<http://www.diva-portal.org>

This is the published version of a paper published in *Image and Vision Computing*.

Citation for the original published paper (version of record):

Brunnström, K., Eklundh, J-O., Lindeberg, T. (1990)
On scale and resolution in active analysis of local image structure
Image and Vision Computing, 8(4): 289-296
[https://doi.org/10.1016/0262-8856\(90\)80005-E](https://doi.org/10.1016/0262-8856(90)80005-E)

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:miun:diva-24526>

On scale and resolution in active analysis of local image structure

Kjell Brunnström, Jan-Olof Eklundh and Tony Lindeberg

Focus-of-attention is extremely important in human visual perception. If computer vision systems are to perform tasks in a complex, dynamic world they will have to be able to control processing in a way that is analogous to visual attention in humans. Problems connected to foveation (examination of selected regions of the world at high resolution) are examined. In particular, the problem of finding and classifying junctions from this aspect is considered. It is shown that foveation as simulated by controlled, active zooming in conjunction with scale-space techniques allows for robust detection and classification of junctions.

Keywords: foveation, scale, resolution, junction classification, scale-space techniques, focus-of-attention, active vision, histogram analysis

It is often argued that human vision provides a two-legged proof of the feasibility of general purpose machine vision. Human vision is a highly dynamic process working over time. Its input data changes continuously as well as its system parameters, which vary due to eye movements, fixation, focusing, etc. In fact, this is also the case when we look at static and monocular scenes such as images. Therefore, it is not clear that we generally pose the machine vision problems in a comparable context, or, put differently, that we give machine vision a fair chance.

A central theme of computational vision is the derivation of low-level cues from local features of image brightness. We may for instance try to extract (all the) edges from a picture like that shown in Figure 1. However, the results obtained even from a first rate method are usually quite inferior to those predicted by a human looking at the corresponding scene. Hence, subsequent high-level interpretation tasks often face difficult problems due to shortcomings in the low-level features extracted.

Referring to our initial remarks, one can point out two reasons for this. One is that algorithms like the Canny-Deriche edge detector applied in Figure 1 are usually based on fixed models (step edge with noise), and have no means of detecting that another model may be more appropriate, e.g. that of a corner. The other reason is that a performance corresponding to a human looking at the actual scene can hardly be expected simply due to limits in resolution.

Standard camera systems usually have a visual angle of about 50° and give image matrices of say 500 × 500 up to 1000 × 1000 pixels. This should be compared to the 2° in foveal vision, which in view of visual acuity can be said to correspond to an image resolution of about 200 × 200 pixels. Obviously, this difference implies that multi-resolution processing, like that done in pyramids, cannot be seen as analogous to foveation. The images coming out of a standard camera system are limited too much by resolution.

In this paper we consider such tasks using a framework that does not have these limitations. We argue that the processing tasks can be made more tractable by using *focus-of-attention* and *active focusing*, i.e. by doing controlled zooming and windowing in regions where we have indications of an interesting structure. In other words, what we aim at is to acquire images with a sufficiently fine resolution such that the phenomena we intend to study become *clearly resolved*.

Figure 1. (Left) image of a wooden house (the Godthem Inn at Djurgården, Stockholm); (right) edges obtained from the Canny-Deriche edge detector

This has the implication that we ought to have mechanisms for controlling the image acquisition process in a closed loop. In this way we can simulate a system that performs foveation, i.e. *takes a closer look* at interesting regions in the field view and tries to detect a set of properties (not necessarily all) that can be safely and precisely inferred.

An approach like this raises computational problems different from those appearing, for example, in edge detection or other general methods for searching for local structure in normal *overview* pictures. Obviously, we need a method for deciding where to focus our attention. This is discussed below. Moreover, the increased resolution is likely to enhance the noise, at least relative to the prominence of the structures we are looking for. This may call for different types of algorithms for detecting local structure. In fact, we show how very simple local statistical techniques can be used to obtain robust detection and classification of junctions and corners or high-curvature points.

In that context it should be mentioned that we also try to simplify the tasks by detecting features at proper scales*, in a manner similar to the guidance of early visual processing with qualitative scale and region information described by Lindeberg and Eklundh¹.

Finally, we have the problem of *detecting* structure when we dynamically vary resolution and window size. What we propose is a method of stability of responses, or what can more formally be called *transformational invariance*. The rationale for this approach is the observation that local structure, in the highly resolved foveated images, will generally belong to one of a small set of generic types. Stable responses will occur for the models that best fit the data. In the work reported we use techniques quite simple to implement. More sophisticated approaches exist to determine which case holds, i.e. which model is correct, but they are only briefly touched upon.

DETECTING LOCAL STRUCTURE

In this paper we apply our general arguments, and especially our idea of simulating foveation by active focusing, to the detection and analysis of junctions of two or more boundary segments in grey-level images. We are particularly trying to find points giving cues to 3D structure, such as *T*-junctions and projections of 3D corners. For instance, the *T*-junctions (generically) indicate interposition and hence relative depth.

It is well-known that elaborate edge detection methods like those proposed by Hückel², Marr and Hildreth³, Canny⁴ or Bergholm⁵, have problems at junctions. Zero-crossings of the Laplacian, that is boundaries between positive and negative regions, will of course not correctly divide three regions in a junction. If first order derivatives are used other problems arise, e.g. at the computation of the gradient direction.

*The scale and the resolution concepts are sometimes used interchangeably in the vision literature, and their precise meaning becomes unclear. Worth emphasizing in this context is that with resolution in this paper we mean just the spatial density of grid points used in the sampling of the image. Scale, on the other hand, stands for the characteristic length over which variations in the image take place and/or the operator size used for processing the image data.

To overcome such problems direct methods for junction detection have been proposed, e.g. by Moravec⁶, Nagel⁷ and Kitchen and Rosenfeld⁸. The results have been applied to matching for stereo and motion analysis, but it is not clear that these approaches give any precise information about the nature of the junction. On the other hand, explicit use of the feature as a cue to scene structure would require some sort of classification or description of the junction.

Proposed edge and junction detectors like those cited above are based on precise assumptions about the local image structure. Usually, one assumes two or three regions, each one with locally constant or linearly varying intensity; regions which in turn are clearly separated. It is obvious that these models are difficult to assess from noisy data with very few samples from each population. Marr and Hildreth³ pointed out that while the smallest receptive fields according to, for example, Wilson⁹, contain about 500 cones, most edge and corner detectors work with a few tens of samples. This indicates that one should indeed work with large operators, as some authors also do (e.g. see Canny⁴ and Bergholm⁵). However, the problem is not solved with this. First, we still have the relation to the size of the structure we are looking for. We can't assess, say, a *three regions* model by simply using a large operator if the window contains many different instances of it, just because the window is large. Hence, the idea of using large operators should only be expected to give *good* (see above) performance if it could be applied so that, at least ideally, no more than one feature occurred in each field. Now this poses a second problem. Since we can only assume coarse knowledge or expectancy about the scale of the features, how do we find a reasonable operator size? Especially, how do we detect if two or a few nearby features appear in the same field? This can happen even if the resolution is increased. The proposed answer to this is to use the stability of the computed measures as the parameters, field size, resolution and location, are varied. More precisely, we have addressed the problem of finding the initial hypotheses about existing structures and their scales using scale-space representations. We briefly review this theory below. From these initial hypotheses zooming is performed and stability is verified or rejected.

DETERMINING SIGNIFICANT IMAGE STRUCTURES AND THEIR SCALES

The prominent structures in the grey-level image can be determined using the scale-space primal sketch approach developed by Lindeberg and Eklundh^{1,10} on the basis of the scale-space theory of Witkin¹¹ and Koenderink and van Doorn¹².

The main idea is to build a scale-space representation of blobs, i.e. peaks and valleys in the grey-level landscape, and their *extent*. It is based on a conservative notion of a grey-level blob, which implies an inherent competition between parts (see Figure 2). These blobs are then *linked* between scales into higher order objects, called scale-space blobs, which have extent not only in space and grey-level but also in scale. The bifurcations occurring in scale-space are registered, and a hierarchical data structure is created.

Figure 2. (a) Example illustrating the definition of a grey-level blob; (b) by linking grey-level blocks between scales one obtains (c) scale-space blobs, which are objects having extent both in space, grey-level and scale. These objects are the fundamental primitives in the representation

There are several computational problems to solve in building this representation (see Lindeberg and Eklundh¹⁰ concerning the details).

From measurements of stability of scale and significance in scale-space the representation gives a qualitative description of the image structure, which enables determination of approximate location and extent of relevant regions in the image, as well as appropriate levels of scale for treating those regions (see Figure 3 for an example). The significance value of a scale-space blob is essentially its volume in the four-dimensional scale-space given by the spatial, grey-level and scale coordinates.

Experimentally it has been demonstrated that the extracted blobs with their associated significance measures correspond to a perceptually reasonable ranking of relevant regions in the image. It has also been shown that the related qualitative scale and region information is useful for simplifying other processing tasks such as edge detection and histogram analysis (see Lindeberg and Eklundh¹ for a detailed description). Hence, we suggest the use of this output for generating a hypothesis for the focusing procedure as well as for controlling the setting of initial values for window size and position.

It should be stressed that the intrinsic shape of the grey-level landscape is extracted in a completely bottom-up data-driven way without any *a priori* information about either spatial location, scale or shape of the primitives.

An important problem in deriving this representation concerns measuring significance and scale in such a way

Figure 3. Illustration of the scale-space primal sketch output for the lower right toy block image in Figure 5. The dark regions correspond to significant blobs in the scale-space primal sketch. Note that they are not just plain regions in the image, but entities with well-defined scale information and an associated significance measure determined from the behaviour of the blob in scale-space. Observe that the individual blocks are extracted as single units and that the adjacent blocks become grouped into coarser scale objects

that comparisons between significance values can be made for blobs existing at different levels of scale. This issue must be explicitly dealt with. For instance, a concept called 'effective scale', which is the natural unit for measurements of scale-space lifetime, needs to be introduced. It is argued, that when the effective scale is increased with a small increment, then the expected amount of image structure to be destroyed as well as the probability that an extremum point disappears should be independent of the current scale and the current amount of structure. From these requirements we have shown that this effective scale can be defined in (in principle) one way only. The resulting effective scale is essentially the logarithm of the expected number of local extrema at the given level of scale (see Lindeberg and Eklundh¹⁰ for details, as well as the difficult question about how to estimate the expected amount of structure in and an image).

The significance values obtained from scale-space primal sketch also induce a well-defined ranking between the blobs, which empirically has been found to give an intuitively reasonable ordering of the events. Thus, we believe that such a module really can serve as a guide to the focus-of-attention.

IMAGE STRUCTURE AT JUNCTIONS

We now turn to our problem of finding the classifying junctions. When detecting such candidate junction points, or interest points, it turns out that most methods working on standard imagery will have thresholding problems. However, various additions to the basic algorithm can be used to limit the number of falsely detected corners, with a certain risk of losing some of the true feature points. However, this trade-off seems inevitable and tends to occur in all proposed approaches. In particular, false responses tend to occur along edges. Also noise spikes, if they exist, will according to the model, give responses that are strong local maxima.

There are, in fact, five situations in which strong responses mainly occur:

- at noise spikes
- along edges
- at 2-junctions
- at 3-junctions
- at n -junctions, $n > 3$

Let us consider an ideal case in which the resolution is infinite and each region corresponds to a smoothly varying surface. Let us also assume that the illumination varies smoothly over each region. In that situation we could easily discriminate between the different cases by considering the distributions of grey-levels and edge directions in a sufficiently small neighbourhood of the possible (image) corner. In fact, this would be possible also if the surfaces contained some non-infinitesimal surface markings as well. The classification would be as shown in Table 1.

Experiments by Nagel and his co-workers¹³ and our previous work¹⁴ show that high quality intensity images indeed satisfy such conditions just as a straightforward model would predict.

The important question now is what happens in a

Table 1. Local intensity and directional distributions for the ideal case (* = inconclusive)

Case	Intensity	Edge direction
At noise spikes	uniform	*
Along edges	bimodal	unimodal
At 2-junctions	bimodal	bimodal
At T-junctions	trimodal	bimodal
At 3-junctions	trimodal	trimodal
At n -junctions	*	*

realistic case, e.g. with direct and indirect illumination and noise and with finite resolution. One would still expect, as is indeed the basis for most of the low-level methods proposed, that the given classification would be valid anyway. However, establishing this classification requires sufficiently many samples of the different distribution. Moreover, there is a need for a method of classification that is robust with respect to noise and variations due to the imaging process.

What this means in practice is that the resolution has to be high enough and that the window size should be appropriately chosen for correct classification. With a precise model for the imaging process as well as the noise characteristics one could conceive deriving bounds on the resolution, at least in some simple cases. However, as developed further below, direct setting of a window size immediately valid for correct histogram classification seems to be a very difficult or even impossible task. What we propose instead is to use the process of focusing. Focusing means that we increase the resolution locally in a *continuous* manner (even though we still have to sample at discrete resolutions). We can simultaneously either keep the window size constant with respect to the original image, or vary it independently. The contention we make about the focusing approach is that it avoids the problem of selecting an appropriate pair of resolution/window size, since the simple features we are looking for will show up in a stable manner during the variation of the parameters. An important reason behind this argument is that the classification can be based on simple features (the number of peaks of a histogram) of simple statistics. We can summarize the main principles of the approach as follows:

- take a closer look at candidate interest points by increasing the resolution, i.e. by acquiring a new image with higher sampling density, and varying the window size;
- detect stable classifications based on the simple features in Table 1.

So far we have used no objective method to assess what we mean by *detecting the stable classifications*. However, we have shown in a number of experiments (see also Brunnström *et al.*¹⁴) demonstrating that our computational model indeed predicts what happens. Moreover, the scale-space approach described above can also be applied to this problem.

EXPERIMENTAL TECHNIQUE

We demonstrate our approach on a set of images at different resolutions of scenes containing simple man-

Figure 4. Schematic view over the junction classification cycle. A pair consisting of an interest point matched to a blob from the scale-space primal sketch gives rise to a series of windows of different size and resolution over which histograms are accumulated. The modalities of these histograms are compared to the generic cases and their stability measured. This gives a hypothesis about the possible nature of the junction candidate, which is tested by backprojection into the original image. If it cannot be verified then additional data are acquired, invoking a new analysis cycle in a closed-loop fashion

made objects. Let us describe the various processing steps (see Figure 4 for an overview).

We first perform a bottom-up preprocessing step, in which a set of candidate-points is determined as described above. These points are tentative junctions, i.e. points where several regions meet. A crucial issue is to find a set of such points that is limited in size and is likely to contain true and significant junctions in the image.

The algorithm we use is based on the method of computing directional variances, suggested by Moravec⁶ and contains some additional constraints of *cornerity*. In this way significant junctions can be automatically obtained without any strong dependence on thresholding. Finding these parameters seems difficult in Moravec's original algorithm. We do not discuss this further since we are not trying to analyse all junctions here (see Brunnström *et al.*¹⁴ for details).

In addition to a fixation point the focusing module needs an estimation of the local scale. For a sequence of windows of varying resolution and size we compute the distribution of grey-levels and edge directions. The decision to be made concerns only the number of peaks in the histograms and their stability over focusing. If the classification does not fit any of the generic cases, or if it cannot be verified by backprojection into actual data, there is a need for changing the fixation point. How this step should be fully automated is not very clear, but we demonstrate with experiments below that this closed-loop approach seems to provide a powerful tool for resolving unclear cases.

We now develop in finer detail the individual actions of the different modules, and show how they communicate.

Setting window size from blob information

Having found a point of interest we invoke the focusing procedure analysing local histograms. This calls for some mechanism for actually setting an initial value for

window size, since the size of a suitable neighbourhood region around a junction candidate will in general vary both within and between images. To select a single value directly valid for precise classification seems to be an ill-posed or impossible task. If the window is too large then other structures than the actual corner region around the point of interest might be included in the window, and the histogram modality would be affected. Conversely, if it is too small then the directional histogram could be severely biased and deviate far from the ideal appearance in case the physical corner is slightly rounded – a (scale) phenomenon that seems to be commonly occurring in realistic scenes. A too small window might also fall outside the actual corner if the interest points are associated with localization error.

Therefore, the methodology we have adopted is to use the context information from the blobs for setting just coarse approximate values giving (generous) upper and lower bounds on a focusing interval. Then, the intention is that the systematic variation of window size combined with the consistency check over parameter variation should allow for a more robust modality determination.

A matching step has to be performed between the blobs and the interest points. The method we use in our current implementation is similar to the matching between blobs and edges performed by Lindeberg and Eklundh¹. For an interest point to be related to a blob, two conditions about spatial coincidence must be satisfied:

- 1 The interest point should be at least coarsely associated with the blob, without being too strongly associated with other blobs. We compute a Voronoi diagram of the grey-level blob image at the given level of scale, using a distance transformation, and demand the interest point to be included in the Voronoi region associated with the support region of the grey-level blob. This condition prevents points, which are closely related to one particular blob, from being associated to other blobs.
- 2 The interest point should not be located too far from the blob boundary. We require the minimum distance between the interest point and the blob not to exceed a typical spatial length at the current level of scale. Here, we have set this parameter to the square root of an experimentally determined typical blob area $A_m(t)$ at that scale level (see Lindeberg and Eklundh¹⁰ for further details).

A weakness with this blob-point matching compared to the earlier mentioned blob-edge matching is that the scale information from the blobs is not used as extensively for the interest points as for the edges. In the case with edges, the detection step was performed at the same scale as the blob manifested itself, which meant that interfering structures at finer levels of scale had to a large extent been suppressed by the scale-space smoothing, something that simplified the matching problem considerably. But here the interest points are detected directly from the raw grey-level image at the finest level of scale without any use of scale knowledge. However, we are currently investigating different promising approaches for including this information already in the phase of detecting the interest

Figure 5. Illustration of matching procedure between blobs and interest points for the toy block image. (a) The 50 most significant interest points obtained with the Moravec interest point operator; (b) a significant blob extracted by the scale-space primal sketch (black) and the other grey-level blobs (grey) at the scale level given by the scale-space blob; (c) Voronoi diagram of the previous grey-level blob image (Criterion 1). The boundaries between different regions are marked with black; (d) stripe around the boundary of the blob. The width of the stripe has been set to a characteristic length at the current level of scale (Criterion 2); (e) the interest points matched to the blob; (f) resulting minimum and maximum window size around two of the interest points as set from the data given by the blob information

points, with the purpose of reducing the number of false blob-point matches and also, possibly, obtaining safer points of interest.

Once we have a relation between a blob and an interest point we fit an ellipse to the blob, extract the shorter one of the two semi-axes to get a characteristic length of the blob, and set the minimum and maximum window widths to some constants ($\approx 1/3$ and 3, respectively) times this value (see Figure 5 for an example). This information constitutes the input data for the next module.

Classification by histogram modality

Using the computational model described above the neighbourhood of a junction can be categorized into a number of different classes, given by the modalities of the local grey-level and edge direction histograms of that region. The idea is to use this property as the basis of a rough junction classification scheme. Thus, in a robust manner, we need to find the modalities of a number of histograms, a problem which in general is difficult. However, in our case we just have to establish the exact distribution up to the trimodal case, which still has its complications, but is feasible in most cases.

Accumulating histograms

Obviously, it is desirable to have histograms that are as easy to analyse as possible. The number of peaks in the grey-level histograms correspond, according to our model, to the number of regions, with distinct intensity values, meeting at the junction. Since there will be a transition band between such regions, these edge pixels can be excluded from the grey-level histograms.

Similarly, the edge direction statistics are obtained for pixels labelled as edge elements only. Hence, some edge detector has to be applied first. A simple detector like Roberts' cross might do, but the choice is of little importance. Something that may be important, though, is the effect of the point of interest itself. As discussed above, it is at this point that most low-level algorithms will have problems. Assuming that the resolution is high enough, so that we can appropriately sample the postulated simple underlying structure in the neighbourhood, we can treat the point as being a removable discontinuity. Hence, the statistics are computed in all but a subwindow around the point. Since we vary both the resolution and the window size and look for stability, we have chosen to let the subwindow cover a fixed part of the total window, say $1/9$ or $1/16$ of the area.

Finding histogram modality

The modality of a histogram could be said to correspond to its number of prominent peaks. Peaks are, of course, as in the 2D case scale phenomena. We are only interested in the large scale peaks, so what we would like to do, in principle, is to blur the histogram and then detect the peaks. It is difficult, however, to determine the exact degree of blurring. For 2D data it has been demonstrated that a proper amount of smoothing can be obtained from the scale-space primal sketch. This technique should apply also to one-dimensional signals, but since this case has not yet been implemented, we have instead been using a method of iteratively sharpening the peaks into spikes, which will have a similar effect as smoothing followed by peak detection. This method was first suggested by Rosenfeld and Davis¹⁵, but our implementation is taken from Peleg¹⁶.

The idea is to let large bins grow at the expense of smaller ones. The algorithm is simple and could be described as follows: compare the value of an accumulator with the average over a small neighbourhood on one side of the accumulator. If the value is greater move a fraction of the values of each accumulator in the neighbourhood one step towards the accumulator. Repeat the procedure for the other side of the accumulator, and iterate until the histogram just consists of spikes.

There are two parameters which affect the result; the number of bins and the size of the neighbourhood. They are coupled in a way that is hard to analyse exactly, but intuitively you would expect that decreasing the number of bins has the same effect as increasing the size of the neighbourhood. This seems empirically to be the case.

Of course the result should not depend too much on what particular combination of neighbourhood size and number of bins that has been used. Hence, the number of bins is fixed and the size of the neighbourhood varied from a low value up to a high value. The spikes which are stable under this variation are detected. This turns out to be less parameter sensitive at least if the goal is to detect prominent peaks. Incidentally, this is another example of using stability of responses as an indication of what information is significant.

Verifying classifications by back projection

The classification obtained so far is, however, only a

hypothesis, so there is a need for some verification with the actual data. There are several ways of doing this, for instance fitting a parametric model, changing the fixation point in a pattern given by the classification or analysing how the local edges meet. We have only implemented the last method. Moreover, there is still no automatic decision on the correctness of the classification.

The verification module used makes linear approximations of the thin edges in the neighbourhood around the junction point, analyses how they meet and from this decides if the classification was correct or not. It ought to be pointed out that what we are doing is not equivalent to classifying the junctions just by looking at how the edges join, since this module only has to verify the hypothesis given by the histogram analysis. Furthermore, the answer just needs to be yes or no.

The way we make the linear edge approximations is as follows: thin edges are extracted locally with the Canny-Deriche edge detector for a window, selected as representative of all the windows used by the classification step, which usually is one with high resolution. A Hough transform of the edge elements is computed. The histogram analysis of the edge directions together with the assumption that an edge that forms a junction should pass close to the candidate interest point, give positions in Hough-space. We take as lines the maxima found in a small neighbourhood around these positions. To find the endpoints of the lines, they are checked against linked-lists of edge elements. From this, it can be tested whether the lines meet in a way consistent with the current hypothesis or not.

Figure 6. Illustration of the stability of the local statistics under resolution variations. In the left column, where the coarsest resolution is shown, the window is too large, which here shows up as an extra peak in the edge direction histogram

Results

In the toy block image in Figure 6, we focus on a corner of the bright block to the left. The possible existence of a corner has previously been established as described above. At coarse resolution three grey-levels and three directions are detected. This is because the outer boundary on the left is visible. As the focusing procedure continues, there is a stable response of three grey-levels and two directions. The vertical edge is very weak and cannot be detected, not even by, for example, Canny-Deriche's edge detector. However, backprojection into the image shows that this response is not generic. Therefore, it is assumed that some information is missing or incorrect. The point focused at is therefore shifted up and down the bisector of the detected L-junction, in a simulated eye-movement. The correct structure can then be assessed, as is illustrated in Figure 7. Note that in all cases both the window size and the resolution are varied, but only very few pictures can be shown here.

It should be stressed that the results shown in Figures 6 and 7 are obtained with simple tests in a case when, for instance, sophisticated edge detectors have failed.

Figure 8 shows some illustrative output in the case of a true T-junction, and when one of the objects is curved. We are showing the final stable results which indicate a T-junction. Notably, a third but weakly supported direction is detected. This direction is, in fact, due to the lower part of the curved edge. However, backprojection of the predicted edges can again be used to characterize this type of response (for more examples see Brunnström *et al.*)¹⁴.

Figure 7. Analysis of the fixation point shown in the top left image, with a neighbourhood size shown in the top middle window, gives the result displayed at the top right; i.e. three grey-levels and two edge directions (indicating an L-junction). This is not a generic case. For resolving the situation two new fixation points are analysed, has shown in the middle and in the bottom rows. This results in a correct classification

Figure 8. T-junction between a slightly curved contour and a straight one. The left image shows the overview of the scene, the middle one indicates the local neighbourhood which has been processed and the right one displays the result. The histograms are from left to right; grey-level; edge direction; and peak sharpened edge direction

CONCLUSIONS

We have argued that dynamic mechanisms like focus-of-attention are necessary in computer vision systems, if they are to perform their tasks in a complex, dynamic world. We discussed how foveation (examining selected regions of the visual world at high resolution) can be incorporated in an active vision system. We studied the use of this capability in the task of finding and classifying junctions of edges. It was shown that the junctions in this context could be labelled into a small set of generic cases, and that these cases could be established computationally by robust and simple algorithms.

It has also been an intention to stress that low-level feature extraction tasks can be simplified by use of a sufficiently fine resolution and applying operations at a proper level of scale.

ACKNOWLEDGEMENT

The support from the National Swedish Board for Technical Development (STU) is gratefully acknowledged.

REFERENCES

- 1 Lindeberg, T and Eklundh, J O 'Guiding early visual processing with a scale-space primal sketch' *Proc. 3rd ICCV Osaka, Japan (1990: to appear)*
- 2 Häckel, M 'Operator which locates edges in digitized pictures' *J. ACM* Vol 18 (1971) pp 113-125
- 3 Marr, D and Hildreth, E 'Theory of edge detection' *Proc. Roy. Soc. Lond. B* Vol 207 (1980) pp 187-217
- 4 Canny, J F 'A computational approach to edge detection' *IEEE Trans. PAMI* Vol 8 No 6 (1986) pp 679-698
- 5 Bergholm, F 'Edge focusing' *IEEE Trans. PAMI* Vol 9 No 6 (1987) pp 726-741
- 6 Moravec, H P *Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover*

- Stanford AIM-340, USA (1977)
- 7 **Nagel, H H** 'Image sequences – ten (octal) years – from phenomenology towards a theoretical foundation' *Proc. 8th ICPR* (1986) pp 1174–1185
 - 8 **Kitchen, L and Rosenfeld, A** 'Gray-level corner detection' *Patt. Recogn. Lett.* Vol 1 No 2 (1982) pp 95–102
 - 9 **Wilson, H R** 'Psychophysical evidence for spatial channels' in **Braddick, O J and Sleight, A C** (eds) *Physical and Biological Proc. of Images* Springer Verlag, New York, USA (1983)
 - 10 **Lindeberg, T and Eklundh, J O** 'Computation of a scale-space primal sketch' *J. Visual Commun. & Image Represent.* (to appear)
 - 11 **Witkin, A P** 'Scale-space filtering' *Proc. 8th IJCAI* (1983) pp 1019–1021
 - 12 **Koenderink, J J and van Doorn, A J** 'The structure of images' *Biol. Cybern.* Vol 50 (1984) pp 363–370
 - 13 **Nagel, H H** Personal communication (1989)
 - 14 **Brunnström, K, Eklundh, J O and Kakimoto, A** 'On Focus-of-Attention by Active Focusing' *Proc. NATO ASI on Robotics and Active Computer Vision* Springer Verlag, New York, USA (1989)
 - 15 **Rosenfeld, A and Davis, L S** 'Iterative histogram modification' University of Maryland Computer Science Center TR-519, MD, USA (1977)
 - 16 **Peleg, H H** 'Iterative histogram modification 2' *IEEE SMC* Vol 8 (1978) pp 555–556