International Journal of Computer Vision, 11:3, 283–318 (1993)

Detecting Salient Blob-Like Image Structures and Their Scales with a Scale-Space Primal Sketch: A Method for Focus-of-Attention

TONY LINDEBERG
Computational Vision and Active Perception Laboratory (CVAP), Department of Numerical Analysis and Computing Science, Royal Institute of Technology, S-100 44 Stockholm, Sweden

Received July 15, 1991; Revised April 27, 1993.

Abstract
This article presents: (i) a multiscale representation of grey-level shape called the scale-space primal sketch, which makes explicit both features in scale-space and the relations between structures at different scales, (ii) a methodology for extracting significant blob-like image structures from this representation, and (iii) applications to edge detection, histogram analysis, and junction classification demonstrating how the proposed method can be used for guiding later-stage visual processes.

The representation gives a qualitative description of image structure, which allows for detection of stable scales and associated regions of interest in a solely bottom-up data-driven way. In other words, it generates coarse segmentation cues, and can hence be seen as preceding further processing, which can then be properly tuned. It is argued that once such information is available, many other processing tasks can become much simpler. Experiments on real imagery demonstrate that the proposed theory gives intuitive results.

1 Introduction

Scale-space representation introduced by Witkin [70] and Koenderink [33] provides a well-founded framework for dealing with image structures, which naturally occur at different scales. According to the theory of this representation, one can from a given signal generate a family of derived signals by successively removing features when moving from fine to coarse scales. In contrast to other multiscale or multiresolution representations, scale-space is based on a precise mathematical definition of causality, or scale invariance (Florack et al. [21]), and the behavior of structure as scale changes can be analytically described. However, the information in the scale-space embedding is only *implicit* in the grey-level values. The smoothed images in the raw scale-space representation contain no explicit information about the features in them or the relations between features at different levels of scale.

The goal of this presentation is to present such an explicit representation, called the *scale-space primal sketch*, and to demonstrate that it enables extraction of significant image structures in such a way that the output can be used for guiding later stage processes in early vision.

The treatment will be concerned with grey-level images, and the chosen features will be blobs, that is, bright regions on dark backgrounds or vice versa. However, the methodology applies to any bounded function and is therefore useful in many tasks occurring in computer vision, such as the study of level curves and spatial derivatives in general, depth maps, etc., and also, histograms, point clustering and grouping, in one or in several dimensions.

1.1 Scale and Segmentation

Many methods in computer vision and image analysis implicitly assume that the problems of scale detection and initial segmentation have already been solved. One example is edge detection, where the selection of step size in gradient computations leads to a well-known trade-off problem. A small step size leads to a small

truncation error in the discrete approximation, but the noise sensitivity might be severe. Conversely, a large step size will in general reduce the noise sensitivity, but at the cost of an increased truncation error. In the worst case, a slope of interest can be missed and meaningless results be obtained, if the difference quotient approximating the gradient is formed over a wider distance than the object considered in the image. Although here we shall mainly be concerned with static images, the same kind of problem arises when dealing with image sequences. Similarly, models based on spatial derivatives ultimately rely on the computation of difference approximations from measured data. This implies that they always fall back to the basic scale problem that objects in the world and features in images only exist as meaningful entities only over limited ranges of scale.

A commonly used technique to improve the results obtained in computer vision and other fields related to numerical analysis is by preprocessing the input data with some amount of smoothing and/or careful tuning of the operator size or some other parameters. In some situations the output may depend strongly on these processing steps. In certain algorithms these so-called tuning parameters can be estimated; in other cases they are set manually. A robust image analysis method intended to work in an autonomous robot situation must, however, be able to make such decisions automatically. How should this be done? I contend that these problems are in many situations nothing but disguised scale problems.

Also, in order to apply a refined mathematical model like a differential equation or some kind of deformable template, it is necessary to have some kind of qualitative initial information, that is, a domain where the differential equation is (assumed to be) valid or an initial region for application of the raw deformable template. Examples can be obtained from many "shape-from-X" methods, which in general assume that the underlying assumptions are valid in the image domain in which the method is applied. A commonly used assumption is that of smoothness implying that the region in the image, to which the model is applied, must correspond to, say, one physical object or one facet of a surface, etc. How should such regions be selected *automatically*? Many methods cannot be used unless this nontrivial part of the problem has been solved.

How to detect appropriate scales and regions of interest when there is no a priori information available. In other words, how to determine the scale of an object and where to search for it before knowing what kind of object we are studying and before knowing where it is located. Clearly, this problem is intractable if treated as a pure mathematical problem. Nevertheless, it arises implicitly in many kinds of processes, for example, dealing with texture, contour, etc., and seems to boil down to an intractable chicken-or-the-egg problem. The solution of the pre-attentive recognition problem seems to require the solution of the scale and region problems and vice versa.

The goal of this presentation is to demonstrate that such preattentive groupings can be performed in a bottom-up manner, and that it is possible to generate initial hypotheses about blob-like regions of interest as well as to give coarse indications about the scales at which the regions manifest themselves. The basic tools for the analysis will be scale-space theory, and a heuristic principle stating that blob-like structures that are stable in scale-space are likely candidates to correspond to significant structures in the image. Concerning scale selection, scale levels will be selected that correspond to local maxima over scales of a measure of blob response strength. (Precise definitions of these notions will be given later.) It will be argued that once such scale information is available, and once regions of interest have been extracted, later-stage processing tasks can be simplified. This claim is supported by experiments on edge detection and classification based on local features.

1.2 Detection of Image Structure

The main features that arise in the (zero-order) scale-space representation of an image are smooth regions that are brighter or darker than the background and stand out from their surroundings. These will be termed blobs (a precise definition will be given later). The purpose of the suggested representation is to make these blobs, as well as their relations across scales, explicit. The idea is also that the representation should reflect the intrinsic shape of the grey-level landscape—it should not be an effect of some externally chosen criteria or tuning parameters. The theory should in a bottom-up fashion allow for a data-driven detection of significant structures, their relations, and the scales at which they occur. It will, indeed, be experimentally shown that the proposed representation gives intuitively reasonable results, in which salient structures are (coarsely) segmented out. Hence, this representation can serve as a

guide to subsequent, more finely tuned processing, which requires knowledge about the scales at which structures occur. In this respect it can serve as a mechanism for focus-of-attention.

Since the representation tries to capture important image structures with a small set of primitives, it bears some similarity to the *primal sketch* proposed by Marr [53, 54], although fewer primitives are used. The central issue here, however, is to represent explicitly the scales at which different events occur. In this respect the work addresses problems similar to those studied by Bischof and Caelli [8]. They tried to parse scale-space by defining a measure of stability. Their work, however, was focused on zero-crossings of the Laplacian. Moreover, they overlooked the fact that the scale parameter must be properly treated when measuring significance or stability. Here, the behavior of structures over scale will be analyzed in order to give the basis of such measurements. Of course, several other representations of the grey-level landscape have been proposed without relying on scale-space theory. Let us also note that Lifshitz and Pizer [40] have studied the behavior of local extrema in scale-space. However, we shall defer discussing relations to other work until the suggested methodology has been described.

The idea of scale-space representation of images, suggested by Witkin [70], has, in particular, been developed by Koenderink and van Doorn [33, 34, 36], Babaud et al. [3], Yuille and Poggio [71], Hummel [28], Lindeberg [41, 47], and Florack et al. [21]. This work is intended to serve as a complement addressing computational aspects, and adding means of *making significant structures and scales explicit.* The main idea of the approach is to *link* features at different scales in scale-space into higher-order objects, called scale-space blobs, and to extract significant image features based on the appearance and stability of these objects over scales.

As a guide to the reader, it should be remarked that certain subsets of this material have been presented in other articles (see the references). The aim with this presentation is to provide a coherent overview, including precise definitions and applications that have not been presented elsewhere. An extensive treatment with algorithmic details can be found in the author's thesis (Lindeberg [43]).

2 The Scale-Space Primal Sketch

The scale-space representation of a signal is an embedding of the original data into a derived one-parameter family of successively smoothed signals, intended to represent the original data at multiple scales. Given a two-dimensional continuous signal $f : \mathbb{R}^2 \to \mathbb{R}$, the scale-space $L : \mathbb{R}^2 \times \mathbb{R}_+ \to \mathbb{R}$ is defined as the solution to the diffusion equation

$$\partial_t L = \frac{1}{2} \nabla^2 L = \frac{1}{2} (\partial_{x_1 x_1} + \partial_{x_2 x_2}) L \qquad (1)$$

with initial condition $L(\cdot; 0) = f$, or equivalently by convolution with the Gaussian kernel $g : \mathbb{R}^2 \times \mathbb{R}_+ \backslash \{0\} \to \mathbb{R}$

$$L(\cdot; t) = g(\cdot; t) * f, \qquad (2)$$

where

$$g(x; t) = \frac{1}{2\pi t} e^{-(x_1^2 + x_2^2)/(2t)} \qquad (3)$$

and $x = (x_1, x_2) \in \mathbb{R}^2$. The parameter $t \in \mathbb{R}_+$ is denoted scale parameter, and corresponds to the square of the standard deviation of the Guassian kernel $t = \sigma^2$.

From experiments one can (visually and subjectively) observe that the main features that arise in this scale-space representation seem to be blob-like, that is, they are regions either brighter or darker than the background (see figure 8). Especially, such regions that appear to stand out from the surroundings in the original image seem to be further enhanced by scale-space smoothing. The suggested scale-space primal sketch focuses on this aspect of image structure. The purpose is to build a representation for making such information in scale-space explicit. Therefore, there is a need to formalize what should be meant by a "blob."

2.1 Grey-Level Blob

What properties should be required from a blob definition? It is clear that a blob should be a region associated with (at least) one local extremum. However, it is also essential to define the spatial extent of the region around the blob, and to associate a significance measure with it. Ehrich and Lai [20] considered the extent problem. They allowed peaks to extend to valleys, a definition that will give nonintuitive results, for example, for small peaks on large slopes. Koenderink and van Doorn [33] briefly touched upon the problem with reference to work by Maxwell [55] concerning level curves and critical points. The definition proposed here is related to those arguments.

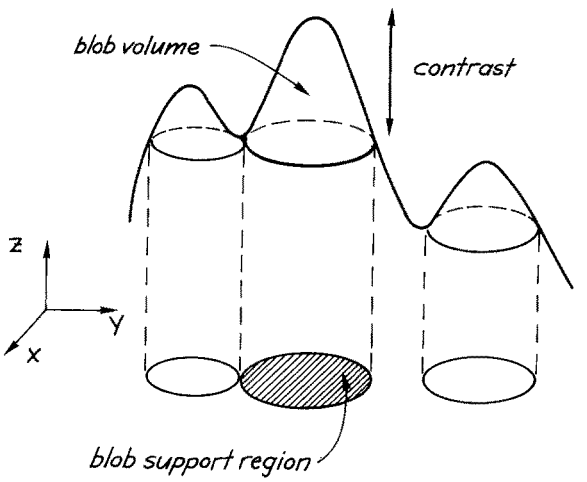

Fig. 1. Illustration of the grey-level blob definition for a two-dimensional signal, with some descriptive quantities of a grey-level blob; volume, area, and contrast. This figure shows bright blobs on a dark background. Generically, a grey-level blob is given by a pair consisting of one extremum and one saddle point, denoted delimiting saddle point.

The blob definition on which this work is based should be evident from figure 1. The basic idea is to let the blob extend "until it would merge with another blob." To illustrate this notion, consider a grey-level image at a fixed level of scale, and study the case with bright blobs on a dark background. Imagine the image function as a flooded grey-level landscape. If the water level sinks gradually, peaks will appear. At some instances two different peaks become connected. The corresponding elevation levels or grey-levels are called the base-levels of the blobs, and are used for delimiting the spatial extent of the blobs. The support region of the blob is defined to consist of those points that have a grey-level exceeding the base-level and can be reached from the local maximum point without descending below the base-level of the blob.

Hence, a bright blob will grow and include points having lower grey-levels until it would meet with another blob. In this sense the blob definition can be regarded as conservative, since no attempt is made to include points in other directions. From this construction, the *grey-level blob* is defined as the three-dimensional volume delimited by the grey-level surface and the base-level. The three-dimensional grey-level blob volume constitutes a combined measure of the contrast and the spatial extent (area) of the blob.

2.1.1 Grey-Level Blob Definition. A precise mathematical definition of the grey-level blob concept can be stated as follows: Consider again the case with bright blobs on dark background, and assume a continuous nondegenerate[1] grey-level function $f : \mathbb{R}^2 \to \mathbb{R}$ at a fixed level of scale. Consider a local maximum $A \in \mathbb{R}^2$. For any grey-level $z < f(A)$ let

$$X_z^{(A)} = \Big\{\text{the connected component of } [(x, \zeta) \in \mathbb{R}^2 \times \mathbb{R} : z \leq \zeta \leq f(x)] \text{ which contains } \big(A, f(A)\big)\Big\} \tag{4}$$

and define the sets $G_z^{(A)}$ and $H_z^{(A)}$ as follows: A point $(B, \zeta_0) \in X_z^{(A)}$ belongs to $G_z^{(A)}$ ($H_z^{(A)}$) if and only if there exists a path $p_{(A,f(A)),(B,\zeta_0)}$ from $(A, f(A))$ to (B, ζ_0) such that (i) every point on the path belongs to $X_z^{(A)}$, and (ii) the derivative of ζ along this path $\zeta'|_{p_{(A,f(A)),(B,\zeta_0)}} < 0$ ($\zeta'|_{p_{(A,f(A)),(B,\zeta_0)}} \leq 0$). The base-level of the blob $z_{\text{base}}(A)$ is then defined as the maximum value of z such that

$$z_{\text{base}}(A) = \max_{z<f(A)} z : \overline{G_z^{(A)}} \neq H_z^{(A)}, \tag{5}$$

where the notation $\bar{C}$ stands for the closure of a set C. ($z_{\text{base}}(A)$ is the grey-level value of the *delimiting saddle point* $S = S_{\text{delimit}}(A)$ associated with A). The grey-level blob associated with the local maximum A is the set of points

$$G_{\text{blob}}(A) = \overline{G_{z_{\text{base}}(A)}^{(A)}} \tag{6}$$

with the (three-dimensional) *grey-level blob volume*

$$G_{\text{vol}}(A) = \int_{(x,\, z) \in G_{\text{blob}}(A)} dx\, dz. \tag{7}$$

The projection of this region onto the spatial plane is called the *support region*,

$$D_{\text{support}}(A) = \{x \in \mathbb{R}^2 : (x, \zeta) \in G_{\text{blob}}(A) \text{ for some } \zeta\}, \tag{8}$$

and the difference in grey-level between the extremum point and the base-level gives the *blob contrast*

$$C_{\text{blob}}(A) = f(A) - z_{\text{base}}(A). \tag{9}$$

It is worth stressing that the grey-level blob is treated as an object with extent both in space and grey-level. The definition is expressed in terms of two-dimensional continuous signals, but can be extended to arbitrary dimensions as well as to discrete grids, if the paths are given by a suitable connectivity concept (e.g., eight-connectivity for a two-dimensional square grid), and

the derivative condition $f|'_{p_{A,B}} < 0$ is replaced by a difference condition $f(x^{(k+1)}) - f(x^{(k)}) < 0$ along the path $\{x^{(k)}\}$. Local minima can be treated analogously, and every local minimum point gives rise to a dark blob on a bright background.

2.1.2 Properties. It can be easily verified that a blob will be connected. Moreover, the base-level of a bright blob is in one dimension attained at a minimum point, in two dimensions at a saddle point. Consequently, the blobs are directly determined from topological properties of the grey-level landscape, namely the first-order singularities.

These blobs are not purely local features, as are extrema, but regional. An inherent property of the stated definition is that it leads to a competition between parts; the presence of another nearby blob might neutralize a blob or reduce its size. In other words, features manifest themselves only relative to the background. These aspects reflect important principles of the approach.

Note that this definition leads to separate systems for bright and dark blobs. This implies that some points may be left unclassified. Consequently, the given definition will—in contrast to, for example, the sign of the Laplacian of the Gaussian—only attempt to make a partial (and hopefully safer) classification of the grey-level landscape.

In one dimension the bright and dark blobs of a signal will be closely related, since a minimum point which delimits the extent of a bright blob will also constitute the seed of a dark blob. In two dimensions the situation is slightly different, since a saddle point that delimits the extent of a bright blob will not delimit the extent of any dark blob, unless the signal is degenerate. Therefore, in two dimensions, a point in a blob will in general belong to either a dark blob or a bright blob. In certain types of situations, however, it may indeed happen that some points are classified as belonging to both a dark blob and a bright blob, see figure 2 for an example. If for some reason this type of phenomenon is not desired, then it can be easily prevented by modifying the blob definition slightly so that the blob is allowed to "delimit its own extent."

2.1.3 Grey-Level Blob Tree. If the imaginary water level used for constructing grey-level blobs in figure 1 is allowed to decrease below the base-level of a blob, then the grey-level blob will merge with the adjacent region sharing the same saddle point. By considering all such events under variations of the water level, a tree-like structure can be defined with successive inclusion relations. Every arc corresponds to a range in grey-level where the topology is locally the same, and the grey-level blobs constitute the leaves.

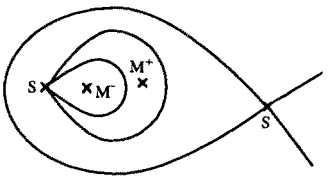

Fig. 2. Example with a dark blob contained in a bright blob. This phenomenon can be avoided if the blob definition is modified such that a blob is allowed to delimit its own extent in such situations. Then, it will be guaranteed that no point belongs to both a dark and a bright blob. (M^+ = maximum point, M^- minimum point, S = saddle point).

This representation, termed *grey-level blob tree*, has a qualitative similarity with the relational tree studied by Ehrich and Lai [20]. Simple self-explanatory examples demonstrating its construction are given in figure 3. Similarly to a leaf, every arc in the tree can be associated with a three-dimensional volume, as well as a support region and an area measure. The formal procedure for defining such relations is by treating every encountered delimiting saddle point as the seed of a new region, and then proceeding with the successive construction of grey-level blobs and arcs with decreasing water level. In this approach, every delimiting saddle point representing two merging regions is treated in the same way as an ordinary maximum, once a merge between the two regions at the saddle point has been registered.

By simultaneously considering the bright and dark grey-level blob trees of a signal, it is possible to express formal relations between blobs of reverse polarity. The saddle points and the level curves through those constitute the links, since these are the only entities occurring in both systems.

Finally, it should be pointed out that what has been defined here is a grey-level blob (and a corresponding tree) at one level of scale. When such objects are linked across scales, they result in scale-space blobs (and corresponding trees), which will be described in section 2.3.

2.2 Motivation for a Multiscale Hierarchy

It is easy to realize that the concept of a grey-level blob at a single level of scale is not powerful enough for stable extraction of image structures. It leads to an

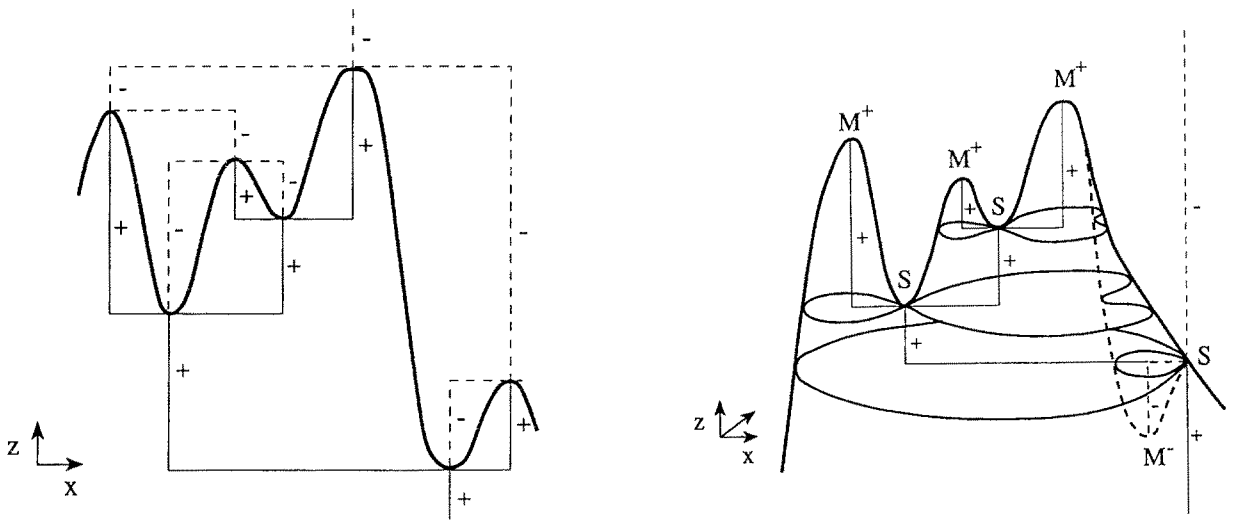

Fig. 3. Examples of grey-level blob trees: (left) for a one-dimensional signal, and (right) for a two-dimensional signal. The right figure shows a mountain-like grey-level landscape with three main peaks (marked by M^+), and one hole (marked by M^-). Arcs originating from bright blobs are drawn with filled lines (marked by '+'), and arcs corresponding to dark blobs are drawn by dashed lines (marked by '−').

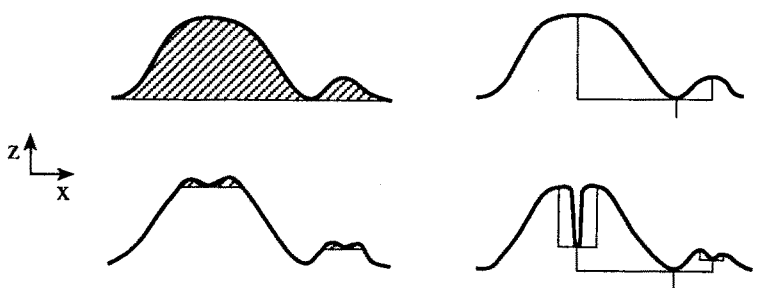

Fig. 4. (Left) A high-contrast large peak with two superimposed low-contrast fine-scale peaks will not be detected as a grey-level blob if the signal is considered at one scale only. (Right) A single noise spike can also substantially affect the relational tree.

extreme degree of noise sensitivity, since two closely situated local extrema will neutralize each other. This means that a large peak distorted by a few superimposed local extrema of low amplitude will not be detected as one unit; only the fine-scale blobs will be found (see figure 4).

Also the grey-level blob tree (and the relational tree) will be noise sensitive when considered at a single level of scale, since the hierarchical relations between different blobs are determined directly by the grey-levels in the valleys of the original signal. For example, thin elongated structures superimposed onto the data may completely change the topological relations (see figure 4).

In order to obtain more stable descriptors, it is natural to consider the behavior of the grey-level blobs and the grey-level blob tree in scale-space. Since no a priori information can be expected about what scales are relevant, the only reasonable approach is to consider *all* scales simultaneously.

2.3 Scale-Space Blobs

Given a grey-level blob existing at some level of scale, there will in general be a corresponding blob at both a slightly finer scale and a slightly coarser scale. Linking such grey-level blobs across scales gives four-dimensional objects, called *scale-space blobs* (see figure 5). (A formal definition of how the linking is performed is given in appendix A.1.)

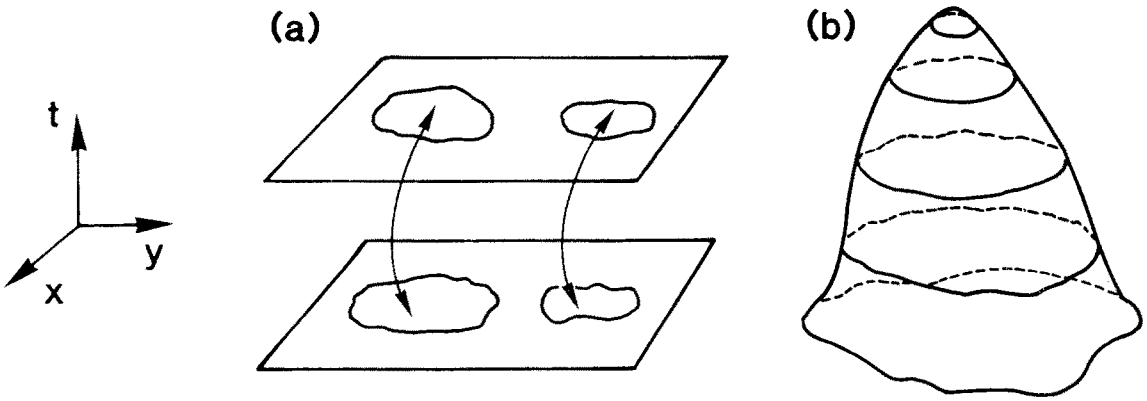

Fig. 5. (a) Linking similar grey-level at adjacent levels of scale gives (b) scale-space blobs, that are objects with extent both in space, grey-level, and scale. (In this figure the grey-level coordinate has been omitted. The slices illustrate the support regions of the grey-level blobs.)

At some level of scale it might be impossible to accomplish a plain link between a grey-level blob at that scale and a corresponding blob at a slightly coarser or finer scale. A *blob event* has ocurred affecting the connectivity of the blobs. According to a classification in appendix A.2, there are four possible types of blob events with increasing scale:

— *annihilation*: one blob disappears,
— *merge*: two blobs merge into one,
— *split*: one blob splits into two,
— *creation*: one new blob appears.

In summary, the classification of blob events means that each blob event corresponds to an annihilation or a creation of a pair consisting of one saddle point and one extremum point. For example, the difference between a blob annihilation and a blob merge, is that in the first case, the delimiting saddle point of the blob is contained in only one grey-level blob, while in the second case, it is part of two different grey-level blobs (see figure 6).

The scale levels where these singularities take place delimit the extent of the scale-space blobs in the scale direction. Consequently, every scale-space blob will be associated with a minimum scale, denoted appearance scale t_A, and a maximum scale, denoted disappearance scale t_D. The difference between the disappearance scale and the appearance scale gives the *scale-space lifetime* of the blob.[2]

In merge and split situations the grey-level blobs existing before the bifurcation are regarded as belonging to different scale-space blobs than the grey-level blobs existing after the bifurcation.

In special configurations it may happen that a blob without a hole forms a torus, or that a torus fills in its hole. These events are also stable in the sense that a small disturbance of the original signal will not affect

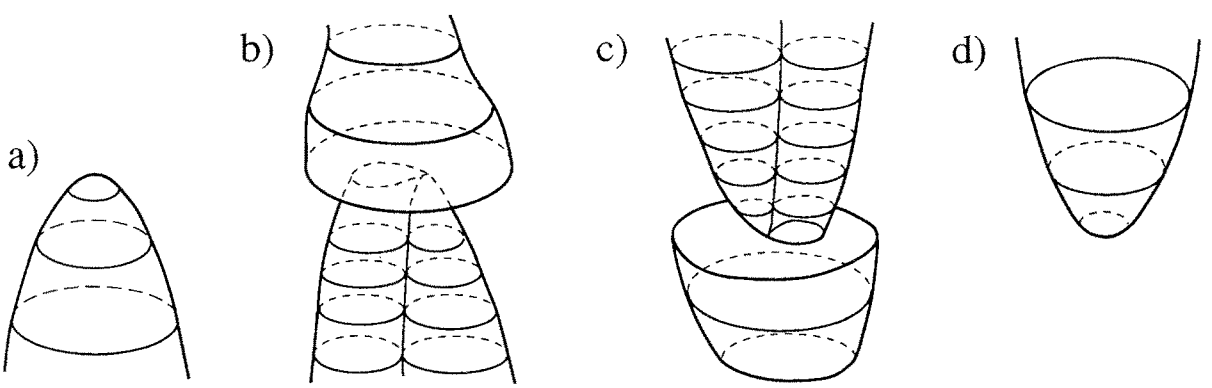

Fig. 6. Generic blob events in scale-space: (a) annihilation, (b) merge, (c) split, (d) creation.

the qualitative behavior. Here, such events will be considered not to affect the scale-space blobs; the grey-level blobs existing before such an event will be regarded as belonging to the same scale-space blob as the grey-level blobs existing after.

2.4 Scale-Space Blob Tree

Similar considerations can be applied to the evolution properties over scales of the grey-level blob tree. This gives a scale-space blob tree. Interestingly, the only bifurcations that can occur in scale-space are those that affect the leaves, that is, the grey-level blobs and the scale-space blobs. This is a direct consequence of the fact that both an extremum point and a saddle point must be involved in a bifurcation (appendix A.2), and every extremum point corresponds to a unique grey-level blob, and hence a unique scale-space blob.

The additional complexity that arises when considering grey-level blob trees over scales compared to grey-level blobs is that a structural event called *reordering* may occur. It is the result of a relative change in grey-level between two different saddle points, which directly determine the ordering relations in the grey-level blob tree, see figure 7 for an example.

2.5 Grey-Level Blob Extraction: Experimental Results

Figure 8 displays an example of extracting (dark) grey-level blobs at different scales in scale-space.[3] It can be seen that at fine scales mainly small blobs due to noise and surface texture are detected. When the scale parameter increases, the noise blobs disappear gradually although much faster in regions near steep gradients. Notable in this context is that blobs due to noise can survive for a long time in scale-space if located in regions with slowly varying grey-level intensity. This observation shows that *scale-space lifetime alone cannot be used as the basis for a significance measure*, since it would substantially overestimate the significance of blobs due to noise.[4]

The buttons on the keyboard manifest themselves as blobs after a small amount of smoothing. At coarser levels of scale, they merge into one unit (the keyboard). One can also observe that some other dark details in the image, the calculator, the cord, and the receiver, appear as single blobs at coarser levels of scale.

This example demonstrates that, as anticipated, the grey-level blob concept shows an extreme degree of noise sensitivity, which can be circumvented by the scale-space smoothing. But it is certainly far from a trivial problem to determine the proper amount of smoothing *automatically*, based on previous conventional methods.

The aim with the suggested blob linking across scales is to determine which blobs in the scale-space representation can be regarded as significant, without any a priori information about either scale, spatial location, or the shape of the primitives.[5] As we shall see later, the output from the linking procedure also enables determination of a suitable scale level for handling each *individual* blob.[6]

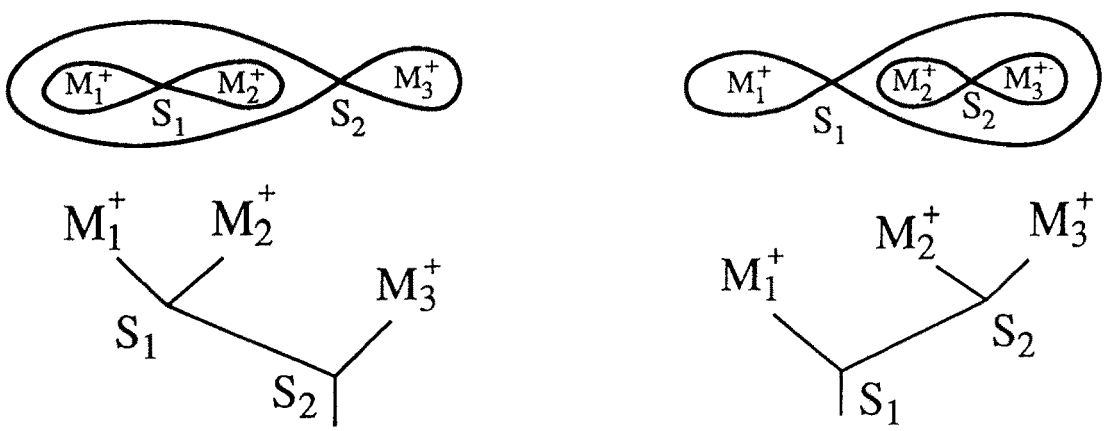

Fig. 7. The additional complexity that occurs when considering grey-level blob trees is the introduction of reorderings, which are changes in the ordering relations in the grey-level blob tree resulting from relative changes in grey-level between saddle points. This figure shows a simple example with bright blobs; in the left case $z(S_1) > z(S_2)$, while in the right case $z(S_1) < z(S_2)$.

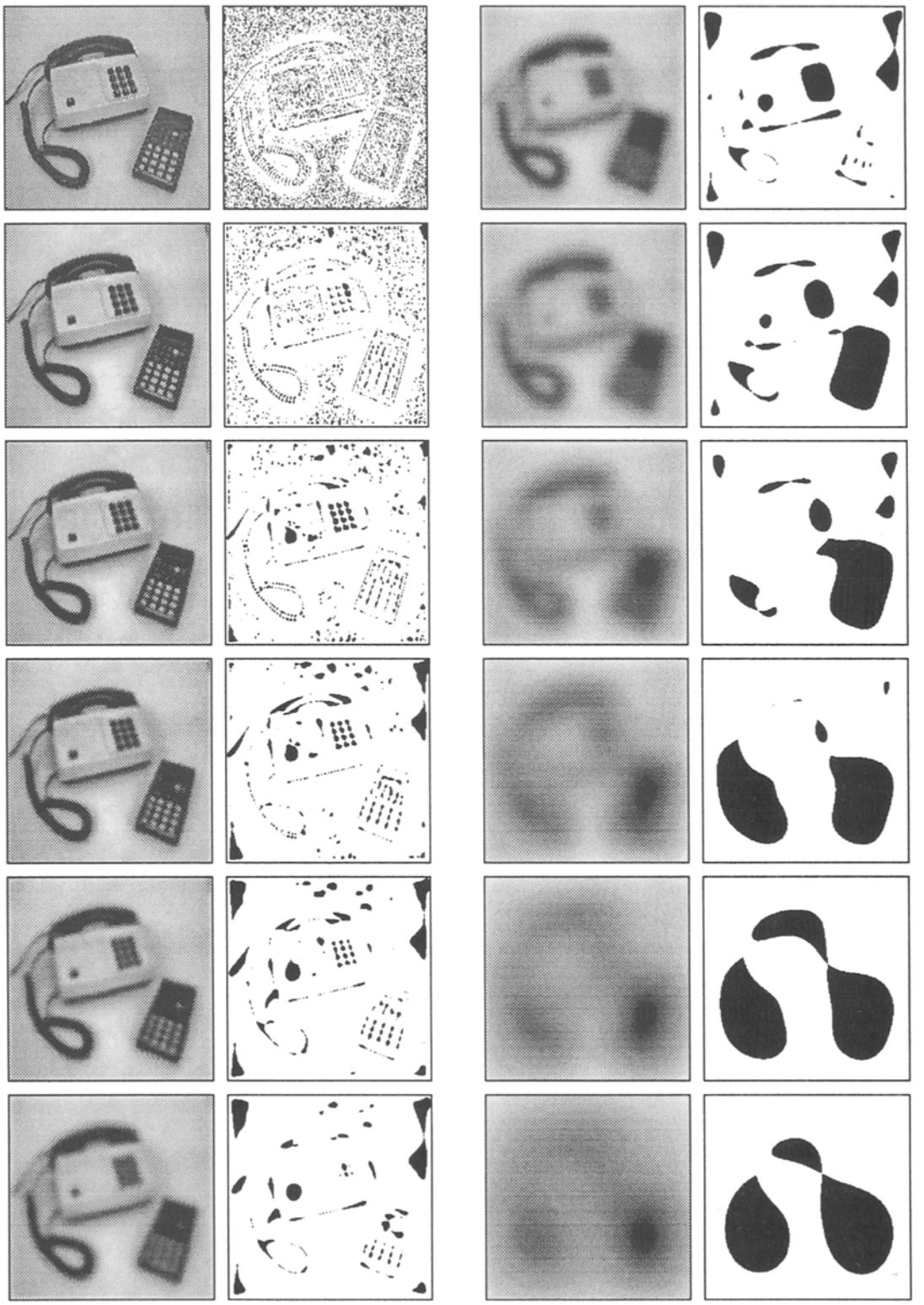

Fig. 8. Grey-level and (dark) grey-level blob images of a telephone and calculator image at scale levels $t = 0$, 1, 2, 4, 8, 16, 32, 64, 128, 256, 512, and 1024 (from top left to bottom right).

2.6 Measuring Blob Significance

Since the ultimate goal of this analysis is to extract important structures in the image based on the appearance and significance of scale-space blobs in the scale-space representation, there is an absolute need for some methodology for comparing significance *between* different levels of scale. In other words, what is desired is a mechanism for judging whether a blob existing only at coarse levels of scale can be regarded as more significant or less significant than a blob with extent primarily at fine levels of scale.

The approach proposed here is to use the four-dimensional volumes of the scale-space blobs in scale-space (defined in appendix A.1.3). It is suggested that it is a useful entity for a significance measure, since it comprises both the spatial extent, the contrast, and the lifetime of the blob. Qualitative motivations for incorporating these entities into the significance measure can be summarized as follows:

spatial extent x: In the absence of further information, a blob having large spatial extent may be treated as more significant than a corresponding smaller blob.

contrast z: In the absence of further information, a high-contast blob may be treated as more significant than a similar blob with lower contrast.

lifetime t: In the absence of further information, a blob having a long lifetime in scale-space may be treated as more significant than a corresponding blob having a shorter lifetime. In general, a blob B_1 far away from another blob B_2 will survive longer in scale-space than a blob B_3, similar to B_1, but nearer to B_2. Moreover, the lifetime of a blob will in general be longer if there are no spatially coincident structures at other scales. Hence, two special cases implied by this heuristic principle are that

(1) a blob B_1 *far away* from another blob B_2 will be treated as more significant than a blob B_3 similar to B_1, but nearer to B_2,[7]
(2) a blob, for which there are no spatially overlapping finer or coarser scale structures, will be treated as more significant than a similar blob, for which such *interfering structures* at nearby scales exist.

If the significance measure, however, is to be based on the scale-space blob volume, it is of crucial importance that the coordinates are measured in proper units, since in principle they could be transformed by arbitrary monotone functions.

2.6.1 Measuring Scale-Space Lifetime. Consider first the measurement of scale-space lifetime. A natural choice of scale parameter for a continuous signal case is the logarithm of the ordinary scale parameter. Based on this idea, one could be inspired to define scale-space lifetime as $\log t_D - \log t_A$, where t_D and t_A denote the disappearance and appearance scales of the scale-space blob respectively. It seems reasonable that this would give a good description at coarse scales, since it is well known that changes in scale-space occur logarithmically with scale. For example, the scale parameter is usually sampled such that the ratio between successive scale values is constant.

Such an approach would, however, lead to unreasonable results for discrete signals at fine levels of scale, since then a blob existing in the original signal (at $t = 0$) would be assigned an infinite lifetime.[8] Similarly, it can be observed that $t_D - t_A$ does not work either, since then the lifetime of blobs at coarse scales in scale-space would be substantially overestimated.

Consequently, there is a need for introducing a transformed scale parameter $\tau = \tau_{\text{eff}}(t)$ such that scale-space lifetime measured by

$$\tau_{\text{life}} = \tau_D - \tau_A = \tau_{\text{eff}}(t_D) - \tau_{\text{eff}}(t_A) \qquad (10)$$

gives a proper description of the behavior in scale-space also for discrete signals. This scale parameter should neither favor or not favor fine scales over coarse scales.

At first glance the problem of transforming the scale parameter may seem somewhat ad hoc. What properties are required from an effective scale parameter? The approach that will be adopted here is to assume that the expected remaining lifetime of a local extremum should not vary with scale. More precisely, it will be assumed that the probability that a certain local extremum disappears after a small amount of smoothing $\Delta\tau$, expressed in effective scale, should be constant over scales, that is, the *relative decay rate* should be *independent of scale.*[9]

Assume that it is known how the expected number of local extrema per unit area varies with scale. In other words, assume that

$$p(t) = \{\text{expected density of extrema at scale } t\} \qquad (11)$$

is known. The relative decay rate requirement can be stated as

$$\frac{\partial_\tau p}{p} = \partial_\tau (\log p) = C = \text{constant}. \qquad (12)$$

Integration and introduction of new arbitrary constants C_1 and C_2 gives

$$\tau = \tau_{\text{eff}}(t) = C_1 + C_2 \log p(t). \quad (13)$$

Without loss of generality, C_1 can be set to zero. It is just an offset coordinate, and cancels in the scale-space lifetime. Similarly, C_2 just corresponds to an arbitrary but unessential linear rescaling of the effective scale parameter.

So far no assumptions have been made about the dimensionality of the signal, or whether it is continuous or discrete. What is left to determine is how the density of local extrema can be expected to vary with scale.

For a large class of continuous signals, the number of local extrema decreases with scale approximately as t^{α} (Lindeberg [48]). This result can be derived from a one-dimensional stationary, normal white-noise process as well as a corresponding process with a spectral density of the form $\omega^{-\beta}$ with $0 \leq \beta < 3$. It can also be shown from dimensional analysis that in arbitrary dimensions N, the density of local extrema can be expected to decrease with scale as $t^{-N/2}$. Under these conditions, the effective scale is given by a logarithmic transformation

$$\tau_{\text{eff}}(t) = - C_2 \alpha \log t. \quad (14)$$

For discrete signals, the density of local extrema can be expected to show the same qualitative behavior at coarse scales, where the grid effects are negligible. At fine scales, however, the $t^{-\alpha}$ behavior cannot hold, since it is based on the assumption that the original signal contains equal amounts of structure at all scales. The discrete signal is limited by its finite sampling density.

These ideas are illustrated in figure 9, which shows the logarithm of the number of local extrema in a finite image as a function of the logarithm of the ordinary scale parameter t. The left diagram shows simulated results for a large number of white-noise images generated from three different distributions, normal, rectangle, and exponential distribution. The right diagram shows the average of these results. Note that a straight-line approximation is valid only in an interior scale interval. At fine scales there is interference with the *inner scale* of the image given by its sampling density, and at coarse scales there is interference with the *outer scale* of the image given by its finite size.

The notion of effective scale takes this notion of inner scale into account, and enables a precise definition of scale-space lifetime also at fine levels of scale. Combined with the concept of scale-space for discrete signals it provides the necessary tool for investigating fine scale structures. For implementational purpose, $p(t)$ is estimated from synthetic simulation results for a set of reference data. Then, the transformation function is determined by

$$\tau_{\text{eff}}(t) = \log \frac{p_{\text{ref}}(0)}{p_{\text{ref}}(t)}, \quad (15)$$

where $p_{\text{ref}}(t)$ denotes the average density of local extrema in the simulations on the reference data. In the current implementation, the reference data is selected as a large set ($\approx 10^2$) of white-noise images. A motivation for this choice is given in next section.

2.6.2 Transformed Grey-Level Blob Volumes. Similarly, the grey-level blob volumes need to be transformed, since the average volume can be expected to vary substantially with scale. When the scale parameter increases, the average contrast can be expected to decrease, and the average area to increase. What about the grey-level blob volume? Experimental results demonstrate that it actually decreases at fine scales and increases at coarser scales, see figure 10.

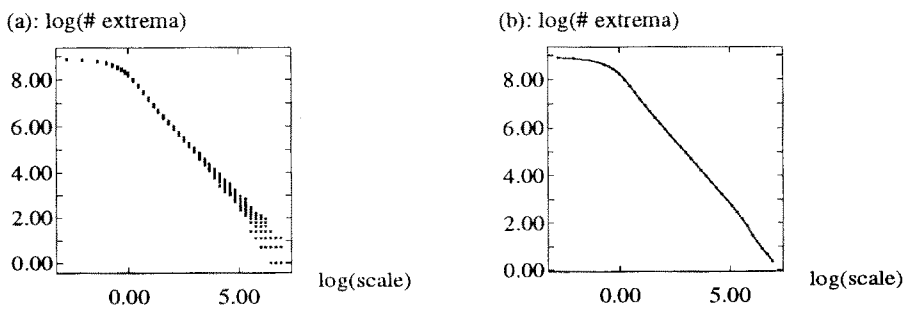

Fig. 9. Experimental results showing the number of local extrema as function of the scale parameter t in log-log scale: (a) measured values, (b) accumulated mean values. Note that a straight-line approximation is valid only over a limited range of scales.

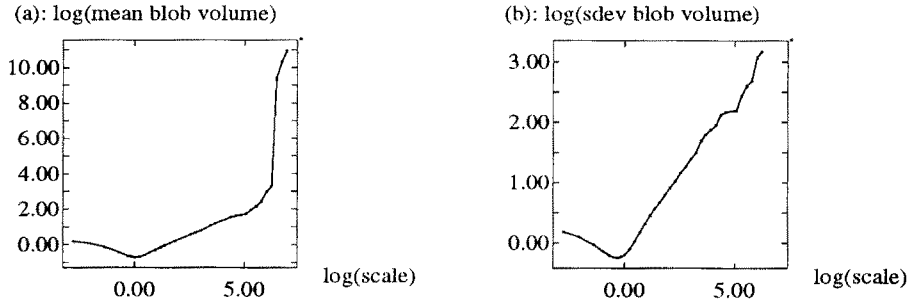

Fig. 10. Experimental results showing (a) the mean value, and (b) the standard deviation of the grey-level blob volumes as a function of scale for white-noise images of different distribution.

Within the parts of the graphs where a linear approximation is valid, the mean value, $V_m(t)$, and the standard deviation of the grey-level blob volume, $V_\sigma(t)$, vary with scale approximately as

$$V_m(t) \sim \sqrt{t}, \qquad V_\sigma(t) \sim \sqrt{t}, \tag{16}$$

while corresponding experiments demonstrate that the variation of the area and contrast is of the form

$$A(t) \sim t, \qquad C(t) \sim \frac{1}{\sqrt{t}}, \tag{17}$$

as can be expected from dimensional analysis or a study of a single Gaussian blob.

If these effects are not taken into account, the significance of coarse-scale blobs will be substantially overestimated. It is clear that the blob behavior depends strongly upon the image (since we actually want to use it for segmentation). Is it then possible to talk about expected behavior?

A conservative approach is to consider white-noise data, that is, images without any structured relations between adjacent pixels. If statistics are accumulated on how blobs can be expected to behave in such images, then the result will be an estimate of to what extent accidental groupings can be expected to occur in scale-space.

If a grey-level blob at some level of scale has a volume smaller than the expected volume for white-noise data, then it can hardly be regarded as significant. On the other hand, if at some level of scale the blob volume is much larger than the expected blob volume, and in addition, the difference in blob volume is much greater than the expected variation around the average value, then it may be reasonable to treat the blob as significant.

A natural normalization is performed by subtracting a measured grey-level blob volume G_{vol} by the mean value, $V_m(t)$, and dividing by the standard deviation, $V_\sigma(t)$. This gives a transformed grey-level blob volume

$$V_{\text{prel}} = \frac{G_{\text{vol}} - V_m(t)}{V_\sigma(t)}. \tag{18}$$

Since, however, such a quantity may take negative values, it is not suitable for integration (which is a necessary step in the computation of the scale-space blob volume). Therefore, in the current implementation, the *effective grey-level blob volume* is defined in the following way, which empirically turns out to give reasonable results,

$$V_{\text{eff}} = V_{\text{trans}}(G_{\text{vol}}; t) = \begin{cases} 1 + V_{\text{prel}} & \text{if } V_{\text{prel}} \geq 0, \\ e^{V_{\text{prel}}} & \text{otherwise.} \end{cases} \tag{19}$$

With this definition, the effective volume of the mean value is one. For larger values it grows linearly with V_{prel}. Thus, V_{eff} and V_{prel} show the same qualitiative behavior for the significant grey-level blobs. For smaller values of V_{prel}, V_{eff} decreases to zero, and the qualitative difference between V_{prel} and V_{eff}, increases as the significance decreases.

Hence, a qualitatively correct behavior is obtained for the important blobs, and it can be expected that this solution should not affect the result too seriously. It should also be mentioned, that in order to adapt the amplitude of the signal to the reference data, V_m and V_σ are rescaled linearly from a least-squares fit between the actual and the expected behavior of these entities.

Finally, these transformed grey-level blob volumes are integrated over the scale-space blob according to (31) in appendix A.1.3. A discussion about other possible approaches to normalizing the scale-space blob volume is given in section 8.4.

2.7 Resulting Representation

To summarize, the data structure proposed is a tree-like multiscale representation of blobs at all levels of scale in scale-space *including* the relations between blobs at different scales. Grey-level blobs should be extracted at all levels of scale; the bifurcations in scale-space should be registered; grey-level blobs stable over scales should be linked across scales into scale-space blobs; and the normalized scale-space blob volumes should be computed.[10]

Since the representation tries to capture significant features and events in scale-space using a small set of primitives, it is called a *scale-space primal sketch*. In the resulting data structure constructed according to this description, every scale-space blob contains explicit information about the grey-level blobs it consists of. The grey-level blobs are detected at (sampled) scale levels obtained from an adaptive scale linking and refinement algorithm outlined in appendix A.4. Further, the normalized scale-space blob volumes have been computed, and the scale-space blobs "know" about the type of bifurcations that have taken place at the appearance and disappearance scales. There are also links to the other scale-space blobs involved in the bifurcations. Hence, the representation explicitly describes the hierarchical relations between blobs at different scales, see figure 11 for a schematic illustration.

In the next section it will be shown how some directly available information from this representation can be used for extracting significant image structures. Then, applications will be given of how such output from the scale-space primal sketch can be used for tuning later-stage processes and guiding the focus-of-attention.

3 Detecting Salient Blob Structures and Their Scales

A major motivation for this research has been investigation of whether the scale-space model allows for determination and detection of stable phenomena. In this section it will be demonstrated that this is indeed possible, and that the suggested representation can be used for extracting regions of interest with associated stable scales from an image in a solely bottom-up data-driven way. The treatment is based on the *assumption* that:

> Structures, that are significant in scale-space, are likely to correspond to significant structures in the image.

This statement has been expressed on a general form, since it can be speculated that the approach can be applied also to structure types other than the blobs considered here.[11] More precisely, since the primitives that will be used are scale-space blobs, the heuristic selection method is formulated as follows:

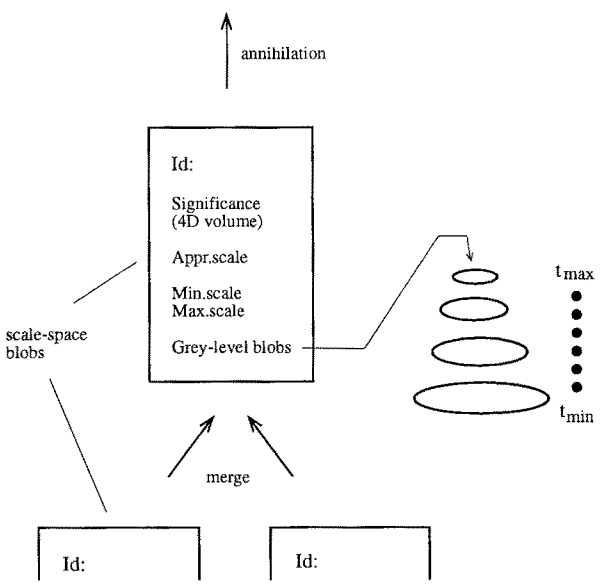

Fig. 11. The scale-space primal sketch based on grey-level blobs can be seen as a tree-like multiscale representation of blobs with the scale-space blobs as basic primitives (vertexes) and the relations (bifurcations) between scale-space blobs at different levels of scale as branches. (When grey-level blob trees are considered as well, the scale-space primal sketch becomes a three-dimensional graph with hierarchical relations along the grey-level dimension as well.)

Assumption 1: Ranking of blob structures on significance: In the absence of other evidence, a scale-space blob having a large normalized scale-space blob volume in scale-space is likely to correspond to a relevant blob-like region in the image.

A scale-space blob will, in general, exist over some range of scales in scale-space. When there is a need for reducing the amount of data represented, and to select a single scale and region as representative of the scale-space blob, the following postulates are suggested.

Assumption 2: Scale selection: Maximum response over scales: In the absence of other evidence, the scale at which the scale-space blob assumes its maximum normalized grey-level blob volume over scales is likely to be a relevant scale for representing the blob.

Assumption 3: Selection of spatial representative: In the absence of other evidence, the spatial extent of a scale-space blob can be represented by the support region of the blob at the scale level selected according to assumption 2.

The ranking on significance depends on the actual scaling of the four coordinates in the scale-space representation. Therefore, the extraction method implicitly relies upon the assumption that it should be sufficient to transform the coordinates once and for all as was done in section 2.6.

Assumption 4: Normalization with respect to reference data: The coordinate axes in the scale-space representation can be normalized on the basis of the behavior in scale-space of reference data.

Below, experimental results will be given demonstrating how these assumptions can be used for segmenting out intuitively reasonable regions from various types of imagery. First, however, motivations will be given to why these assumptions have been stated.

3.1 Motivations for the Assumptions

A central problem in low-level vision concerns what should be meant by image structure. In other words, which features in an image should be regarded as significant, and which ones can be rejected as insignificant or as due to noise. As we discussed in the introduction, this problem seems impossible if stated as a pure mathematical problem, as is the segmentation problem if seen in isolation. Nevertheless, since there are indications that biological vision systems are able to perform natural preattentive groupings in images (Witkin and Tenenbaum [69], Lowe [51]), one may speculate whether there are any inherent properties in data that can be used for defining such groupings.

The scale-space primal sketch constitutes an attempt to express such groupings for blob-like structures by a formal mathematical framework.[12] The technique consists of constructing primitives from the scale-space representation, which are defined solely in terms of the singularities that occur in scale-space.

3.1.1 Stability in Scale-Space: Salience. When Witkin [70] coined the term *scale-space*, he observed a marked correspondence between perceptual salience and stability in scale-space:

> . . . intervals that survive over a broad range of scales tend to leap out to the eye. . .

Assumption 1 constitutes an extension of this observation to a heuristic principle for extracting blob-like image structures. The significance measure is, however, not based on the scale-space lifetime alone, since as mentioned in section 2.5, blobs due to noise can survive over large ranges of scale, if located in regions with slowly varying grey-level.

Observe how this measure of significance relates to a definition of structure in terms of *transformational invariance*. If a feature is to be useful for recognition, it must necessarily be stable with respect to small disturbances. Otherwise it would hardly be useful, since then it would be impossible to compute it accurately. Here, this stability requirement is used for actually formulating an operational method for detecting image structures—by subjecting the image to *systematic parameter variations*, and explicitly measuring the stability of the structures (here, blobs) under parameter variations (here, scale variations).

In line with this idea, assumption 1 states that the scale-space blobs that are the most stable ones under variations of the scale parameter in scale-space, are the most likely ones to correspond to significant image structures. Of course, the reverse statement does not hold. There are many other sources of information, for example, lines in line drawings, which are not captured by a blob concept, and scale-space smoothing.

Note, that this use of transformational invariance is different than what is usually meant by invariance in an algebraic or a geometric sense; the transformational

invariance of the scale-space blobs concerns local topological properties which are stable over finite intervals of parameter variations.

3.1.2 Reduction of the Representation: Abstraction. Because of complexity arguments, the entire parameter variation information from the low-level modules cannot be transferred to the higher-level modules in a vision system.

Assumption 2 and assumption 3 express such a desire to represent a scale-space blob with a grey-level blob at a single level of scale, in order to give a more compressed representation—an *abstraction*—for further processing.

The motivation for selecting the scale at which the maximum of the normalized grey-level blob volume is assumed is that it should reflect the scale where the blob response is maximally strong. It turns out that this scale will often be close to the appearance scale of the scale-space blob, except at blob splits and blob creations, for which the grey-level blob volume may be zero at the appearance scale.[13]

It is worth noting that assumption 2 implies a projection from a four-dimensional scale-space blob to a three-dimensional grey-level blob, and that assumption 3 implies a projection from that grey-level blob to its two-dimensional support region.

3.2 Basic Method for Extracting Blob Structures

The basic methodology for extracting significant blobs from an image should now be obvious from the above presentation.

- Generate the suggested scale-space primal sketch, where blobs are extracted at all levels of scale, and linked across scales into scale-space blobs.
- Compute the normalized scale-space volume for each scale-space blob based on the notion of effective scale and effective grey-level blob volumes.
- Sort the scale-space blobs in descending significance order, i.e., with respect to their normalized scale-space blob volumes.
- For each scale-space blob determine the scale where it assumes its maximum grey-level blob volume, and extract the support region of the grey-level blob at that scale.

3.3 Experimental Results

Figures 12–13 show the result of applying this procedure to two different images, one with a telephone and a calculator, and one with a set of toy blocks.

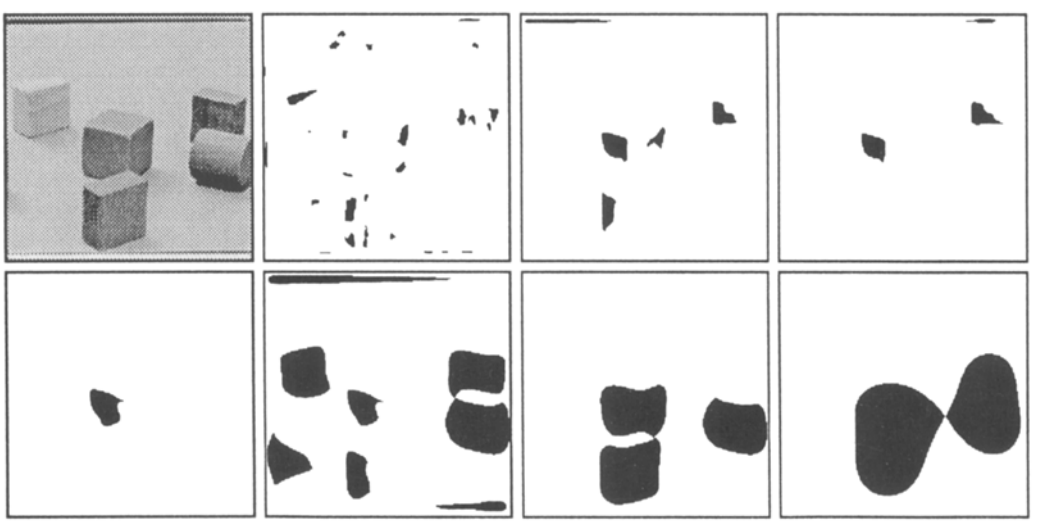

Fig. 12. The 50 most significant dark blobs from a toy block image. (Note how these images have been produced—they are not just blob images at a few levels of scale. Instead every blob has been marked at its representative scale. Finally, the blobs have been drawn in different images so as to avoid overlap.)

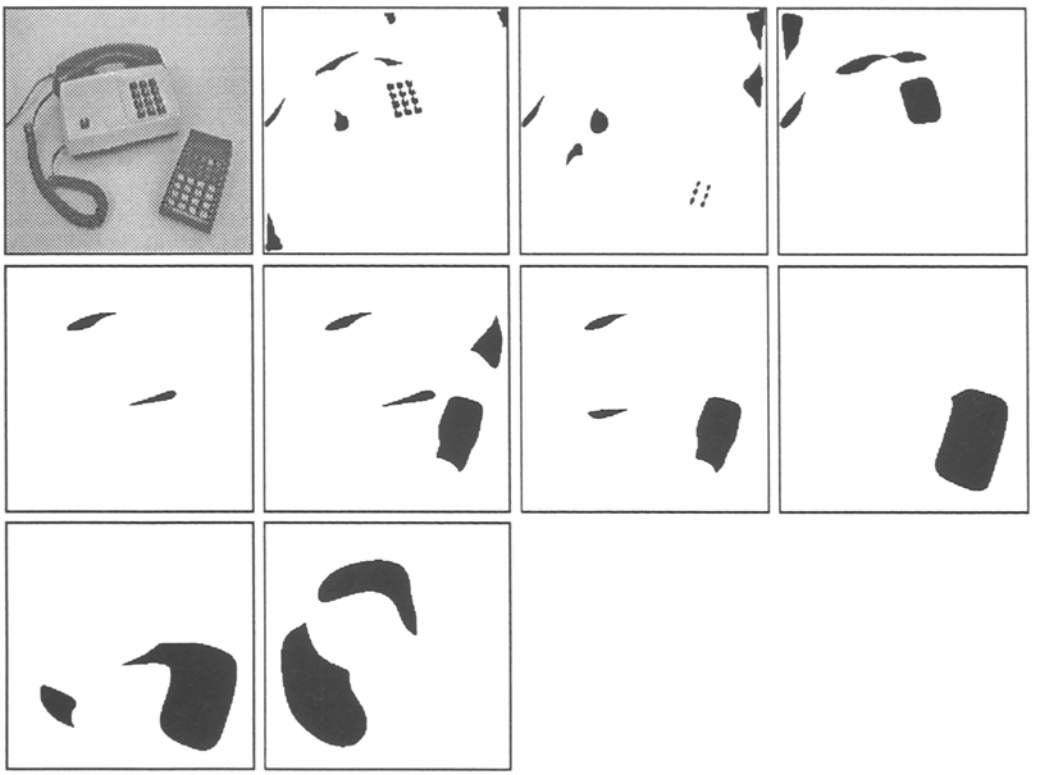

Fig. 13. The 50 most significant dark blobs from a telephone and calculator image.

For display purpose the N most significant dark scale-space blobs have been extracted. Each blob is displayed at its representative scale, that is the previously mentioned scale at which the scale-space blob assumes its maximum grey-level blob volume. The spatial representative of each blob (which is the blob support region of the grey-level blob at the representative scale) is marked in a binary image, where black indicates the existence of a significant dark blob, and white represents background. In order to avoid overlap, the display routine shifts to a new fresh image each time the addition of a new blob would imply overlap between two different blobs.

We can see that in the toy block image, the individual blocks are extracted. Also, at coarser scales, adjacent blocks are grouped into coarser scale units, and the imperfections in the image acquisition near the boundaries are pointed out. In the telephone scene, the buttons, the keyboard, the calculator, the cord, and the receiver are detected as single units. In order to show the spatial relations between the blobs at different scales, figures 14 and 15 show the blob boundaries superimposed onto each other. More experimental results, including bright blobs, are presented in following sections; see also (Lindeberg [43], Lindeberg and Eklundh [46]).

3.4 Further Treatment of the Generated Blob Hypotheses

The number of scale-space blobs selected for display above is, of course, rather arbitrary. Note, however, that there is a well-defined ranking between the blobs. If one studies their significance values (see table 1 and figures 14–17), one can observe that those blobs we regard as the most significant ones have significance values standing out from the significance values of the other ones. Hence, it seems plausible that a few regions can be extracted just based on this observation. In more general situations, there is a need for feedback and reasoning.

The output information from this representation should not be over-estimated. Since it is a low-level processing module, the output should be interpreted mainly as indicators signaling that "there might be something there of about that size—now some other

Table 1. Significance values and selected scale levels for the 20 most significant scale-space blobs computed from the toy block image. Note that a few blobs have significance values clearly standing out from the other.

Significance	Scale	Blob label
1450.55	32.00	1760
1266.43	64.00	1767
1030.53	50.80	1764
591.16	80.60	1768
297.60	812.90	1770
284.72	645.10	1769
150.64	45.25	1761
131.99	28.51	1758
73.69	45.25	1763
63.51	35.91	1065
35.92	28.51	1759
35.42	22.65	1753
20.45	8.00	1703
17.43	8.99	1702
12.84	11.99	1723
9.94	28.51	1757
6.84	4.00	1256
6.20	9.53	1708
5.33	14.25	1725
5.10	2.00	1440

Fig. 14. Boundaries of the dark scale-space blobs extracted from the toy block image. (Left) Original image. (Middle left) The 50 most significant dark blobs. (Middle right) Low threshold on the significance measure set in one of the "gaps" in the sequence of significance values (between 74 and 131). (Right) High threshold on the significance measure set in another "gap" (between 298 and 591). (The significance values are shown in table 1.)

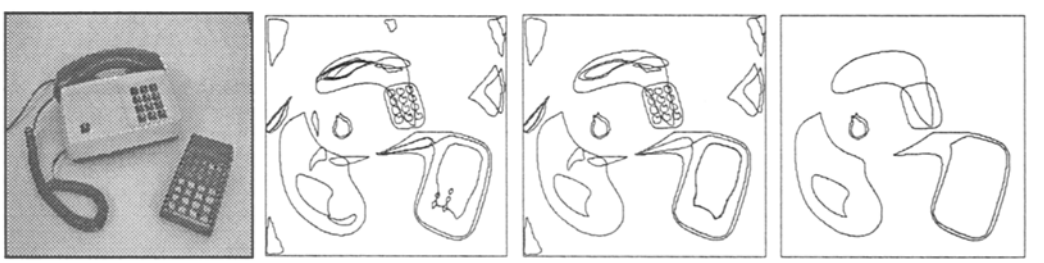

Fig. 15. Corresponding blob boundaries from the telephone and calculator image.

Fig. 16. Significance values of the 50 most significant blobs from the toy block image. The significance value of each blob has been marked with an "x" along a horizontal logarithmic scale. The vertical lines indicate the manually selected thresholds used in figure 14.

Fig. 17. Corresponding significance values and manually selected thresholds for the blobs from the telephone and calculator image in figure 15.

module should take a closer look." From this viewpoint it can be noted how well the extracted blobs describe blob-like features in the previous images, considering that the blobs have been extracted almost without any a priori information.

In principle, a reasoning process working on the output from the scale-space primal sketch can operate in either of two possible modes:

— Use a threshold on the significance measure. In a real system, such a threshold may—in some situations—be set from given context information and expectations.
— Evaluate the generated hypotheses in decreasing order of significance, i.e., first try to interpret the first hypothesis in a feedback loop, then consider the second one, etc. Continue as long as the hypotheses deliver meaningful interpretations for the higher-level modules.

An inherent property of this representation is that it does not have any limiting requirement that there is just one possible interpretation of a situation. Instead it generates a variety of hypotheses. Given some region in space, several hypotheses may be active for it (or parts of it) concerning different structures at different scales.

A principle advocated in this work is that the *qualitative scale and region information* extracted from the scale-space primal sketch can be useful for *guiding other visual processes*, and will *simplify their tasks*. Now, examples will be given of how such integrations of the scale-space primal sketch with other processing modules can be performed.

4 Guiding Edge Detection with Blob Information

As a first application of the suggested methodology, an integration with an edge-detection method known as edge focusing (Bergholm [5]) will be described.

The main idea is to use the scale and region information for guiding an edge-detection scheme working at an adaptively determined level of scale. It will be demonstrated that this task can be simplified, and that thresholding on gradient magnitude can be avoided. Given a significant scale-space blob, edge detection is performed at the appropriate scale of the scale-space blob. Then a matching step is carried out between the support region of the blob and the edges in a neighborhood of the blob. Finally, the matched edges are tracked to finer scales in order to improve the localization; see figure 18 for a schematic overview.

It is not maintained that the approach to be presented describes any "optimal way" to solve every occurring subproblem. Instead, the intention is to exemplify how a connection between the scale-space primal sketch and other processing modules can be done.

4.1 Edge Detection at a Coarse Scale

A rather simple edge detector is used deliberately. The image is smoothed to the scale associated with the scale-space blob. Then, derivatives along the two coordinate directions are estimated by difference approximations, and a nonmaximum suppression step (Canny [14], Korn [38]) is performed (without thresholding on gradient magnitude) in order to obtain thin edges.[14] In order to suppress spurious noise points, only edge segments of length exceeding, say 2 pixels, are accepted.

4.2 Matching Blobs to Edges

Associating blobs with edges leads to a matching situation. The matching procedure used for associating blobs with edges is based on the following criteria:

Geometric Coincidence. The edge segments should "encircle," "be included in," or "intersect" the blob. A convenient way to express such a criterion is by requiring

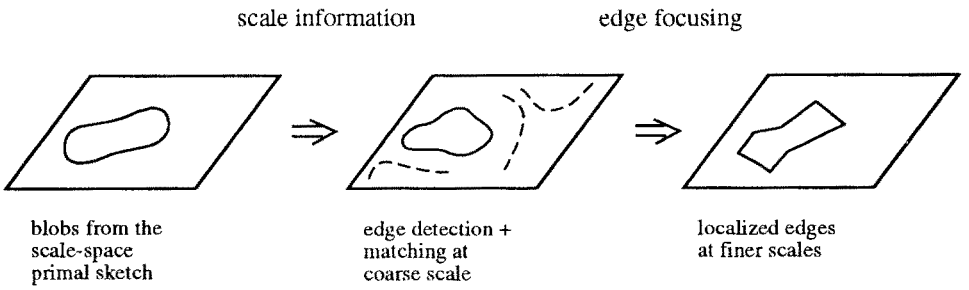

Fig. 18. Schematic view over the proposed integration of the scale-space primal sketch module with edge focusing.

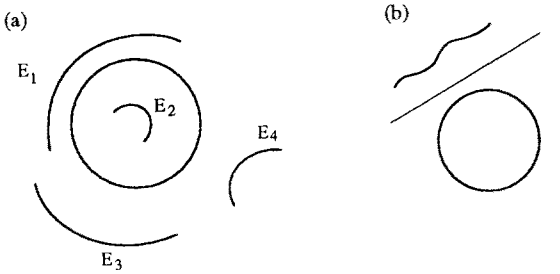

Fig. 19. The geometric coincidence condition means that the edge should surround the blob, be included in it, or intersect it—it should be impossible to draw a straight line separating the edge from the blob (b). In example (a) edges E_1 and E_2 are treated as matching candidates of the blob, while edges E_3 and E_4 are not.

it to be *impossible to draw a straight line separating the edge from the blob*, see figure 19.

A simple way to approximate this criterion computationally is as follows: Let $B \subset \mathbb{R}^2$ be the set of points contained in the support region of a blob, and let $E \subset \mathbb{R}^2$ be the set of points covered by an edge segment. Further, given any region R and any arbitrarily rotated coordinate system $(\zeta, \eta) \in \mathbb{R}^2$, define the extreme coordinate values $\zeta_{\min}, \zeta_{\max} \in \mathbb{R}$ by

$$\zeta_{\min}(R) = \min_{(\zeta,\eta)\in R} \zeta, \qquad \zeta_{\max}(R) = \max_{(\zeta,\eta)\in R} \zeta. \quad (20)$$

Now, an edge segment E is regarded as a matching candidate of a blob B if

$$\zeta_{\min}(E) \leq \zeta_{\max}(B) \qquad \zeta_{\max}(E) \geq \zeta_{\min}(B) \quad (21)$$

hold in a sufficiently large number of directions. For practical implementation, this condition is required to hold along both the coordinate directions of a standard Cartesian coordinate system, as well as a corresponding coordinate system rotated by 45 degrees.

Proximity. The edge segment should *not be too far away from the blob boundary.* In other words, the edge segment should comprise at least some point located near the boundary of the blob. This condition can be stated as

$$\min_{x_E\in E; x_B\in B} \|x_E - x_B\|_2 \leq \frac{d(t)}{2}, \quad (22)$$

where $d(t)$ represents a characteristic length at scale t.[15] The purpose of this criterion is to prevent (interior and exterior) edges far away from the blob boundary from being associated with the blob; see figure 20(a).

Voronoi Diagram of the Grey-Level Blob Image: The edge segment should not be strongly associated with other blobs. A natural way to express such a criterion is in terms of a Voronoi diagram of the grey-level blob image at the selected scale. An edge segment is regarded as a Voronoi matching candidate of a blob if it has at least one point in common with the Voronoi region associated with the grey-level blob; see figure 20(b). In fact, this type of criterion turns out to be useful also to other matching problems.

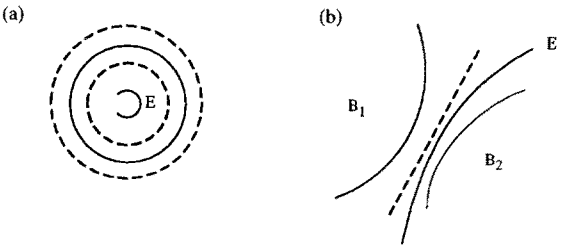

Fig. 20. (a) The purpose of the proximity criterion is to prevent edges far away from the blob boundary from being associated with the blob. (b) The purpose of the Voronoi region matching is to prevent edges strongly related to one blob from being associated with other nearby blobs.

Composed Matching Procedure: For an edge segment to be accepted as a matching candidate of a blob, it must satisfy all these criteria; see figure 21 for an illustration. Hence, the matching is relatively restrictive. It is also improved by the fact that it is performed at a scale at which a blob has manifested itself. Once it is known that a spatial region has given rise to a large blob at some level of scale, it seems unlikely that conflicting edges should appear at the same scale, since most interfering structures ought to be suppressed by the scale-space smoothing.

The main problem with this matching procedure is that it does not include any mechanism for splitting long edge segments into shorter ones. Hence, certain edge segments can be very long and spread far away from the blob boundary; see the example in figure 22.

4.3 Blob-Initiated Edge Focusing

Edge focusing (Bergholm [5]) is a method for following edges through scale-space. The basic principle is to detect edges at a coarse scale, where the detection problem can be expected to be easier, and then track the edges to a finer scale, in order to improve the localization, which can be very poor at coarse scales.

In this application the focusing procedure is initiated from several scale levels, since the significant blobs

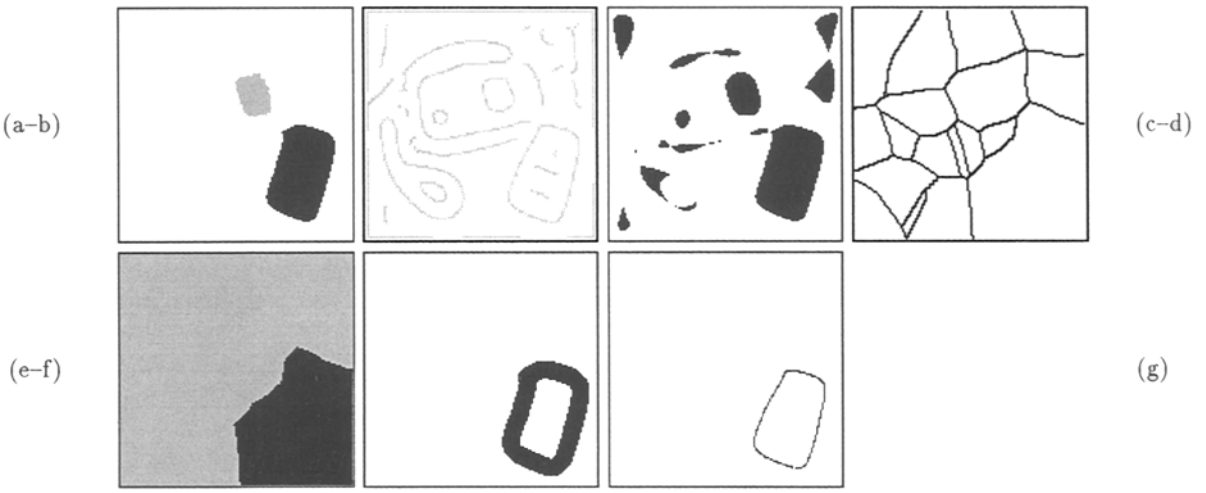

Fig. 21. The matching procedure between blobs and edges for one blob from the telephone and calculator image. (a) The support region of a dark scale-space blob (black). (b) Edges detected at the scale given by the blob. (c) All grey-level blobs at the same level of scale. (d) Voronoi diagram of the grey-level blob image. (e) The Voronoi region corresponding to the given blob. (f) The proximity stripe around the blob edge. (g) Resulting edges matched to the blob.

from the scale-space primal sketch manifest themselves at different scales. Hence, the blobs are first presorted in decreasing scale order. The procedure starts with the coarsest scale blob, detects edges at that scale, and matches those to the blob. This gives the input data for the edge-focusing procedure, which follows these edges to the scale given by the second blob. Then, the edge-detection and matching steps are repeated, etc.

Figure 22 illustrates some steps from this procedure applied to the telephone and calculator image. In order to reduce the number of blob hypotheses treated, a threshold has been introduced on the significance value. The "final result" is shown in the lower right corner.

Observe that this method, called *blob-initiated edge focusing*, is not just another edge detector, but that the edge elements obtained in this way are more likely to correspond to meaningful entities, since they are explicitly grouped into edge segments, and are associated with blobs and explicit scale information. Note that label information for the edge segments can be easily inherited during the edge-focusing process.[16]

With this integration experiment, two of the tuning parameters in the edge-focusing algorithm have been eliminated. What remains undetermined is the stop scale down to which the edge focusing procedure should be performed. In this work it has been throughout set to $t = 1$, a scale where the sampling effects due to the discrete grid start to become important. It seems plausible that some further guidance for this selection could be obtained by studying the behavior of the focused edges in scale-space as done by (Sjöberg and Bergholm [64]; Zhang and Bergholm [72]; and Lindeberg [49]).

The integration of the two algorithms exemplifies the previously mentioned guidance of focus-of-attention. Note that the processing initiated by the scale-space primal sketch is performed only for a small subset of the image data. In this sense the approach bears similarity with the idea of a "focused beam" derived by Tsotsos [66] from complexity arguments; see also the experimental work by Culhane and Tsotsos [18].

5 Automatic Peak Detection in Histograms

The scale-space primal sketch is well suited for automatic cluster detection, since it is designed for detection of bright blobs on dark background and vice versa. Hence, it lends itself as a natural module for peak detection in algorithms based on histogramming techniques. Although it is well known that histogram-based segmentation hardly can be expected to work globally on entire images (due to illumination variations, interference because of many regions, etc.), such methods can often give useful results *locally* in small windows, where only a few regions of distinctly different characteristics (e.g., color or grey-level) are present.

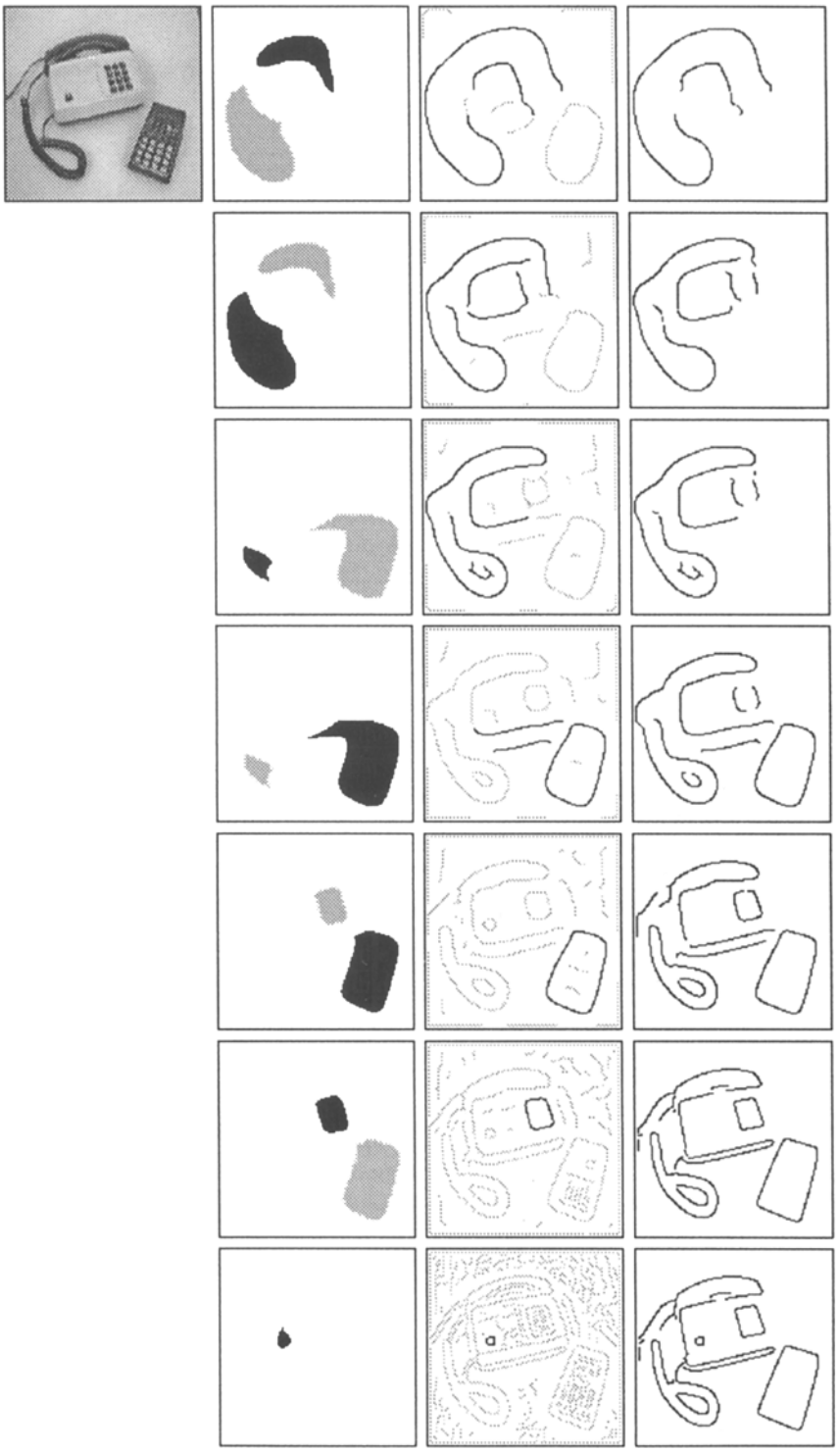

Fig. 22. Illustration of the composed blob-edge focusing procedure for the telephone and calculator image. The left column shows the active blob hypothesis; its blob support region has been marked with black. The middle column shows the edge image at the level of scale given by the previous blob; matched edge segments are drawn black while the other edge pixels are grey. The right column shows the result after focusing, just before the new blob is considered. The image in the lower right corner displays "the final result," that is, the edges that are related to the dark blobs in the image. The scale and significance values for the different blobs are from top to bottom (101.6, 14.1), (50.8, 252.8), (32.0, 11.4), (25.4, 660.9), (14.3, 40.8), (6.4, 63.6), and (1.3, 13.2) respectively.

5.1 Experimental Results: Histogram-Based Color Segmentation

Figure 23 and figure 24 illustrate how the scale-space primal sketch can consitute a helpful tool in such histogram modality analysis of multispectral data. It shows histograms of the (two-dimensional) chroma information, together with blobs detected by the scale-space primal sketch, and backprojections of the blobs.[17] We see that the extracted blobs induce a meaningful partitioning of the histogram, corresponding to regions in the image with distinctly different colors.

Of course, there is a decision finally to be made about which peaks in the histogram should be counted as being significant. However, it seems plausible that the significance values given by the scale-space blob volumes reflect the situation in a manner useful for such reasoning, especially since the regions around the peaks are extracted automatically. In these examples, (single) thresholds have been set manually in "gaps" in the sequences of significance values; see the captions of figures 23–24.

It can also be noted that this peak-detection concept will be less sensitive to quantization effects in the histogram acquisition than many traditional peak-detection methods. The problems due to too fine a quantization in the accumulator space will be substantially reduced, since the scale-space blurring will leak to a propagation of information between different accumulator cells. Thus, even though the original histogram may have been acquired using "too many and too small" accumulator cells, large-scale peaks will be detected anyway, since the contents of their accumulator cells will merge to large-scale blobs in scale-space after sufficient amounts of blurring.

Finding peaks in histograms is a problem that arises in many contexts. Let us point out that the case with color-based histogram segmentation has been

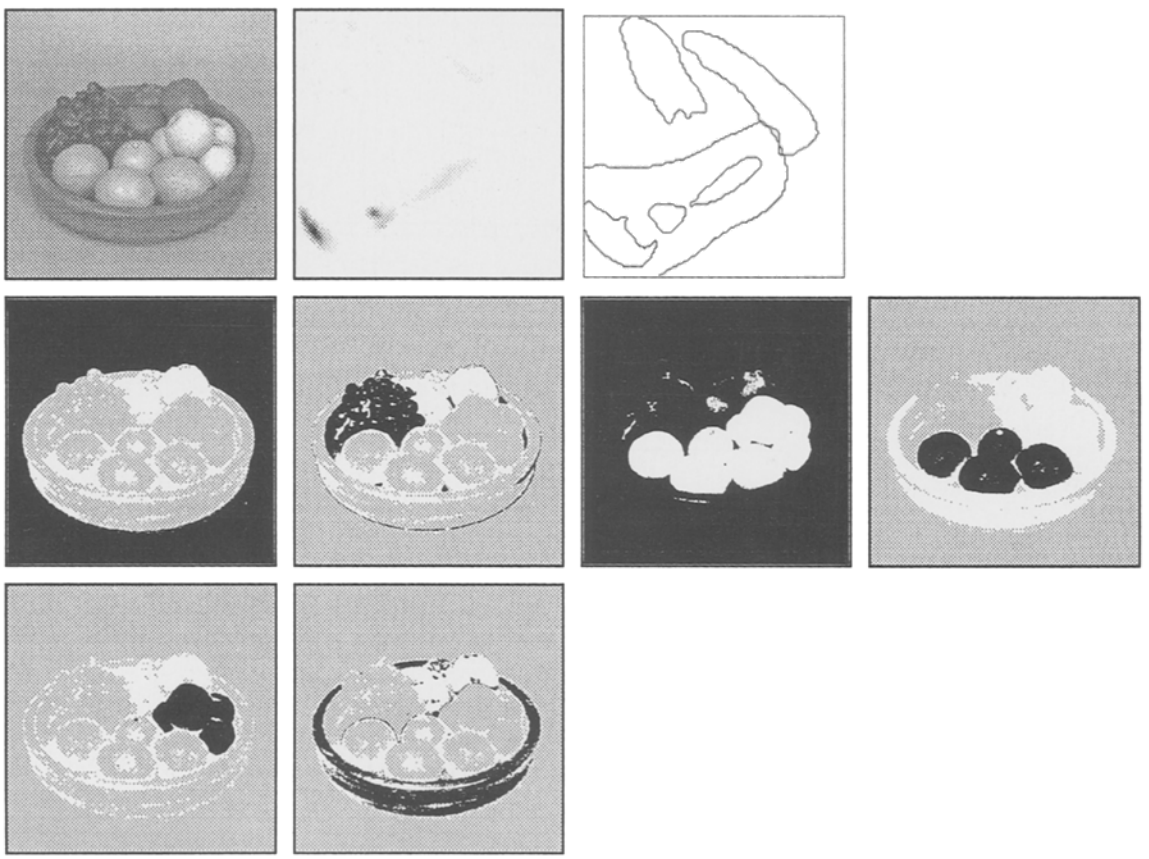

Fig. 23. Histogram-based color segmentation of a fruit bowl image: (a) Grey-level image. (b) Histogram over the chroma information. (c) Boundaries of the 6 most significant blobs detected by the scale-space primal sketch. (d)–(i) Backprojections of the different histogram blobs (in decreasing order of significance). The pixels corresponding to the various blobs have been marked in black. (The region in figure (f) is the union of the regions in figures (d), (e), and (i).) The significance values of the accepted blobs were 42.6 (background), 8.3 (grapes), 3.6 (oranges), 3.1 (apples), 3.0 (bowl), and for the rejected blobs 2.0 and less (2.0, 1.9, 1.8, 1.4, 1.3, 1.1, 1.1, 1.1, ...).

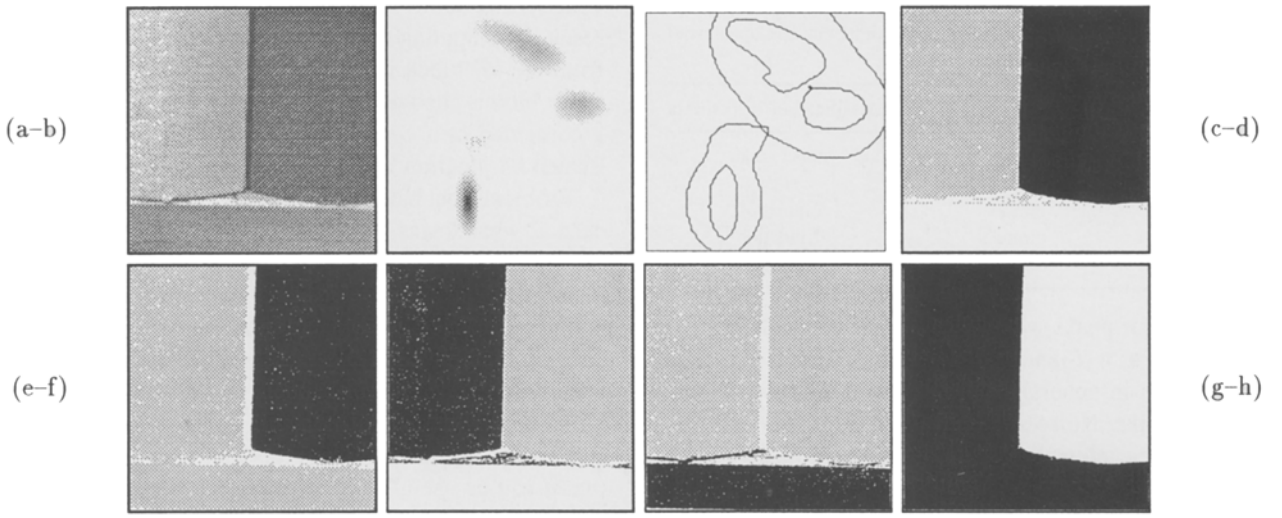

Fig. 24. Similar histogram-based color segmentation of a detail from an office scene. The image shows a small window from a bookcase with two binders (yellow and blue) on a shelf made of (yellowish) wood. The displayed blobs have significance 187.9 (blue binder, large blob), 173.7 (blue binder, small blob), 170.1 (yellow binder), 80.6 (shelf), and 66.7 (yellow binder and shelf). As can be seen, two blobs corresponding to the blue binder have been detected. This is a common phenomenon in the scale-space primal sketch, that arises because a large blob merges with a small (insignificant) blob and forms a new scale-space blob. Two such duplicate blobs corresponding to the yellow binder (significance 18.0) and the shelf (significance 17.9) have been suppressed. The remaining blobs had significance 2.5, 2.0, 2.0, 2.0, 1.2, 1.2, 1.2, 1.1, and less.

considered just as one possible application of the scale-space primal sketch to histogram analysis. Because of the general-purpose nature of this tool, there are potential applications for similar techniques such as Hough transforms, texture classification, etc. in two as well as other dimensions. For related work, see Carlotto [15], and Mokhtarian and Mackworth [56].

6 Junction Classification: Focus-of-Attention

More generally, the scale-space primal sketch can serve as a primitive mechanism for focus-of-attention. As an illustration of this, an experimental work will be briefly described, where the scale-space primal sketch has been used for guiding the focus-of-attention of an active head-eye system applied to a specific test problem of classifying junctions.

The presentation is aimed at showing how the suggested approach can be used when addressing some of the most fundamental problems in active analysis: (i) how to generate hypotheses about the existence of objects, (ii) how to determine where to look, and (iii) at what scale(s) to analyze image structures.

6.1 Background: Junction Classification by Active Focusing

It is well known that junctions provide important cues to three-dimensional structure (Malik [52]). Since most edge detectors cannot be expected to give accurate results at junctions, direct methods for junction detection have been proposed.

Brunnström et al. [11] have demonstrated that a reliable classification of junctions can be performed by analyzing the modalities of local intensity and directional histograms during an active focusing process. The basic principle of the method is to accumulate local histograms over the grey-level values and the directional information around candidate junction points, which are assumed to be given by some interest-point operator. Then, the numbers of peaks in the histograms can be related to the type of junction according to table 2.

The motivation for this scheme is that in the neighborhood of a point where three edges join, there will generically be three dominant grey-level peaks corresponding to the three surfaces that meet. If the point is a 3-junction (an arrow-junction, or a *Y*-junction) then the edge-direction histogram will (generically) contain

Table 2. Basic classification scheme for local intensity and directional distributions around a candidate junction point. (Adapted from Brunnström et al. [11].)

Intensity	Edge Direction	Classification Hypothesis
unimodal	any	noise spike
bimodal	unimodal	edge
bimodal	bimodal	*L*-junction
trimodal	bimodal	*T*-junction
trimodal	trimodal	3-junction

three major peaks, while two directional peaks can be expected at a *T*-junction. Similarly, at an *L*-junction there will in general be two intensity and two directional peaks. Noise spikes and edges must be considered, since interest-point operators tend to give false alarms at such points. Situations with more than three peaks in either the intensity or the directional histogram are treated as nongeneric or as corresponding to surface markings.

6.2 Required Context Information

Taking such local histogram properties as the basis for a classification scheme leads to two obvious questions: Where should the window for accumulating the statistics be located and how large should it be?

The problem of detecting candidate junctions has been extensively studied in the literature (Kitchen and Rosenfeld [32]; Förstner and Gülch [22]; Koenderink and Richards [35]; Florack et al. [21]). A useful entity for junction detection is the curvature of level curves. In order to give a stronger response near edges, this entity is usually multiplied by the gradient magnitude raised to some power k. A natural choice is $k = 3$. This gives a polynomial expression

$$\tilde{\kappa} = |L_{xx}L_y^2 + L_{yy}L_x^2 - 2\,L_{xy}L_xL_y| \qquad (23)$$

which turns out to be skew invariant (Blom [9]). The result of computing this *rescaled level-curve curvature* from the toy block image at a scale given by a scale-space blob is shown in figure 25(b). Local maxima in $\tilde{\kappa}$ computed at a certain scale in scale-space can be treated as junction candidates at that scale.

Problems that have not been very much treated concern at what scales the junctions should be detected, and how to determine regions of interest around those. Corners are usually regarded as pointwise properties, and thereby treated as very-fine-scale features. Realistic corners from man-made environments are, however, usually rounded. This means that small-size operators will find them noisy images.

In order to *extract* junction candidates, it is proposed that it can be useful to perform blob detection on the level-curve curvature data. Such blobs are termed *curvature blobs*. Figure 25(c) shows the result of applying this operation to the data in figure 25(b). Note that a set of regions is detected corresponding to the major corners of the toy block. Note also that the support regions of the blobs serve as natural descriptors for a characteristic size of a region around the candidate junction, which can be used for setting the window size for the histogram classification step.

Of course, direct setting of a window size immediately valid for correct histogram localization seems to be a very difficult, or even an impossible, task. If the window is too large, then other structures than the corner of interest may be included. Conversely, if it is too small, the histograms could be severely biased, and deviate far from the ideal appearance if the physical corner is rounded. A too-small window may also fall outside the actual corner if the interest point is associated with a localization error.

Therefore, the proposed approach is to use the support region of the curvature blob for determining

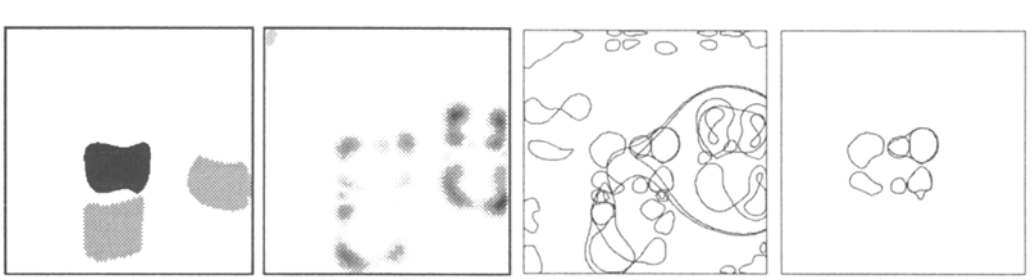

Fig. 25. Blob-initiated detection of junction candidates in the toy block image. (a) Support region of a scale-space blob (marked with black). (b) The rescaled level-curve curvature computed at the scale of the scale-space blob. (c) Boundaries of the 50 most significant curvature blobs detected by applying the scale-space primal sketch to the curvature data. (d) Curvature blobs matched to the original scale-space blob by spatial overlap, under the additional condition that the scale of the curvature blob must not exceed the scale of the original scale-space blob that invoked the analysis.

generous upper and lower bounds on an interval of window sizes, and then applying the focusing procedure described by Brunnström et al. [11]. The intention is that a systematic variation of window size combined with a consistency check over parameter variation should allow for a more robust modality determination. The method is based on the assumption that stable responses will occur for the model that best fits the data, which closely relates to the scale variation principle in section 3.1.1.

A trade-off with this approach is that the localization of the corner will in general be affected by the smoothing operation. Therefore, it should be emphasized that the main goal of this first step is to *detect* candidate junctions at the possible cost of poor localization. Then, if improved localization is needed, it can be obtained from a separate process, using edge and curvature information at finer scales (Lindeberg [49]).

6.3 Experimental Technique

Figures 26–29 illustrate some of the main processing steps in the composed classification method (Brunnström et al. [12]), which is integrated with an active head-eye system (Pahlavan and Eklundh [57]) allowing for algorithmic control of the image acquisition. Figure 26 shows an overview image of a scene under study together with the 20 most significant dark and bright scale-space blobs. Each such region constitutes a hypothesis about the existence of an object, a facet of an object, or an illumination phenomenon in the scene.

In figure 27 the cameras of the head-eye system have been redirected toward one of the dark blobs corresponding to the central toy block, and a new image of higher resolution has been acquired around the region of interest. This step simulates foveation. At the scale of the scale-space blob (transformed with respect to the increased sampling density), the level-curve curvature is computed, and curvature blobs are detected using the scale-space primal sketch; see figure 27(c).

In figure 28 the algorithm has zoomed in further to one of the curvature blobs, and invoked a histogram classification procedure tuned to the size of the curvature blob. This junction was classified as a 3-junction based on three peaks stable with respect to variations in window size, detected in the grey-level and directional histograms respectively. Figure 29 shows corresponding results for an L-junction.

To summarize, this experiment indicates how the scale-space primal sketch can be used in dynamic situations like focus-of-attention. Such mechanisms are necessary if computer vision systems are to perform their tasks in a complex and dynamic world. It should be emphasized that the treatment here describes on-going experimental work, and that there is still more work to be done concerning control strategies of the reasoning process. Nevertheless, the presentation illustrates some basic ideas of how the suggested approach can be used in an active vision situation, and specifically, how qualitative scale and region information can be used for guiding a junction detection module by detecting curvature blobs from grey-level data.

7 Other Possible Applications

Let us finally mention a few other problem areas where the approach can be applicable.

Texture Analysis

A basic problem in many shape-from-texture algorithms concerns how to detect texture elements (Julesz and Bergen [30]; Blostein and Ahuja [10]; Vorhees and Poggio [67]; Gårding [23]). Since the scale-space primal sketch does not require any prior scale information, and scale levels can be automatically adapted to size variations in image data, it is a useful tool for such analysis.

Figures 30–31 show experimental results for one synthetic and two realistic images. Note that in both cases a set of blobs is extracted with a size gradient that can be used as a cue to the three-dimensional structure.[18]

Perceptual Grouping

In the presented experiments we have seen that the blobs extracted from the scale-space primal sketch often induce intuitively reasonable groupings of various patterns. For example, in figure 31(a–c) in principle only the individual squares were ranked as important, while in figure 31(d–f) vertical stripes were also found. See also the dot pattern example in figure 32. Note that the grouping process is not given by any of the prespecified logical rules, but by a differential equation combined with a set of geometric constructions.

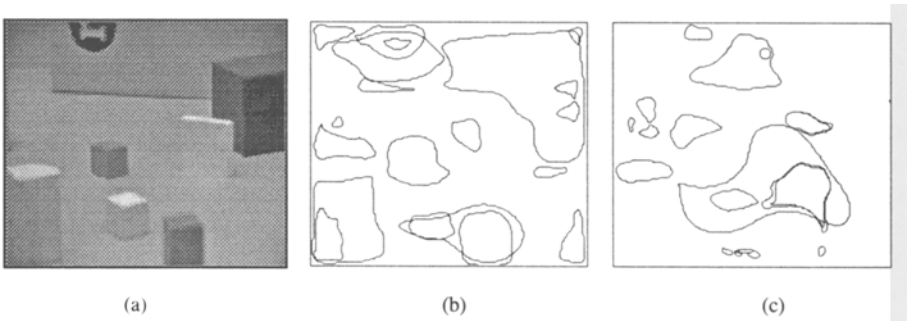

Fig. 26. (a) Overview image of a scene under study. (b–c) Boundaries of the 20 most significant dark and bright scale-space blobs respectively.

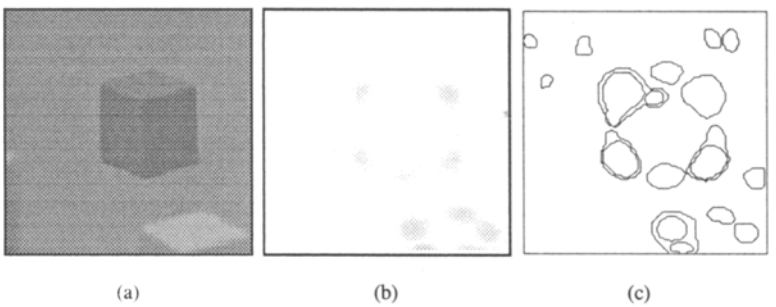

Fig. 27. Zooming in to a region of interest given by a scale-space blob from the previous processing step. (a) A window around the region of interest, set from the location and the size of the blob. (b) The rescaled level-curve curvature computed at the scale of the blob. (c) The boundaries of the 20 most significant curvature blobs obtained from blob detection in the curvature data.

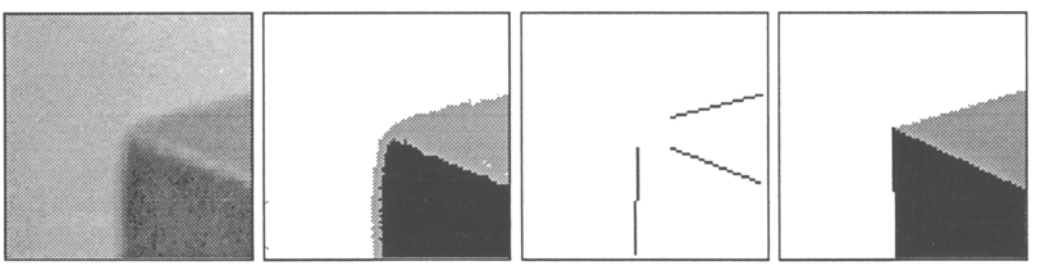

Fig. 28. Zooming in to a junction candidate given by a curvature blob. (Left) Maximum window size for the focusing procedure set from the size of the curvature blob. (Middle left) Backprojected peaks from the intensity histogram. (Middle right) Lines computed from the backprojected peaks from the directional histogram. (Right) Schematic illustration of the classification result in which a simple junction model has been adjusted to the data. (This junction candidate was classified as a 3-junction.)

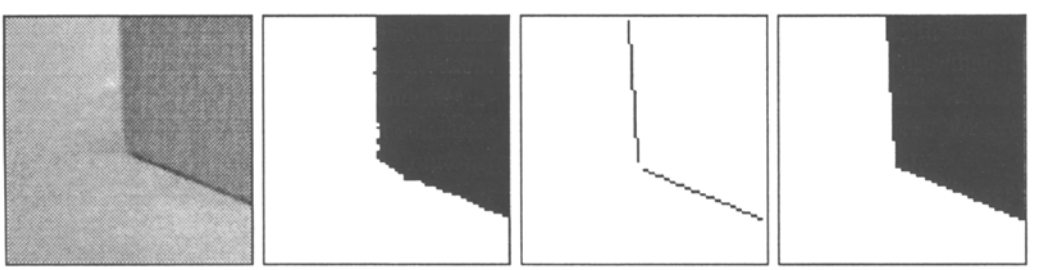

Fig. 29. Similar classification result for an *L*-junction.

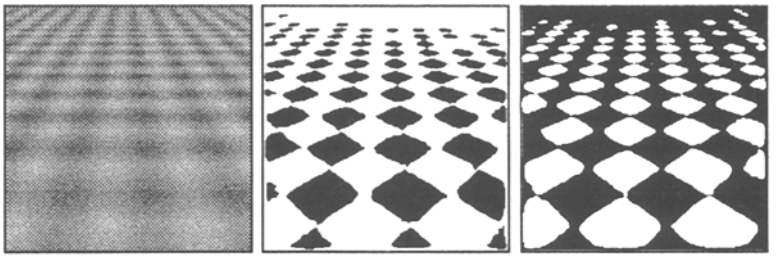

Fig. 30. Multiscale blob detection on a synthetic-texture image generated from perspective projection of a planar surface with a sinusoidal grey-level pattern. (a) Grey-level image with added white Gaussian noise with standard deviation 10% of grey-level range. (b) The 75 most significant dark blobs. (c) The 75 most significant bright blobs.

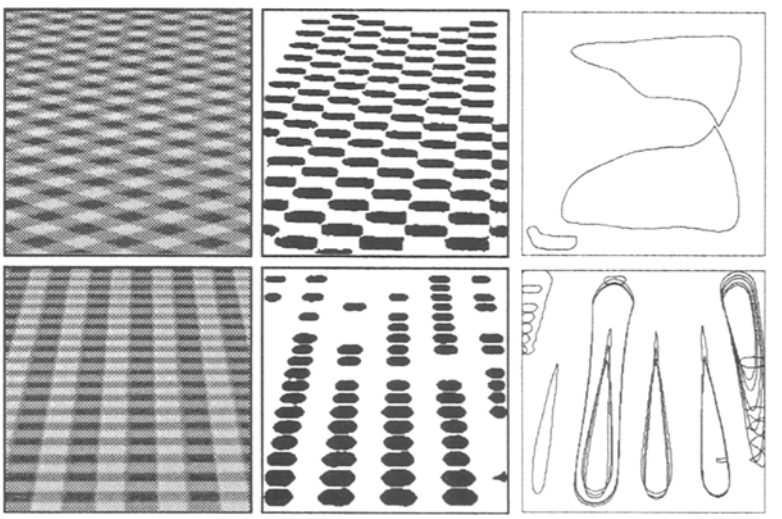

Fig. 31. Multiscale blob detection applied to two different views of a real-world surface texture. (Left) Grey-level image. (Middle and right) The 100 most significant dark blobs (marked either as blob regions or blob boundaries).

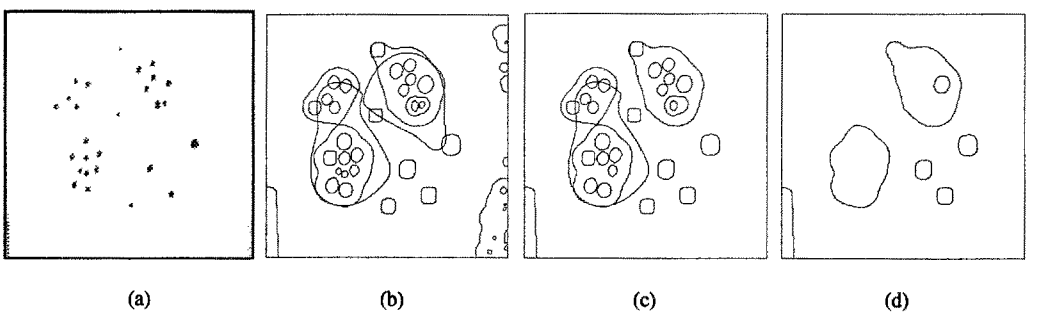

Fig. 32. Multiscale blob detection on a dot pattern image. (a) Original grey-level image. (b) Boundaries of the 50 most significant blobs. (c) Low threshold on the significance measure. (d) High threshold on the significance measure. Note that all dots are detected and that a number of intuitive groupings are performed.

Object Detection and Matching

The blobs delivered from the scale-space primal sketch can serve as coarse landmarks for different types of matching purposes. The relation given by, say, matches between a blob and a set of edges and junctions, provides a sparse set of features, which could be used, for example, for delimiting the search space in model matching. Another possible application is to use the blobs for initiating object models, like deformable models (Kass et al. [31]; Terzopoulos et al. [65]; Pentland [58]), or geon primitives (Biedermann [6]; Dickinson et al. [19]). Experimental work indicates that the approach may be useful for establishing coarse correspondences in sequence data; see also (Koller et al. [37]). Blobs are conceptually easy to match over time based on spatial overlap.

8 Summary and Discussion

The proposed representation is similar to the primal sketch suggested by Marr [53, 54], in the sense that it is a two-dimensional representation of the significant grey-level structures in the image. It is also computed under extremely weak assumptions. And, although it is a region-based and not an edge-based representation, it is more *qualitative*, without strong assumptions about the shapes of the primitives. It consists of blobs (extremum regions) at multiple scales in scale-space, and allows for

- automatic detection of salient (stable) scales, if they exist,
- ranking of blob-like structures in order of significance, and
- generation of hypotheses for grouping and segmentation.

This implies that candidate regions are generated for further processing, as well as information about the scale. We have seen that the proposed representation gives clues to subsequent analysis, and that it can guide focus-of-attention mechanisms. At the same time it is obtained with no a priori assumptions, and, in principle without tuning parameters. The only free parameter is the number of blobs to be selected for further analysis (or display).

The underlying principle used for extracting blob structures is that structure should be invariant under transformations in parameter space. The suggested method consists of three steps; (i) vary the parameter systematically, (ii) detect (locally) stable states (intervals), (iii) choose a representative descriptor as an abstraction of each stable interval, and pass only this information on to the higher-level modules. In this specific case, the parameter that is varied is the scale parameter in the scale-space representation, and the significance measure is defined in terms of a four-dimensional volume in scale-space. The methodology can, however, be applicable also in other types of situations. One example, concerning junction classification, is described in section 6.

8.1 Scale-Space Experiences

Let us finally point out a few aspects of scale-space representation that have been given little or insufficient attention in the literature, and have to be dealt with when building a representation of the type proposed.

Suppression of Local Extrema Due to Noise. First, it is noteworthy, that the amount of noise in real images usually leads to a large number of local extrema. These extrema may disappear rather early if they are subsumed by some more prominent extremum. However,

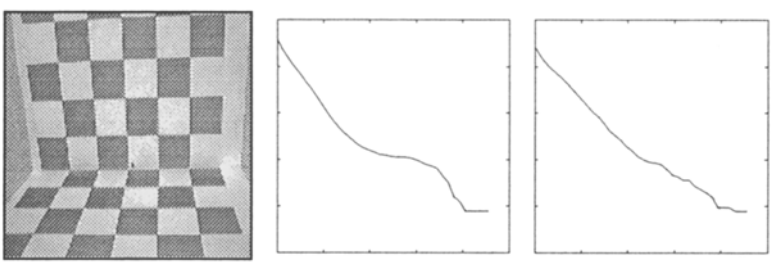

Fig. 33. (a) An unusual situation, where one could possibly talk about a global stable scale for a whole image. (b) This property manifests itself as a plateau in a graph showing the logarithm of the number of local extrema as a function of (effective) scale. (c) For realistic images of moderate complexity it will, however, usually not be possible to find such globally stable states. Even if there were a number of prominent plateaus corresponding to locally stable structures at different scales, adding a large number of such profiles would give a relatively uniformly decreasing curve.

if they are located regions with smoothly varying grey-level, they will exist over a large range of scale. This effect is alleviated, but not remedied, by annihilation between nearby noise extrema. Even though the amplitude can be expected to decrease rapidly, it is not clear that a globally valid threshold can be set on objective grounds. This problem is related to the issue of estimating the noise level in an image, which hardly can be addressed without some constraining assumptions, as in Voorhees and Poggio [67].

Stable Scale Is a Local Property. Another property, indicated in section 2.5, is that images of scenes of even moderate complexity rarely have a global scale, at which all structure above the noise level is present (figure 33). Stable scales are local properties associated with objects, not with entire images. This aspect is explictly dealt with in the proposed representation.

Stable Scale Is a Multivalued Function. Moreover, given some region in space there may be several stable scales associated with that region, corresponding to structures at different scales. Therefore, the task of finding "a best scale" for treating a certain *point* is in general an impossible problem (except for very simple images, for which there is only one such stable scale associated with each point in the image).

In this context it should be remarked that the scale value given by assumption 2 does not necessarily reflect the size of the corresponding blob region in the image. Although, in general, large values of the scale parameter can be expected to correspond to large-scale structures, there is no *direct* relationship. Under certain conditions (typically when there are no superimposed finer-scale structures) a large-scale structure may, in fact, be assigned a small-scale value. Therefore, the scale value obtained from assumption 2 should rather be interpreted as an abstract scale parameter, *reflecting the smallest amount of smoothing for which the blob manifests itself as a single blob entity.*

Decreasing Amplitude of Feature Points. The behavior of local extrema in scale-space has been studied also by Lifshitz and Pizer [40]. They link points across scales based on *iso-intensity*, using integral paths of the vector field $\left(L_{x_1} L_t, L_{x_2} L_t, -(L_{x_1}^2 + L_{x_2}^2)\right)$, and construct a "stack" representation, in which the grey-level at which an extremum disappears is used for defining a region in the original image by local thresholding at that grey-level. The representation is demonstrated to be applicable for certain segmentation problems in medical image analysis. However, Lifshitz and Pizer observe the serious problem of *noncontainment*. It essentially means that a point, which at one scale has been classified as belonging to a certain region (associated with a local maximum), can escape from the region when the scale parameter increases. Moreover, such paths can be intertwined in quite a complicated way.

The main cause of the problem in iso-intensity linking is that grey-levels, corresponding to features tracked over scales, will *change* under scale-space smoothing.[19] For example, concerning a local extremum, it is a necessary consequence of the diffusion equation that the grey-level at a maximum point must decrease with scale. This problem is avoided in the scale-space primal sketch, in which the linking is explicitly based on qualitative feature points (here, local extrema).

8.2 Relations to Previous Work

There are earlier attempts to derive similar representations of the grey-level landscape. Rosenfeld and his coworkers (Gross [26]; Sher and Rosenfeld [63]) have studied blob detection in pyramids, for example, using relaxation methods. Blostein and Ahuja [10] detect texture elements based on zero-crossings at multiple scales and a significance measure based on a background noise assumption. There is also a wealth of literature on pyramids (see, e.g., Levine [39]; Burt [13]; Crowley and Parker [16]; Crowley and Sanderson [17]). The texton theory (Julesz and Bergen [30]; Voorhees and Poggio [67]) essentially also treats the blob-detection problem. There are finally a number of representations based on intensity changes (Marr [54]; Bergholm [5]; Watt [68]; Baker [4]); and approaches working at higher levels, like the token based-symbolic grouping by Saund [62]. Also of interest is the approach by Haralick et al. [27], which allows a more detailed representation, but only at a single spatial scale.

The suggested approach differs from these in three important aspects. Firstly, it can be seen as preceding the edge-based schemes in that it selects the appropriate scales and regions, intrinsically defined by the image itself, in a complementary data-driven manner. Secondly, it is a hierarchical representation of image structures at *all scales* with explicit information about their significance and relations, and a competition between parts at different locations and scales. Finally, it is

derived in a formal way using the well-defined notion of scale-space, which allows a precise study of events at different scales.

One can ask more generally, what is the relation between the suggested representation and the zero-crossings of the Laplacian. Given a function $f: \mathbb{R}^2 \rightarrow \mathbb{R}$, define a bright (dark) *Laplacian sign blob* as a connected region satisfying $\nabla^2 L < 0$ (> 0). Since at any local maximum (minimum) it holds that $\nabla^2 L < 0$ (> 0), it follows that to every grey-level blob there is a unique Laplacian sign blob of the same polarity. However, the reverse relation is not valid; given a bright (dark) Laplacian sign blob, there may be one maximum (minimum), several maxima (minima), or even no maximum (minimum) in that region.

In fact, a representation similar to the grey-level-blob tree at a single scale has been studied independently by Blom [9], who considers the nesting structure of level curves through critical points. The major differences are that Blom considers a degenerate (non-Morse) critical point, and that he points out that a hexagonal discrete grid has certain theoretical advantages. In this work, grey-level volumes are associated with the different arcs of the nesting tree, and the representation is embedded in scale-space.

8.3 Invariance Properties

Since the scale-space primal sketch is defined solely in terms of topological properties as local extrema, level curves through saddle points, and bifurcations between critical points, it obeys a number of natural invariance properties. Invariance with respect to *translations* and *rotations* of the spatial domain is trivial. Further, given a certain scale level, the topological relations of the grey-level blob tree are preserved under arbitrary monotone intensity transformations. Under evolution in scale-space, the invariance of the hierarchical relations is restricted to *linear intensity transformations*. Such transformations also leave the relative ranking of blobs on significance unaffected. Trivially, under *uniform rescalings* of the spatial coordinates, $x \mapsto sx$ ($s \in \mathbb{R}_+$), a singularity at a point $(x_0; t_0)$ in the scale-space representation of the original signal is transferred to a new point $(sx_0; s^2 t_0)$. This means that the hierarchical relations are preserved, and the appearance and disappearance scales of the scale-space blobs are multiplied by constant factors. Concerning the ranking on significance, it is clear that the logarithmic measure τ_{eff} is invariant to uniform rescalings, and hence the scale-space lifetime. The intention with the transformation function V_{trans} is that also the integrand should be well behaved under this operation.[20]

8.4 Alternative Approaches and Further Work

Let us finally mention a few issues that are subject to future work:

Normalization. In the current implementation, the normalization of the scale parameter and the grey-level blob volume has been based on white noise data. The reason for this is that it constitutes a conservative choice, and makes theoretical analysis simple. If statistics are accumulated on how blobs in such data can be expected to behave over scales, then the result is an estimate of to how large extent accidental groupings take place in scale-space. By experiments, this normalization based on white-noise images has been demonstrated to give reasonable results. Moreover, concerning scale-space lifetime, it has been theoretically shown, that for continuous signals such reference data gives rise to the same transformation function ($\tau \sim \log t$) as other self-similar distributions (section 2.6.1). This selection should, however, not be interpreted as excluding that other approaches, which are equivalent in the continuous case, may lead to different results for discrete signals. For example, some interesting alternatives to consider would be (i) to let the vision system accumulate statistics for a large (representative) selection of different types of realistic imagery, or, (ii) if possible, consider some discrete analogue of colored noise with a (scale invariant) Fourier spectrum of the form $|\omega|^{-N}$, where N denotes the dimension.

A possible problem with the subtraction of the mean grey-level blob volume (18) is that it makes the normalized grey-level blob volume sensitive to the actual scaling of the data. Therefore, in order to reduce this sensitivity, the tabulated values are rescaled linearily from least-squares fit between the real and the tabulated values. A possible way to avoid this problem, and also to avoid the heuristically chosen transformation function in (19), is by redefining the normalized grey-level blob volume as

$$V_{\text{eff}} = \frac{G_{\text{vol}}}{V_m(t)} \tag{24}$$

and then taking as normalized significance values

$$S_{\text{eff}} = \frac{S_{\text{vol}} - S_m(t)}{S_\sigma(t)}, \tag{25}$$

where S_m and S_σ denote mean values and standard deviations of scale-space blob volumes computed from reference data. This method has not yet been implemented, mainly because the simulation work for building the normalization tables is much larger.

Multiple Blob Instances. The scale-space primal sketch leads to separate systems of bright and dark scale-space blobs. Moreover, a spatial region may give rise to multiple blob responses; typically as the result of a large blob merging with a smaller blob and forming a new scale-space blob. An obvious problem concerns how to integrate blobs at different polarity and at different scales. In general, it is argued that this problem can hardly be addressed in isolation, but has to be related to a visual task. The following are some basic properties that can be used by a reasoning system.

Given a fixed level of scale, the problem of integrating bright and dark blobs can be approached by considering the grey-level blob tree, which constitutes the natural link between grey-level blobs of reverse polarity; see also Blom [9]. Concerning the behavior over scales, it is clear that a tree describing bifurcations between scale-space blobs will be strongly coupled to the grey-level blob tree. For example, for the simple noise-free pattern in figure 1, a tree describing the bifurcations between scale-space blobs can be expected to be identical to the grey-level blob tree of the original signal. In the presence of noise, however, the hierarchical relations will be different. More generally, blob splits and blob creations are blob events without correspondences in the grey-level blob tree.

Other natural descriptors to define between different blobs are (i) whether two blobs overlap, and (ii) whether one blob is completely contained in another one. In this way, the problem with multiple responses from a single region may be approached. This is, however, a subject for further analysis and experimentation.

9 Conclusions

A multiscale representation of grey-level image structure has been presented similar to the primal sketch idea. It can be used for extraction of important blob-like regions from an image in a solely bottom-up data-driven way, without any a priori assumptions about the shape of the primitives. The representation, which is essentially free from tuning parameters and ad hoc error critieria, gives a qualitative description of the grey-level landscape with information about *approximate locations*, *spatial extent*, and an *appropriate scale* for relevant regions in the image. In other words, it generates coarse but safe segmentation cues, and can serve as a hypothesis generator for higher-level processes. It has been demonstrated how such information can serve as a guide to an edge-detection scheme working at a locally adapted level of scale and that it is applicable for automatic cluster detection, modality analysis of histograms, as well as junction detection and junction classification. More generally, the approach provides a mechanism for focus-of-attention and for guiding other low-level processes.

The methodology is based on a number of postulates (assumptions 1–4 in section 3) stated without proof. Their interpretation essentially is that in the absence of further information (i) significance blob-like structures in scale-space are likely to correspond to significant regions in the image, and (ii) that scale levels should be selected where the normalized blob response is maximal. Starting from these assumptions and scale-space theory, several theoretical results have been obtained. Moreover, by integrating the scale-space primal sketch with several other visual modules and by applying the methodology to different types of images, it has been experimentally demonstrated that the proposed methodology gives intuitive results, and that it generates highly useful results for further processing. This is the main support for the validity of the approach.

Acknowledgments

I would very much like to thank Jan-Olof Eklundh for excellent supervision during the initial development of this work, and Kjell Brunnström for the enjoyable collaboration on the junction classification problem. I would also like to thank Harald Winroth for several valuable discussions, Fredrik Bergholm for useful comments when implementing the edge focusing algorithm, Demetrios Betsis for valuable suggestions about this manuscript, as well as Jonas Gårding and Hans Tråven for providing the texture and fruit bowl images respectively.

This work was partially performed under the ESPRIT-BRA project InSight. The support from the Swedish National Board for Industrial and Technical Development, NUTEK, is gratefully acknowledged.

Notes

1. Unless otherwise stated, the signals are throughout assumed to be *Morse*, i.e., all critical points are assumed to nondegenerate, and all critical values are assumed to be distinct.
2. It turns out that some transformation of the scale parameters is necessary in order to capture the concept of scale-space lifetime properly (see section 2.6).
3. This behavior of the grey-level blobs over scales may be regarded as somewhat complex by a reader unfamiliar with these concepts. A detailed theoretical analysis is given by Lindeberg [45].
4. Of course, the contrast of such blobs decreases, but it is far from clear that it is possible to set a threshold on objective grounds.
5. Except for the fact that scale-space smoothing favors blob-like bell-shaped objects.
6. The word *individual* is emphasized here, since stable scales when they exist are, in general, local properties associated with objects (or parts of objects)—not with entire images. However, the assumption of a globally stable scale is sometimes used implicitly in computer vision algorithms, for example, when edge detection is performed using uniform smoothing all over an image. Instead, it is argued that better performance can be obtained by adapting the scale levels to the local image structure, see sections 4–6 for examples.
7. Note in this context that if two blobs (B_2 and B_3 above) are closely located, there will in general be a large blob corresponding to the union of these two blobs at coarser scales in scale-space. Hence, although the smaller one of these blobs (B_3 above) may be assigned a small significance value, the union of these two blobs will be assigned a larger significance value, and hence attract the focus-of-attention to the union of the two adjacent structures.
8. A brief review of the scale-space concept for discrete signals is given in appendix A.3.
9. For one-dimensional signals, the number of local extrema in a signal is guaranteed to decrease monotonically with scale. In two and higher dimensions the situation is more complicated, since the number of local extrema can in fact increase locally with scale-space smoothing due to creations of saddle-extremum pairs. However, the expected number of local extrema, treated as an average over many signals can always be expected to decrease.
10. As explained in previous sections, grey-level blob trees can be treated in a similar way. Since, however, the problem of normalizing the spatial and grey-level coordinates has so far been studied only concerning the grey-level blobs and the scale-space blobs, the remaining part of this presentation will be concerned with these objects.
11. For example, it seems plausible that the lifetime of an edge in scale-space is an important property for measuring significance. As will be demonstrated in section 6, a multiscale blob-detection approach can be useful in junction detection, provided that a proper transformation of the intensity domain is performed as preprocessing.
12. It should be stressed that no claims are made here that the proposed approach is an appropriate description of the mechanisms in biological vision. When relations to biological vision are discussed, it is only as a source of inspiration.
13. More precisely, at blob creations the grey-level blob volume of the new blob is always zero. At blob splits, the grey-level blob volume of the blob associated with the new local extremum is zero, while the volume of the other blob may be nonzero (see the polynomial representative of the fold singularity (32)).
14. Edges are defined as the ridges of the gradient magnitude map, i.e., the points for which the gradient magnitude assumes a maximum in the gradient direction.
15. For implementational purpose, this characteristic length is determined as the square root of an experimentally determined blob area $A_m(t)$ at scale t. It is accumulated in the same way as the statistics of the grey-level blob volume $V_m(t)$.
16. Clearly, the problem of relating a blob to edges becomes trivial if a separate focusing process is invoked for each scale-space blob.
17. The color images have been converted from the usual RGB format to the CIEu^*v^* 1976 format (Billmeyer and Saltzman [7]), which separates the intensity and the chroma information. The histograms are formed only over the (two-dimensional) chroma information, ignoring the (one-dimensional) intensity information.
18. A shape-from-texture method using a (simplified) blob-detection method of this type is presented by Lindeberg and Gårding [50].
19. A similar problem arises in the motion-constraint equation for optical flow, where it is usually assumed that the intensity value of a physical point is preserved under motion. However, as Pentland [59] has demonstrated, the photometric distortions can under certain conditions be much larger than the geometric effects due to motion.
20. If a perfectly scale-invariant reference signal could be determined, then a scale-invariant normalization would be trivially obtained. This is, however, very hard to accomplish on a discrete grid, which has a certain preferred scale given by the distance between grid points.
21. Generically, these events occur at isolated scales, and only two different critical points have the same critical values.
22. In one dimension, the only possible events are annihilations of pairs consisting of one maximum point and one minimum point—the number of local extrema in the scale-space of a one-dimensional signal is always guaranteed to decrease with scale.

Appendix

A.1 Definition of Scale-Space Blob

This appendix section describes how a scale-space blob is formally defined from a two-dimensional continuous signal.

A.1.1 Extremum Path and Saddle Path. Consider a critical point $x_0 \in \mathbb{R}^2$ at some scale $t_0 \in \mathbb{R}_+$ in scale-space. It is given by

$$\nabla L|_{(x_0;t_0)} = \left(\begin{array}{c} \partial_{x_1} L \\ \partial_{x_2} L \end{array} \right) \Bigg|_{(x_0;t_0)} = \left(\begin{array}{c} 0 \\ 0 \end{array} \right) \qquad (26)$$

The implicit function theorem ensures that if the Hessian matrix

$$\mathcal{H}L|_{(x_0;t_0)} = \begin{pmatrix} \partial_{x_1x_1} L & \partial_{x_1x_2} L \\ \partial_{x_1x_2} L & \partial_{x_2x_2} L \end{pmatrix} \Bigg|_{(x_0;t_0)} \tag{27}$$

is nonsingular at this point, then there exists some smooth function $r_0 : I_{t_0} \rightarrow \mathbb{R}^2$

$$x = r_0(t) \tag{28}$$

such that $x_0 = r_0(t_0)$, and for every t in some neighborhood I_{t_0} of t_0 the point $(r_0(t); t)$ is a critical point for the mapping $x \mapsto L(x; t)$. By continuation, such local paths can be extended to curves as long as the Hessian matrix remains nonsingular. It can be easily shown that the type of critical point remains the same as long as the Hessian matrix is nonsingular.

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

A.1.2 Scale-Space Blob. Concerning grey-level blobs, this result means that a unique linking of grey-level blobs across scales can be performed as long as both the extremum point and the saddle point determining the extent of the grey-level blob remains nonsingular. In summary, a scale-space blob is defined as the union of all grey-level blobs associated with the extremum points along a segment of an extremum path where such a unique linking can be performed.

In order to express this statement precisely, let $[t'_{\min}, t'_{\max}] \in [t_{\min}, t_{\max}]$ be a (maximal) subset of an extremum path, along which the delimiting saddle point $S_{\text{delim}}(r_0(t))$ associated with the extremum point $r_0(t)$ is always nondegenerate. At some distinct scales it may happen that the delimiting saddle point jumps from one saddle path to another. In such non-Morse situations, when two saddle points have the same grey-level, both saddle points are required to be nondegenerate.[21] At the end points, either of $r(t'_{\min})$ and $S_{\text{delim}}r(t'_{\min})$ and also either of $r(t'_{\max})$ and $S_{\text{delim}}r(t'_{\max})$ are degenerate critical points (unless $t_{\min} = 0$ or $t_{\max} = \infty$).

Then, the scale-space blob associated with this segment $r'_0 : [t'_{\min}, t'_{\max}] \rightarrow \mathbb{R}^2$ is the set

$$\begin{aligned} S_{\text{blob}}(r'_0) = \overline{\{(x, z; t) \in \mathbb{R}^2 \times \mathbb{R} \times \mathbb{R}_+} \\ \overline{: (t'_{\min} < t < t'_{\max})} \\ \overline{\wedge [(x, z) \in G_{\text{blob}}(r'_0(t))]\}} \end{aligned} \tag{29}$$

where $G_{\text{blob}}(r'_0(t))$ is the grey-level blob associated with the extremum point $r'_0(t)$ in the scale-space representation L at scale t.

A.1.3 Scale-Space Blob Volume. Strictly, in this coordinate system the scale-space blob volume is

$$\begin{aligned} S_{\text{vol}}(r'_0) &= \int_{(x,z;t) \in S_{\text{blob}}(r_0)} dx\, dz\, dt \\ &= \int_{t \in [t'_{\min}, t'_{\max}]} G_{\text{vol}}(r'_0(t))\, dt \end{aligned} \tag{30}$$

However, when the scale-space blob volume is to be used as a significance measure in the scale-space primal sketch, it turns out that some transformations must be performed. One would like structures at different scales to be treated uniformly, such that the significance measure neither favors fine scales over coarse scales nor the opposite. Therefore, the *normalized scale-space blob volume* is defined by

$$\begin{aligned} &S_{\text{vol,norm}}(r'_0) \\ &\quad = \int_{t \in [t'_{\min}, t'_{\max}]} V_{\text{trans}}[G_{\text{vol}}(r'_0(t)); t]\, d(\tau_{\text{eff}}(t)) \end{aligned} \tag{31}$$

where $\tau_{\text{eff}} : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is a transformation function mapping the ordinary scale parameter t to a transformed scale parameter τ called *effective scale* (see section 2.6.1), and $V_{\text{trans}} : \mathbb{R} \times \mathbb{R}_+ \rightarrow \mathbb{R}$ is a corresponding transformation function normalizing the variations of the grey-level blob volumes into a more uniform behavior over scales (see section 2.6.2).

A.2 Scale-Space Blob Events

The implicit function theorem used in a previous appendix section guarantees that linking of nondegenerate critical points is a well-defined operation. When the Hessian matrix becomes singular, *bifurcations* may

occur. Useful techniques for analyzing the behavior at such points can be obtained from a branch of mathematics known as *singularity theory*; see Poston and Stewart [60] and Gibson [24] for application-oriented introductions, and Arnold et al. [2] and Golubitsky and Schaeffer [25] for more rigorous treatments of the subject.

In summary, the following result holds concerning the behavior of critical points in scale-space. In two dimensions, the only generic (structurally stable) bifurcations are annihilations and creations of pairs consisting of one extremum point and one saddle point (Koenderink and van Doorn [34]; Lifshitz and Pizer [40]; Lindeberg [45]; Johansen [29]).[22] A natural model of this so-called *fold singularity* is the polynomial

$$x_1^3 + 3x_1(t - t_0) \pm (x_2^2 + t - t_0), \qquad (32)$$

which also satisfies the diffusion equation. The positions of the critical points are given by

$$(x_1(t), x_2(t)) = \pm(\sqrt{t_0 - t}, 0) \qquad (t \leq t_0) \qquad (33)$$

that is, the critical points merge along a parabola. At the bifurcation point, the drift velocity tends to infinity. This property demonstrates that any algorithm for following extrema over scales needs a mechanism for adaptive scale sampling.

Concerning scale-space blobs, this classification means that two distinct types of cases can be distinguished, depending on whether the saddle point involved in the bifurcation is part of one or two grey-level blobs. A saddle point delimiting the extent of only one grey-level blob is said to be *nonshared*, while a saddle point belonging to two grey-level blobs is said to be *shared*. Hence, in the generic case, the following four cases are possible at a structurally stable bifurcation (see figure 6 for an illustration, and Lindeberg [43, 45] for a more extensive description):

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

A.3 Scale-Space for Discrete Signals

Given a discrete signal $f: \mathbb{Z}^2 \to \mathbb{R}$, the scale-space $L: \mathbb{Z}^2 \times \mathbb{R}_+ \to \mathbb{R}$ is for some $\gamma \in [0, 1]$ defined as the solution to the semidiscretized diffusion equation

$$\partial_t L = \frac{1}{2} \nabla_\gamma^2 L = \frac{1}{2} [(1 - \gamma) \nabla_5^2 L + \gamma \nabla_\times^2 L)] \qquad (34)$$

where the five-point operator ∇_5^2 and the cross-operator $\nabla_\times^2$ are two common discrete approximations to the Laplacian operator given by (below the notation $f_{-1,1}$ stands for $f(x_1 - 1, x_2 + 1)$ etc.):

$$(\nabla_5^2 f)_{0,0} = f_{-1,0} + f_{+1,0} + f_{0,-1} + f_{0,+1} - 4 f_{0,0},$$

$$(\nabla_{\times^2}^2 f)_{0,0}$$

$$= \frac{1}{2} (f_{-1,-1} + f_{-1,+1} + f_{+1,-1} + f_{+1,+1} - 4 f_{0,0}).$$

In the special case when $\gamma = 0$, the two-dimensional scale-space is given by separable convolution with the one-dimensional discrete analogue of the Gaussian kernel

$$T(n; t) = e^{-t} I_n(t) \qquad (35)$$

where $I_n(t)$ denotes the modified Bessel functions of integer order (Abramowitz and Stegun [1]). This is the scale-space concept (Lindeberg [41, 47]) that underlies all implementations described in this presentation.

A.4 Algorithmic Aspects

When building the scale-space primal sketch representation of an image, there are several computational aspects that need to be treated. The algorithm for building the suggested data structure consists of two major modules; an algorithm for finding grey-level blobs (or grey-level blob trees) at a single level of scale, and an adaptive scale linking and refinement procedure. Because of lack of space, the algorithm cannot be described here. An early description can be found in Lindeberg and Eklundh [44], while a more complete treatment is given in Lindeberg [43].

References

1. M. Abramowitz and I.A. Stegun, eds., *Handbook of Mathematical Functions*. Applied Mathematics Series, National Bureau of Standards, 55 ed., 1964.
2. V.I. Arnold, S.M. Gusein-Zade, and A.N. Varchenko, *Singularities of Smooth Maps*, vol. I, Birkhäuser: Boston, 1985.

3. J. Babaud, A.P. Witkin, M. Baudin, and R.O. Duda, Uniqueness of the Guassian kernel for scale-space filtering, *IEEE Trans. Patt. Anal. Mach. Intell.* 8(1):26–33, 1986.
4. H.H. Baker, Surface reconstruction from image sequences, *Proc. 2nd Intern. Conf. Comput. Vis.* Tampa, FL, pp. 334–343, 1988.
5. F. Bergholm, Edge focusing, *IEEE Trans. Patt. Anal. Mach. Intell.* 9:726–741, November 1987.
6. I. Biederman, Human image understanding: Recent research and a theory, in *Human and Machine Vision II*, Academic Press: San Diego, pp. 13–57, 1985.
7. F. Billmeyer and M. Saltzman, *Principles of Colour Technology*, Wiley: New York, 1982.
8. W.F. Bischof and T. Caelli, Parsing scale-space and spatial stability analysis, *Comput. Vis. Graph. Image Process.* 42:192–205, 1988.
9. J. Blom, Topological and Geometrical Aspects of Image Structure, Ph.D. thesis, Dept. Med. Phys. Physics, Univ. Utrecht, NL-3508 Utrecht, Netherlands, 1992.
10. D. Blostein and N. Ahuja, Shape from texture: integrating texture element extraction and surface estimation, *IEEE Trans. Patt. Anal. Mach. Intell.*, 11:1233–1251, 1989.
11. K. Brunnström, J.O. Eklundh, and T. Lindeberg, Scale and resolution in active analysis of local image structure, *Image Vis. Comput.* 8:289–296, 1990.
12. K. Brunnström, T. Lindeberg, and J.O. Eklundh, Active detection and classification of junctions by foveation with a head-eye system guided by the scale-space primal sketch, *Proc. 2nd Europ. Conf. Comput. Vis.* Santa Margherita Ligure, Italy, May 1992. In G. Sandini, ed., vol. 588 of *Lecture Notes in Computer Science*, pp. 701–709, Springer-Verlag.
13. P.J. Burt, Fast filter transforms for image processing, *Comput. Graphics, Image Proc.* 16:20–51, 1981.
14. J. Canny, A computational approach to edge detection, *IEEE Trans. Patt. Anal. Mach. Intell.* 8(6):679–698, 1986.
15. M.J. Carlotto, Histogram analysis using a scale-space approach, *IEEE Trans. Patt. Anal. Mach. Intell.* 9:121–129, 1987.
16. J.L. Crowley and A.C. Parker, A representation for shape based on peaks and ridges in the Difference of Low-Pass Transform, *IEEE Trans. Patt. Anal. Mach. Intell.* 6(2):156–170, 1984.
17. J.L. Crowley and A.C. Sanderson, Multiple resolution representation and probabilistic matching of 2-D grey-scale shape, *IEEE Trans. Patt. Anal. Mach. Intell.*, 9(1):113–121, 1987.
18. S.M. Culhane and J.K. Tsotsos, An attentional prototype for early vision, *Proc. 2nd Europ. Conf. Comput. Vis.*, Santa Margherita Ligure, Italy, 1992, pp. 551–562.
19. S.J. Dickinson, A.P. Pendland, and A. Rosenfeld, "Qualitative 3-D reconstruction using distributed aspect graph matching," in *Proc. 3rd Intern. Conf. Comput. Vis.*, Osaka, Japan, 1990, pp. 257–262.
20. R.W. Ehrich and P.F. Lai, Elements of a structural model of texture, *Proc. PRIP*, pp. 319–326, IEEE Computer Society Press: Los Alamitos, CA, 1978.
21. L.M.J. Florack, B.M. ter Haar Romeny, J.J. Koenderink, and M.A. Viergever, Scale and the differential structure of images, *Image Vis. Comput.* 10:376–388, 1992.
22. M.A. Förstner and E. Gülch, A fast operator for detection and precise location of distinct points, corners and centers of circular features, in *ISPRS Intercommission Workshop*, 1987.
23. J. Gårding, Shape from surface markings. Ph.D. thesis, Dept. of Numerical Analysis and Computing Science, Royal Institute of Technology, Stockholm, 1991.
24. C.G. Gibson, *Singular Points of Smooth Mappings*, Research Notes in Mathematics, Pitman Publishing: London, 1979.
25. M. Golubitsky and D.G. Schaeffer, *Singularities and Groups in Bifurcation Theory I*, vol. 51 of *Applied Mathematical Sciences*, Springer-Verlag: New York, 1985.
26. A.D. Gross, Multiresolution object detection and delineation, Tech. Rept. TR-1613, Computer Vision Laboratory, University of Maryland, 1986.
27. R.M. Haralick, L.T. Watson, and T.J. Laffey, The topographic primal sketch, *Intern. J. Robotics Res.* 2(1):50–72, 1983.
28. R. Hummel, The scale-space formulation of pyramid data structures, *Parallel Computer Vision*, L. Uhr, ed., pp. 187–223, Academic Press: New York, 1987.
29. P. Johansen, On the classification of toppoints in scale space, *J. Math. Imag. Vis.* (to appear), 1993.
30. B. Julesz and J.R. Bergen, Textons, the fundamental elements in preattentive vision and perception of textures, *Bell Syst. Tech. J.* 62(6):1619–1645, 1983.
31. M. Kass, A. Witkin, and D. Terzopoulos, Snakes: Active contour models, *Intern. J. Comput. Vis.* 1:321–331, 1987.
32. L. Kitchen and A. Rosenfeld, Gray-level corner detection, *Patt. Recog. Lett.* 1(2):95–102, 1982.
33. J.J. Koenderink, The structure of images, *Biological Cybernetics* 50:363–370, 1984.
34. J.J. Koenderink and A.J. van Doorn, Dynamic shape, *Biological Cybernetics* 53:383–396, 1986.
35. J.J. Koenderink and W. Richards, Two-dimensional curvature operators, *J. Opt. Soc. Amer.* 5(7):1136–1141, 1988.
36. J.J. Koenderink and A.J. van Doorn, Generic neighborhood operators, *IEEE Trans. Patt. Anal. Mach. Intell.* 14:597–605, June 1992.
37. D. Koller, K. Daniilides, T. Thórhallson, and H.-H. Nagel, Model-based object tracking in traffic scenes, *Proc. 2nd Europ. Conf. Comput. Vis.*, Santa Margherita Ligure, Italy, 1992.
38. A.F. Korn, Toward a symbolic representation of intensity changes in images, *IEEE Trans. Patt. Anal. Mach. Intell.* 10(5):610–625, 1988.
39. M.D. Levine, Region analysis using a pyramid data structure. In *Structured Computer Vision*, S. Tanimoto and A. Klinger, eds., pp. 57–100, Academic Press: New York, 1980.
40. L.M. Lifshitz and S.M. Pizer, A multiresolution hierarchical approach to image segmentation based on intensity extrema, *IEEE Trans. Patt. Anal. Mach. Intell.* 12(6):529–541, 1990.
41. T. Lindeberg, Scale-space for discrete signals, *IEEE Trans. Patt. Anal. Mach. Intell.*, 12:234–254, 1990.
42. T. Lindeberg and J.O. Eklundh, Scale detection and region extraction from a scale-space primal sketch, *Proc. 3rd Intern. Conf. Comput. Vis.*, Osaka, Japan, pp. 416–426, December 1990.
43. T. Lindeberg, Discrete scale space theory and the scale space primal sketch, Ph.D. thesis, Dept. of Numerical Analysis and Computing Science, Royal Institute of Technology, Stockholm, May 1991. A revised and extended version to appear as book *Scale-Space Theory in Early Vision* in Kluwer International Series in Engineering and Computer Science.
44. T. Lindeberg and J.O. Eklundh, On the computation of a scale-space primal sketch, *J. Visual Commun. Image Represent.* 2(1):55–78, 1991.
45. T. Lindeberg, Scale-space behavior of local extrema and blobs, *J. Math. Imag. Vis.* 1:65–99, March 1992.

46. T. Lindeberg and J.O. Eklundh, The scale-space primal sketch: Construction and experiments, *Image Vis. Comput.* 10:3–18, 1992.
47. T. Lindeberg, Discrete derivative approximations with scale-space properties: A basis for low-level feature extraction, *J. Math. Imag. Vis.*, 1993 (in press).
48. T. Lindeberg, Effective scale: A natural unit for measuring scale-space lifetime, *IEEE Trans. Patt. Anal. Mach. Intell.*, 1993 (in press).
49. T. Lindeberg, On scale selection for differential operators, *Proc. 8th Scandinavian Confer. Image Anal.*, Tromsö, Norway, pp. 857–866, May 1993.
50. T. Lindeberg and J. Gårding, Shape from texture from a multiscale perspective, *Proc. 4th Intern. Conf. Comput. Vis.*, Berlin, pp. 683–691, May 1993.
51. D.G. Lowe, *Perceptual Organization and Visual Recognition*, Boston: Kluwer Academic Publishers, 1985.
52. J. Malik, Interpreting line drawings of curved objects, *Intern. J. Comput. Vis.* 1:73–104, 1987.
53. D. Marr, Early processing of visual information, *Phil. Trans. Royal Soc. London (B)*, 273:483–524, 1976.
54. D. Marr, *Vision*, W.H. Freeman: New York, 1982.
55. J.C. Maxwell, On hills and dales," *The London, Edinburgh and Dublin Philosophical Magazine and J. of Science*, 40(269): 421–425, 1870. Reprinted in W.D. Niven, *The Scientific Papers of James Clark Maxwell*, vol II, Dover Publications: New York, 1956.
56. F. Mokhtarian and A. Mackworth, Scale-based description and recognition of planar curves and two-dimensional objects, *IEEE Trans. Patt. Anal. Mach. Intell.* 8:34–43, 1986.
57. K. Pahlavan and J.O. Eklundh, A head-eye system—analysis and design, *Comput. Vis. Graph. Image Process.* 56(1):41–56, 1992.
58. A.P. Pentland, Extraction of deformable part models, *Proc. Ist Europ. Conf. Comput. Vis.*, Antibes, France, pp. 397–401, 1990.
59. A.P. Pentland, Photometric motion, *IEEE Trans. Patt. Anal. Mach. Intell.*, 13(9):879–890, 1991.
60. T. Poston and I. Steward, *Catastrophe Theory and Its Applications*, Pitman: London, 1978.
61. K. Rohr, Modelling and identification of characteristic intensity variations, *Image Vis. Comput.* 10(2):66–76, 1992.
62. E. Saund, Symbolic construction of a 2-D scale-space image, *IEEE Trans. Patt. Anal. Mach. Intell.* 12(8):817–831, 1990.
63. A.C. Sher and A. Rosenfeld, Detecting and extracting compact textured objects using pyramids, Tech. Rept. TR-1789, Computer Vision Laboratory, University of Maryland, Maryland, 1987.
64. F. Sjöberg and F. Bergholm, Extraction of diffuse edges by edge focusing, *Patt. Recog. Letts.* 7:181–190, 1988.
65. D. Terzopoulos, A. Witkin, and M. Kass, Constraints on deformable models: Recovering 3-D shape and nonrigid motion, *Artificial Intelligence* 36:91–123, 1988.
66. J.K. Tsotsos, Analyzing vision at the complexity level, *Behav. Brain Sci.* 13:423–469, 1990.
67. H. Voorhees and T. Poggio, Detecting textons and texture boundaries in natural images, *Proc. Ist Int. Conf. Comput. Vis.*, London, 1987.
68. R. Watt, *Visual Processing: Computational, Psychophysical and Cognitive Research*, Lawrence Erlbaum Associates: London, 1988.
69. A.P. Witkin and J.M. Tenenbaum, On the role of structure in vision. In *Human and Machine Vision*, J. Beck, B. Hope, and A. Rosenfeld, eds., Academic Press: New York, 1983.
70. A.P. Witkin, Scale-space filtering, *Proc. 8th Intern. Joint Conf. Artif. Intell.*, Karlsruhe, West Germany, 1983, pp. 1019–1022.
71. A.L. Yuille and T.A. Poggio, Scaling theorems for zero-crossings, *IEEE Trans. Patt. Anal. Mach. Intell.*, 8:15–25, 1986.
72. W. Zhang and F. Bergholm, An extension of Marr's signature based edge classification and other methods for determination of diffuseness and height of edges, as well as line width, *Proc. 4th Intern. Conf. Comput. Vis.*, Berlin, pp. 183–191, May 1993.