linear differentials, since that is the line we will follow in this thesis.

2.6.1 Non–linear diffusion

The idea of non–linear diffusion schemes (Perona *et al.* [48]) is to forget about the homogeneity and isotropy requirements and introduce requirements like immediate localization of region boundaries, and intra–regional smoothing should occur prior to inter–regional smoothing. A now classical example is the anisotropic diffusion scheme presented by Perona *et al.* [48]:

$$I_t = \nabla(c(x,y,t)\nabla I) = c(x,y,t)\nabla^2 I + \nabla c(x,y,t)\nabla I \tag{20}$$

where $c(x,y,t)$ denotes the conduction coefficient. In the Perona–Malik scheme the conduction c is a smooth, bounded, monotonically decreasing function depending solely on the gradient magnitude. The advantages of non–linear diffusion like the Perona–Malik scheme are clear from the requirements: region boundaries stay fixed and regions on coarse "scale" consist of fine scale regions; consequently you get a nice tree structure of regions, a hierarchy. The main problem though is the lack of connection to the measurement formalism. There is no obvious and simple relation between the number of time steps used to iterate the non-linear diffusion scheme and the resulting scale of the structures. Related to this, is when to stop the diffusion. In the linear case, statements like "we will now study the structure on this or that scale" make sense, but the equivalent is not obvious in the non–linear diffusion scheme. Based on these considerations we shall not consider non–linear schemes further here. That is not to say that a front end representation based on non–linear scheme cannot be interesting and useful.

2.6.2 Non–linear differentials

Numerous authors have studied non–linear differentials for image description often in terms of differential geometric descriptors. Differential geometric descriptors and differential Equations form a natural framework for expressing local geometric properties as well as physical processes and this is one of the most profound reasons to express image properties in the language of differential geometry. Non–linearity has to come in to play when describing geometric properties since a single partial derivative is inherently tied to a particular coordinate system. A non–linear differential is a non-linear combination of linear differentials. The choice of coordinate system should be irrelevant when describing the structure itself. For instance horizontal edge detection needs a reference frame to make sense, but more often it is of much greater interest to study image properties independent of the selected reference frame. In general it is more interesting to detect edges of any orientation than to detect horizontal edges. What is needed here is a description invariant under rotation of the coordinate system. In general it is of interest to study differentials invariant under simple transformations of the coordinate system since they often capture intrinsic geometric properties of the images. In table 1 are listed some non–linear differentials invariant to translation and rotation of the coordinate system. We take the opportunity to introduce some notation.

Name	Tensor	Cartesian	Gauge
Luminance	L	L	L
Gradient squared	$L_i L_i$	$L_x^2 + L_y^2$	L_w^2
Laplacian	L_{ii}	$L_{xx} + L_{yy}$	$L_{ww} + L_{vv}$
Flowline curvature	$\frac{L_i \epsilon_{ij} L_{jk} \delta_{kl} L_l}{(L_m L_m)^{3/2}}$	$\frac{L_x L_y (L_{yy} - L_{xx}) + L_{xy}(L_x^2 - L_y^2)}{(Lx^2 + Ly^2)^{3/2}}$	$\mu = \frac{-L_{wv}}{L_w}$
Isophote curvature	$\frac{L_i \epsilon_{ij} L_{jk} \epsilon_{kl} L_l}{(L_m L_m)^{3/2}}$	$\frac{2 L_x L_y L_{xy} - L_{yy} L_x^2 - L_{xx} L_y^2}{(L_x^2 + L_y^2)^{3/2}}$	$\kappa = \frac{-L_{vv}}{L_w}$

Table 1: Non–linear invariants expressed in tensor, Cartesian and Gauge notation

Tensor notation is often a convenient tool for writing differential invariants in a compact manner. The Einstein summation convention apply. That is, if an index appears twice in the same term a summation for this index over the dimensions is silently assumed. The useful symmetric Kronecker tensor δ_{ij} is defined by $\delta_{ii} = 1$ and $\delta_{ij} = 0$ for $i \neq j$. The 2–dimensional, anti–symmetric Levi–Civita tensor ϵ_{ij} is defined by $\epsilon_{ii} = 0$, $\epsilon_{12} = -\epsilon_{21} = 1$. In d dimensions the Levi–Civita tensor is defined like this:

$$\epsilon_{i_1 \ldots i_d} = \begin{cases} 1 & \text{, if } (i_1 \ldots i_d) \text{ is even} \\ -1 & \text{, if } (i_1 \ldots i_d) \text{ is odd} \\ 0 & \text{,otherwise} \end{cases} \tag{21}$$

Actually, tensor calculus provides an even more powerful tool than just notation since differential invariants can be automatically constructed using group index formalism [19, 17]. Tensor calculus describe the transformation behaviour of objects like derivatives L_x of the image, such objects are called tensor components. A tensor is a coordinate independent object consisting of tensor components. After a coordinate transformation the new value of a tensor component can be expressed as a function of the old values. For an object to be a tensor (coordinate independent) the function may only depend on the transformation parameters.

Example 2 (The tensor L_i in 2D)
The tensor components L_x and L_y form a tensor since their values after a transformation (for instance a rotation) can be expressed as a function of the old values, and the function only depends on the transformation parameter (the angle).

Partial derivatives of order n form a n–tensor. New invariants can be constructed from a given set of tensors by means of full contractions and alternations of indices in a tensor product. A contraction reduces pair–wise the number of free indices in a tensor product. For instance the contraction between the Hessian, a 2–tensor, and the Kronecker tensor yields the Laplacian: $L_{ii} = L_{ij} \delta_{ji}$. An alternation of d number of tensor indices is defined as a full contraction of these indices onto the Levi–Civita tensor ϵ. An example in 2D is: $L_i \epsilon_{ij} L_{jk} L_k$. Functions of invariants are themselves invariant, examples of this was given in Table 1. A quick introduction to tensor calculus applied in scale–space framework is given by Florack *et al.* [19] and a more extensive treatment is given by Florack [17, 18] and references therein.

The Gauge co–ordinant system (v,w) is an orthogonal co–ordinant system with frame vectors (**v**,**w**) defined in the following way:

$$v_i = \frac{\epsilon_{ij} L_j}{\sqrt{L_k L_k}} \tag{22}$$

$$w_i = \frac{\delta_{ij} L_j}{\sqrt{L_k L_k}} \tag{23}$$

Hence the w-direction is tied to the direction of the gradient in every point with a non–vanishing gradient. Consequently the first order derivative along the w–direction equals the gradient magnitude and the first order derivative in the perpendicular v–direction equals zero. Expressing differential invariants in coordinate systems fixed by the local image structure insures, that the expression is invariant under the transformation of the coordinate systems. In the case of the (v,w) system (called Gauge coordinates) the expressions are invariant under rotation. Differentials expressed in a (v,w) system or another locally fitted system are often easier to interpret directly geometrically than in the Cartesian system. Another convenient system often called (p,q) is fixed by the local second order structure. In this (p,q) system the mixed second order derivative is zero.

For actually computing these expressions it is necessary to express them in a system (the Cartesian) where the derivatives are known. Here the tensor calculus can help. If one can find the tensor expression which reduces to the wanted expression in the locally fitted system then the problem is solved. The wanted expression in Cartesian coordinates is obtain by expressing the tensor expression in the Cartesian frame.

2.6.3 Evolution of non–linear differentials

In contrast to linear differentials non–linear differentials do not evolve according to the diffusion Equation. We shall later study the gradient squared, so let us study the evolution of the gradient squared as an example.

The evolution of the square of the gradient magnitude does not obey the principle of causality. An example of this is the creation of an extremum/saddle pair in the image L, two corresponding zeros will arise in the square of the gradient magnitude.

When the image evolves according to the diffusion Equation $L_t = L_{ii}$ the gradient squared follows this evolution scheme using tensor notation:

$$\frac{\partial (L_i L_i)}{\partial t} = 2 L_i L_{it} = 2 L_i L_{ti} = 2 L_i L_{kki} \tag{24}$$

Let us calculate the Laplacian $\triangle = \partial_{jj}$ of the gradient magnitude to see the difference:

$$\triangle (L_i L_i) = \partial_{jj} (L_i L_i) = \partial_j (2 L_i L_{ij}) = 2 L_{ij} L_{ij} + 2 L_i L_{ijj} \tag{25}$$

The two expressions are not the same. For clarification let us write Equation 24 and 25 in the 2D case and in a Cartesian (x,y) coordinate system:

$$(L_x^2 + L_y^2)_t = 2 L_x (L_{xxx} + L_{xyy}) + 2 L_y (L_{yxx} + L_{yyy}) \tag{26}$$

$$\begin{aligned}(\partial_{xx}+\partial_{yy})(L_y^2+L_x^2) &= 4L_{xy}^2+2L_{xx}^2+2L_{yy}^2+2L_x(L_{xxx}+L_{xyy})+2L_y(L_{yxx}+L_{yyy}) \\ &= 4L_{xy}^2+2L_{xx}^2+2L_{yy}^2+(L_x^2+L_y^2)_t \end{aligned} \quad (27)$$

The difference from the diffusion Equation is expressed in the three first terms on the right hand side of Equation 27. Since the diffusion Equation is not fulfilled, the evolution is not causal. The difference, the three first terms, can be expressed in the following way:

$$2L_{ij}L_{ij} = 2(L_{pp}^2+L_{qq}^2) \quad (28)$$

The square root of the irreducible invariant $L_{ij}L_{ji}$ is an absolute measure for the total deviation from flatness. So the evolution scheme for the gradient squared deviates from the diffusion equation as the structure deviates from flatness.

3 The notion of an image

In the previous section scale was explicitly introduced in the representation of image as a consequence of the finite aperture size in the detector. It was noted that each member $L(.;t)$ of the family $L(.;.)$ of Gaussian blurred versions of the original image was in fact infinitely differentiable $L(.;t) \in C^\infty$. In this section we shall limit our image model even further based on the requirement of typicality implied by the property of finite precision of our measurements themselves, the scalar values. In the introduction it was argued that the selected model should be typical for the subset of scalar functions specified by the measurements. The notion of typicality have been studied in statistics, and in topology, where the equivalent notion is genericity. Typicality is defined in the statical framework in this way: Let a set of items with varying properties be given. A property is called typical if the process of picking an item at random yields an item with the property with probability one.

3.1 Generic properties

To define the topological concept genericity we need the basic concepts of open set and dense set:

Definition 2 (Open)
A subset X is open if and only if $\forall a \in X\ \exists \epsilon > 0 : N_\epsilon(a) \subset X$, where N_ϵ is the neighbourhood.

Definition 3 (Dense)
A subset X is dense in Y if and only if $\forall b \in Y\ \forall \epsilon > 0 : N_\epsilon(b) \cap X \neq \emptyset$

Definition 4 (Genericity)
Let a space of possible items be given. A property is generic if the items with this property form an open and dense subset of the space.

Definition 5 (Morse function)
A Morse function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is infinitely differentiable and have no degenerated critical points:

$$\{x \in \mathbb{R}^d \mid \nabla f(x) = 0 \wedge detH_f(x) = 0\} = \emptyset \tag{29}$$

The critical points for Morse functions are isolated and are finite in number [25].
Being a Morse function is generic in C^∞ [22]. This property have made the Morse functions an evident basis for the notion of image and have been suggested by many authors [25, 40, 13].

That the subset of Morse functions in the space of scalar functions is open and dense implies[6] :

1. An infinitesimal perturbation can transform a given function into a Morse function
2. Picking a scalar function at random yields a Morse function with probability one.

[6]This implication holds for whatever type of item is studied. It is the property of openness and denseness being unfolded here, with the set of Morse functions as a useful and highly relevant example

3. A perturbation of a Morse function yields a Morse function with probability one.

One realizes that the three properties mentioned above capture the notion of typicality desired in the introduction very well. Property 2 arising from the denseness say that if we pick functions at random we will get Morse functions. Property 1 and 3 state that wherever we are in function space a Morse function is only an infinitesimal perturbation away. Property 3 gives a warranty for stability. One stays in the set, when the function is slightly perturbed.

Note that genericity is a relative property, namely relative to the base set. One can make any property generic by restricting the base set. Since we work in the scale–space framework the most general set is the set of infinitely differentiable functions C^∞; consequently this class form a natural base class.

The justification for using Morse functions in the analysis is that the difference between the real measurements and a Morse function is never more than infinitesimal. The result of any analysis should change continuously with the input, i.e. an infinitesimal change of input should give an infinitesimal change of the result. Hence, it is reasonable to perturb a function into the Morse class (if for some surprising reason it is not already in the class).

3.2 Structural Stable functions

What we want to deduce from the measurements is the structure of the function. When we redo the experiment under similar conditions we expect the result to be similar. We will never be able to reconstruct the exact same conditions (the moon has moved a bit etc.) but we can make the introduced perturbation very small. What is desirable here is that a small perturbation yields a result with the same structure. We wanted the model to be structurally stable. The concept of structural stability is closely connected to the concept of local equivalence:

Definition 6 (Local equivalence)
Two functions $f, g : \mathbb{R}^d \to \mathbb{R}$ are said to be locally equivalent around 0 if there exists a local diffeomorphism $e : \mathbb{R}^d \to \mathbb{R}^d$ around 0 and a constant $\gamma \in \mathbb{R}$ such that

$$g(x) = f(e(x)) + \gamma \tag{30}$$

in some neighbourhood around the point 0.

Definition 7 (Structural stability)
Let functions $f : \mathbb{R}^d \to \mathbb{R}$, $g : \mathbb{R}^d \to \mathbb{R}$ be given as well as $\epsilon > 0$. We perturbed the function f with ϵg:

$$\tilde{f} = f + \epsilon g \tag{31}$$

The function f is locally structural stable (or just locally stable) if the functions $\tilde{f}$ and f are equivalent for a sufficiently small ϵ. A function is globally structural stable (or globally stable or stable) if the function is stable in all points in the domain.

Proposition 1
Morse functions are globally structural stable. [22]

Hence it is generic to be structural stable in the set of infinite differentiable functions C^∞.

3.3 The Topology

In order to refer to the concepts of open, dense and perturbation it is necessary to define a neighbourhood relation, that is a topology. The obvious way to do this is via a distance measure. The ϵ–neighbourhood for a is simply all points within the distance ϵ. We will use "Taylor series topology" [22]. The Taylor series of $f(x)$ at $x_0 \in \mathbb{R}^n$ is given by this expression:

$$f(x) = f + f_i(x_i - x_{0_i}) + \frac{1}{2!} f_{ij}(x_i - x_{0_i})(x_j - x_{0_j}) + \ldots \tag{32}$$

where the lower index on f indicate differentiation with respect to the i variable and the lower index of x picks elements in the vector x. Hence x_{0_i} means the ith element of the constant vector x_0. All derivatives of f are evaluated at x_0. If we consider terms of up to kth order of the Taylor coefficients $f, f_i, f_{ij}, \ldots, f_{ij\ldots k}$ then the function can be represented by a point in Euclidean Space $\mathbb{R}^D$ where D is given by:

$$D = 1 + n + \frac{n(n+1)}{2!} + \ldots + \frac{(n+k-1)!}{k!(n-1)!} = \frac{(n+k)!}{n!k!} \equiv \left(\begin{array}{c} n+k \\ n \end{array} \right) \tag{33}$$

Topologies on $\mathbb{R}^D$ are easy to work with. The distance between two functions f and g at x^0 may be defined as:

$$\begin{aligned} \|f - g\|_{x_0}^p &= \mid f(x_0) - g(x_0) \mid^p + \sum_i \mid f_i - g_i \mid^p + \sum_{ij} \mid f_{ij} - g_{ij} \mid^p \\ &+ \ldots + \sum_{ij\ldots k} \mid f_{ij\ldots k} - g_{ij\ldots k} \mid^p, p \geq 1 \end{aligned} \tag{34}$$

For $k \to \infty, D \to \infty$ the different distance measures $\|f - g\|_{x_0}^p$ induce different topologies. The distance measure for $p = 1$ is in this limit the most useful [22]. This distance measure $p = 1$ induces a topology denoted by $C^k(x_0, \mathbb{R}^1)$. For $k = \infty$ the topology is denoted by $C^\infty(x_0, \mathbb{R}^1)$. Two functions, which are close in sense of the $C^k(x_0, \mathbb{R}^1)$ topology, have the property of being almost equal and so are the derivatives of functions up to and including k'th order.

The function given by the terms up to k order of the Taylor series is called the k–jet for the function and is denoted $j^k f(x)$. If two functions are equal in the $C^k(x_0, \mathbb{R}^1)$ topology then they have the same k–jet.

3.4 What we gained

To summarize Morse functions yield a reasonable model for an image, since the class captures the property of being typical (dense) in the function space and the class is stable (open) against infinitesimal perturbations. A Morse function is globally structural stable; a very important property for a robust model.

The analysis is greatly simplified when restricting the class of permitted functions to the Morse functions. Morse functions exhibit two kinds of structures namely regular points where first order structure is present and critical points where the first order structure

vanish. The critical points of Morse functions are finite in number, isolated and non–degenerate. This property is interesting since it is the topological structure of the function we want to describe. A series of theorems in Catastrophe theory (singularity theory) tells how the local form of the function can be brought to a canonical form by a smooth change of coordinates and thereby ease the classification of the topology.

The implicit function theorem guarantees that a coordinate transform exists so the local form at regular points can be brought to a form with a non-vanishing first order term in only one direction. The Morse lemma states that the qualitative structure of a function can locally be described by its second order derivatives. That is , the Morse lemma guarantees the existence of a transformation at critical points so the local canonical form is merely made of second order terms.

4 Segmentation and Superficial Structure

It is now established how the image at a particular scale is modelled namely with a Morse function. It is now possible to define objects of interest in the image, the segments. Having a segmentation corresponding to semantic entities in the scene is a basic and unsolved problem in computer vision. It has been argued [37] that the reason for this state of affairs is that having such a segmentation will be almost identical to having a solution to the image understanding problem. A segmentation corresponding to semantic entities is referred to as the scene partitioning problem. This is not what we try to solve in this work. We will suggest a solution to what could be called the image partitioning problem, that is, a partitioning of the image into coherent regions. Under simplifying conditions on the world, the underlying physical scene objects will project to coherent regions in the image. This leads to the conclusion that a segmentation should include a description of the discontinuities in the image. In this way the segmentation can at least handle the simplified problem.

A description of the superficial structure in form of discontinuities of the structure can be viewed as a way of throwing away the irrelevant information and keep the interesting changes[7] . In the same sense, segmentation can be considered as data reduction without loss of essential data. The quality of a description can be evaluated by reconstructing the original data from the available information in the description. If the reconstruction is close to the original data the description has captured the essential of the original data. An example of very good reconstructions obtained from the knowledge of the discontinuities, their intensity value and scale has been presented by Elder and Zucker [15]. These reconstruction results support descriptions based on discontinuities.

In this section we shall introduce often used elements for describing the superficial structure such as critical points, isophotes and flowlines. To give a short overview, a taxonomy for superficial descriptions as well as algorithms producing those descriptions is given. Finally the chosen description, watersheds of the gradient squared, and the concepts within it are presented. The structure of the resulting segmentation is analysed.

4.1 Superficial structure

Often used elements in superficial analysis are isophotes, flow lines and critical points. Descriptions based on these elements are presented in the following sections.

4.1.1 Critical points, isophotes and flow lines

The points in an image consist of regular points with nonvanishing gradient and critical points where the gradient vanishes. The structure at regular points can be sufficiently described locally with a first order description. In critical points the structure has to be described with second order terms. The critical points can be further classified into

[7] Koenderink [34] p.276 notes the relation between smoothness and singularities in the following way: "...In real life most of the interest and charm of any structure is in its singular "skeleton". The smooth parts tend to be relatively boring "filling in".

maxima, minima and saddles according to the local second order structure. The corresponding principal curvatures for the points are respectively both negative, both positive and one positive and one negative. The second order structure is captured in the Hessian H_L:

$$H_L = \begin{bmatrix} L_{xx} & L_{xy} \\ L_{xy} & L_{yy} \end{bmatrix} \tag{35}$$

The principal curvatures are found as the eigenvalues of the Hessian and the corresponding eigenvectors are the principal curvature directions.

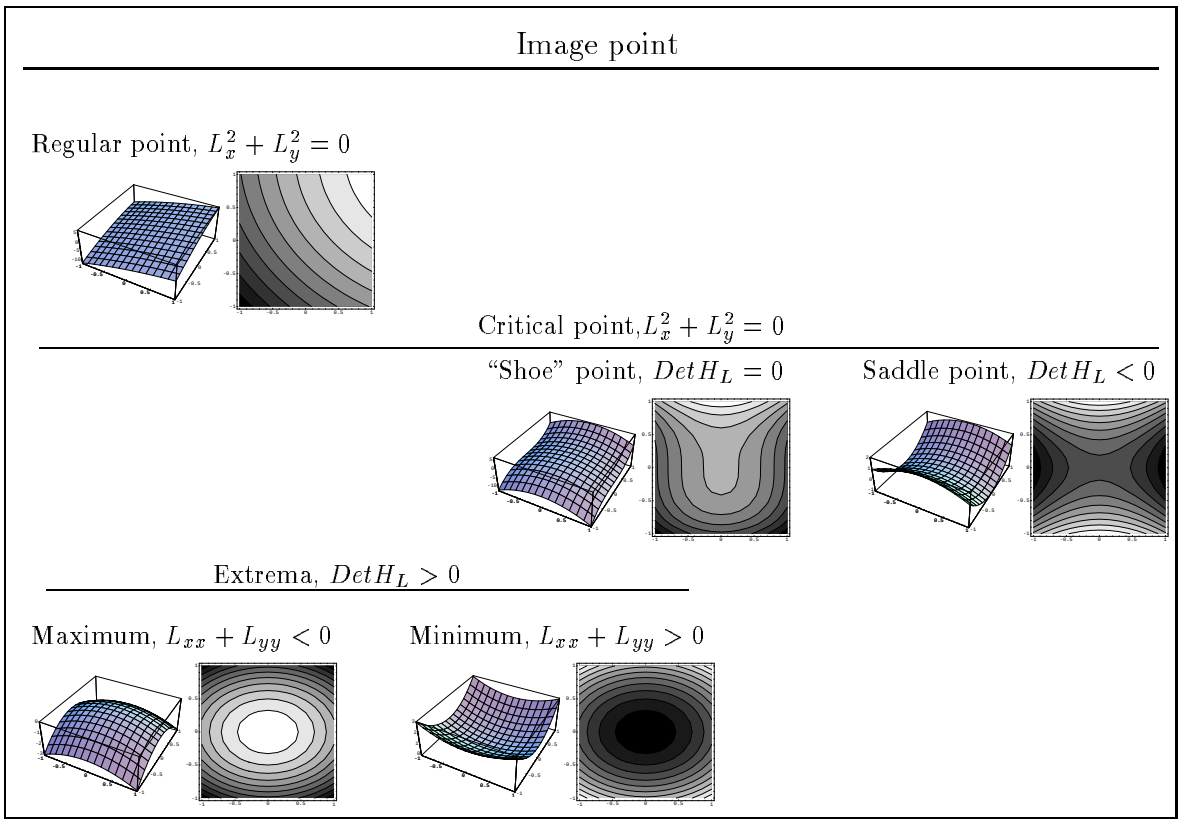


Figure 4: Classification of distinct types of local structure. For each point type is plotted the local form as a surface and as a contour plot

The two rotational invariants, the trace and the determinant, of the Hessian, give an alternative decision scheme for the same classification.

$$Det \;\; H_L \;\; = \;\; L_{xx}L_{yy} - L_{xy}^2 \tag{36}$$

$$Trace \;\; H_L \;\; = \;\; L_{xx} + L_{yy} \tag{37}$$

A negative determinant corresponds to a saddle, a positive to an extremum and zero determinant is a degenerated surface with shape as the tip of a shoe (see Figure 4. The trace (the Laplacian) splits the extrema into minima (positive trace) and maxima (negative trace). A trace equal to zero for an extremum is not possible since it implies that the Hessian equals zero. If desired the trace can divide the saddles into three categories:

ridge–like for negative trace, valley–like (called through saddle by Griffin [26]) for positive trace and the degenerated (unstable) balanced saddle for zero trace (see [25, 26] for more discussion on classification of saddles). In Figure 4 is presented an overview of the classification of the local structure based on the three rotational invariants: the gradient magnitude, the determinant of the Hessian matrix and the trace of the Hessian matrix. The gradient magnitude divides the image points in regular and critical points. The determinant subdivides the critical points in extrema, "shoe" points and saddles. The trace (the Laplacian) can further subclassify the extrema into maxima and minima. The Laplacian can also subdivide the class of saddles, this is not shown in Figure 4. The displayed saddle in Figure 4 is a valley–saddle.

The isophotes are curves with the same scalar value. The local generic appearances of these level curves are a point (corresponding to an extrema) , a curve (of regular points) and a cross (corresponding to a saddle). The global appearances are three distinct possible configurations of a closed curve: The common one is a simple (non intersecting curve); the two uncommon ones are one with one self-intersection (a cross) where one loop is within the other, and one with one self-intersection where the two loops are lying next two each other (like in the number eight: 8). A contour plot illustrates several isophotes in the same plot, so the contour plots in Figure 4 can be used to get an idea of the described appearances. The contour plot in Figure 10 (b) in section 5.1 would for slightly different isophote values have the small loop within a big loop.

The flow lines are the dual to the isophotes. They run between singularities along the gradient direction. The common (infinitely many) flow line connects a minimum and a maximum. The uncommon (finite in number) connects a saddle and an extremum. The latter uncommon flow lines are known as the separatrices and link all the critical points into a network.

4.1.2 Critical points, isophotes and flow lines in 3D

For three dimensional images the same kind of entities (critical points, isophotes and slope lines) are interesting. The same type of classification of critical points is possible. For regular points the local form can be described by a first order description, the gradient. For critical points the principal curvatures classify the image point. In the maxima all principal curvatures are negative, minima have all principal curvatures greater than zero. The hyper-saddles have one positive and two negative principal curvatures or vice versa. We denote them by the sign of the curvatures with these symbols C_{---}, C_{+++}, C_{+--} and C_{++-}.

The visualization of the local structure in 3D is a bit more tricky than in the 2D. One visualisation technique is slicing the 3D dimensional domain by a plane and colour this plane according to the value of the function. Another visualisation method is an iso-surface of the function, the term isophote will also be used. Both techniques have been used in Figure 5 to present the local structure of a minimum and the two hypersaddles.

Three orthogonal slices of the cubic domain are shown coloured according to value in the left column. The bottom row in Figure 5 presents a local minimum. The three orthogonal planes in (a) have been coloured slightly grey to indicate their location in 3D space. The function values have been added to this grey value. Reversing the function

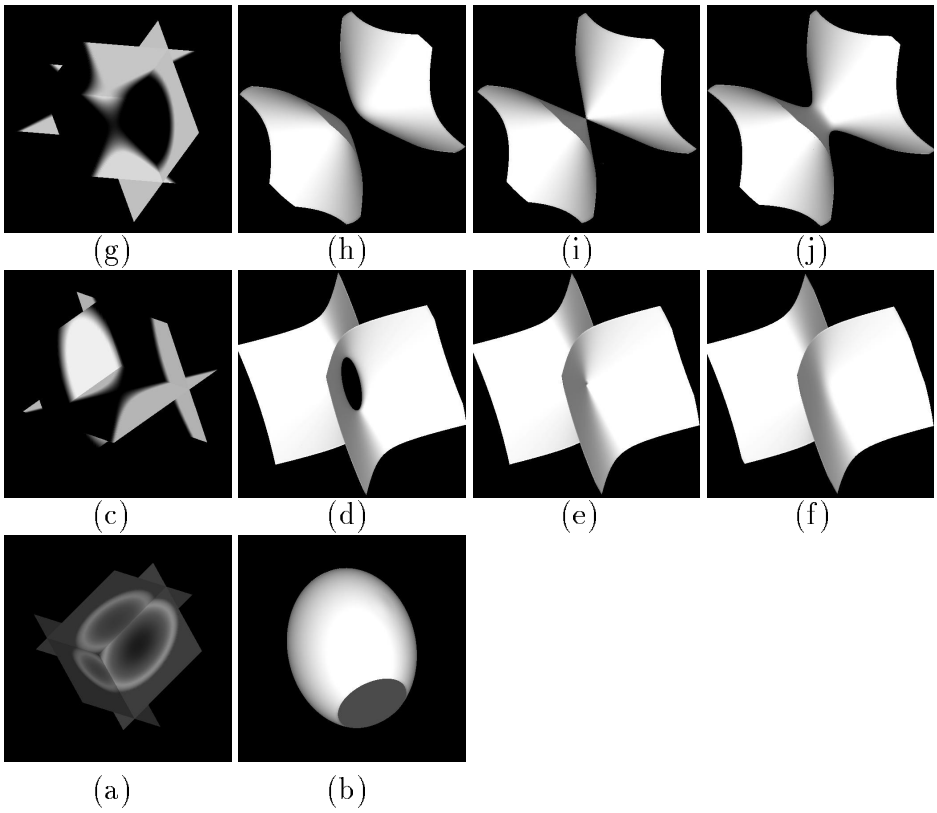


Figure 5: Local forms for a scalar function of three variables. Lower row is for a minimum (C_{+++}, $f = ax^2 + by^2 + cz^2$); second row is a hypersaddle of the type (C_{+--}, $f = ax^2 - by^2 - cz^2$); top row is a hypersaddle of type (C_{++-} ,$f = ax^2 + by^2 - cz^2$). The coefficients a, b and c are positive. Left column presents three orthogonal slices of the functions. The rest of the columns are isophotes. (b) is a positive isophote for the function with a minimum. (d) and (h) are isophotes for a negative value for the hypersaddles in (c) and (g), respectively. (e) and (i) are the zero isophotes for the hypersaddles. (f) and (j) are positive isophotes

value will of course produce a maximum and the isophote will have the same shape, an ellipsoid. One end of the ellipsoid in 5 (b) has been chopped off to give the viewer a better feeling for the shape. Hence at the bottom of the ellipsoid one sees the inside of closed isophote surface. In 3D the isophotes are in general simple closed surfaces. If one looks at the isophotes for a scalar value which equals the value in a critical point then one will see the uncommon structures. For an extremum value a point isophote will exist. For a hyper-saddle value there will be two blobs touching in exactly one point as in Figure 5 (e) and (i). In (h),(i),(j) is displayed a sequence of three isophotes for a negative, zero and positive value respectively. Only for the critical value do the two isophotes touch in one point. For negative value they are separated and for positive they are melted into one surface. A similar result can be studied in the second row for the other hypersaddle (C_{+--}). The ordering of the separation and reunion of the two isophotes is reversed since one of the principal curvature has changed sign.

The dual of the isophotes are the flow lines which are curves since they are perpendicular to surfaces. The common flowlines run between a maximum and a minimum, this kind of flowlines fill volumes. Between maxima and hyper-saddles run flowlines of type C_{++-} and similarly between minima and hyper-saddles of type C_{+--}. These two kinds of flowlines span bounded surfaces. The boundaries of these surfaces are the rare (finite in number) flowlines between critical points with only one sign difference in the principal curvature, like from maxima C_{---} to hyper-saddles C_{+--} and C_{+--} to C_{++-} and finally from C_{++-} to C_{+++}

4.2 Superficial descriptions

In this section different approaches to describe the superficial structure of an image are presented. A list of some descriptions of superficial structure will be presented and examples of the algorithms producing the presented descriptions are given. The listing is made to give the reader a reference frame for the description selected in this work. The focus will be on descriptions including discontinuities. An alternative classification of segmentation methods are in region–based, edge–based and multi resolution methods. This alternative is often used but do not focus on the primary target in this section, the description. Griffin [24] has taken a similar approach. In Figure 6 a schematic overview of common descriptions is given. The most constrained descriptions are placed at the top and the descriptions become less and less constrained until the general edge map is reached. The lines give a partial ordering of the descriptions the most general are located at the bottom of the Figure. A line between two boxes indicates that descriptions in the top box are a subset of the descriptions in the lower box. The ordering is transitive. This implies for instance that the general edge-map description can describe any descriptions above it.

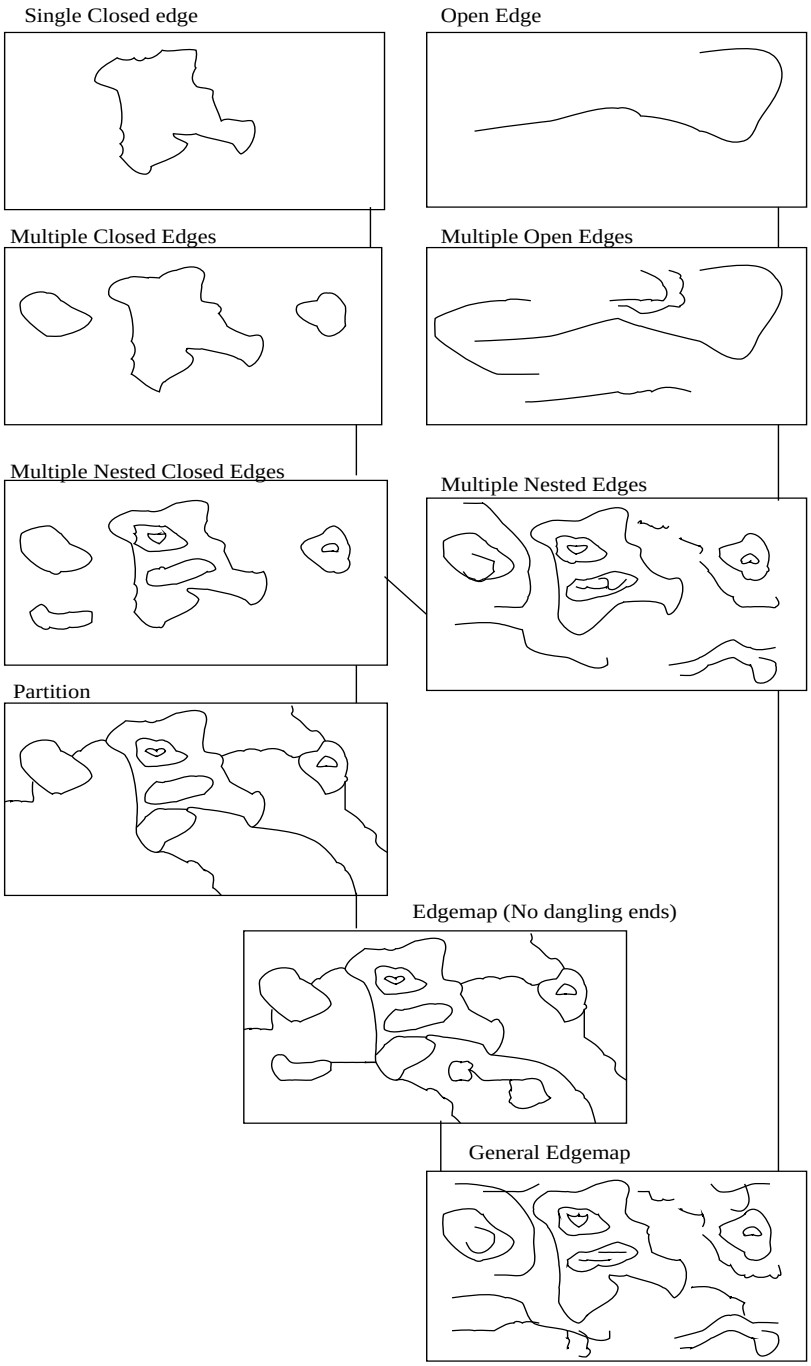


Figure 6: Taxonomy of descriptions for discontinuities in superficial structures. Inspired by Lewis Griffin

4.3 Algorithms producing the descriptions

Examples are given on algorithms producing descriptions from the taxonomy in Figure 6.

4.3.1 The simplest descriptions

The four top categories, single(multiple) closed/open edge(s) in Figure 6 lack the ability to represent holes and are therefore fairly constrained in their power of expression.

Snake algorithms can produce these four kind of descriptions. A single closed edge was the output from the original snake algorithm introduced by Kass *et al.* [31]. Later development of the idea has introduced algorithms for finding a single and multiple open edge(s) [3]. The idea of snake algorithms is to fabricate an energy function (a potential). By differentiating the energy function with respect to position one can derive Equations of motion. Placing the snake in this motion field will drag it to the equilibrium of the potential. The art is to design the energy function so the snake ends up where the designer want it. In a simple form the equilibrium is a place with high gradient. The snake will be twisting seeking to follow the zero–crossing for L_{ww} since this is an equilibrium. If the scale is near pixel width then the snake will twist more than desirable since it will twist around every little noise bump. Since scale is usually not considered in snake algorithms, the desired smoothing is accomplished by adding bending energy terms to the energy function. A lot of energy terms have been invented to avoid or introduce different behaviour; the weighting of the coefficients of the different terms depends on the task and is a non–trivial problem. Snake algorithms formulate a functional and minimize it using gradient descent, therefore the property of getting caught in local minima always exists.

Seed growing algorithms with hole filling as a post processing step produce a single closed edge as output. A seed is interactively marked and pixels are iteratively grouped from this seed according to some grouping criteria [23].

Straight line[8] detection come in the category of multiple open edges and a good example is "The straight edge extraction" by Burns *et al.* [5]. The algorithm computes the gradient in each point, groups points of similar gradient into "edge support regions", fits a rectangle to each group and finally intersects each rectangle with the mean luminance for the rectangle to produce a straight line.

The detection of blobs and anti-blobs as defined by Lindeberg [39, 40] gives a description of multiple closed edges. To each maxima corresponds a blob. The extend of the blob is given by a saddle point. Regions are built by the union of isophotes. A blob is defined as the largest region containing a maximum and no other critical points. Hence, the isophotes with intersections are a superset of the boundaries. If one of the loops in such self intersecting isophotes contains exactly one maximum then the loop is the boundary of a blob. Anti–blobs are defined in a similar manner based on minima and saddles.

4.3.2 Descriptions including nested edges

The next level of descriptions includes nested closed edges and the more general nested edges. These two types of descriptions plus the previous four can not describe edge–

[8] Straight lines occur in tremendous amount in man–made scenes and can therefore be very useful as a scene description, for instance for indoor robot navigation

junctions. Hence, the edges must be disjoint, open or closed curves.

Examples of algorithms producing nested closed edges are thresholding and zero–crossings of the Laplacian introduced by Haralick [27]. The former is simply a division of the pixels into two groups according to their intensity value above or below the threshold. The latter is defined by:

$$\triangle L = L_{xx} + L_{yy} = L_{ww} + L_{vv} = L_{ww} + \kappa L_w = 0 \tag{38}$$

Expressed in gauge coordinates the difference between a commonly used edge detection, zero–crossings in L_{ww}, and zero–crossings in the Laplacian is easily seen from Equation 38. For straight edges $\kappa = 0$ they are identical. Zero–crossings for the Laplacian tends to round off when the isophotes are curved. Comparisons like this are greatly simplified when the detector can be expressed in terms of differentials. Zero–crossings for L_{ww} have been computed in several ways. Haralick [27] did it by local surface fitting follow by explicit calculating the gradient. Another way is to express L_{ww} in terms of Gaussian linear filters $L_x, L_y, L_{xx}, L_{yy}, L_{xy}$.

Edge map can be further pruned by thresholding the gradient magnitude. All edge points with gradient magnitude larger than the threshold survive. An alternative is thresholding with hysterese. For edges found by zero–crossings of L_{ww} an edge point would be defined like this:

$$L_{ww} = 0 \wedge L_w > T_1 \wedge \text{somewhere on the } L_{ww} \text{ zero–crossing is } L_w > T_2 \tag{39}$$

where T_1 and T_2 are locally or globally determined thresholds. Canny [8] introduce the idea of thresholding with hysterese. Both methods with and without hysterese would produce multiple nested edges.

Multiple nested edges are also the output from the Canny edge detector [8]. Canny derived an optimal filter which maximised the following criteria:

- Signal to Noise ratio
- Localization of an ideal step edge
- Uniqueness of the edge marking

The optimal filter was approximated with the first derivative of a Gaussian based on similarity in shape between the optimal filter and the first derivative of the Gaussian. Deriche [14] took Canny's criteria as starting point, but assumed a filter with infinite support. The resulting filter has a better performance than the filter derived by Canny, and an implementation using recursive filtering is possible which is not the case for the first derivative of the Gaussian. Both the Canny detector and the Deriche detector produce multiple nested edges.

4.3.3 Descriptions with edge junctions

The remaining four descriptions can describe edge junctions. It is worth noting, that detectors based on zero–crossings of a differential operator can not detect T–junctions (zero–crossings are closed curves or end at image boundaries [53]), but separatrices of a differential operator can [47, 26]!

Partitions

A partition divides the domain into distinct regions. A path from one side of a boundary to the other can not be constructed without crossing at least one boundary, namely the one under consideration, or several other boundaries. Partitions are a very good way of describing boundaries of objects. The boundaries of an object must be a closed curve for the term object to make sense. Closed boundaries are so desirable, that output including dangling edge ends often are post processed in order to close the boundaries [2, 45]. The idea is that the dangling ends arise from an algorithm unable to detect the real closed curves. Consequently the gaps are closed in a post–processing step. One way is to connect nearest loose ends. Another possibility of closing gaps is to use the distance map of the edge map as energy function for a snake (active contour).

There are several approaches to a more direct construction of partitions, for instance multiple thresholding [23], minimum description length (MDL) [37] and watersheds of the gradient magnitude [42, 54, 26, 25, 44, 47]

Multiple thresholding divides the intensity range in a number of intervals [23]. Spatial neighbouring pixels in the image are grouped if they belong to the same interval. The threshold values determine the cut between intervals. The thresholds could for instance be determined by a cluster analysis on the histogram of the image. The idea is to make a cluster for each important mode in the histogram, a non trivial task.

LeClerc [37] has formulated the partitioning problem in the MDL frame work. The idea is to choose an optimal description language for the chosen model for the observation. The cost in bits to encode the model can be calculated. The residuals between the description and the observation can also be encoded with a certain amount of bits. The best description is according to the MDL principle the one for which the sum of the two encodings is minimal. The important steps in working with MDL are to define the descriptive language for the model and the descriptive language for the residuals. On top of this some kind of optimization scheme has to be introduced to find the right description among all possible within the language. LeClerc [37] models the image as made of piecewise smooth regions. A region can be encoded by a chain code for the boundaries plus a start position for the chain plus a constant term to describe the smooth surface. The residuals in each point are encoded separately. The resulting functional to be minimised exhibits many local minima. In order to solve the optimization problem the functional is embedded in a family of smooth versions of the original functionals. In the most smooth version the global minimum can be detected with standard gradient descent method. This solution can be iteratively improved as the minimum is tracked during deblurring[9].

The watersheds of a function form per definition closed curves and consequently a partitioning of the domain. The watersheds of the gradient magnitude image have proven to give good intuitive segmentations [42, 54, 47, 44, 25]. The idea is to flood the gradient magnitude landscape, the mountain ranges in this landscape are selected as boundaries. This thesis will extend the classical watershed segmentation to 3D images and we will embed the segmentation scheme in a multi–scale framework. In section 5 the chosen

[9]Smoothing the functional, finding a solution and then iteratively improve the solution by sharping is also known as graduated non-convexity

segmentation scheme is further discussed.

4.3.4 The most general descriptions

The separatrices of the gradient magnitude can only be described by a general edge map, since both junctions and edge terminations can occur. The algorithm by Geman [21] also produces a general edge map description as output using a "stochastic model for boundary detection". Mumford and shah [46] proposed a functional minimization procedure for computing a general edge map. Mumford and Shah requires that the total length of the edges is low, the simplified image should be similar to the original and the simplified image should be relative constant between edges. The weighting between the requirements are chosen in ad hoc manner. The optimization uses a graduated non–convexity method to handle the problem of local minima.

Griffin *et al.* [26] have produced an edge map without edge terminations by a technique called "gradient magnitude ridge with closing". The method prunes the set of separatrices of the gradient magnitude in such a way that the resulting edge map is without terminations. The output from this method can be described by an edge map without dangling ends.

4.4 3D Segmentation methods

Segmentation in 3D is very similar to the classical problem of segmentation in 2D. All the presented 2D algorithms can be extended to 3D and many actually have been. The equivalent 3D version of a snake is often referred to as a balloon [9, 10] or adaptive models. Deriche Edge detector has been extended to 3D by Monga *et al.*[45], using recursive filtering[10]. Watersheds in 3D have been used by Vincent *et al* [54] in medical imaging. Watersheds in 3D will be defined in section 5.4 and have been implemented as a 3D segmentation tool. Results from this implementation will be presented in section 10.2.3.

In the 3D descriptions are edges simply replaced with surfaces. So the general idea is to describe discontinuities with $(n-1)$–manifolds embedded in $\mathbb{R}^n$ which is taken to be equivalent with the image domain.

[10]Recursive implementation of filters yields very fast algorithms especially for filters with large support. The reason is that the filter response is computed from a very small support plus previous calculated filter responses.

5 The Chosen Segmentation

In the previous sections nine different ways of describing superficial structure were presented and sixteen algorithms producing these descriptions. In this section we will motivate the choice of watersheds of the gradient squared as the basis for the non committed segmentation tool.

The definition of segments must be powerful enough to express segmentations which are useful in a range of situations. The definition must be formulated mathematical clean to enable analysis of which kind of superficial structure gives rise to a segment and to enable analysis of the structure across scale as well. A common approach to achieve a clean definition is to state it in terms of non–linear differentials. It is well if the resulting segmentation gives good intuitive results, in the sense that the segmentation corresponds well with a segmentation made by a human observer. Since the scale, position and shape characteristics of the object(s) of interest are unknown, it is reasonable to loosen the latter requirement to the following weaker requirement: The resulting segmentation should explicitly contain a segmentation which corresponds well with man made segmentation. In other words, a description (the segmentation) is desired which captures the essential structures over all reasonable scales (between inner and outer scale of the image) and for all reasonable positions (within the image boundaries). This description should explicitly contain a segmentation close to the one found by a human. A simple thing like T–junctions must be expressible for the description to be powerful enough. A definition defining the segmentation to be a partition agrees well the intuitive idea of what a segmentation is. For instance in medical imaging a segmentation usually means finding segments corresponding to anatomical objects.

We realized that the duality between catchment basins and minima in the gradient squared makes possible a segment definition which suits the above considerations well. Based on this observation we define our segments:

Definition 8 (Segment)
A segment is the catchment basin for a minimum of a dissimilarity measure.

In this thesis the dissimilarity measure is taken to be the squared of the gradient magnitude. This dissimilarity measure is used because it has been reported to give good intuitive segmentations. Many authors [42, 54, 47, 44] have reported that the images were over–segmented by the watershed method. Different approaches have been used for merging the regions into large regions more appropriate for the task. Maes *et al.* [42] used a minimum description length scheme where the basic segments were merged if such an operation would result in a reduction of total code length. In morphological segmentation a common approach for reducing the over–segmentation is to mark points in regions of interest [47]. The image landscape is then flooded from these points. Watersheds separating a region without a marker and one with a marker are ignored. This results in a segmentation consisting of a subset of the watersheds, the ones that separates water from marked regions.

This thesis takes a different approach: the watershed segmentation is taken to give the correct result which describes the superficial structure of the image at the selected scale. The real problem with over–segmentation is that the image structure is extracted

at a total different scale than the scale range for the regions of interest. Consequently as in any[33] image analysis situation the scale must be taken explicitly into account. The following subsections analyse the definition of segments in the scale–space framework and the type of segmentation it induces.

5.1 Watershed segmentation

The notions of watersheds and catchment basins [11] come from the analysis of landscapes. The image is viewed as a landscape where the intensity is the elevation and the terminology is transferred from the topographical landscape to the image. Catchment basins are areas draining to the same minimum. The watersheds are the lines which separate different catchment basins. A well known watershed is the great dividing range in the US diving the Pacific and Atlantic. An 1D example is shown in Figure 7.

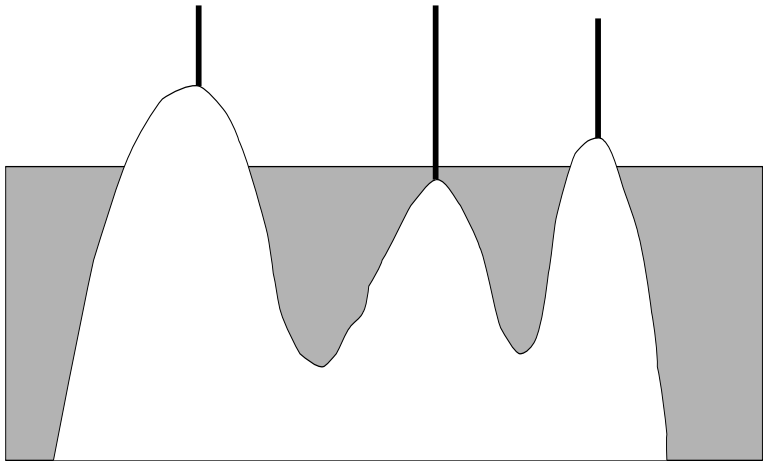


Figure 7: Flooding of landscape. Black borders correspond to watersheds

In the one dimensional case the watersheds are simply all the maxima of the image and the catchment basins are the regions in between, see Figure 7. It gets a bit more complicated in higher dimensions since the water can run separate ways and still end up in the same lake.

In Figure 8(a) is presented a checker board image with noise at scale $\sigma = 1.43$ pixels. The surface plot is shown in (b) and the watersheds of this function is presented in (c). The watersheds separate each of the minima in the function and run from maximum to saddle to maximum, etc. The border of the image is treated as an infinitely high. This example is intended to demonstrate what watersheds are, it does not illustrate the segmentation technique itself.

The watersheds of the image itself do not in general provide a good segmentation. This is clearly seen in Figure 8. What gives a good segmentation are watersheds of the *gradient magnitude*, a non-linear differential of the image. The segmentation technique is one of the key tools within the framework of morphological segmentation [44]. The idea of this

[11] Catchment basins are also known as regions of influence

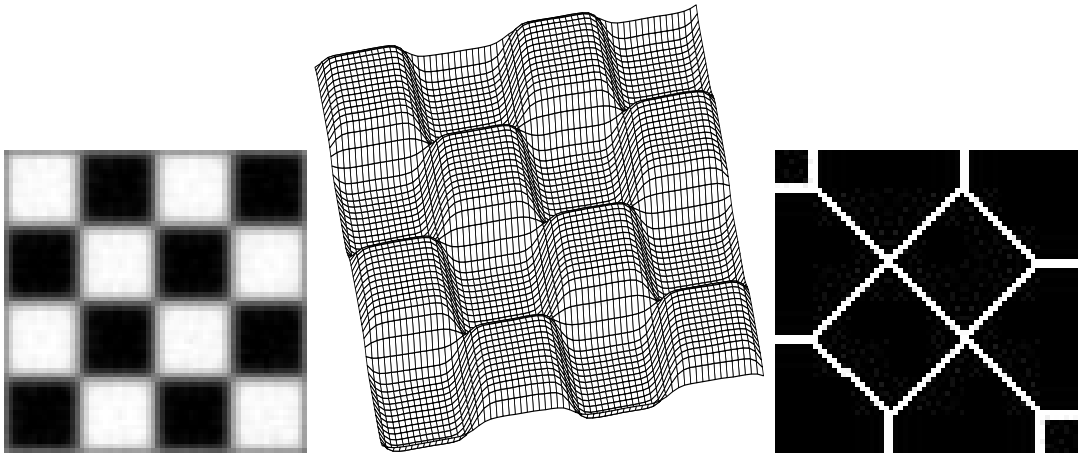


Figure 8: (a) Checker board with noise at scale 1.43 pixel. In (b) is a mesh plot, the landscape, of (a). Watersheds of the image is presented in (c). This example is intended to demonstrate what watersheds are, it does not illustrate the segmentation technique itself.

segmentation is that a large change in image value corresponds to an important change in the scene and therefore the measure of change, the gradient magnitude, is a good one. The "mountain ranges" of the gradient magnitude are detected by watersheds and correspond to the object boundaries. An example of this method on the checker board in Figure 8 is given in Figure 9. Figure 9(a) is the checker board, (b) is the gradient magnitude squared presented as an image and in (c) as surface plot. In (d) are the watersheds of the gradient squared which provide a good intuitive segmentation of the checker board image (a). The saddle regions in the image L give rise to regions but the extend of these regions are for the presented scale below the resolution of the segmentation. The watersheds of the image (Figure 8) and of the gradient squared (Figure 9) are presented for the coarse scale structures.

Hitherto only examples where the watersheds equal the separatrices have been presented. An example where the two sets are unequal is presented in Figure 10. Figure 10 (a) and (b) are the surface plot and contour plot of a function, respectively. Viewed as a landscape the function resembles a big crater with a small peak on the inside of the crater rim. There is exactly one watershed following the circular ridge. There is *not* a watershed from the peak to the ridge structure since the water falling one either side of the peak end up in the same minimum. Hence, even though the water is parted by the line, it is not a watershed since two water-drops falling one on each side end up at the same minimum. This is a clear example of why watersheds can not be determined by local structure. The global topology is needed.

The use of the global structure make it possible to determine wether or not an edge separate two distinct regions. The use of global topology also helps when detecting edge junctions. The watersheds are basically a limiting set of the slopelines: where two limiting sets meet there is a junction. In edge detection with zero–crossings, the edges so to

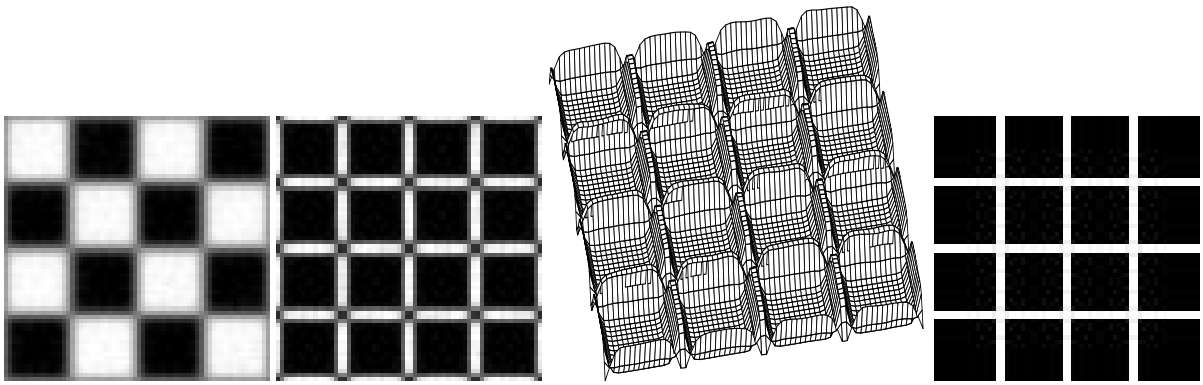


Figure 9: (a) Checker board with noise. The gradient squared of the checker board (b) and the surface plot of the gradient squared (c). The watersheds of the gradient squared is shown in (d), this subset of the separatrices yields in general a good segmentation.

speak fight for the right to the junction and the strongest wins. The need for the global information is on the other hand a disadvantage for the method simply because the feature is not locally detectable. This means that critical points are needed as an anchor points when constructing the watersheds.

5.2 Watersheds and catchment basins

The notions of watersheds and catchment basins arise when a function is viewed as a topographic relief with the height set to the scalar value. The watersheds are the boundaries between areas that drain to different local minima. The area draining to one minimum is referred to as the catchment basin.

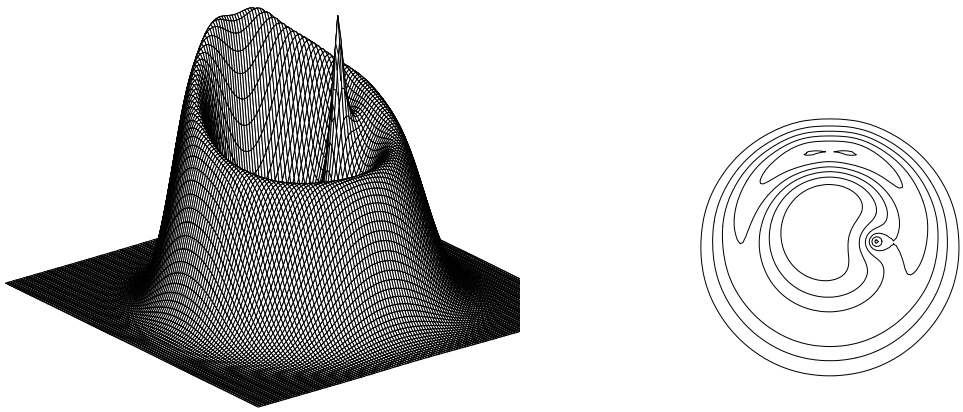
In the following watersheds and catchment basins will be defined on the basis of a partial ordering of the critical points. A similar approach is taken by Najman and Schmitt [47]. An interesting alternative approach is to define a metric such that a point belongs to a catchment basins if the distance to the corresponding minimum is the shortest among all distances to minima. An approach along this line is presented by Meyer[44], in the article a distance measure, the topographical distance, is defined similar to this: if there exits a line of steepest slope between the two points the distance is the height difference between the two points, if such a line does not exist the distance is infinite.

Now for the definitions using partial ordering. In the following we study functions f sufficiently regular to be differentiated (lie in C^2).

Definition 9 (Slope line)
A path $\gamma : \mathbb{R} \to \mathbb{R}^d$ *is called a slope line if*

$$\begin{aligned} \forall s \in \mathbb{R}, \gamma_s(s) = \pm\nabla f(\gamma(s)) &\neq 0 \\ \lim_{s\to-\infty} \gamma_s(s) = \lim_{s\to+\infty} \gamma_s(s) &= 0 \end{aligned}$$

A slope line is descending if $\forall s \in \mathbb{R}, \gamma_s(s) = -\nabla f(\gamma(s))$



(a) (b)

Figure 10: (a)A two dimensional landscape. Only the ring structure produces a watershed, not the peak. (b) Contours of the plot in (a).

In general these lines are the well known steepest descent lines.

Definition 10 (Ordering of critical points)
Let a and b be two critical points of f. We define a partial ordering of the critical points of f by saying that a is above b if there exits a descending slope line linking a to b. If there is a sequence of critical points a_i such that a_i is above a_{i+1} we say that a_0 is above a_n for $n > 0$. A partial ordering of the critical points is established by this.

Definition 11 (Catchment basins)
A point belongs to a catchment basin for a minimum M if one of the following three conditions are fulfilled.

1. *The point is on a slope line which is connected to M.*
2. *The point is a critical point which is above exactly one minimum, namely M.*
3. *The point is on a slope line which is connected to one of the critical points fulfilling condition 2.*

Definition 12 (Watersheds)
The watersheds form the boundaries between catchment basins. Let $P(f)$ be the subset of the critical points which is above more than one local minimum of f. Let $S(f)$ be all slope lines which connect two points from $P(f)$. The watersheds for f are the union of the points in $P(f)$ and the points on a slope line in $S(f)$.

The critical points outside $P(f)$ consist of two groups: the local minima of f and the saddles and maxima which are only above one minimum. The separatrices of a function f

are the slope lines linking the maxima and saddles. From definition 12 it is seen that the watersheds form a subset of the separatrices. An important property of the watersheds is the representation of global structure with the implication that watersheds are not in general locally detectable. Formally stated:

Proposition 2
Adapted from Najman and Schmitt [47]. Let a be a point of the domain of f such that $\nabla f(a) \neq 0$. *Let* N_a *be a neighbourhood of a which does not contain any critical point. Let* γ *be a path containing a and parallel to the gradient of f on* N_a. *Then there exists a function* f_0, *equal to f on* N_a *such that* γ *is in the watershed of* f_0.

Hence the watersheds are not locally detectable. Watersheds are only detectable if the global topology is probed. This latter property is one of the reasons that watersheds of the gradient magnitude handle junctions fairly gracefully. Another important property especially interesting for segmentation is the fact that:

Proposition 3
Watersheds form closed curves for Morse functions in 2D. In 3D watersheds form closed surfaces.

Finding the watersheds of a function gives a full partitioning of the domain; there is no need for closing or connecting edges to get a partition. The closing of edges in a consistent way is one of the problems which has to be solved when segmenting based on edge detection [2].

5.3 Watersheds for the gradient magnitude

The idea of segmenting by watersheds of the gradient magnitude image is well known from the field of morphological segmentation [24]. There is a segment for each minimum in the gradient magnitude. The gradient magnitude is not differentiable in critical points of the image. To ease the analysis we will studied the square of the gradient magnitude which of course has the same extrema and saddles the same place as the gradient magnitude. All analyses and all experiments are carried out on the gradient magnitude squared which has the same steepest descent lines and thereby the the same watersheds as the gradient magnitude itself.

The singularities (and hence the minima) of the gradient squared come in two flavours formally stated in 2D in the Equations 40 and 41:

$$L_x = 0 \wedge L_y = 0 \tag{40}$$

$$L_x^2 + L_y^2 \neq 0 \wedge L_{ww} = 0 \wedge L_{wv} = 0. \tag{41}$$

Where the gauge coordinate system (w,v) is used, w is in the direction of the gradient, v is perpendicular to w. The latter conditions imply that the slope line curvature $\mu = \frac{-L_{wv}}{L_w}$ and the relative variation of the scalar field L_w along slope lines $\delta = \frac{-L_{ww}}{L_w}$ must both equal zero. In other words either there is a singularity, or the slope lines are locally straight and the isophotes are locally equally spaced along the slope lines.

5.4 Watersheds and catchment basins in 3D

The families of slope lines in 3D that run into a single extremum are separated by the three-dimensional analogous to watersheds which are singular surfaces. The surfaces divide space into distinct cells labelled by the extremum. Koenderink [34, page 516] touch these structures briefly. Formally we can define watersheds for scalar function over a three dimensional domain exactly as in the 2D case. The definitions 9, 10, 11 and 12 are immediately applicable.

Let us have a closer look at $P(f)$ the critical points above more than one minimum and at $S(f)$ the slope lines connecting the critical points in $P(f)$. The points in $P(f)$ are the maxima and hyper-saddles connected to two distinct minima. These points have in one of the three principal curvature directions slopelines descending to the distinct minima; the two slopelines run in opposite directions along the principal curvature direction. These points make the anchor points for a watershed surface defined by these points and the slope lines $S(f)$ connecting them. Let us now restrict ourselves to this surface manifold and consider the 3D scalar function restricted to this manifold. The critical points of this function are the points in $P(f)$. The maxima in 3D are maxima here as well. The two kinds of hyper-saddles are respectively saddles (C_{+--}) and minima (C_{++-}) on the manifold. The surface is spanned by slopelines connecting the critical points in $P(f)$. The common slopelines run between maxima C_{---} and minima on the manifold (hypersaddles of the type C_{++-}). These common slopelines may be regarded as the faces of cells. The faces are separated by the uncommon slopelines from $S(f)$ running from maxima C_{---} to hypersaddles C_{--+} and from hypersaddles C_{-++} to minima C_{+++}. If the scale function on the two dimensional manifold is considered alone then the watersheds for this 2D function will be the latter described uncommon slopelines in $S(f)$.

5.5 Singularities for the gradient squared in 3D

The singularities of the gradient squared and with them the basis of segments occur in the critical points of the image (see Equation 42) and in the points where the second order structure of the image vanishes in one direction (see Equation 43), just as in 2D:

$$L_x^2 + L_y^2 + L_z^2 = 0 \tag{42}$$

$$L_x^2 + L_y^2 + L_z^2 \neq 0 \wedge L_{ww} = 0 \wedge L_{wv} = 0 \wedge L_{wu} = 0 \tag{43}$$

where (w, v, u) forms are Cartesian co-ordinant system with w in the gradient direction and (u, v) in the perpendicular plane to w (the tangent plane to the isophote).

5.6 Summary

A catchment basin is the region which drains to exactly one minimum. The watersheds form the boundaries between catchment basins. Watersheds are a subset of the separatrices for the considered function and form closed curves in 2D (surfaces in 3D). Catchment basins can be formally defined in arbitrary finite dimension by construction a partial

ordering of the critical points. The partial ordering is based on slopelines (lines of steepest descent). Watersheds can not be locally detected, and consequently capture global topology.

In this and previous sections a segmentation for a single scale was defined and analysed. The relation between segments at different scales can be analysed and a linking of the segments can be based on this. Segmentation with watersheds introduces a duality between the segment, the catchment basin, and the singularities of the dissimilarity measure. This duality allows the analysis of the inter scale relation between the segments to be based on catastrophe theory.

6 Catastrophe theory

This section present the basic concepts and results from catastrophe theory needed for the following analysis. The literature used mainly for the study of catastrophes and singularities has been Gilmore, Saunders, Poston and Stewart, Lindeberg, Johansen, Damon [22, 50, 49, 39, 28, 12, 13]. Other references, commenting on Morse or catastrophe theory or using it in their analysis, are Koenderink, Florack, Griffin, Blom, Lifshitz and Pizer [33, 34, 18, 25, 38].

Families of functions with members with degenerate critical points are generic in the set of families of scalar functions. Put differently, when the parameter for a family is changed smoothly, it is expected to encounter members with degenerate critical points. The point is that the qualitative structure (the singularities) of the function can change abruptly when the control parameters are changed smoothly. Catastrophe theory studies this change. It is obvious to view the scale–space representation of an image as a family of blurred functions with the scale parameter as a control parameter, consequently the study of catastrophe theory is a tool for studying the change of structure across scale. The family can be studied as a family.

Section 3, the notion of an image, defined and discussed the concepts of genericity, local equivalence of functions and structural stable functions. The concept of genericity were used to establish when a property is typical for a scalar function. When families of functions are studied, it is useful to establish similar concepts for families as was done for single functions.

First local equivalence is defined for families of functions.

Definition 13 (Local equivalence for families of functions)
Two r–parameter families of functions $f(x;s), g(x;s) : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}$ *are equivalent if there exist*

- a *diffeomorphism* [12] $e : \mathbb{R}^r \to \mathbb{R}^r$
- a *smooth map* $y : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}^n$ *such that for each* $s \in \mathbb{R}^r$ *the map* $y_s : \mathbb{R}^n \to \mathbb{R}^n$ *defined by* $y_s(x) = y(x;s)$ *is a diffeomorphism.*
- a *smooth map* $\gamma : \mathbb{R}^r \to \mathbb{R}$

defined in a neighbourhood around the point 0 such that

$$g(x;s) = f(y_s(x); e(s)) + \gamma(s) \tag{44}$$

for all $(x;s) \in \mathbb{R}^n \times \mathbb{R}^r$ *in that neighbourhood.*

Hence if one family can be smoothly transformed into another, they are said to be equivalent. This notion of equivalence implies that the singularity set of the two families will be equivalent sets up to a diffeomorphic transformation. The singularity set is the set of points where the first differential is zero:

[12] A diffeomorphism is a differentiable, isomorphic mapping. An isomorphism is a one to one mapping.

Definition 14 (Singularity set)

$$S_f = \{(x;u) \in \mathbb{R}^n \times \mathbb{R}^r : (\nabla f)(x;u) = 0\} \tag{45}$$

A diffeomorphic transformation do not change the topology of the singularity set. This is basically why it is interesting to study diffeomorphic transformations. The topology for the transformed family is the same as for the original family. Hence, a horrible formed family can be transformed by a diffeomorphism into a family on a simple form. The topology of the simple form is easy to analyse and the obtained results are directly applicable to the original function. Even though diffeomorphisms preserve the topology they can make tremendous changes to the landscape, the limitations are that the transformation has to be smooth and is not allowed to introduce singularities. Another explanation of diffeomorphic transformation is this: imagine a network where the nodes are the singularities for a function. The nodes are connected by the separatrices. A diffeomorphism can stretch this network wildly by dragging the nodes around, but nodes can not be dragged across the connections and nodes can not be removed or introduced. This is so because the set of allowed transformations is limited to the diffeomorphic transformations. The notion of equivalence is the basic way of comparing families of functions. Structural stability can be expressed in terms of equivalence.

Definition 15 (Structural stability for families)
Let $f(x;a), g(x;a) : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}$ be families of functions and $\epsilon \in \mathbb{R}$. The family f is locally stable if every perturbation $\tilde{f} = f + \epsilon g$ for a sufficiently small ϵ is equivalent to f. The family is globally stable if it is locally stable in all points $(x;a)$ in the product space $\mathbb{R}^n \times \mathbb{R}^r$ between the spatial space and the parameter space.

Hence a family is structural stable if its local form over the product space is essential the same after a small perturbation. Only a local neighbourhood is considered in the given definitions. It is of course always possible to perturb a function or a family of functions to such a degree that the local form changes.

A fundamental theorem in catastrophe theory is due to Thom. The theorem states that the local qualitative behaviour of families indexed by a small number of parameters can be classified into a small *finite* set of canonical families. A summarising result is given by Poston and Stewart [49] up til and including five parameters. Four parameters are more than enough for the application in this thesis, and the original theorem by Thom was only for four parameters. Hence here is given a summarised version of the summarised result:

Theorem 1 (Thom's Classification Theorem)
Typically an r–parameter family $\mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}$ of smooth functions $\mathbb{R}^n \to \mathbb{R}$, for any n and $r \leq 4$ is structurally stable and is in every point locally equivalent to one of the following forms:

- *Regular point:* x_1
- *Non–degenerate critical point (Morse)*
  $x_1^2 + \ldots + x_i^2 - x_{i+1}^2 \ldots - x_n^2, (0 \leq i \leq n)$

- *Degenerate critical point (Catastrophe)*

Name	r	*Germ*	$+$ *unfolding*	$+$ *Morse*
fold (A_2)	1	x_1^3	$+ u_1x_1$	$+ M$
cusp $(A^{\pm}{}_3)$	2	$\pm x_1^4$	$+ u_2x_1^2 + u_1x_1$	$+ M$
swallowtail (A_4)	3	x_1^5	$+ u_3x_1^3 + u_2x_1^2 + u_1x_1$	$+ M$
butterfly $(A^{\pm}{}_5)$	4	$\pm x_1^6$	$+ u_4x_1^4 + u_3x_1^3 + u_2x_1^2 + u_1x_1$	$+ M$
elliptic umbilic $(D^-{}_4)$	3	$x_1^2x_2 - x_2^3$	$+ u_2x_1^2 + v_2x_2 + u_1x_1$	$+ N$
hyperbolic umbilic $(D^+{}_4)$	3	$x_1^2x_2 + x_2^3$	$+ u_2x_1^2 + v_1x_2 + u_1x_1$	$+ N$
parabolic umbilic (D_5)	4	$x_1^2x_2 + x_2^4$	$+ v_2x_2^2 + u_2x_1^2 + v_1x_2 + u_1x_1$	$+ N$

where M and N are Morse functions on the forms:

$$M = x_2^2 + \ldots + x_i^2 - x_{i+1}^2 \ldots - x_n^2, (2 \leq i \leq n) \quad (46)$$

$$N = x_3^2 + \ldots + x_i^2 - x_{i+1}^2 \ldots - x_n^2, (3 \leq i \leq n) \quad (47)$$

The first two itemised forms, regular point and non–degenerate critical point, do not correspond to catastrophe forms. All the forms in the table do correspond to catastrophe forms, and they are a complete list of possible canonical forms for the given number of parameters. The results in the Table hold directly for families without no special and symmetric conditions [22]. Extension to five parameters introduce three new catastrophes. Further extension makes the number of catastrophes grows fast. A treatment of this is given by Gilmore [22].

The commonly used pet names (like fold) for the catastrophes in the list 1 refer to the shape of bifurcation set in the space of control parameters (the u and v parameters). The given pet names will be used as reference in this work, and usually the same names are used in literature although ambiguities occur. There are nice illustrations of the geometry for all seven catastrophes in Saunders [50]. The symbols D_5, etc. refer to a classification scheme often often used within the theory. The symbols are not important in this context, and are only given for the sake of precision.

Hence if the conditions hold and the number of control parameters for a system is sufficiently small then the expected topological changes for the system can be found merely by a table lookup!

The sum of the first two terms, the germ and the unfolding, is called the catastrophe function or the universal unfolding of the germ.

Unfoldings

Given a family of functions, for instance $f(x, u) = x^4 + ux^2$, it is of interest to know if it is locally stable. If it is not locally stable, it is of interest to know how many and which terms are needed to stabilize it. The process of adding terms to a family of functions to make it structural stable is called unfolding. If it is locally stable then a perturbation can not alter the local topological structure. If the highest order of the function is k then a sufficient unfolding is the one that includes all terms with order less than k. In case of $f(x, u) =$

$x^4 + ux^2$, this total unfolding would be $f(x, u, u_1, u_2, u_3) = x^4 + u_1x^3 + ux^2 + u_2x + u_3$. A sufficient unfolding for a family is called a versal unfolding. Now it is interesting to know the smallest amount of terms needed for the family to be stable. This is interesting since it reflects the number of essential free degrees of freedom in the system. Such a minimal unfolding is called a universal unfolding. The universal unfoldings are listed in Table 1 for the canonical germs.

The perturbation formulation

Perturbations are often used as an alternative way of formulating the canonical forms. Here one focus on the degenerate member of the family, that is a particular function consisting of the sum of a germ and a Morse function. Hence the function consists of a non-Morse part and a Morse part. Perturbing this function will not influence the qualitative structure of the Morse part, but in general the qualitative structure of the non–Morse part, the germ, will be influenced. The germ split up into Morse parts. This process is called morsification. The non–morse part hold the name, germ, since it is the part that determine the qualitative structure. The listed unfoldings in Table 1 are the perturbations that will describe the morsification. Hence the perturbation approach studies non–morse function by embedding them in a family of functions where the family is given by the perturbation.

The splitting lemma

Thom's splitting lemma is based on the idea of splitting the local form into a Morse and non–Morse part. The lemma is closely related to Thom's classification theorem.

Lemma 1 (Thom's Splitting lemma)
Let $f : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}$ be smooth, and let $(x; u) = (x_1, \ldots, x_n; u_1, \ldots, u_r)$ be a point in the product space $\mathbb{R}^n \times \mathbb{R}^r$. Suppose the Hessian matrix has corank l at $(x; u)$. Then f is locally equivalent to a family g of the form:

$$\begin{aligned} g &= \tilde{f}(y_1(x;u), \ldots, y_l(x;u);\ u) + \grave{f}(y_{l+1}, \ldots, y_n) \\ &= \tilde{f}(y_1(x;u), \ldots, y_l(x;u);\ u) \pm y_{l+1}^2 \pm \ldots \pm y_n^2 \end{aligned} \tag{48}$$

where y is a diffeomorphic coordinate transformation

The splitting lemma states that a coordinate transformation can bring the function on a form where the "good" and the "bad" variables are separated. It corresponds to a transformation that brings the Hessian on the following form:

$$H_g = \left[\begin{array}{c|c} \frac{\partial^2 \tilde{f}}{\partial y_i \partial y_j} & 0 \\ \hline 0 & \frac{\partial^2 \grave{f}}{\partial y_r \partial y_s} \end{array}\right] \tag{49}$$

where $1 \leq i, j \leq l$ and $l + 1 \leq r, s \leq n$. The lower right submatrix is a diagonal matrix which only depend on the $n - l$ well behaved variables. The upper left sub matrix depend on the l first variables and the control parameters u. The latter dependency on the

parameters u implies that the sub matrix can have less than full rank for specific values of u.

This strong lemma states that only the essential l variables need to be study for changes in structure the remaining variables correspond to stable directions which are not tied to the control parameters.

6.1 Manifolds in the jet space

A k–jet for a function is basically the truncated Taylor series as presented in Section 3 on Morse functions. Equal functions in the sense of $C^k(x_0, \mathbb{R}^1)$ topology have the same k–jet. A particular k-jet can be represented by a point in Euclidean Space $\mathbb{R}^D$ where D is given by

$$\begin{pmatrix} n+k \\ n \end{pmatrix}$$

where n are the dimension of the space of the state variables. The Euclidean space can in same sense be interpreted as a space of jets.

Let us look at a function $f : \mathbb{R}^n \to \mathbb{R}$. A mapping $\mathcal{F} : \mathbb{R}^n \to \mathbb{R}^D$ based on f can be defined between the domain $\mathbb{R}^n$ and the space of jets $\mathbb{R}^D$ in a simple manner: Each point $x_0 \in \mathbb{R}^n$ maps to the point representing the k–jet for the function f: $j^k f(x_0)$. Similarly can a family of functions $g : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}$ induce a mapping $\mathcal{G}$ between the product space $\mathbb{R}^n \times \mathbb{R}^r$ and the jet space $\mathbb{R}^D$. The introduced mappings are in general well behaved, and will be a parametric representation of a $n+r$–dimensional manifold in $\mathbb{R}^D$ [22].

Now other interesting manifolds in the space of jets $\mathbb{R}^D$ exist. The jets for which the first order terms vanish correspond to jets expanded in critical points. This subspace Λ of R^D is a linear subspace of $\mathbb{R}^D$ with dimension $D - n$.

Based on the idea of transversality of mappings to manifolds, it can be stated that the typical critical points for the function f are isolated! In the following we describe how this is possible.

6.2 Transversality

Transverse means cross. The entities crossing are manifolds and mappings:

- A manifold is transverse to another manifold
- A mapping is transverse to a manifold
- A mapping is transverse to a mapping.

The notion of transversality can be made mathematically precise:

Definition 16 (Pointwise transversality)
Let $\mathcal{P}$ and $\mathcal{L}$ be manifolds embedded in $\mathbb{R}^m$. The manifolds are said to intersect transversally at x_0, or are transverse at x_0, if their tangent spaces span the tangent space $\mathbb{R}^m$ at x_0

A very simple example of manifolds intersecting transversally is two perpendicular planes embedded in $\mathbb{R}^3$.

Definition 17 (Transversality for manifolds)
Two manifolds $\mathcal{P}$ and $\mathcal{L}$ in $\mathbb{R}^m$ intersect transversally, or are transverse,if one of two conditions are fulfilled:

- *They intersect transversally at all points of the intersection.*
- *They do not intersect at all.*

Proposition 4
If $\mathcal{P}$ is transverse to $\mathcal{L}$ in $\mathbb{R}^m$ then

- *if $dim\mathcal{P} + dim\mathcal{L} < m$ then $\mathcal{P}$ and $\mathcal{L}$ do not intersect at all*
- *if $dim\mathcal{P} + dim\mathcal{L} \geq m$ then $\mathcal{P}$ and $\mathcal{L}$ either do not intersect or intersect in a manifold of dimension $dim\mathcal{P} + dim\mathcal{L} - m \geq 0$*

Proposition 5 (Stability)
Suppose $\mathcal{P}$ and $\mathcal{L}$ are perturbed slightly to $\tilde{\mathcal{P}}$ and $\tilde{\mathcal{L}}$. If $\mathcal{P}$ is transverse to $\mathcal{L}$ then $\tilde{\mathcal{P}}$ is transverse to $\tilde{\mathcal{L}}$ in the same way that $\mathcal{P}$ is transverse to $\mathcal{L}$.

Proposition 6 (denseness)
Suppose $\mathcal{P}$ and $\mathcal{L}$ are perturbed slightly to $\tilde{\mathcal{P}}$ and $\tilde{\mathcal{L}}$. If $\mathcal{P}$ is not transverse to $\mathcal{L}$ then $\tilde{\mathcal{P}}$ will be transverse to $\tilde{\mathcal{L}}$.

Such transversality is a generic property of manifolds according to the above propositions. Transversality involving mappings is based on the same principle with focus on the image $\mathcal{F}(\mathcal{P})$ of the mapping, a manifold in $\mathbb{R}^m$

Definition 18 (Transversality for mappings)
A mapping $\mathcal{F}$ is transverse to a manifold $\mathcal{L}$ if $\mathcal{F}(\mathcal{P})$ is transverse to $\mathcal{L}$.

The above propositions hold for manifolds $\mathcal{F}(\mathcal{P})$ which are the image of a mapping.

The promised result can now be established: Critical points are isolated. The k-jets corresponding to a critical point form a $D - n$ dimensional manifold $\mathcal{L}$ in $\mathbb{R}^D$. The mapping $\mathcal{F} : \mathbb{R}^n \to \mathbb{R}^D$ induce by a function f intersects the "critical point" manifold $\mathcal{L}$ transversally in $\mathbb{R}^D$. According to Proposition 4 will the two manifolds $\mathcal{L}$ and $\mathcal{F}(\mathbb{R}^n)$ intersect in a manifold of dimension $(D - n) + n - D = 0$. A manifold of dimension zero consist of isolated points. Further, by the properties for transversality: if the function f is perturbed slightly then the intersection will be perturbed slightly but the intersection is still transverse. Hence, the critical points just move a bit, when the function is perturbed, so perturbation of a Morse function leads to a Morse function.

The idea of transversality is very powerful for answering questions about stability and genericity. This is so because the set of mappings $\mathcal{F}$ of $\mathcal{P}$ into $\mathbb{R}^n$ for which $\mathcal{F}(\mathcal{P})$ is transversal to $\mathcal{L}$ is open and dense in the set of all mappings of $\mathcal{P}$ into $\mathbb{R}$. The property is generic.

In the sections to come families of functions will be studied. The idea of transversality is immediately applicable to these kind of mappings. The position in the jet space now depends on the n state variables and r control parameters. The family introduce a mapping $\mathcal{F} : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}^D$. The image of this mapping can transversally intersect a manifold of dimension $D - (n + r)$. We use up n degrees of freedom in locating the jets corresponding to critical points in the jet space. The remaining nonzero coefficients for the jets on this "critical points" submanifold depends on the r control parameters. In particular the matrix elements of the Hessian. There are $n(n+1)/2$ independent matrix elements in the Hessian. The image of the mapping $\mathcal{F}$ induce by the family will intersect transversally with a manifold consisting of jets corresponding to critical points where r of $n(n+1)/2$ Hessian matrix elements are degenerate. For l eigenvalues of the Hessian to vanish the number of degenerate matrix elements must be $l(l+1)/2$. Hence the largest l for which $r \geq l(l+1)/2$ holds tell the corank for the Hessian for r degenerate matrix elements. So for one and two control parameters one only sees corank of 1 for the Hessian, but for three control parameters corank of 2 for Hessian is possible.

The idea of transversality gives an alternative way of understanding the results in Thom's classification theorem besides the purely algebraic. Thom's original theorem is actually stated in terms of manifolds, stability of mappings and equivalence, but this is beyond the scope of this thesis.

7 Multi–scale segmentation

Multi–resolution segmentation techniques attempt to gain a global view of the image structure by examining it at many different resolutions. The coarse scale view gives the global structure, the fine scale view provides the details. For the ideas of coarse and fine scale structures to make sense it is expected that the number of segments decreases with increasing scale. Multi–scale techniques use a fine sampling in the scale direction to make it practical to link structures across scale. The term "tracking of structures" is also used since a motion of the image structure takes place over scale. In this section we will present some terminology and present multi–scale segmentation methods which are related to the chosen one. The design of a multi–scale segmentation method must consider which image structures are of interest, analyse these structures and their expected transformations with scale and determine how to link them across scale. Note that this process does not necessarily involve catastrophe theory at all. If the segmentation is not based on singularities, it is not possible to use catastrophe theory; and even if the segmentation scheme is based on singularities, the linking can be determined in an ad hoc manner based on reasonable heuristics. But since the scale–space basis invites to use differential geometry and study singularities of differential descriptors, the ideas of the multi–scale methods are often highly inspired by singularity theory. The terms cause and effects will be used. A cause is an entity at some fine scale causing an entity, an effect, at a coarser scale.

7.1 Multi–scale segmentation methods

This section presents four distinct approaches to produce a segmentation based on a multi–scale framework. In the following section the four approaches will be compared and discussed.

Lifshitz and Pizer [38] based their segmentation on the stack proposed by Koenderink [33]. The idea is to divide the image into black and white spots called extremal regions. Extrema in the image are linked across scale in the following way. Start with an extremum at the finest scale, then search at the coarser scale in a fixed sized neighbourhood around the location for the fine scale extremum. If an extremum with similar intensity and type is found then the two extrema are linked. In the case of multiple matches, the fine scale extremum closest in position and intensity is marked as the cause of the coarse scale extremum. The possible remaining links from fine scale extrema to the coarse scale extrema are marked as isointensity links. The linked extrema form extremal paths. Generic points are linked in so called isointensity paths. The points location is taken as starting point for the search for a match, then the gradient direction is searched, and the first hit of a pixel with intensity close enough to the desired is chosen as match. All iso intensity paths are continued until an extremal path is reached. The pixels belonging to these isointensity paths form the extremal region for the extremum in the extremal path.

The 3D segmentation schemes by Vincken [55] and Koster [35] build on a generalization of the stack introduced by Koenderink [33] for image segmentation. The stack samples the scale–space representation resulting in a stack of increasingly blurred versions of the image. The generalization to higher dimensions is the hyperstack. The hyperstack is a "pile" of

three dimensional images, hence the term hyperstack. The linking of voxels between scales is based on a measure of affection. The measure is a weighted sum of different measures: intensity proximity, ground volume size and proximity to ground volume mean intensity. A ground volume is the finest scale slice of a 4D scale–space segment. The links depend tremendously on the local image structure and the relative weights in the sum. Vincken [55] states that general and robust sum weights can be determined from experiments.

Lindeberg [40, 39] links blobs and anti–blobs across scale. We discuss blobs in the following, similar results and ideas apply for anti–blobs. Since a blob corresponds uniquely to a maximum, the possible events can be determined with arguments from catastrophe theory. The possible events for a maximum in the image are no interaction with anything, creation with a saddle and annihilation with a saddle. Recall that a blob was delimited by an isophote containing a saddle point. If the involved saddle in the events is shared with a neighbouring blob then it means that the two segments interact. In the case of maximum creation it means that a segment splits up, in the case of maximum annihilation it means that two segments merge. If the saddle is not shared the singularity event corresponds to annihilation and creation, respectively, for a segment. This can be implemented with a region based tracking scheme. The correspondence between two blobs is simply established if they overlap. Region based singularity tracking is a robust way of tracking [40], we shall return to this issue in the Sections 9 and 9.2.

Griffin *et al.* [36] make a multi–scale n–ary hierarchy. The basic idea is to make a hierarchical description for each sample value of the scale and then link these hierarchies across scale. In a hierarchical description of the structure the segments are ordered in a tree structure. A segment is either the sum of its sub objects or an atom, the smallest possible part, which is taken to be a pixel. Hierarchical descriptions can in principle be formed from any superficial description, but some descriptions are more natural to extend. Particularly descriptions where the edges separate distinct regions are conceptually easy to extend, because the notion of object and sub objects is most reasonable for objects with closed boundaries. Vincent *et al.* [54] suggest to build such a hierarchy using the idea of flooding the landscape for the gradient squared; when two bodies of water meet the regions are merged. Griffin *et al.* [36] build the hierarchy by iteratively merging adjacent objects. The order of the merging is done in a manner similar to watershed techniques. The linking of the hierarchies proceeds from coarse scale to fine scale and from the top of the hierarchies to the bottom. First the roots in the hierarchies are linked, then the subobjects of the roots are matched, etc. The result is a multi–scale n–ary hierarchy.

7.2 Properties of the multi–scale methods

The four presented algorithms (Lifshitz and Pizer, Vincken, Lindeberg and Griffin *et al.* [38, 40, 55, 36]) can be compared in several ways. The two first track points through scale–space to make the linking, Lindeberg tracks regions and Griffin *et al* track hierarchies.

All schemes except Lindebergs have used a logarithmic sampling of the scale, Lindeberg uses an adaptive sampling which is also scale invariant in the sense that no scale is preferred over another. This sampling scheme is an interesting alternative, since the sampling can be adapted to the place where the topological changes take place and thereby optimise the number of necessary scale samplings.

The linking has been processed fine to coarse by Lifshitz and Vincken, coarse to fine by Griffin *et al.* and pairwise by Lindeberg. The latter makes it theoretically possible to process the linking in parallel. Often, and especially in a segmentation framework coarse–to–fine is preferable to fine–to–coarse. This is so since the structures of interest are usually not the very fine details. If the fine details were the structures of interest, one is better of by taking an image with better resolution to gain robustness. A coarse–to–fine construction allows coarse scale information to be used before details and it might be possible to get the desired answers without going into details. Pair–wise construction can mimic both coarse–to–fine and fine–to–coarse constructions.

Another comparison of methods require the introduction of some terminology: The algorithms produce segments extending over scale. In other words, the linked segments will form a scale–space segment extending both in the spatial as well as in the scale direction. Taking a slice of the scale–space segment corresponding to a particular scale produces the segmentation at that scale. The linking scheme determines the possible topological form of the scale–space segments. A multi–scale segmentation algorithm allows:

- disappearing, (annihilation)
  if a scale–space segment may exist on a fine scale but not on a coarser scale.

- creation,
  if a scale–space segment may exist on a coarse scale but not on a finer scale.

- merging,
  if a scale–space segment is disconnected at a fine scale and connected at a coarse scale.

- splitting,
  if a scale–space segment is connected at a fine scale and disconnected at a coarse scale.

A good figure to keep in mind in the following is Figure 3 in section 2.3 illustrating an isophote of the scale–space image. A scale–space segment defined by one isophote surface allows disappearing (the surface is closed on top), merging (a bridge) and splitting (extrema creation).

A slice of a scale–space object/segment for a particular scale will give the object/segment at that particular scale. For a 3D scale–space segmentation the slice will be a segmentation of the domain for a 2D blurred version of the original image where the degree of blurring corresponds to the chosen scale. For example slicing the surface isophote in Figure 3 will give 2D isophotes.

All the described multi–scale algorithms allow objects to disappear and merge with increasing scale. The creation of objects is by definition not possible in the multi–scale n–ary hierarchy neither can creation occur in the hyperstack segmentation by Vincken [55] and Koster [35] by construction. But both Lindeberg and Lifshitz implement creation events. Splitting events are only allowed by Griffin *et al.* and Lindeberg.

Another point of interest is that the hyperstack and the extremal region methods need the scale–space representation to produce a segmentation but the n–ary hierarchy and the blobs can form a segmentation on a single scale without knowledge of other scales.

Lifshitz and Pizer, and Lindeberg base the segmentation directly on differential singularities, namely the critical points in the image. Such schemes induce a natural linking scheme based on the motion of the critical points in scale–space. This thesis defines the segmentation based on the minima for the gradient magnitude.

8 Catastrophe theory applied to the gradient squared

In this section results from catastrophe theory will be applied to study the possible changes of topology for the gradient squared when the image is blurred.

It is analysed how the singularities of the magnitude of the gradient can be created, annihilated, split and merged as scale changes. When one has a family of functions over a multi–dimensional domain with possibly degenerate points then often the analysis starts by extracting the degenerate directions and ignoring the non-degenerate using the splitting lemma [50]. For simplicity we will take a different approach and only study the one dimensional case from the beginning, but note that adding Morse functions in perpendicular directions will bring the catastrophe into a higher dimensional space. Hence results from the one–dimensional analysis is applicable in higher dimensions. The question arise if higher order catastrophes are missed by this one–dimensional analysis. This is not a trivial question since the umbilic catastrophes are not present in the one–dimensional case. The splitting lemma can solve this problem. If the Hessian matrix for the analysed problem has corank 1 then an analysis for one dimension is sufficient. This will be analysed in section 8.3.

8.1 Possible catastrophes for the gradient squared in 1D

The image L is expanded in the heat polynomials[13] [13] to fourth order:

$$^4\tilde{L} = L + L_x x + \frac{1}{2!}L_{xx}(x^2+2t) + \frac{1}{3!}L_{xxx}(x^3+6xt) + \frac{1}{4!}L_{xxxx}(x^4+12x^2t+24t^2) \quad (50)$$

Equation 50 is our local model of the image written in an arbitrary coordinate system. The coefficients, the L's, are image derivatives evaluated at $(x=0, t=0)$. The derivatives and the gradient squared can easily be calculated as:

$$^4\tilde{L}_x = L_x + L_{xx}x + \frac{1}{2!}L_{xxx}(x^2+2t) + \frac{1}{3!}L_{xxxx}(x^3+6xt) \quad (51)$$

$$\begin{aligned} ^4\tilde{L}_x^2(x,t) \;=\; & L_x^2 + 2L_xL_{xxx}t + L_{xxx}^2t^2 + \\ & (2L_xL_{xx} + L_{xx}L_{xxx}t + 2L_xL_{xxxx}t + 2L_{xxx}L_{xxxx}t^2)x \\ & (L_{xx}^2 + L_xL_{xxx} + L_{xxx}^2t + 2L_{xx}L_{xxxx}t + L_{xxxx}^2t^2)x^2 \\ & (L_{xx}L_{xxx} + L_xL_{xxxx}/3 + 4L_{xxx}L_{xxxx}t/3)x^3 \\ & (L_{xxx}^2/4 + L_{xx}L_{xxxx}/3 + L_{xxxx}^2t/3)x^4 \\ & (L_{xxx}L_{xxxx}/6)x^5 + (L_{xxxx}^2/36)x^6 \end{aligned} \quad (52)$$

Formula 52 defines a one–parameter family of 1D functions. As we saw in the Section on catastrophe theory it is expected to encounter degenerate critical points in such a

[13]Heat polynomials form a basis for the solutions to the heat equation. Hence expanding in this basis insures that the expansion is a solution to the heat equation. The heat polynomials can be construct by taken a standard monomial and adding terms until the heat equation is fulfilled, for instance $x^4 + 12x^2t + 24t^2$.

family. In more general terms it is expected to encounter points defined by two constraints. In the case of degenerate critical points one of the constraints is "used" to constrain attention to a "critical point" manifold in the space of Taylor coefficient. We start by imposing such a constraint. In other words we assume having a critical point and since critical points are generic it is in fact reasonable to make such an assumption.

Assuming a critical point, a simple transformation $x = z + u(t)$ of x can eliminate the first order term in Equation 52. After the transformation the coefficient to the first order term in z is a fifth order polynomial in u. Setting the coefficient equal to zero and solving for u yield the wanted transformation. Since the order of the polynomial (the coefficient) is odd, there is always a real solution. The Equation was solved for u using Mathematica[14]. In the z coordinate system we stay on the singularity with changing parameter t:

$$^4\tilde{L}_x^2(z,t) = f_0(t) + f_2(t)z^2 + f_3(t)z^3 + f_4(t)z^4 + f_5(t)z^5 + f_6(t)z^6 \tag{53}$$

The coefficient $f_0(t)$ can be removed by a mere translation of the axis for $^4\tilde{L}_x^2(z,t)$ and is not considered further. If the coefficient $f_2(t)$ equals zero for a real value of t then a catastrophe will occur for this t-value. Hence, it is interesting to solve:

$$f_2(t) = 0 \tag{54}$$

Letting Mathematica solve Equation 54 for t yield four solutions: two real valued solutions denoted t_1 and t_2, and two complex valued solutions. Only the real valued solutions are meaningful in this context. Equation 53 takes the following form for $t = t_1$:

$$^3\tilde{L}_x^2(z,t_1) = f_3(t_1)z^3 + f_4(t_1)z^4 + f_5(t_1)z^5 + f_6(t_1)z^6 \tag{55}$$

The second order structure disappear leaving only third order structure. This is known as the fold catastrophe where a maximum/minimum pair is created or annihilated. Inserting the t_2 solution in Equation 53 yields a different result:

$$^3\tilde{L}_x^2(z,t_2) = f_4(t_2)z^4 + f_5(t_2)z^5 + f_6(t_2)z^6 \tag{56}$$

Solving $f_3(t) = 0$ yields three solutions and the solutions coincide with the three of the four solutions to the Equation 54; one of the real ones t_2 and the two complex ones. This is the algebraic explanation for the event. In catastrophe theory the event is known as the cusp catastrophe where all terms up til fourth order disappear. The cause of a cusp catastrophe in a 1–parameter family is a special symmetry of the family. Here the symmetry arises from squaring the first order derivative. It is not because of the derivation, it is squaring that induces the symmetry. Squaring the function itself will give the same result.

Actually the total reduced form of Equation 55 is:

$$^3\tilde{L}_x^2(z,t_1) = f_3(t_1)z^3 + f_6(t_1)z^6 = (\frac{-L_{xx}L_{xxx}}{3} + \frac{L_{xxx}^3}{9L_{xxxx}} + \frac{L_xL_{xxxx}}{3})z^3 + \frac{L_{xxxx}^2}{36}z^6 \tag{57}$$

Only the lowest order term z^3 is significant. Consequently will the disappearing of higher order terms not influence the local behaviour.

[14]Mathematica is a registered trademark of Wolfram Research, Inc.

8.2 Expansion at the catastrophes

We have proved that two kinds of catastrophes occur generically. Now let us make the Taylor expansion in a point where a catastrophe takes place. This trick of expanding at the catastrophe is general applicable [50, 22] and usually result in a tremendous simplification of the expressions. The general expansion in Equation 52 is used as a starting point. So we assume a catastrophe at $(x,t) = (0,0)$. Since the point is critical and degenerated the first and second order derivatives must vanish:

$$
\begin{aligned}
{}^4\tilde{L}^2_{x_x}(0,0) &= 2L_xL_{xx} = 0 && (58)\\
{}^4\tilde{L}^2_{x_{xx}}(0,0) &= L^2_{xx} + L_xL_{xxx} = 0 && (59)
\end{aligned}
$$

We have two solutions $L_{xx} = 0 \wedge L_x = 0$ and $L_{xx} = 0 \wedge L_{xxx} = 0$ corresponding to a degeneration of the singularities described in section 5.3, Equation 40 and 41. Each is considered separately in the following:

- $L_{xx} = 0 \wedge L_x = 0$
  Equation 52 reduces in this case to:

$$
\begin{aligned}
{}^4\tilde{L}^2_{xL_{xx}=0\wedge L_x=0}(x,t) &= L^2_{xxx}t^2 + (2L_{xxx}L_{xxxx}t^2)x && (60)\\
&\quad +(L^2_{xxx}t + L^2_{xxxx}t^2)x^2 + (4L_{xxx}L_{xxxx}t/3)x^3\\
&\quad +(L^2_{xxx}/4 + L^2_{xxxx}t/3)x^4 + (L_{xxx}L_{xxxx}/6)x^5 + (L^2_{xxxx}/36)x^6
\end{aligned}
$$

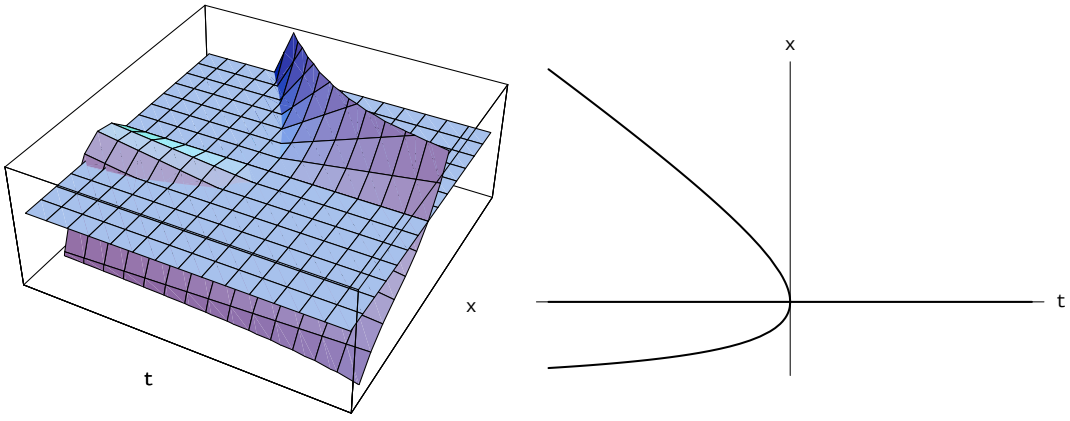
  For $t = 0$ the cusp catastrophe occur. It is the cusp catastrophe because all term until fourth order vanish. The topological structure (the singularities) for this function in the neighbourhood of $(x,t) = (0,0)$ is determined by differentiating Equation 60 and keeping significant terms:

$$
({}^4\tilde{L}^2_{xL_{xx}=0\wedge L_x=0})_x(x,t) = 2L^2_{xxx}tx + 4L_{xxx}L_{xxxx}tx^2 + L^2_{xxx}x^3 \qquad (61)
$$

  The significance of the terms is determined by the order of the term. The third order term in x stays for $t = 0$, so higher order terms in x are irrelevant. Knowledge of the topology in the immediate neighbour of the catastrophe in $(x,t) = (0,0)$ is sufficient. This implies that second order terms in t is not important. The resulting expression was given in formula 61.

  In Figure 11 (a) is plotted two functions: the function given by Equation 61 and the constant function zero. So the intersection of the two surfaces corresponds to solving function 61 equal zero. The zeros of function 61 is plotted in (b). Hence Figure 11 (b) is the traces of the singularities in the x–t plane. It is a bit hard to see but there is a singularity trace for $x = 0$. Figure 11 (b) illustrates how three singularities merge into one.

  The position of the singularities can be determined as a function of the parameter t by solving Equation 61 equal zero. The set of solutions is called the singularity set. For $t \leq 0$ there are three real–valued solutions; they coincide for $t = 0$. For positive t values there is only one real–valued solutions. Hence, for $t = 0$ we pass a catastrophe, namely from a minimum/maximum/minimum configuration to one



(a) (b)

Figure 11: (a) The function $({}^4\tilde{L}^2_{xL_{xx}=0 \wedge L_x=0})_x(x,t)$ defined in Equation 61 is plotted as a surface in (a) for the values $L_{xxx} = 2$ and $L_{xxxx} = 3$ in the intervals $t \in (-0.1, 0.1)$ and $x \in (-0.5, 0.5)$. Also in (a) is the constant function zero. In (b) is shown the intersection between the two surfaces in (a). The intersection corresponds to the singularity set.

minimum. When an extra dimension is added the reverse event is also possible. The set of degenerate singularities are called the bifurcation set. In this case the set contains one point $(0, 0)$

A cusp catastrophe in the gradient squared happens when a fold catastrophe happens in the image due to blurring. Figure 12 pictures the fold catastrophe for the image (top row) and the corresponding cusp catastrophe in the gradient magnitude. From left to right there is an annihilation of two singularities in the image and the corresponding transition in the gradient magnitude from three singularities (d) over three coinciding singularities (e) to one singularity (f). The function in (f) is above the horizontal axis.

- $L_{xx} = 0 \wedge L_{xxx} = 0$
  In this case Equation 52 reduces to:

$$\begin{aligned} {}^4\tilde{L}^2_x(x,t) &= L_x^2 + (2L_xL_{xxxx}t + L_{xxxx}t^2)x + (L^2_{xxxx}t^2)x^2 + (L_xL_{xxxx}/3)x^3 \\ & \quad (L^2_{xxxx}t/3)x^4 + (L^2_{xxxx}/36)x^6 \end{aligned} \tag{62}$$

  The same kind of strategi as in the previous case can be applied.

  We can study the structure arbitrary close to the catastrophe, that is, for arbitrary small scale values t and therefore disregard terms with t^2. Ignoring the zeroth order term and terms with t^2 formula 62 reduces to:

$${}^4\tilde{L}^2_x(z,t) = 2L_xtz + \frac{L_xz^3}{3L^2_{xxxx}} + \text{Higher order terms} \tag{63}$$

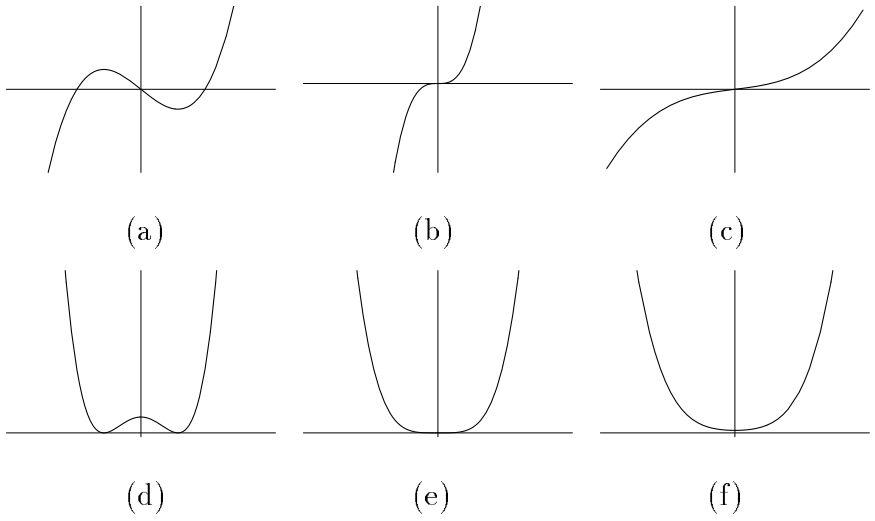


Figure 12: Top row is the function $f(x,v) = x^3 - vx$ for $v = 1$ (a), $v = 0$ (b) and $v = -1$ (c). Bottom row is gradient magnitude of f squared: $(f_x(x,v))^2 = (3x^2 - v)^2$ for the same values of v. Each column has the same value for v

> For t–values less than zero the function above has a maximum and a minimum. For t–values greater than zero the function is without critical points.

The analysis has shown two types of events: (1) a degenerated critical point in the image corresponds to the cusp catastrophe in the gradient magnitude and (2) vanishing second and third order structure in the same direction in the image corresponds to the fold catastrophe in the gradient magnitude. For higher dimensions the image events, fold catastrophe and vanishing second and third order, can be reversed. The genericity of the reverse events is well known [12, 38]. This implies that the reverse events for the gradient squared are also generic.

8.3 Splitting the dimensions

In this section we use the splitting lemma stated in equation 1 to show that the one dimensional analysis actually captures all the possible catastrophes. Again we make an expansion in heat polynomials up to fourth order, but this time in 2D:

$$\begin{aligned} {}^4\tilde{L}(x,y;t) &= L + L_x x + L_y y + \frac{1}{2!}L_{xx}(x^2+2t) + L_{xy}xy + \frac{1}{2!}L_{yy}(y^2+2t) \\ &+ \quad \text{terms up to and including fourth order heat polynomials} \end{aligned} \tag{64}$$

We want to study the corank for the Hessian matrix for the gradient squared since this determines the essential number of variables. We know that the case of corank 1 occurs. We would like to show that the case of corank 2 do not occur. Corank 2 means that the analysis have to be done on two "bad" variables instead of only one. We can safely assume the expansion to be at a catastrophe since it is proved that catastrophes occur.

Hence a catastrophe is assumed at $(x, y; t) = (0, 0; 0)$, this means in particularly that several singularities coincide at the point $(x, y; t) = (0, 0; 0)$. Mathematica has been used for the following derivation. The expansion in Equation 64 is differentiated with respect to x and y, and the gradient squared is combined from them:

$$(\nabla({}^4\tilde{L}))^2 = ({}^4\tilde{L}_x(x, y; t))^2 + ({}^4\tilde{L}_y(x, y; t))^2 \tag{65}$$

The second order derivatives are calculated and the Hessian is evaluated at $(x, y; t) = (0, 0; 0)$:

$$H_{(\nabla^4\tilde{L})^2} =$$
$$2\begin{bmatrix} L_{xx}^2 + L_x L_{xxx} + L_{xy}^2 + L_{xxy} L_y & L_x L_{xxy} + L_{xx} L_{xy} + L_{xyy} L_y + L_{xy} L_{yy} \\ L_x L_{xxy} + L_{xx} L_{xy} + L_{xyy} L_y + L_{xy} L_{yy} & L_{xy}^2 + L_x L_{xyy} + L_{yy}^2 + L_y L_{yyy} \end{bmatrix} \tag{66}$$

There are two kinds of singularities for the gradient squared: the critical points in the image and vanishing second order structure. For critical points in the image the Hessian in equation 66 reduces to:

$$2\begin{bmatrix} L_{xx}^2 + L_{xy}^2 & L_{xx} L_{xy} + L_{xy} L_{yy} \\ L_{xx} L_{xy} + L_{xy} L_{yy} & L_{xy}^2 + L_{yy}^2 \end{bmatrix} \tag{67}$$

The determinant for this matrix equals $4(L_{xx} L_{yy} - L_{xy}^2)^2$. It is not generic for the image to have a total degeneration of the second order structure in a critical point, it is only generic to have a degeneracy in one direction. Consequently the Hessian matrix for the gradient squared must have corank 1 in a catastrophe for this type of singularity.

The other type of singularity in the gradient squared is given by $L_{ww} = 0$ and $L_{wv} = 0$ in the (v, w) coordinate system, the Hessian in Equation 66 will reduce to:

$$2\begin{bmatrix} L_{vv}^2 + L_{vvw} L_w & L_{vww} L_w \\ L_{vww} L_w & L_w L_{www} \end{bmatrix} \tag{68}$$

The determinant of the Hessian equals $4L_w(L_{vv}^2 L_{www} + L_{vvw} L_w L_{www} - L_{vww}^2 L_w)$. There is no special dependency among the elements that reduces the corank from zero. The only way to get the corank up is to impose conditions on the coefficients. Imposing the condition on the coefficient that the determinant should equal zero, will impose a corank 1 on the Hessian matrix. Imposing corank 2 on the Hessian matrix will require two extra conditions. So corank 2 requires a total of five conditions, namely two conditions for the singularity, and three for the Hessian. Five conditions correspond to an object of codimension five. It is *not* generic to meet an object of codimension five in a one–parameter family of 2D functions.

A similar argument can be made for higher dimensions, including the special case of three dimensions. The conclusion is that a one dimensional analysis gives the exhaustive result for the possible catastrophe events for the gradient squared.

8.4 Two catastrophes in a row, an example

An example[15] might clarify and illustrated a bit more. In Figure 13 a 1D signal is embedded in a family of blurred versions of the signal. The original signal (a) is blurred

[15]The Matlab system made these graphs and the Matlab source code for these graphs has been cut and pasted from source code of Peter Johansen.

to such a degree that fold catastrophe takes place in (c) where a minimum and maximum annihilates. Further blurring smoothes away the second order structure remaining after the fold catastrophe (c–e). In Figure 14 is displayed the squared first order derivative of the 1D signal from Figure 13. From (a) to (c) is seen how the two minima and a maximum merge into one minimum corresponding to the cusp catastrophe. Further from (c) to (e) the remaining minimum annihilates with the neighbouring maximum. So from (a) to (e) two catastrophe take place, first a cusp catastrophe (a–c) secondly a fold catastrophe (c–e).

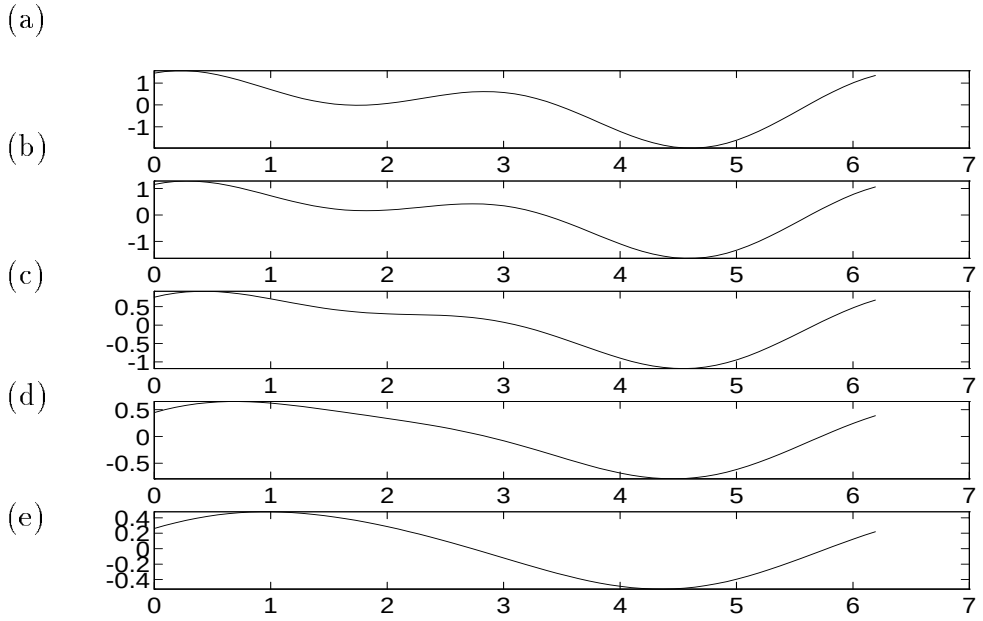


Figure 13: A 1D signal has been embedded in a family of Gaussian blurred versions of the signal. (a) is the original signal and the blurring increases from frame to frame with the must blurred version presented in (e). From (a) to (c) a fold catastrophe takes place approximately for $x = 2.3$, the minimum at $x = 1.7$ and the maximum at $x = 2.7$ in frame (a) annihilates in frame (c). From (c) to (e) the concavity at approximately $x = 2.3$ disappears.

A watershed segmentation of the gradient squared for the signal presented in Figure 13 and 14 would in frame (a) have 5 segments. The intervals for the segments would be located approximately at $(0, 0.9)$,$(0.9, 2.3)$,$(2.3, 3.7)$,$(3.7, 5.6)$ and $(5.6, 6.2)$. In frame (c) the two segments will have merge into one leaving four segments. One additional merge would take place before frame (e) where only three segments would remain. In (e) the intervals would be $(0, 3)$, $(3, 5.6)$ and $(5.6, 6.2)$. The above described segmentation assume that the image boundaries are borders. Actually the signal is periodic, hence it would also make sense to wrap the image around and let the two outer most segments become one.

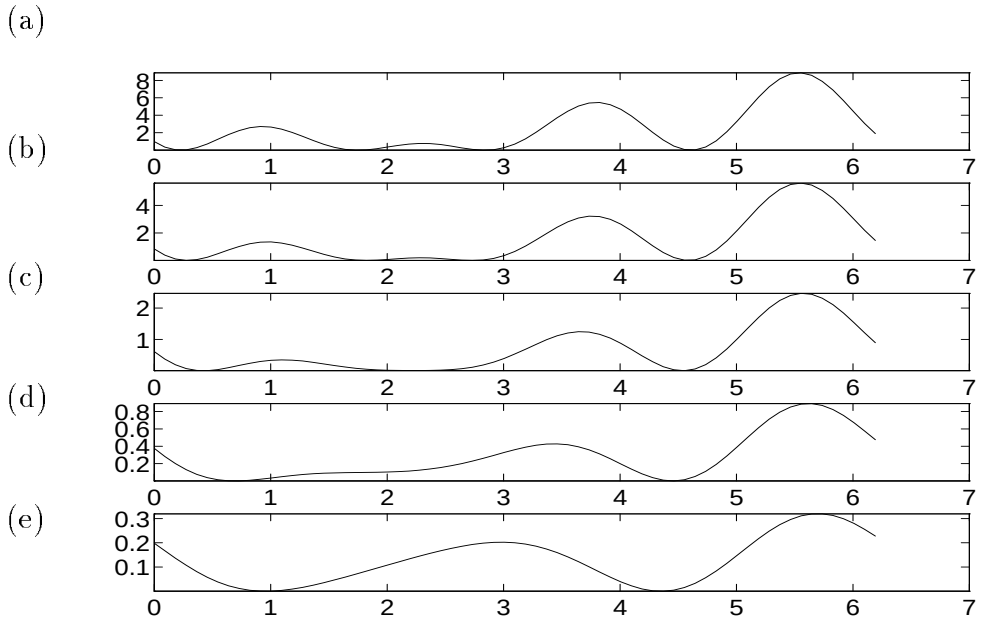


Figure 14: The squared first order derivative of the signal in Figure 13 is displayed at the same scales as the function itself. From (a) to (c) the minima at $x = 1.7$ and $x = 2.7$ and the maximum at $x = 2.3$ meet in a cusp catastrophe. In (c) the singularities have merge into one minimum at $x = 2.3$. From (c) to (e) the fold catastrophe happens. The minimum at $x = 2.3$ now meet the maximum around $x = 1.4$ and annihilates. During this process the minimum at $x = 0.4$ in (c) moves towards $x = 1$ and when the annihilation takes, this latter minimum takes over the catchment basin "owned" by the annihilated minimum.

9 Linking

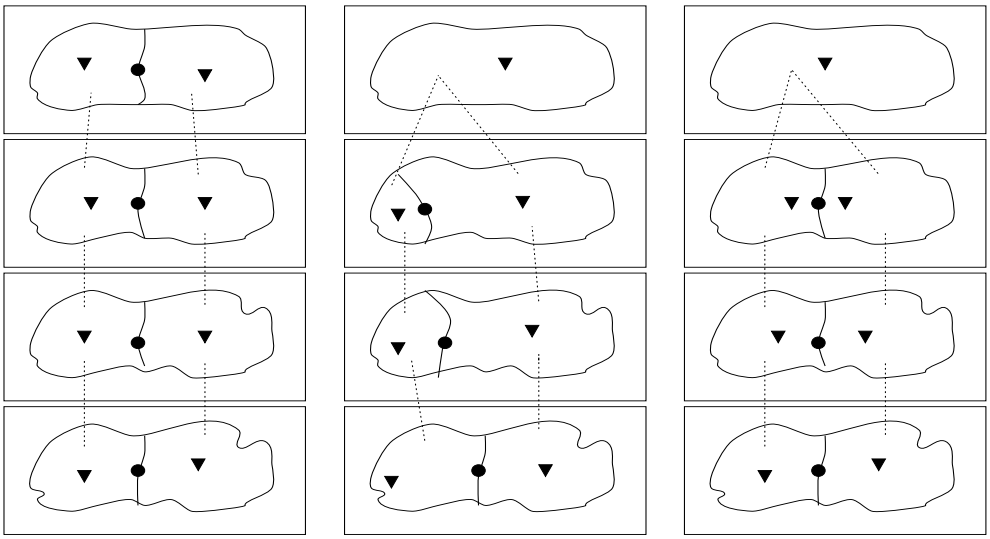
In the previous sections we established the superficial and deep structure of the gradient squared, and in a previous section was presented four different multi–scale linking schemes based either on critical points in the image or on ad hoc determined criteria. In this section are established possible linking schemes for segments defined by local minima for the gradient squared. The linking of catchment basin segments can be based on the analysis of the singularities, due to the duality between segments and the minima of the gradient squared. There are five generic events for minima of the gradient squared:

- No interaction with other singularities
- Creation in a pair with a saddle
- Annihilation with a saddle
- Merging with a saddle and another minimum into one minimum
- Splitting into saddle and two minimum

The above wording with reference to a saddle hold for the 2D case. In the 1D case all occurrences of saddles should be replaced with the word maximum and in 3D the word should be hypersaddle.

A linking establishes the structural relation between segments at different scales. The cause and effect relationship is needed for projecting the coarse scale structure to a sufficiently fine localization scale. The term projection will be used in the sense that a coarse scale structure can be tracked via the linking to a finer scale. If one coarse scale segment links to several fine scale segments then the union of these segments are the projection of the coarse scale segment.

It will be instructive to see the motion of some segments and minima together. In Figure 15 and 16 is presented the relation between segments and singularities based on the gradient squared. The chosen linking scheme is indicated by broken lines; the possible linking schemes will be discussed in section 9.1. In the figures the following symbols are used: A triangle pointing downwards indicates a minimum, a filled circle indicates a saddle, a unbroken line indicates a border between segments and broken lines symbolises a link between segments. The relation between segments and singularities will now be discussed. Scale always increases upwards in the figures, and the description of the events will be from fine to coarse, so bottom up. When there is no interaction between the singularities, it simply implies that the topology for the segments stay the same. This situation is displayed in Figure 15 (a): the minima (the triangles) move about but do not interact. In (b) an annihilation takes place. The left minimum and the saddle annihilate, in the segmentation this implies that the neighbouring minimum (the right one) takes control of the region previously belonging to the annihilated minimum. In (c) the two minima and the saddle move towards each other and interact in a cusp catastrophe this implies for the segments that the dividing border disappears and one segment remains; a merge has taken place. In Figure 16 (a) a minimum and a saddle is created via a fold catastrophe. The created segment conquers more and more of the image domain



(a), No change (b), An annihilation (c), A merge

Figure 15: Scale increases from bottom to top. No change in topological structure is presented in (a). In (b) is a segment annihilated. The corresponding minimum annihilates with the saddle on the boundary to the neighbouring segment. In (c) two segments merge.

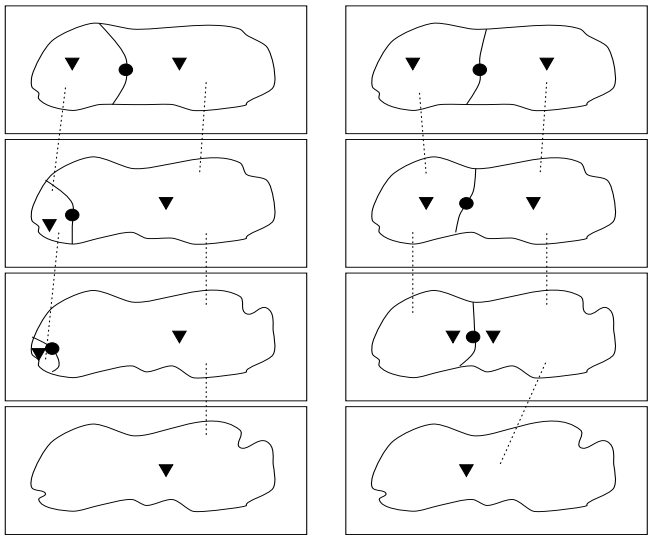
previously owned by the stable minimum to the right. A new segment has emerged. In (b) two minima and a saddle split from one minimum causing the creation of a new segment.

The illustration of the creation and the split are schematic to emphasize the change of structure. In practice emerging segments seem to disappear fairly quickly after appearing. They have a short lifetime over scale in scale–space. The range of scale over which an object exists has been observed by Witkin [56] to be a possible measure for the importance of the object.

9.1 Possible linking schemes

We shall present three linking schemes based on the derived generic events for the singularities. Several considerations have to been taken: How can it be implemented? What will be the projection of a single segment? What will be the projection of a coarse scale segmentation? An end user of the system will typically only be interested in the final result, the projection of a coarse scale segmentation, that is the well localised coarse scale structures.

The three linking schemes are itemised according to the resulting structure of links. The last scheme is the one implemented and the one displayed in Figure 15 and 16. The differences and similarities are discussed with this linking scheme as reference. The differences will arise for the annihilation, creation and split events. The events of no change and merging do not leave much choice for a reasonable linking scheme and are found to be similar for all schemes. Projecting the coarse scale segment after a merge



(a), A creation (b), A split

Figure 16: Scale increases from bottom to top. In (a) is a segment created. In (b) a segment splits into two segments

results in a supersegment consisting of the two merged segments. Projecting the two coarse scale segments after no change of topology simply localises these segments at the finer scales. Projections for the remaining types of events will be discussed for each linking scheme.

A set of undirected graphs.
A theoretical possibility for links are the paths of the minima. This linking requires a robust way of tracking singularities even through catastrophes. Point tracking through catastrophes is not feasible in practice. The tracking through catastrophes is necessary to differentiate between annihilation and merge, and between creation and split. Tracking of singularities will be further discussed in Section 9.2.

The linking in Figure 15 and 16 has to be altered to reflect such a linking. For the annihilation in Figure 15 (b) the left link from the top box has to be removed and for the split in Figure 16 (b) a link should be added from the bottom segment to the left segment above. The remaining links stay. This linking structure will be a set of undirected graphs.

After an annihilation event, the projecting of the coarsest scale segment results in one segment at the finest scale (in the Figure 15 it is the rightmost). After a creation event a projection of the created segment will result in nothing at the finest scale. The neighbouring coarse scale segment projects to the one segment at finest scale. After a split event either coarse scale segments will project to the one segment at finest scale.

A projection of an entire segmentation at a coarse scale will correspond to a segmentation at fine scale: Some regions are not segments because they do not cause any segment at the coarse scale, this is the consequence of an annihilation somewhere between finest and coarsest scale. Some regions will correspond to two or several different segments, this is the consequence of splittings.

Two things suggest that the similar catastrophes should be treated in the same manner, namely (1) the problem of distinguishing between the catastrophes and (2) the events for the segments are similar. Let us elaborate point (2) a bit: an annihilation of singularities does not imply that a segment simply disappears out in the blue, as previously noted it implies that the neighbouring segment moves in and take over the disappearing segment. Hence the same kind of linking should be applied in the event of an annihilation and a merge. Similar should the split and creation be treated equally. The linking of a merge event as presented in Figure 15 seems to be *the* right way to link otherwise the number of segments will increase instead of decrease with increasing scale. Consequently the linking for the annihilation event must follow the merge and not vice versa. The two remaining schemes will use this approach, the linking is presented in Figure 15. For the creation and splitting events there are two possibilities either the creation adapts to the splitting event or vice versa. These two possibilities correspond to the last two linking schemes presented below. The latter scheme is presented in Figure 16.

A set of undirected graphs without ends at inbetween scales.
The linking in the case of creation and splitting is unified by linking the segments according to the involved saddle. For the creation it means that the created segment is linked to the same segment as its neighbour. The neighbourhood relation is given by the created saddle. For the splitting event it implies that the splitting segment links to both subsegments. In Figure 16 two links should be added to illustrate this scheme. One link in (a) between the created segment (the smallest one) and the large segment at the finest scale and in (b) one link between the unlinked, split segment and the finest scale segment.

A projection of one segment will always be at least one segment at the finest scale.

A projection of an entire coarse scale partition will result in a fine scale partition. Because of the linking of splitting and creation event one fine scale segments can corresponds to several coarse scale segments. So one could say that in this partition several coarse segments lay on top of each other or co–exist at the same location.

The resulting linking structure is a set of undirected graphs where all loose ends are located at finest scale and the coarsest scale.

A set of trees, (hierarchies).
The last linking is the one presented in the Figures 15 and 16. Here merge and annihilation are also treated equally; the involved segments are linked to the resulting coarse scale segment. Split and creation is treated in the following way. The split is adjust to the creation. A creation is a dangling end in the linking tree. The two split links are pruned, so only one link remain between the fine scale segment and one of the segments above. A simple pruning method is to keep the link between the segments which have highest spatial correlation.

The linking structure is a set of trees. The size of set reduces when coarser and coarser scales are included. Each segment in the structure is an internal node or an external node (a leave) in the tree. Each internal node has links downwards in scale and these links form a tree structure, a hierarchy.

For this link scheme the projection of one segment results either in non, one or several fine scale segments. The number of segments corresponds to the number of leaves located at the finest scale. No segments results from projecting a segment where the downwards

links end before the finest scale is reached.

The projection of an entire partition will result in a partition. One realises this since each segment at the finest scale has a path of links up to a segment at the coarse scale. So a projection from any coarse scale will result in a partition. Since each segment only have one upward link, a segment will always be the projection of exactly one segment.

The last linking scheme introduces an interesting and pleasant property: The projections to the same fine scale but from increasing coarser scales form by definition a object hierarchy where objects are formed by the merging of sub objects. The sub objects are merged if they are causing the same coarse scale object. This effect can be observed in Figure 17 top row, left to right and in Figure 18 right column, bottom up.

9.2 Tracking singularities

The problem of point tracking singularities through catastrophes has not been solved. The main problem is that features can move arbitrarily fast. Let a feature be defined by:

$$g(x;t) = 0, \quad x \in \mathbb{R}^d, t \in \mathbb{R}_+ \tag{69}$$

The drift velocity of this feature can be estimated [41]:

$$\partial_t x = -(\partial_x^T g)^{-1} \partial_t g \tag{70}$$

This means the drift velocity does not have an upper bound. This should be understood in the following sense: given a scale interval and spatial distance then there exists a signal for which the considered feature moves further than the given spatial distance within the allowed scale interval. In other words, let a fixed cylinder in scale–space be defined by a spatial radius and a scale interval. Let some feature be located in the middle of the cylinder, then there exists a signal such that the path for the feature in scale–space will cross the side of the cylinder.

In practice this means that is hard to distinguish between annihilation and merge, and between creation and split. As an example consider figure 15 (b) and (c): if only the top and bottom slice are available then it is impossible to distinguish the two events. A finer sampling in the scale direction can lighten the problem, but does not solve it. The core of the problem is the quantification of the scale–space. The quantification determines when an event can be classified as one or the other. More concretely, if a toppoint for a fold catastrophe is located in a quantification cube of space through which another singularity path passes then it is not possible to say if the event was a fold catastrophe or a cusp catastrophe. This is possible no matter how small the quantification cube is. Hence, a discrimination between the catastrophes are inherently connected to the quantification.

Stilian Kalitzin [30] has suggested a promising way of detecting singularities using results from integral and differential geometry. The concept of winding number is used. The winding number can be defined and calculated in each point of image domain by integrating the vector field along a closed curve, the curve have to be without singularities. The integral equals an integer called the winding number for the region, in the limit of an infinitely short curve the winding number is for the surrounded point. Each type of

singularity has a specific number. The integral over a closed curve give the sum of the winding numbers for the enclosed region. Flanders [16] has an introduction to winding numbers and further references. The method seems robust and it seems possible to make the linking of slow singularities merely by matching equal winding number across scale. In practice the detection scheme is basically region–based since the curve will have a finite length. The problem of the catastrophes has not been solved by the method. The detection scheme is not mathematically defined for a catastrophe point since it is not possible to enclose the point by a curve without singularities. But the approach seems promising.

Lifshitz and Pizer tracked extrema and regular points in their segmentation scheme. They reported problems with the robustness of this scheme. Especially the serious problem of non-containment: a point which at one scale has been classified to be within a region associated to a local maximum, can wander out of that region when scale increases and the linking paths can be intertwined in a rather complicated way.

Lindeberg [39, 40] suggested a region–based tracking scheme of extrema using the blobs as described in section 7, and reported good results using this scheme.

The conclusion is that the most robust way of tracking singularities is based on some kind of region formulation. An approach this thesis also will take.

9.3 The segment linking scheme

The first presented linking scheme for the gradient squared has to rely on a method for tracking singularities and connecting them in catastrophes. It seems to be quite a fundamental problem to perform such a tracking and connection. We have not found a solution and we are not aware of any robust solution in the literature. Hence, this scheme has not been implemented.

The two other schemes are based on the following observations: an annihilation and a merge correspond to the merging of two segments; a creation and a split correspond to the appearance of a new segment or the splitting of an existing segment.

The focus on segments in the linking scheme allows the implementation to be based on segments. Structures move smoothly with scale, hence if the scale sampling is sufficiently fine then segments can be matched by measuring the spatial overlap with the segments at neighbouring scales. So even though the boundaries might move too fast to be tracked, the entire segment can still be tracked. Another important observation is that the appearing or disappearing border is limited in its motion by the outer borders of the segments. So even when a border is about to disappear and moves fast, the remaining outer borders are fixed relative to the motion of the disappearing border.

Hence, in the limiting case of extremely fine scale sampling: segments are linked correctly if the segments are linked according to spatial overlap with segments at neighbouring scales. The observation on the relative stability of the outer boundaries implies that the subsegments do not move significantly outside the border of the resulting segment during the catastrophe event. This latter observation justifies the method for a fine scale sampling.

The latter of three linking schemes have been implemented by a pair–wise scale linking of the segments using a linking criteria of maximum spatial correlation. A fine scale

segment is linked to the one coarse scale segment with which it overlap the most. A similar idea was used for linking of blobs by Lindeberg [40].

The second of three schemes could have be implemented by regarding maximum overlap both in the bottom–up scale direction and in the up–down scale direction. This has not been implemented. The last presented and simplest linking scheme has been implemented.

A short summary

The linking criteria is in practice implemented by maximum spatial overlap for cause segment to the effect segments. In the limit of very fine scale sampling this corresponds to linking the minima. The linking for minima is based on the appearing or disappearing saddle connect the involved minima. When a saddle appear a new segment is created, when a saddle disappear two segments are merged into one.

9.4 Linking in action

In Figure 17 we try to visualize the ideas and their connection. An artificial made image of a check board plus some low scale noise will be segmented as an example.

In the bottom row are slices from the scale–space image where scale increases left to right. The other rows describe the structure in the bottom row in different ways. Hence each column presents the same structure. The second row show the squared of the gradient magnitude. The third row show the watersheds of the squared of the gradient magnitude. The boundaries are marked by colouring the pixel on each side of the boundary. Note how the structures at low scale include many boundaries which normally are consider to be unwanted. What is usually meant by over-segmentation is that the detected structure is detected at a too fine scale. At the highest scale (column three) the low scale structure is not present. The price paid for this result is a delocalization of the boundaries. The answer to this is to probe the deep structure of the segmentation and determine the fine structure causing these interesting coarse scale structure. In practice this means tracking the structures down to a sufficiently fine scale. The result of this is shown in the fourth (top) row. The segments in row three is tracked down to the scale presented in column one. Hence the sub-pictures in column one, row three and four is identical. In column three is the coarse scale structure of interest and this is tracked to scale of localization which provide the wanted segmentation. The gradient squared at coarsest scale (row two, column three) have some large black regions and some small black regions. Black is low value. The large black regions corresponds to black (minima) and white (maxima) squares in the checker board. The small black regions corresponds to the grey (saddles) regions in the checker board. The watersheds of the gradient squared (column three, row three) divide all regions both the large and small ones. It looks like large and small octagons where the large ones correspond to the extrema in the image and the small ones correspond to the saddles in the image. These watersheds are track down to finer scale and the result is displayed in column three, top row. All the large octagons have been tracked down to the well localised squares (note the number of large octagons equal the number of squares). The small octagons are tracked down to regions which are smaller

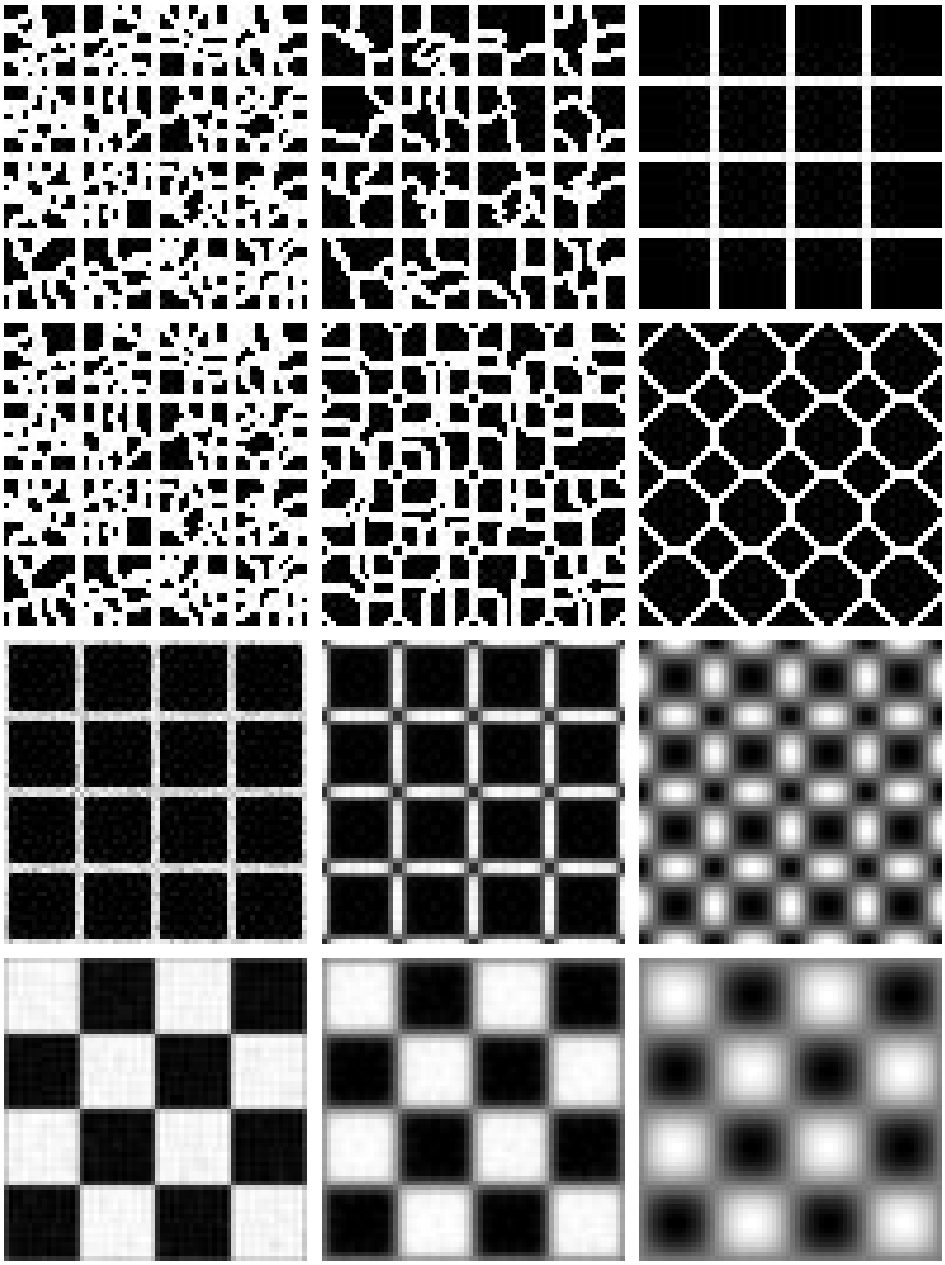


Figure 17: Each column corresponds to the same scale. The sigma values left to right ($\sigma = 0.50, 1.43, 3.70$ pixels) The bottom row presents slices in the scale–space image, the original image is 64x64 pixels and is an artificial check board with fine scale noise added. The row number two from the bottom is the gradient magnitude. Row number three is watersheds of the gradient magnitude. Row number four shows the result of tracking the segments in row three down to the scale in the first column.

than the resolution in the segmentation. The reason is that the saddle regions in the image are very small at fine scale since the corners of the squares are sharp here. When the image is blurred the extension of the saddle regions grow and at the coarsest scale they have grown to a size given be the small octagons.

9.5 The selection of localization and detection scales

In practice it is well known that noise might be amplified to an undesirable degree when taking higher order derivatives. Blom [4] chapter 4 has studied this effect. The second momentum of the noise distribution can be used as a measure of noise influence on the derivatives [4]. We want to put a reasonable criteria for a lower bound on the scale when taking first order derivatives. One way is to demand that the first order derivatives should not be more influenced by noise than zeroth order. Using Blom's approximations for the second momentum for derivatives, one has:

$$Noise \quad momentum_{zeroth \quad order} = \frac{\epsilon^2 < N^2 >}{8\pi} \frac{1}{t} \tag{71}$$

$$Noise \quad momentum_{gradient} = \frac{\epsilon^2 < N^2 >}{8\pi} \frac{1}{4t^2} \tag{72}$$

The gradient expression is derived by Blom [4][p.95] via a generalization of the momentum to a function of several variables.

We want the noise momentum for the gradient to be less than the noise momentum for the zeroth order which gives us the following condition on the scale for first order derivatives $1 \leq 2\sigma^2$. Solving this gives a lower bound for σ: $0.71 \leq \sigma$. This result is necessarily based on a noise model. In this case the model is additive pixel–uncorrelated Gaussian noise with zero mean. Since no task specific knowledge is available, a noise model is out of reach as well, but the result can be taken as a rule of thumb: the structure at scales below 0.7 can be expected to be dominated by noise. The noise model used for derivation by Blom is a very common noise model, so in that sense the calculated lower bound will be applicable in a lot of applications.

10 The segmentation tools

Often only part of the segments in a segmentation is of interest therefore a tool for selection is needed. If knowledge of for instance approximate position or statistic on the intensity values are available, this knowledge can be used to automatize the selection. Since we have set out to make a general framework for segmentation we will let the knowledge come from a user. The segmentation tools shall work on interactive basis. The idea is that the user specifies interactively a localization scale and one or more detection scales. For each detection scale the corresponding partitioning of the domain is presented and the user can interactively select or de-select segments of interest. The result of the selection/de-selection actions is presented on the chosen localization scale.

10.1 Implementation issues

We will in this section briefly present how the segmentation tools have been implemented.

The needed image derivatives were calculated using a image processing package especially developed in the scale space theory framework. The package has been developed at the 3D Computer Vision Research Group, the University of Utrecht. The derivatives of the Gaussian are represented analytically in the Fourier domain. The filtering is performed with a multiplication in the Fourier domain and the result is found by inverse Fourier transformation. It is designed to provide optimal precision and minimum use of memory. The latter is important especially when making scale–space representation of three dimensional images where the result is 4D[16]. An equidistant quantization in the natural scale parameter τ, $\sigma = \sigma_0 exp(\tau), \tau \in (-\infty, \infty)$ is used to obtain scale invariance[19]. The image is made periodical to make Fourier transformation possible by mapping the image domain to a torus. This causes the well known wrap around effects. It implies that the information in a band along the image boundary can not be trusted. The width of the band equals the scale. In any implementation of a filtering the boundary effect is a problem. The common solution is to wrap around or mirror the image and afterwards keep in mind that the results close to the boundary are questionable. An alternative approach is to adapt the spatial kernels for instance by decreasing width near boundaries, chopping of the filter or deforming the filter close to boundary. This approach has been researched by Vincken [55]. His conclusion seems to be that the best result is obtained by ignoring the bad band. This implies for large scale that a substantial part of the image is not trustworthy.

The calculation of the image derivatives is a preprocessing step for the segmentation tools. The segmentation tools were implemented using the Advanced Visual Systems (AVS) as framework. AVS provides tools for basic visualization of three dimensional structures and a possibility for reading mouse activity.

The detection of watersheds, the linking of segments, the interactive tools for selecting and presentation of the selection have been implemented as AVS modules (in C). The segmentation tool has been implemented in the following way. For each scale a segmentation is calculated as the watersheds of the square of the gradient magnitude image. The basic segments of the image are the segments at the localization scale. A segmentation

[16] A typical size for a 4D scale–space in float precision is $40 \times 128 \times 128 \times 128 \times 4 = 336M$

on a coarse scale (detection scale) is tracked to the localization scale using the linking of segments. The user specifies two scale values (called localization and detection) and gets a partitioning of the domain. The segmentation can be specified using a number of different detection scales. The implementation of the linking scheme uses the fact that structure moves smoothly with scale, and therefore a spatial maximum correlation between segments at neighbouring scales can be used as the linking criteria.

One selection is specified by the two scale values, the localization and detections scale, and by a position in the image domain. An entire selection is the union of single selections. In practice, the module for interactive selection of positions and the module for interpreting the two scale values and the position are separate. This means that a user can be allowed to click on any kind of presentation of the domain, for instance the image itself, the segmentation alone, a superposition of the image and the segmentation or a domain with the previous selections. This makes the interactive selection flexible for the user.

The selection in three dimensions has been implemented in the following way: the user selects a direction for slicing and a slice number and a position in the slice as in 2D. The position and the slice number specifies a selection of a segment in 3D. The resulting segmentation can be presented in the many standard visualization forms like orthogonal slices or renderings. A direct selection scheme in 3D would have been preferable; unfortunately the AVS source code revealing the necessary details was not available. In 2D the result of a selection are clusters of pixels and in 3D clusters of voxels.

All rendering and segmentations presented in this thesis have been produced by exporting AVS[17] images. The surface plots and graphs have been produced in Mathematica[18] and Matlab[19]. The hand drawn illustrations have been made using Xfig[20].

10.2 Interactive segmentation

In this section results for the multi–scale segmentation method and the selection schemes are presented. The presented results are on a 2D slice of a 3D MR scan of a head, a 2D image of desk scene and a 3D volume from the abdomen of the visual human project. The 3D volume is made of 2D RGB images stacked on top of each other. The red channel is the one with best contrast and is therefore used here. The volume block is $128 \times 128 \times 112$ voxels á $1.76mm^3$ each. The diversity of modality and dimensions are used to show the generality of the method. The only "knowledge" build into the system is the explicit representation of scale and the gradient squared. The width of the Gaussian is increased by 10 procent between each slice of the scale–space images.

[17] AVS is a trademark of Advanced Visual Systems, Inc.
[18] Mathematica is a registered trademark of Wolfram Research, Inc.
[19] Matlab is a trademark of The MathWorks, Inc.
[20] A drawing program under X11

10.2.1 MR

In Figure 18 is shown a sequence of blurred versions of an image for $\sigma = 0.750, 4.59, 11.9$ and 30.9 pixels. The watersheds of squared of the gradient magnitude of the image are shown in middle column of Figure 18. In the rightmost column, the boundaries of the middle column have be tracked down to the localization scale using the linking.

In Figure 19 is the segmentation at localization scale $\sigma = 0.75$ for two detection scales $\sigma = 4.59$ and 11.9 superimposed on the image at scale $\sigma = 0.75$. The huge over–segmentation at fine scale is reduced tremendously when the scale of interest is raised above noise level. When selecting segments of interest the user could (and did) make the selection by clicking the mouse within the region of interest.

The segmentations from Figure 19 (a,b) have been used for selecting regions of interest by clicking the mouse on the relevant area. The result of the selection is shown in Figure 20 (a). This selection scheme is much faster than drawing the boundaries explicitly and the required user interaction is must simpler than selecting from the overwhelming number of segments at the localization scale. In Figure 20 (b) the same selection as (a) is presented but the selected regions are cropped out of the image.

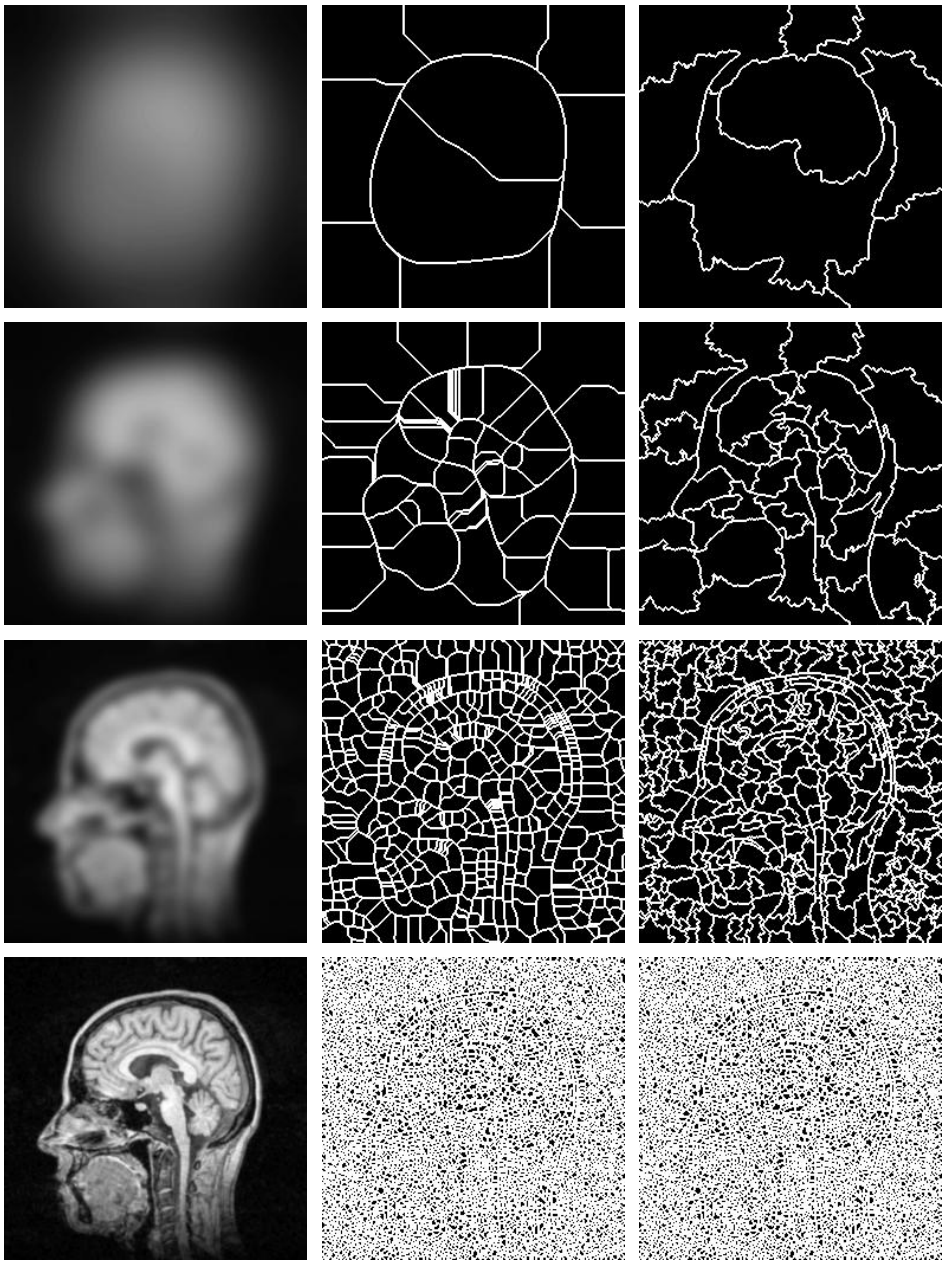
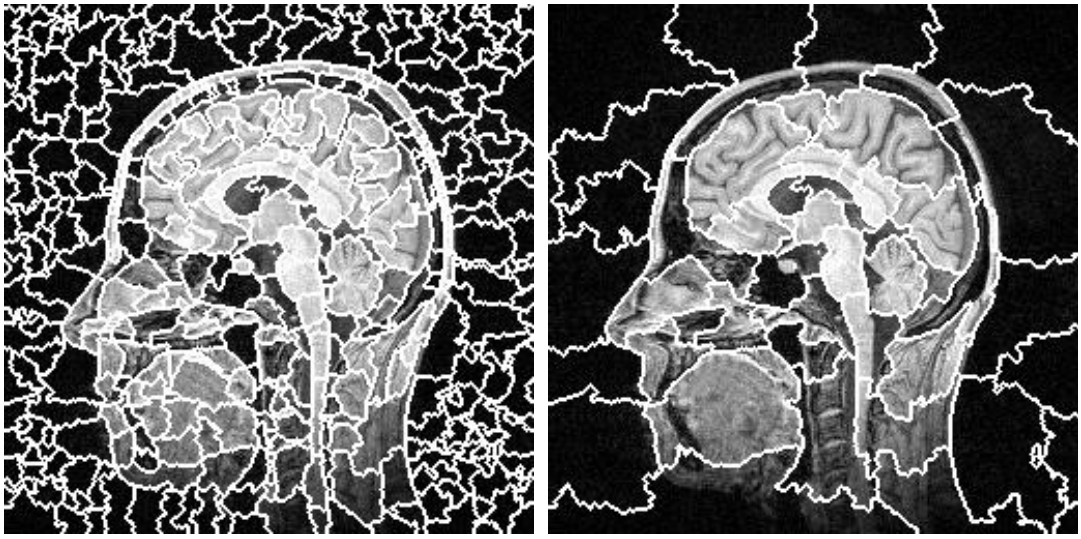
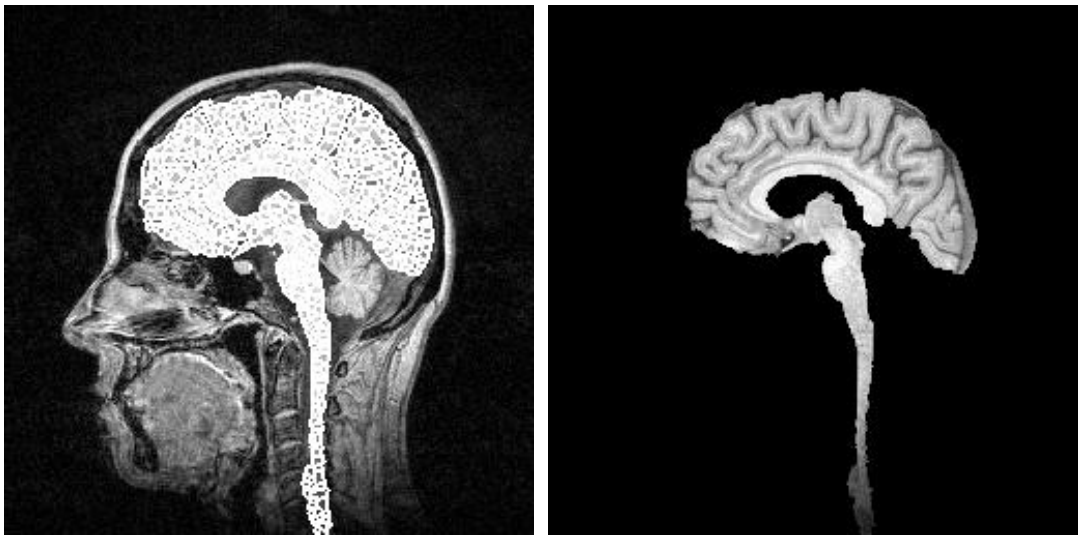


Figure 18: Each row corresponds to the same scale. The sigma values bottom–up ($\sigma = 0.750, 4.59, 11.9, 30.9$ pixels) The leftmost column is slices of the scale space, the original image is 256x256 pixels and is a slice of a 3D MR scan of a head. The middle column is the watershed segmentation of the left column. The right column is the segments at localization scale that caused the segments at the detection scale.



(a) (b)

Figure 19: The image at scale $\sigma = 0.75$ with the projected watershed segmentation from scale (a) $\sigma = 4.59$ and (b) $\sigma = 11.9$ superimposed.

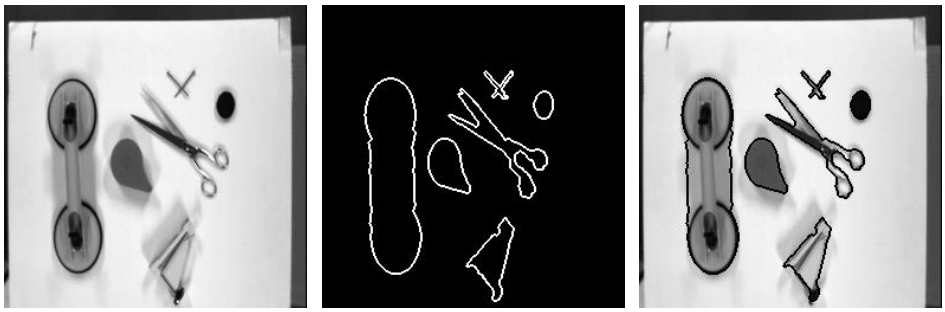


(a) (b)

Figure 20: Selected segments (a) shown at localization scale $\sigma = 0.75$ superimposed on the image. (b) is same selection as in (a)

10.2.2 Desk scene

In this section a interactive segmentation of desk scene. In Figure 21 is presented the image (a) of desk tools, in (b) the selected regions are displayed and in (c) the selected regions have been superimposed on the image to give a context.



(a) (b) (c)

Figure 21: The image to be segmented (a), the selected regions of interest (b), and the same regions superimposed on the image (c)

Three different detection scales (10.6, 4.0, 1.1pixels) were needed for this segmentation. In Figure 22 is presented the segmentations for the three detection scales projected down to the localization scale. In the Subfigures (a),(b) and (c) are presented the boundaries of the segmentations. The boundaries run between the cracks of the pixels and a boundary is presented by marking the two pixel on both side of the boundary crack. In Subfigures (e),(f) and (g) the segmentations are superimposed on the original images. The selection of the objects excepted for the scissor was performed almost entirely on the coarsest scale, a few refinements on the finer scale gave the desired selection. The appropriate scale for the scissor was middle scale segmentation.

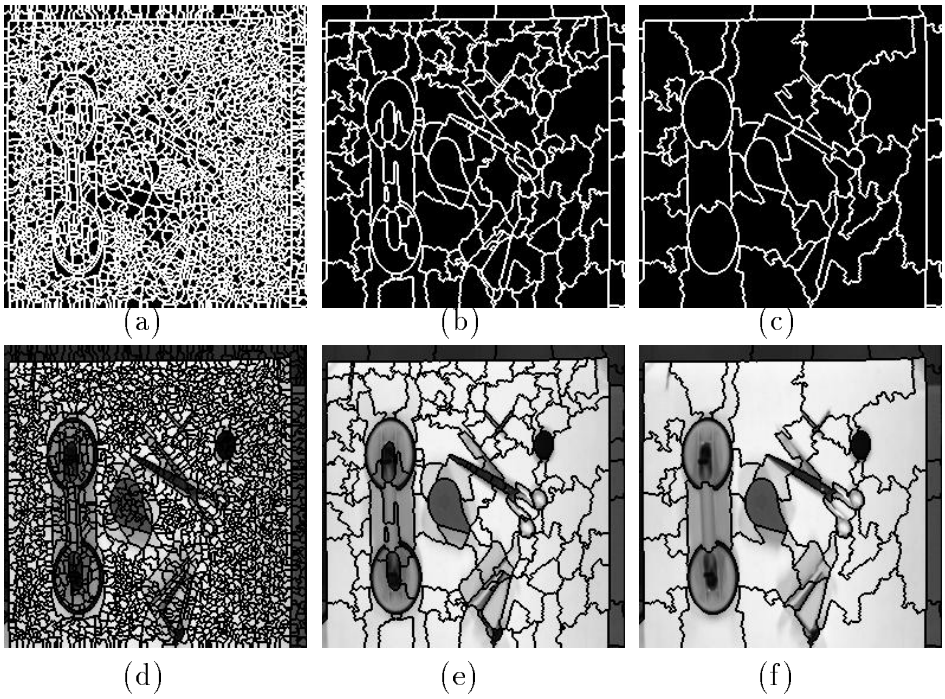
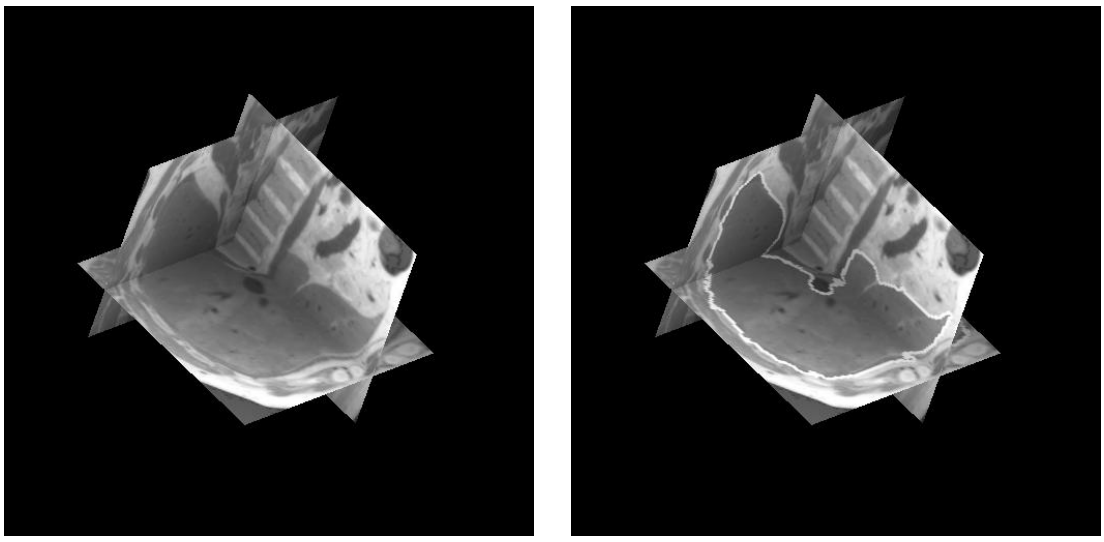


Figure 22: Segmentations of the image in Figure 21. The segmentations have been projected down to a localization scale. The detection scales are $1.1, 4.0$ and 10.6 pixels for (a), (b) respectively.

10.2.3 3D dimensional segmentation of the liver

In Figure 23 (a) is visualised a three dimensional volume from the abdomen. The cube has been cropped from the data set made in Visual Human Project. It is the red channel of the RGB images taken with a CCD camera. Each voxel corresponds to approximately $1.76mm^3$ and the cube is made of $128 \times 128 \times 112$ voxels. The contour of the liver has been superimposed and is presented in Figure 23 (b). Determining the volume corresponding to the liver is consider a hard problem for several reasons: the fine scale texture and color of the liver tissue varies a lot within the liver, the liver is penetrated by a complex network of blood vessel and texture and color of the surrounding tissue are similar to the liver. On top of this the liver consists of two parts, the major one in the right side of the human; the two parts are sparsely connected in the middle.



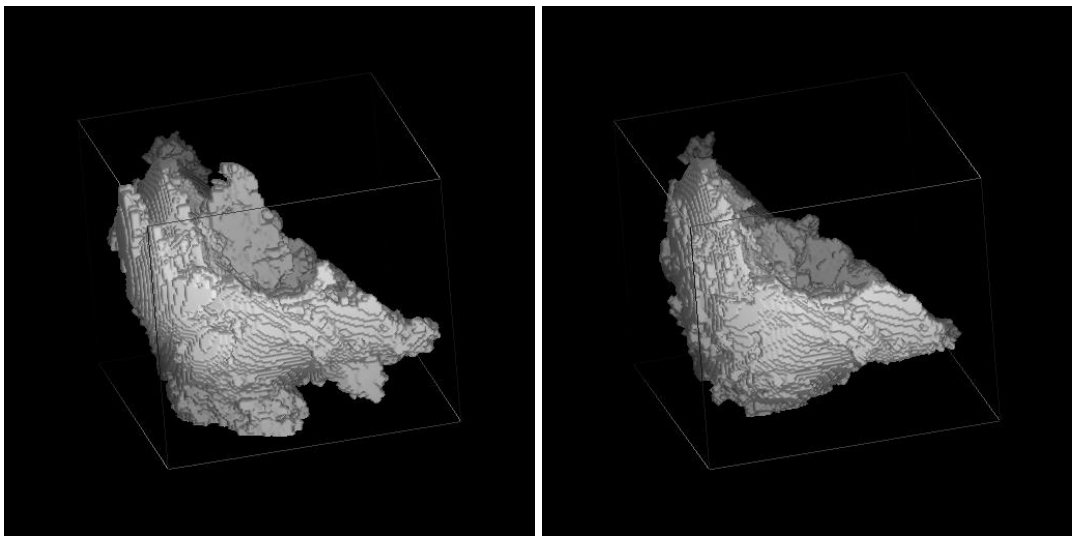
(a) (b)

Figure 23: (a) the original 3D image (b) selected segments superimposed on the original image.

Before presenting the results it is appropriate to point out that this segmentation scheme is not tuned to "liver segmentation" at all. At the finest considered scale (0.75 voxels) the cube is divided in about 29000 cells giving a mean volume for a cell at about 63 voxels (a $4 \times 4 \times 4$). Since the boundaries of these cells follows the gradient magnitude it is possible make a good segmentation merely by "drawing" in 3D using the cells as basic blocks. This drawing approach will not be taken since it basically is boundary drawing at the lowest scale of the image. Instead the user first selects structures at a coarse scale and thereafter refining the selection on lower scales. This latter technique is in the spirit of the idea of segmentation presented in this work.

To give an idea of the concrete selection process in 3D and the structure of the 4D Scale–Space image we have in Figure 24 rendered preliminary results in the interactive selection process. In Figure 24 (a) is a rendering of the result of selecting approximately 50 segments in 9 different slices selected in about 4 minutes at a detection scale of 5.05

pixels. In Figure 24 (b) is presented the rendered result of additionally selecting and de-selecting approximately 125 segments in 13 different slices selected in about 20 minutes at a detection scale of 2.14 pixels. Because the prototype for the 3D segmentation tool is not fast enough for true interactive selecting, the number of clicks with the mouse (the number segments) is probably the best measure of how efficient the interaction is. The turn–around time from a click to the presented result is approximately 10 seconds; way too long for anything but an experimental prototype. The top priority in the implementation was easy programming and ease when making changes. This order of priority and the fact that the AVS system was running on half the recommended amount of memory caused the lack of speed. A reasonable implementation with respect to speed should be possible also for three dimensional images.



(a) (b)

Figure 24: Renderings of coarse selection (a) and refined selection (b)

Even results as coarse as in Figure 24 can be used in tasks like picking regions of interests. In the particular case of the liver, the major task could be to find the network of blood vessels and a big step in solving this task is to reduce the volume of interest from the entire image to the volume roughly corresponding to the liver. In Figure 25 is the result after further refinement of the selections. The view in Figure 24 is closest to the view in Figure 25 (c).

In Figure 25 the resulting segmentation is presented as renderings of the bunch of voxels representing the liver volume. No attempt has been made to smooth the surface, consequently the renderings are angular in appearance but more true in the presentation of the result than a smoothed version. The views give a good impression of the shape of the liver. The views in Figure 25 (a) and (d) from the "back" clearly show hollows where the liver meets other internal organs. The interrelations between internal organs can consequently be established. The hollow for the major blood vessel connected to the liver can be seen in view (a) and (d) as a tube shaped groove in the liver tissue.

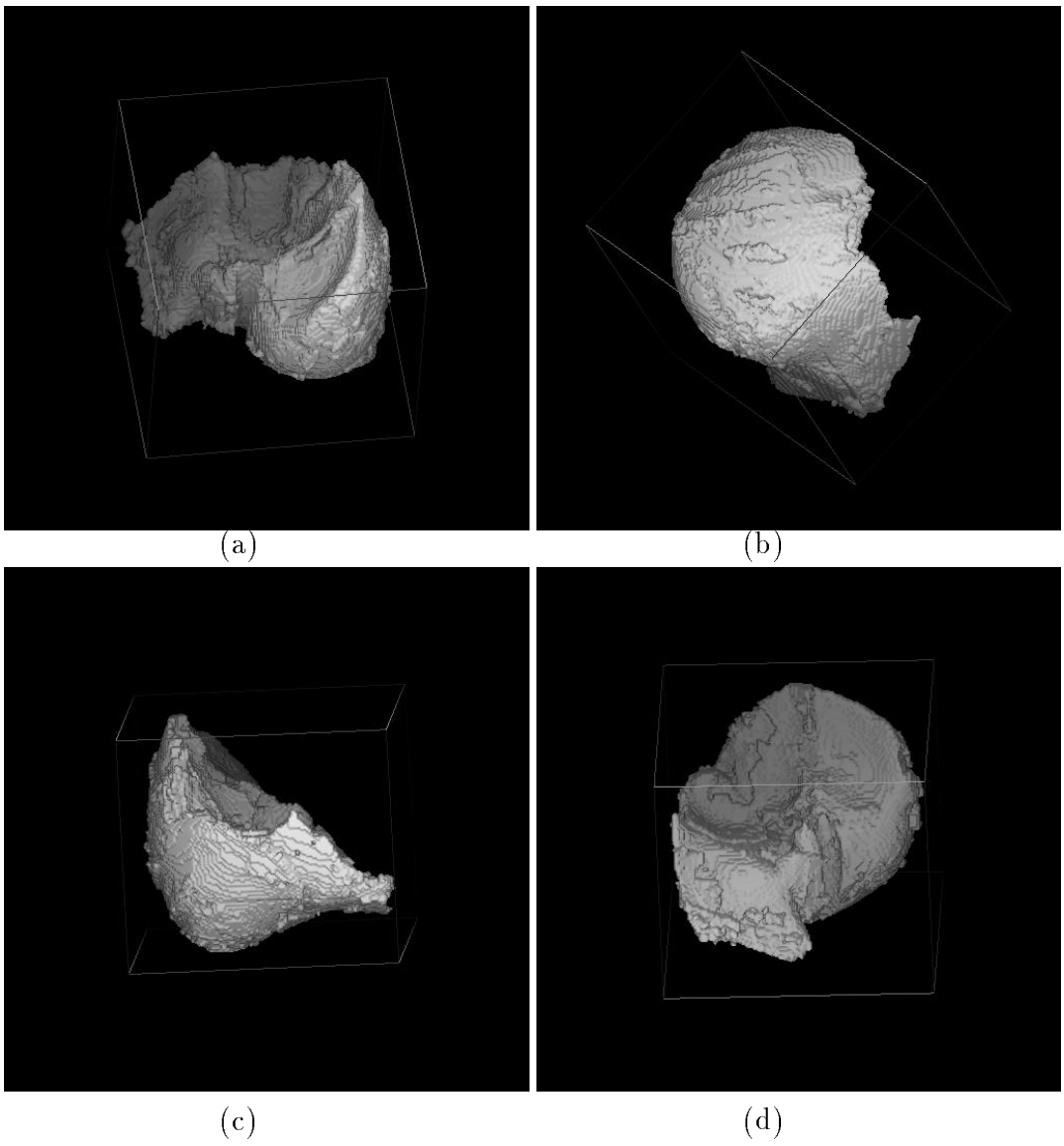


Figure 25: Four different views of the same segmentation of a liver. Positions are relative to a "standing person". Hence, front above means standing in front of the person look down at the liver. Sub-picture (a) is from the back above, (b) looking from below at the front from the right side, (c) is from the front, (d) a view from left side of the person. The sides of the cube are vertical and horizontal with respect to the person.

10.3 Summary of segmentation results

The segmentation at each scale is defined as the watersheds of the gradient squared on the particular scale. At fine scales the detected structure contains an undesirable high degree of details arising from noise in the acquisition and from not interesting low scale texture. Tracking the coarse scale structures of interest to a fine localization scale works well. The segmentations have only been evaluated visually, and this evaluation concludes that the segmentations correspond well with the visual perceived object boundaries. The scheme of accessing the image structures at coarse detection scales and combining the results at a fine localization scales works well. The user can select the structure of interest in a coarse to fine manner, and often only a few coarse scale structures correspond to the desired object of interest. Before using the watersheds of the gradient squared for instance on medical images the method should of course be compared to golden standards. It could be that the gradient information is not the right or sufficient measure of dissimilarity for the specific task.

11 Conclusion

A framework for multi–scale segmentation has been presented. A segmentation at a single scale is defined as the catchment basins for all local minima in a dissimilarity measure. It has been shown how such a segmentation can be embedded in a family of segmentations indexed by scale using the duality between minima and segments. The focus has been on one dissimilarity measure, the gradient squared. This measure is topological equivalent to the gradient magnitude. The gradient magnitude is a natural yard stick of dissimilarity in a differential geometry framework, and this measure has been reported in the literature to give good results. The problem of over–segmentation at fine scale is a typical problem for segmentation methods based on watersheds of the gradient magnitude. This work has shown that explicit probing the structures of interest at an appropriate scale greatly reduces the problem of over–segmentation.

The goal of building an uncommitted segmentation tool based on the front–end linear Gaussian scale–space has been accomplished. The selection mechanism of interactively picking objects at an appropriate scale and combining the results at localization scale provides a fast way of doing semi–automatic segmentation. The obtained results are promising in the sense that the produced segmentations correspond well with the perceived "blobs" in the images. This is true for both two and three dimensional data.

Using the powerful theorems from catastrophe theory it has been derived that the generic catastrophe events for the gradient squared are the fold catastrophe and the cusp catastrophe. This theoretical result has allowed well based considerations on reasonable linking schemes for segments across scale.

The notion of non commitment has been the general idea for the design of this segmentation tool. When no apriori knowledge is available the best way to proceed is to stay uncommitted. Applying this idea to the general segmentation frame work has shown promising results.

In the case of task specific knowledge it would be a plus to incorporate this knowledge into the system. The presented framework can be augumented with semantics in two ways: via the dissimilarity measure and via the diffusion scheme. The semantics of the segmentation can be changed using another measure of dissimilarity instead of the gradient magnitude. This is possible within the same general framework. The result of the singularity analysis will alter and consequently the linking scheme will change. The tools for analysis from catastrophe theory are not confined to the presented scheme but are broadly applicable and useful when studying qualitative structure of function families. The other possibility of imposing semantics on the system is by a different diffusion scheme. This would also require a new analysis to obtain the generic events for the segments, but is theoretically possible.

It is a general approach in linear diffusion to define and study features in terms of singularities of differentials. The region based linking schemes presented in this work can be applied to other differentials where maxima or minima were interesting to link. If the differential has the same generic catastrophe events as the gradient squared then the method is directly applicable, otherwise the relation between the segment events and the generic singularities events have to be analysed.

The force and the weakness of the catchment basins formulation for segmentation is

the need for the global topology. This property insures robustness in the detection of the structures and captures the global topology of the image. On the other hand, it implies that the borders are not locally detectable.

An interesting new aspect to the proposed multi–scale segmentation is the detection of possible interesting roots in the tree of linked segments. This would be a way to use the framework for an automatic segmentation. Root candidates could for instance be segments which are not involved in catastrophes over a range of scale. Such a property would imply some kind of stability, and from the very introduction of scale–space theory it has been proposed that stable structures over scale correspond to important structures. Finding such roots could also be interpreted as automatic selection of the appropriate scale for the present image structures.

Such an automatic detection would be closely connected to describing the entire image in terms of toppoints for the image. Describing images in terms of toppoints (as is possible in the one dimensional case) would be expected to be a robust way of describing and comparing image structures.

References

[1] A.Rosenfeld and M.Thurston. Edge and curve detection for visual scene analysis. *IEEE Trans. Computers*, 20(5):562–569, 1971.

[2] Nicholas Ayache. Medical computer vision, virtual reality and robotics. *Image and Vision Computing*, 13(4), May 1995.

[3] M-O Berger. Snake growing. In Oliver Faugeras, editor, *European Conference in Computer Vision*, pages 570–572. Springer–Verlag, 1990. Lecture Notes in Computer Vision.

[4] Johan Blom. *Topological and Geometrical Aspects of Image Structure.* PhD thesis, Rijksuniversiteit te Utrecht, 1992.

[5] J.Brian Burns, Allen R. Hanson, and Edward M. Riseman. Extracting straight lines. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(4):425–455, July 1986.

[6] Peter J. Burt. Fast filter transforms for image processing. *Computer Graphics, Image Processing*, 16:20–51, 1981.

[7] Peter J. Burt and Edward H.Adelson. The laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 31(4), April 1983.

[8] John Canny. A computiational approach to edge detection. *IEEE Tranactions on pattern analysis and Machine Intelligence*, 8(6):679–698, November 1986.

[9] Isaac Cohen, Laurent D. Cohen, and Nicholas Ayache. Using deformable surfaces to segment 3-d images and infer differential structures. In G.Sandini, editor, *European Conference in Computer Vision*, pages 648–652. Springer–Verlag, 1992. Lecture Notes in Computer Vision nr.588.

[10] Laurent D. Cohen and Isaac Cohen. Finite-element methods for active contour models and balloons for 2-d and 3-d images. *IEEE Tranactions on pattern analysis and Machine Intelligence*, 15(11):1131–1147, November 1993.

[11] J. L. Crowley. *A Representation for Visual Information.* PhD thesis, Carnegie Mellon University, Robotics Institute, Pittsburgh, Pennsylvania, 1981.

[12] James Damon. Local morse theory for solutions to the heat equation and gaussian blurring. *Journal of Differential Equations*, 115(2), January 1995.

[13] James Damon. Local morse theory for gaussian blurred functions. In Sporring et al. [51], chapter by Damon.

[14] Rachid Deriche. Using canny's criteria to derive a recursively implemented optimal edge detector. *International Journal of Computer Vision*, pages 167–187, 1987.

[15] James M Elder and Steven W Zucker. Scale space localization, ,blur, and contour–based image coding. In *Proceedings of Computer Vision and Pattern recognition*, pages 27–34, 1996.

[16] Harley Flanders. *Differential Forms with Applications to the Physical Sciences.* Dover Publications, Inc., New York, 1989. Published by Academic Press in 1963.

[17] Luc Florack. *The Syntactical Structure of Scalar Images.* PhD thesis, Universiteit Utrecht, 1993.

[18] Luc M J Florack. *A definition of images, Lecture notes.* not published yet, 1996. The lecture notes was a preliminary version of a book to appear in 1997.

[19] Luc M J Florack, Bart M ter Haar Romeny, Jan J Koenderink, and Max A viergever. Scale and the differential structure of images. *Image and Vision Computing*, 10(6):376–388, July/August 1992.

[20] J Fourier. *The Analytical Theory of Heat.* Dover Publications, New York, 1955.

[21] Donald Geman. Stochastic model for boundary detection. *Image and Vision Computing*, 5(2):61–65, May 1987.

[22] Robert Gilmore. *Catastrophe Theory for Scientists and Engineers.* Dover, 1981. ISBN 0-486-67539-4.

[23] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing.* Addison–Wesley, 1992. ISBN 0-201-50803-6.

[24] Lewis D. Griffin. *Descriptions of Image Structure.* PhD thesis, University of London, 1995.

[25] Lewis D Griffin and Alan C F Colchester. Superficial and deep structure in linear diffusion scale space: isophotes, critical points and separatices. *Image and Vision Computing*, 13(7):543–557, September 1995.

[26] Lewis D Griffin, Alan C F Colchester, and G P Robinson. Scale and segmentation of grey-level images using maximum gradient paths. *Image and Vision Computing*, 10(6):389–402, July/August 1992.

[27] R M Haralick. Digital step edges from zero crossings of second directional derivatives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(1), 1984.

[28] Peter Johansen. On the classification of toppoints in scale space. *Mathematical Imaging and Vision*, 4, 1994.

[29] Peter Johansen, Stig Skelboe, Klaus Grue, and Jens Damgaard Andersen. Representing signals by their toppoints in scale space. In *Proceedings of the International Conference on Image Analysis and Pattern Recognition*, pages 215–217, 1986.

[30] S. Kalitzin. Topological numbers and singularities in scalar images. scale space evolutio n properties. In Sporring et al. [51], chapter by Kalitzin.

[31] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. In Editor, editor, *First International Conference on Computer Vision*, pages 679–698, 1987. Also in Int. Jour. C.V. 1988.

[32] A. Klinger. Pattern and search statistics. In J.S.Rustagi, editor, *Optimizing Mehtods in Statistics*. Academic Press, New York, 1971.

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

[34] Jan J. Koenderink. *Solid Shape*. The MIT Press, 1990. ISBN 0-262-11139-X.

[35] Andre Koster. *Linking Models for Multiscale Image Segmentation*. PhD thesis, Universiteit Utrecht, 1995.

[36] L.D.Griffin, G.P.Robinson, and A.C.F. Colchester. Multi–scale hierarchical segmentation. In *Proceedings of British Machine Vision Conference*, pages 289–298, 1993.

[37] Yvan G. LeClerc. Constructing simple stable descriptions for image partitioning. *International Journal of Computer Vision*, 3:73–102, 1989.

[38] Lawrence M. Lifshitz and Stephen M. Pizer. A multiresolution hierachical approach to image segmentation based on intensity extrema. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(6):529–540, June 1990. Also in Xth IPMI 1988.

[39] Tony Lindeberg. Scale–space behaviour of local extrema and blobs. *Journal of Mathematical Imaging and Vision*, 1:65–99, 1992.

[40] Tony Lindeberg. *Scale–Space Theory in Computer Vision*. Kluwer Academic Publishers, 1994. ISBN 0-7923-9418-6.

[41] Tony Lindeberg and Bart ter Haar Romeny. Linear scale–space. In ter Haar Romeny [52], chapter 2. ISBN 0-7923-3087-0.

[42] Frederik Maes, Dirk Vandermeulen, Paul Suetens, and Guy Marchal. Computer–aided interactive object delineation using an intelligent paintbrush technique. In Nicholas Ayache, editor, *First International Conference on Computer Vision, Virtual Reality and Robotics in Medicine*, pages 77–83. Springer–Verlag, 1995. Lecture Notes in Computer Vision 905.

[43] D. Marr. *Vision*. W.H.Freeman & Co., 1982.

[44] Fernand Meyer. Topographic distance and watershed lines. *Signal Processing*, 38(6):99–112, July 1994.

[45] Oliver Monga, Rachid Deriche, Gregoire Malandain, and Jean Pierre Cocquerez. Recursive filtering and edge tracking: two primary tools for 3d edge detection. *Image and Vision Computing*, 9(4), August 1991.

[46] D Mumford and J Shah. Boundary detection by minimizing functionals. In *IEEE Conf. on Computer Vision and Pattern Recognition*, San Francisco, 1985.

[47] Laurent Najman and Michel Schmitt. Watershed of a continuos function. *Signal Processing*, 38(6):99–112, July 1994.

[48] Pietro Perona, Takahiro Shiota, and Jitendra Malik. Anisotropic diffusion. In ter Haar Romeny [52], chapter 3. ISBN 0-7923-3087-0.

[49] T Poston and I N Stewart. *Taylor Expansions and Catastrophes.* Pitman, 1976. ISBN 0-273-009-64-8.

[50] P. T. Saunders. *An introduction to Catastrophe Theory.* Cambridge University Press, 1980.

[51] Jon Sporring, Mads Nielsen, Luc Florack, and Peter Johansen, editors. *Gaussian Scale–Space.* Kluwer Academic Publishers, 1996.

[52] Bart M. ter Haar Romeny, editor. *Geometry-Driven Diffusion in Computer Vision*, volume 1. Kluwer Academic Publishers, 1994. ISBN 0-7923-3087-0.

[53] Vincent Torre and Tomaso A. Poggio. On edge detection. *IEEE Pattern Analysis and Machine Intelligence*, 8(2), March 1986.

[54] Luc Vincent and Pierre Soille. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *Pattern Analysis and Machine Intelligence*, 13(6), June 1991.

[55] Koen Vincken. *Probabilistic Multiscale Image Segmentation by the Hyperstack.* PhD thesis, Universiteit Utrecht, 1995.

[56] Andrew P. Witkin. Scale space filtering. In *Proc. of International Joint Conference on Artificial Intelligence (IJCAI)*, Karlsruhe, Germany, 1983.

A Notation

A list of commonly used symbols and their meaning.

C : The set of continuous functions.

C^n : The set of n differentiable functions.

d : The dimension of a space, usually the spatial image domain.

D : The dimension of the space of Taylor coefficients.

$Det\ A$: The determinant of the matrix A.

δ_{ij} : The Kronecker delta, ones one the diagonal, zeros on off diagonal.

ϵ : A small real number.

$\epsilon_{i_1 \ldots i_d}$: The anti-symmetric Levi-Civita tensor.

f : A function or a family of functions.

g : A function or a family of functions.

G : The Gaussian kernel.

H_f : The Hessian matrix for the function f.

I : The original (raw) image/ the measurements.

L : The scale–space image.

$L_{i_1 \ldots i_d}$: A derivative of the scale–space image.

$L(x;\sigma)$: The scale–space image evaluated at $(x;\sigma)$.

$N_\epsilon(a)$: The neighbourhood of a.

n : The dimension of the space of state variables.

r : The dimension of the space of control variables.

$\mathbb{R}$: The set of reals.

σ : The width of the Gaussian/ the aperture.

t : The scale parameter.

τ : The natural scale parameter.

x : A point in the Euclidean space $\mathbb{R}^d$.

x_i : The ith element of x.

x_0 : A particular point in the Euclidean space $\mathbb{R}^d$.

∇f : The gradient of the function f.

$\triangle f$: The Laplace operator on f.

$\emptyset$: The empty set.

$\hat{f}$: The Fourier transformed of the function f.