

Construction of a Scale-Space Primal Sketch

Tony Lindeberg, Jan-Olof Eklundh

Computational Vision and Active Perception Laboratory (CVAP) *

Royal Institute of Technology

S-100 44 Stockholm, Sweden

Email: tony@bion.kth.se, joe@bion.kth.se

We present a multi-scale representation of grey-level shape, called scale-space primal sketch, that makes explicit features in scale-space as well as the relations between features at different levels of scale. The representation gives a qualitative description of the image structure that allows for extraction of significant image structure — stable scales and regions of interest — in a solely bottom-up data-driven manner. Hence, it can be seen as preceding further processing, which can then be properly tuned. Experiments on real imagery demonstrate that the proposed theory gives perceptually intuitive results.

Scale-space theory [11, 3] provides a well-founded framework for dealing with image structures, which naturally occur at different scales. According to this theory one can from a given signal generate a family of derived signals by successively removing features when moving from fine to coarse scale. However, the information in the scale-space embedding constructed from the diffusion equation is only *implicit* in the grey-level values. The smoothed images in the raw scale-space representation contain no explicit information about the features or the relations between features at different levels of scale.

The goal of this paper is to present such an *explicit* representation, called the scale-space primal sketch, and to demonstrate that it enables extraction of significant image structure almost without any a priori information. We shall treat intensity images, the grey-level landscape, and the objects will therefore be blobs, that is bright regions on dark backgrounds or vice versa. However, the theory applies to any bounded function and is therefore useful in many tasks occurring in computer vision, like the study of level curves and spatial derivatives in general, depth maps, colour etc, and also histograms and point clustering and grouping in one or several variables.

Because of lack of space the presentation is heavily condensed. For a more extensive description of this material we refer the reader to [6, 7]. There we also discuss the background to the problem, give the relations to previous work and describe some technical details of the algorithm for constructing the suggested representation.

1 Scale-Space Primal Sketch

From experiments one can (visually and subjectively) observe that the main features arising in the scale-space representation seem to be blob-like, i.e., they are regions ei-

ther brighter or darker than the background. Especially regions which appear to stand out from the surroundings in the original image seem to be further enhanced by the scale-space smoothing. In the suggested scale-space primal sketch we focus on this aspect of image structure with the purpose of building a formal representation to make such information in scale-space explicit. Therefore, there is a need to define what should be meant with a “blob”.

1.1 Grey-Level Blobs

It is clear that a blob should be a region associated with a (at least one) local extremum point. However, it is essential to also define the spatial extent of the blob region around the extremum. Ehrich and Lai [2] considered this problem. They allowed peaks to extend to valleys, a definition that will give unintuitive results e.g. for small peaks on large slopes. Koenderink and van Doorn [3] briefly touch upon the problem and our definition relates to their argument.

The blob definition we base this work on should be evident from Figure 1. The basic idea is to let the blob extend “until it would merge with another blob”. To in-

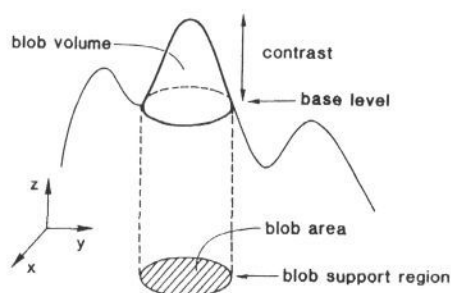


Figure 1: Illustration of the blob definition for a two-dimensional signal. This figure shows bright blobs on dark background. Only one of the three blobs is shown.

tuitively illustrate this notion consider a grey-level image at a fixed level of scale, and study the case with bright blobs on 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 we call the *base-levels* of the blobs and are used to delimit 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. In this sense the blob definition can be regarded as rather conservative, since no attempt is made to include points

*The support from the National Swedish Board for Technical Development, STU, is gratefully acknowledged.

in other directions. Finally, the *grey-level blob* is the region delimited by the grey-level surface and the base-level. Consequently, it is a 3D object with extent both in space and grey-level, whose size is called *grey-level blob volume* and comprises both the amplitude and the spatial extent of the blob. For more precise mathematical definitions of these quantities we refer the reader to [6], Section 2.

Local minima can be treated analogously and every local minimum point will give rise to a dark blob on bright background. Hence, we get *separate* systems for bright blobs on dark background and dark blobs on bright background. This implies that some points will be left unclassified. Consequently, the given definition will, in contrast to, e.g., the sign of the Laplacian of the Gaussian, only attempt to make a partial (and hopefully safer) classification of the grey-level landscape.

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*. This aspect reflects an important principle of the approach.

1.2 Scale-Space Blobs

In general, a grey-level blob existing at one level of scale in scale-space will have a similar blob both at a finer level of scale and at a coarser level of scale. By *linking* together such grey-level blobs across scales we obtain *scale-space blobs*, which are 4D objects having extent both in space, scale and grey-level. At some levels of scale in scale-space

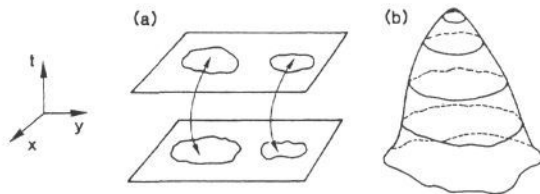


Figure 2: (a) By linking together similar grey-level blobs at adjacent levels of scale we obtain (b) scale-space blobs, which are objects having extent both in space, scale and grey-level. (In this figure we have omitted the grey-level coordinate.)

it might be impossible to accomplish a plain link between a grey-level blob at the current level of scale to a similar grey-level blob at a coarser or finer scale — a catastrophe affecting the connectivity of the blobs has occurred. The scale levels, where such bifurcations (anni-

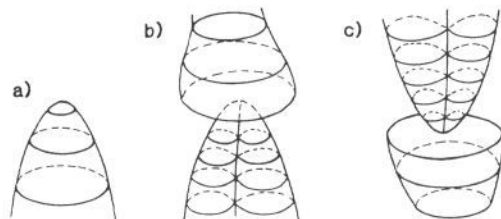


Figure 3: Common blob events in scale-space: (a) annihilation (b) merge (c) split.

hilation, merge, split) take place, are used to delimit the extent in the scale direction of the scale-space blobs. Consequently, every scale-space blob will be associated with a minimum scale and a maximum scale, denoted the *appearance scale* and the *disappearance scale* respectively,

see Figure 2. The difference¹ between the disappearance scale and the appearance scale yields the *scale-space lifetime* of the blob.

1.3 Experimental Results

It is easy to realize that the concept of a blob at a single level of scale is not powerful enough to allow for extraction of relevant image structures. It leads to a strong 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 will be hard to detect as one unit, because at a first attempt only the fine scale blobs will be found.

In Figure 4 we give an example with a toy block image showing how (the support regions of) the extracted grey-level blobs behave with increasing scale together with the raw grey-level images in the scale-space representation. We see that at fine levels of scale mainly small blobs due to noise and surface texture are detected. When the scale 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 survive² for a long time in scale-space if they are located in regions with slowly varying grey-level intensity. This observation shows that *scale-space lifetime alone is not appropriate as a significance measure* for extracting image structure. At coarse levels of scale, the toy blocks appear as single blob objects. Finally, at very coarse levels of scale, adjacent blocks become grouped into larger entities.

One could say that the grey-level blob concept shows an extreme degree of noise sensitivity that can be circumvented by the scale-space smoothing. But, it is certainly a far from trivial problem to determine a proper amount of smoothing *automatically*, based on existing conventional methods. The aim with the suggested construction of scale-space blobs by linking grey-level blobs between scales is to determine which blobs in the scale-space representation can be regarded as significant, *without any a priori information* about neither scale, spatial location nor the shape of the primitives³. As we will see later, the output from the linking procedure also enables determination of a relevant scale for each blob, i.e., a suitable amount of blurring for treating that *individual*⁴ blob. Note that with this definition of a scale-space blob we treat the *scale parameter as equally important as the spatial and grey-level coordinates*, and the *primitives of our representation are objects having extent not only in grey-level and space, but also in scale*.

¹It turns out that some transformation of the scale parameter is necessary in order for the difference between scale values to capture the concept of scale-space lifetime “properly”. This issue is developed in Section 1.4.

²Of course, the contrast of such blobs decreases. But it is far from clear that it is possible to set a globally valid threshold on objective ground. The problem is related to the issue of estimating the noise level in an image, which hardly can be addressed without any constraining assumptions, like e.g. in Voorhees, Poggio [10].

³Except for the fact that the scale-space blurring favours blob-like or Gaussian-shaped objects

⁴We emphasize the word *individual* since we believe that stable scales when they exist are *local* properties associated with objects not with entire images.

1.4 Measuring Significance

Since the ultimate goal of the analysis is to extract important regions in the image based on the significance of scale-space blobs in the scale-space representation, there is an absolute need for some methodology for comparing blob significance *between* different levels of scale. The approach we propose is to use scale-space volume defined in Section 1.2. We suggest that it is a useful quantity for such a significance measure, since it comprises both the grey-level blob volume, which is a combination of the contrast and the spatial extent of the blob, and the lifetime of the blob in scale-space. So far we have no theoretical evidence supporting this claim but state it as an unproven postulate, which is to be supported by experiments.

However, if one is to base a significance measure on the scale-space volume, it is of crucial importance that the scale parameter and the grey-level blob volume are measured in proper units. For instance, measuring scale-space lifetime just by the plain difference or the difference of the logarithms between the disappearance and appearance scales will not work, since then either the significance of coarse scale or fine scale blobs would be substantially overestimated.

Consequently, it is necessary to introduce a transformed scale parameter, which we will denote *effective scale* τ , that captures the scale-space lifetime properly. From a requirement that the amount of structure, which is destroyed if the effective scale parameter is increased with a small increment, should be independent of the current scale and the current amount of structure in the image we have shown [6] that this concept can be defined in essentially one way only. One obtains

$$\tau(t) = \log \frac{p(0)}{p(t)} \quad (1)$$

where the quantity $p(t)$ is the “expected density of local extrema at scale t ”. For implementational purpose $p(t)$ is estimated from simulation results for point noise images.

Similarly, the grey-level blob volumes needs to be transformed, since the size of the grey-level blob volumes will vary substantially with scale. A natural normalization to perform is to subtract the mean value, $V_m(t)$, and divide by the standard deviation, $V_\sigma(t)$.

$$V_{transf,prel}(t) = \frac{V(t) - V_m(t)}{V_\sigma(t)} \quad (2)$$

However, such a quantity is not suited for integration (which is a necessary step in the computation of the scale-space blob volumes), since it will assume negative values. Therefore, in the current implementation we have chosen to define the transformed blob volume as

$$V_{transf}(t) = \begin{cases} 1 + V_{transf,prel} & \text{if } V_{transf,prel} \geq 0 \\ e^{V_{transf,prel}} & \text{otherwise} \end{cases} \quad (3)$$

which empirically turns out to give reasonable results. As in the previous case, $V_m(t)$ and $V_\sigma(t)$ are determined from simulation results for point noise images.

In this context it is of considerable importance to take the discrete aspects of implementation into actual account. Direct application of results obtained from the scale-space theory for continuous signals could give a severe bias at finer levels of scale where the effects from

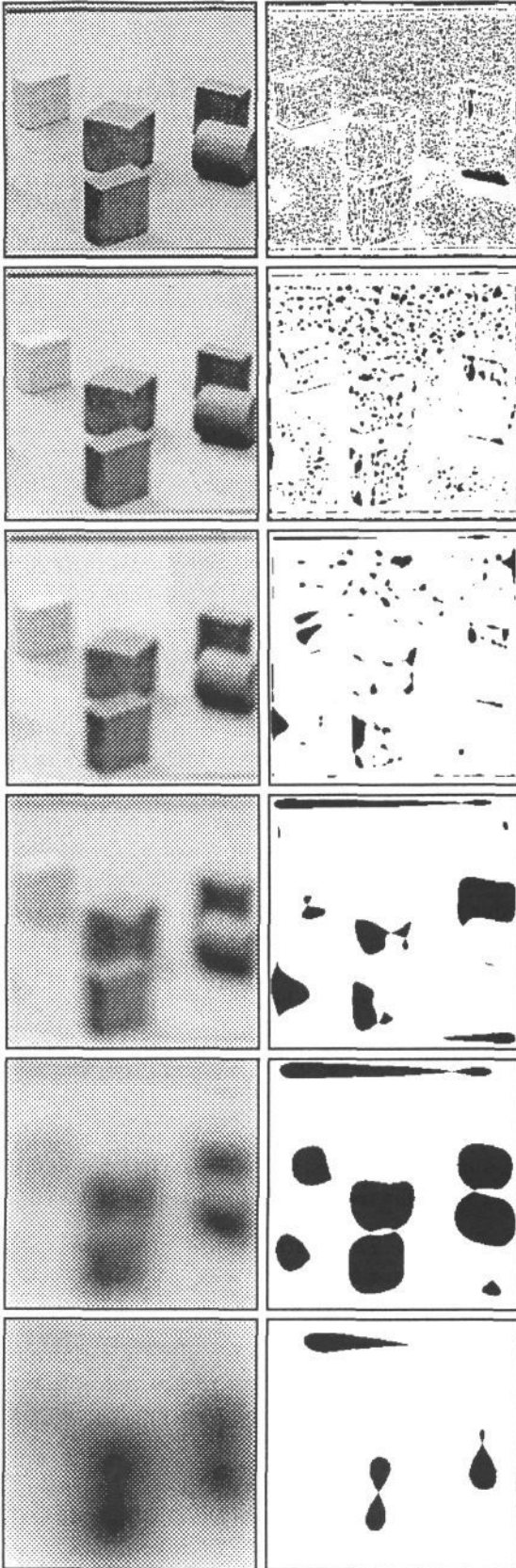


Figure 4: Scale-space representations (left column) and detected dark blobs (right column) of a toy block image at $t = 0$, $t = 2$, $t = 8$, $t = 32$, $t = 128$ and $t = 512$. (Note the noise sensitivity at the finer levels of scale.)

the spatial sampling can be strong. For instance, in continuous theory (for normal processes) it holds that the number of local extrema per unit length varies with scale approximately as t^α [6], which means that the transformation function $\tau(t)$ would be logarithmic. Then, blobs existing at t equals zero would be given infinite lifetime and thereby also infinite significance. By using a scale-space concept especially designed for discrete signals [5] and by simulating the relevant quantities, (for which it is hard to find compact closed-form expressions), we try to overcome such problems⁵.

1.5 Resulting Representation

To summarize, the data structure we propose is a tree-like multi-scale representation of blobs at all levels of scale in scale-space *including* the relations between blobs at different levels of scales. Grey-level blobs should be extracted⁶ at all levels of scale, the bifurcations occurring in scale-space be explicitly registered and grey-level blobs stable over scales be linked across scales into the higher-order objects called scale-space blobs.

The linking process could be a potential source of matching problems, since blobs may move, disappear, merge and split when the scale parameter changes. However, the notion of a scale-space with a *continuous* scale parameter gives us a simple way to circumvent these difficulties in many cases, since problematic matching situations can often be avoided by refinement of the scale sampling. If the scale step is *adaptively* made just fine enough it will be trivial to judge which grey-level blobs belong to the same scale-space blob.

Hence, we start with a relatively fine initial sampling in scale, for each pair of scale levels traverse all pixels to investigate which blobs have a pixel in common. Those blobs are then registered as matching candidates of each other. What remains to decide is when a blob match should be accepted. In our current implementation we perform a refinement in scale each time an unclear matching situation occurs, and accept matches in principle only when all blob events between the two scale levels can be classified as either of the cases: plain link, blob annihilation, blob merge or blob split. Further details about this algorithm can be found in [6] Section 5.

Since this representation tries to capture the significant features and events occurring in scale-space with a small set of primitives we call it a *scale-space primal sketch*. Every scale-space blob contains explicit information about which grey-level blobs it consists of. The grey-level blobs are given at the (sampled) scale levels obtained from the previously outlined adaptive scale linking and refinement procedure. Further, the scale-space blob volume, the appearance scale, the disappearance scale and the scale-space lifetime have been computed (using straightforward numerical techniques). The scale-space blobs “know” about the type of bifurcations (annihilation, split, merge) that have taken place at the appearance and disappearance scale. They also have links to the

⁵ Another way to avoid these problems is of course by starting at coarser levels of scale rejecting structures at the finest levels of scale, but also this “solution” requires a detailed analysis of the situation, since then a threshold in scale needs to be determined.

⁶ An algorithm for detecting grey-level blobs in a grey-level image at a single level of scale is described in [6].

other scale-space blobs involved in the bifurcation processes. Hence, the representation we have computed explicitly describes the hierarchical relations between blobs at different levels of scale.

2 Detecting Image Structures

One motivation for this research was to investigate if the scale-space model really allows for determination and detection of global and stable phenomena. In this section we will demonstrate that this is indeed possible and that the presented representation can be used for extracting stable scales and regions of interest from an image, in a solely data-driven way. The treatment is based on the *assumption* that:

- *Features, which are significant in scale-space, correspond to relevant details in the image.*

More precisely, since the primitives we intend to use are scale-space blobs we formulate the assumption as follows:

Assumption 1 *A scale-space blob having a large scale-space volume in scale-space corresponds to a relevant region in the image.*

The approach is closely related to Witkin’s [11] observation about correspondence between stability in scale-space and perceptual salience. However, here we base the stability measure on the scale-space blob volumes instead of the scale-space lifetime. The intention is that this choice should also reflect the size of the blobs and how strongly they manifest themselves with respect to the background. As mentioned in Section 1.3 we have observed that small blobs due to noise can survive over a large range of scales if they are located in regions with slowly varying grey-level.

A scale-space blob will in general exist over some scale interval in scale-space. When there is a need to reduce the amount of data represented and to select an appropriate scale and a spatial region for a scale-space blob we make use of the following assumptions:

Assumption 2 *The scale-level, at which a scale-space blob assumes its maximum grey-level blob volume, is a relevant scale for treating that individual blob.*

Assumption 3 *The spatial extent of a scale-space blob can be represented by the blob support region corresponding to its grey-level blob at the relevant scale.*

Below, we will give experimental results showing that these assumptions, really segment out perceptually relevant regions in the image. The basic methodology, should be obvious from the previous presentation.

- Generate the suggested multi-scale representation, where blobs are extracted at all levels of scale and linked across scales into scale-space blobs.
- Calculate the scale-space volume for each scale-space blob based on the notion of effective scale and rescaled blob volumes.
- For each scale-space blob determine the scale where it assumes its maximum grey-level blob volume, and extract the blob support region of the grey-level blob at that level of scale.

- Sort the scale-space blob in descending significance order, i.e., with respect to their scale-space blob volumes.

In Figure 5 and Figure 6 we show the results for two different images. For display purpose we have extracted the N scale-space blobs having the largest blob volumes. Each blob is represented at its representative scale, that the previously mentioned scale where the scale-space blob assumes its maximum grey-level blob volume. Finally, the spatial region of each blob (which is the blob support region at its representative scale) has been marked in a binary image, where black indicates the existence of a significant blob and white represents background. In order to avoid overlap in the display, we have shifted to a new fresh layer each time the addition of a new blob would have implied overlap between two different blobs.

As we see, the blocks are extracted from the toy block scene. Also, at coarser scales, adjacent blocks become grouped into coarser scale units and the imperfections of 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.

Let us conclude by stressing that we *extract the intrinsic shape of the grey-level landscape in a completely bottom-up data-driven way without any a priori knowledge and without any assumptions about the shape of the primitives* (except for the fact that the scale-space smoothing favours blob-like objects, since it is equivalent to correlation with a Gaussian-shaped kernel).

We get a segmentation that is coarse in the sense that the localization of object boundaries may be poor, due to the natural distortions of shape which occur in scale-space. However, the segmentation is safe in the sense that those regions, that are given by the scale-space blobs with large scale-space volume, really serve as landmarks of significant structure in the image, with information about

- the *approximate location and extent* of relevant regions in the image.
- an *appropriate scale* for treating those regions.

This is exactly the kind of coarse qualitative information⁷ that is necessary for many other processes in early vision, see e.g. the applications to edge detection, histogram analysis and junction classification in [7].

The number of scale-space blobs N selected for display above is, of course, rather arbitrary. However, note that there is a well-defined ranking between the blobs. If one studies the significance values of the blobs, see [6] Section 7, 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 image regions could be extracted just from the criterion that their significance values should stand out from the significance values of the other ones. However, in more general situations there is a need for feed-back or reasoning, see [7] for further details.

⁷The scale-space primal sketch contains much more information than is presented this rudimentary output. For instance we have not illustrated the registered blob bifurcations in scale-space. Nor have we shown or made use of the hierarchical relations between blobs at different levels of scale induced by the blob events. This information is however explicit in the computed representation.

3 Summary and Discussion

The representation that we build is similar to the primal sketch suggested by Marr [9, 8], 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 almost without any a priori knowledge. However, besides that it is a region-based and not an edge-based representation it is more *qualitative*, without strong assumptions about the shape of the primitives. Moreover, the proposed representation consists of coarse features like blobs represented at multiple scales. We have demonstrated that it allows for

- Automatic detection of salient (stable) scales, if they exist.
- Ranking of events in order of significance.
- Generation of hypotheses for grouping and segmentation (regions of interest).

This implies that candidate regions for further processing are generated, as well as information about the scale. We see that the representation gives clues to subsequent analysis and, hence, it can guide focus-of-attention mechanisms. This issue is developed further in [7], where we demonstrate how this type of qualitative output information can be useful for tuning and simplifying various tasks in early low-level visual processing, see Figure 7 for an example.

References

- [1] Bischof W.F., Caelli T "Parsing Scale-Space and Spatial Stability Analysis", *CVGIP*, **42**, (1988), pp192-205.
- [2] Ehrich R.W., Lai P.F. "Elements of a Structural Model of Texture", *Proc. PRIP*, IEEE CS Press, (1978), pp319-326.
- [3] Koenderink J.J., van Doorn A.J. "The Structure of Images", *Biol. Cyb.*, **50**, (1984), pp363-370.
- [4] Koenderink J.J., van Doorn A.J. "Dynamic Shape", *Biol. Cyb.*, **53**, (1986), pp383-396.
- [5] Lindeberg T.P. "Scale-Space for Discrete Signals", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol PAMI-12, No. 3, (1990), pp234-254.
- [6] Lindeberg T.P., Eklundh J.O. "On the Computation of a Scale-Space Primal Sketch", *Tech. Rep. TRITANA-P9005*, NADA, Royal Inst. of Technology, Stockholm, (1990), *Submitted*.
- [7] Lindeberg T.P., Eklundh J.O. "Guiding Early Visual Processing with a Scale-Space Primal Sketch", *Tech. Rep. TRITANA-P9013*, NADA, Royal Inst. of Technology, Stockholm, (1990), *Submitted*, A short version to appear in *Proc. 3rd ICCV*, Osaka, Japan, (1990).
- [8] Marr D. "Early Processing of Visual Information", *Phil. Trans. of the Royal Soc. (B)*, **27S**, (1976), pp483-524.
- [9] Marr D. *Vision*, Freeman, San Francisco, (1982).
- [10] Voorhees H., Poggio T. "Detecting Textons and Texture Boundaries in Natural Images", *Proc. 1st ICCV*, (1987), pp250-258.
- [11] Witkin A.P. "Scale-Space Filtering", *Proc. 7th IJCAI*, pp1019-1022, (1983).

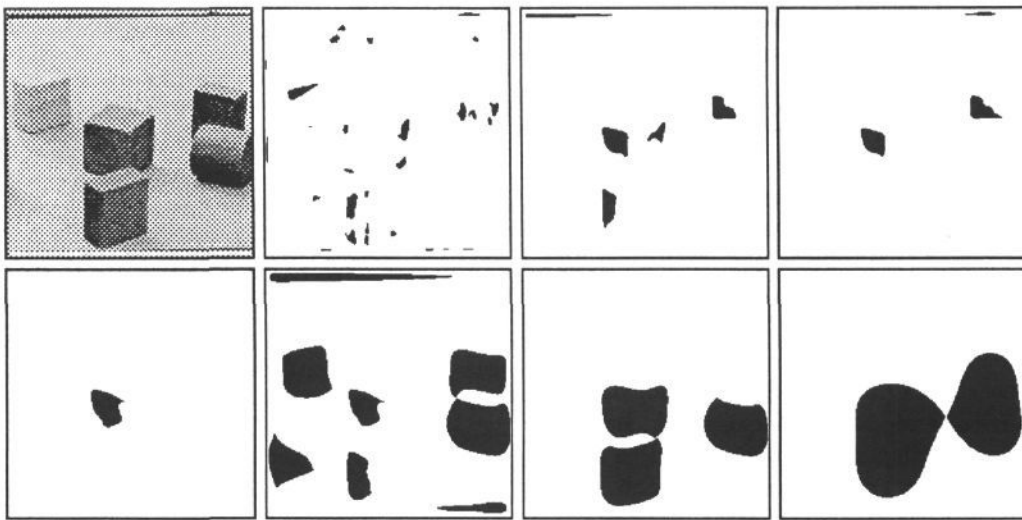


Figure 5: The 50 most significant dark blobs from a toy block image. (Note how these images have been produced — they are not just blob images at a few levels of scale. Instead every blob has been marked at its individual representative scale, that is the previously mentioned scale where the scale-space blob assumes its maximum grey-level blob volume. The blob support region of the grey-level blob at this level of scale has been extracted and, finally, the support regions of the blobs have been drawn in different images as to avoid overlap.)

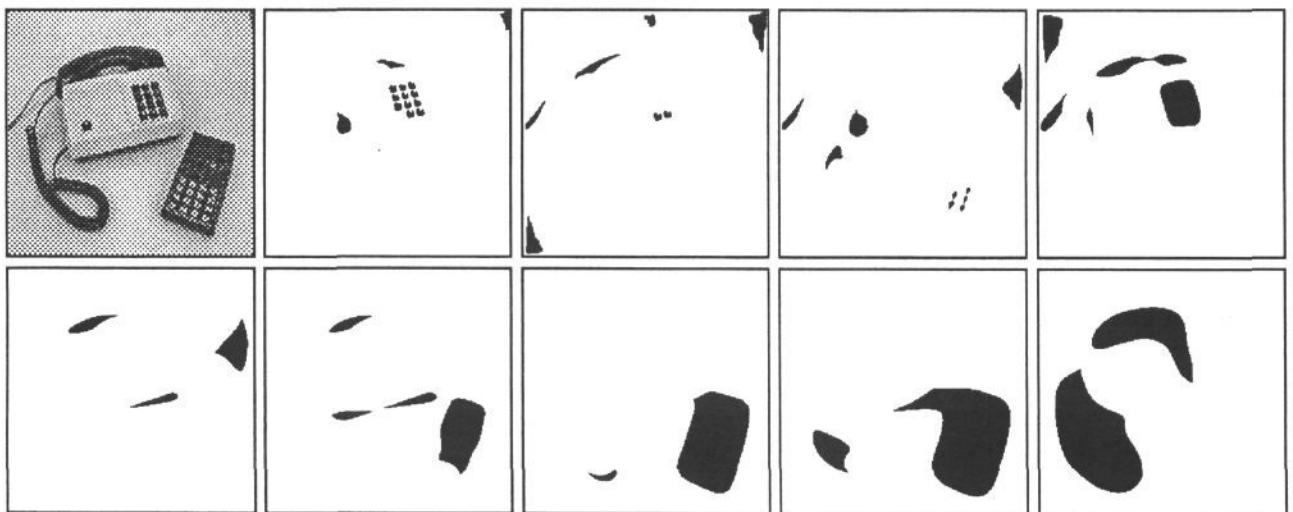


Figure 6: The 50 most significant dark blobs from a telephone and calculator image.

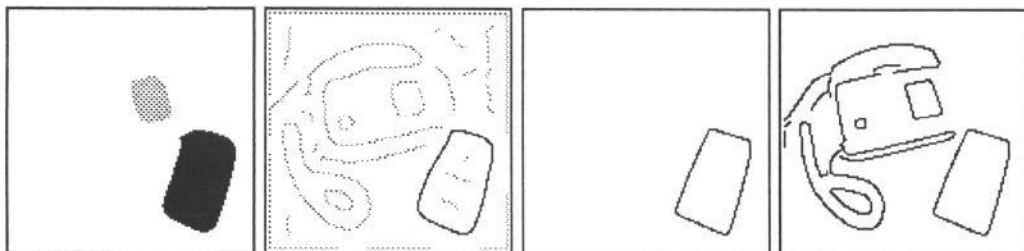


Figure 7: Illustration of how the qualitative scale and region information given by the blobs from the scale-space primal sketch can be used for guiding edge detection. (a) A blob (marked with black) from the telephone and calculator image as obtained from the scale-space primal sketch. (b) Edges (non-maximum suppression without thresholding) detected at the scale given by the previous blob. The detection step will be safe, since the image has been subjected to an appropriately selected amount of blurring. However, the localization might be bad due to the natural shape distortions, that occur at coarse levels of scale in scale-space. (The edges matching matched to the previous blob have been marked with black.) (c) Result after focusing the previous edges down to fine scales using edge focusing, in order to improve the localization. (d) Result after performing the previous steps on the seven most significant blobs from the telephone and calculator image and marking all edges in the same image. Consequently, this image shows edges related to dark blobs in the image. Note, however, that this is not just an ordinary edge image but that these edge segments are more meaningful entities, since a coarse grouping has already been performed and the edge segments are associated with blobs and explicit scale information. (The details of this analysis are described in [7].)