Active Detection and Classification of Junctions by Foveation with a Head-Eye System Guided by the Scale-Space Primal Sketch *

Kjell Brunnström, Tony Lindeberg and Jan-Olof Eklundh

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

Abstract. *We consider how junction detection and classification can be performed in an active visual system. This is to exemplify that feature detection and classification in general can be done by both simple and robust methods, if the vision system is allowed to look at the world rather than at prerecorded images. We address issues on how to attract the attention to salient local image structures, as well as on how to characterize those.*

A prevalent view of low-level visual processing is that it should provide a rich but sparse representation of the image data. Typical features in such representations are edges, lines, bars, endpoints, blobs and junctions. There is a wealth of techniques for deriving such features, some based on firm theoretical grounds, others heuristically motivated. Nevertheless, one may infer from the never-ending interest in e.g. edge detection and junction and corner detection, that current methods still do not supply the representations needed for further processing. The argument we present in this paper is that in an active system, which can focus its attention, these problems become rather simplified and do therefore allow for robust solutions. In particular, simulated foveation[1] can be used for avoiding the difficulties that arise from multiple responses in processing standard pictures, which are fairly wide-angled and usually of an overview nature.

We shall demonstrate this principle in the case of detection and classification of junctions. Junctions and corners provide important cues to object and scene structure (occlusions), but in general cannot be handled by edge detectors, since there will be no unique gradient direction where two or more edges/lines meet. Of course, a number of dedicated junction detectors have been proposed, see e.g. Moravec [15], Dreschler, Nagel [4], Kitchen, Rosenfeld [9], Förstner, Gülch [6], Koenderink, Richards [10], Deriche, Giraudon [3] and ter Haar et al [7]. The approach reported here should not be contrasted to that work. What we suggest is that an active approach using focus-of-attention and foveation allows for both simple and stable detection, localization *and* classification, and in fact algorithms like those cited above can be used selectively in this process.

In earlier work [1] we have demonstrated that a reliable classification of junctions can be performed by analysing the modalities of local intensity and directional histograms during an active focusing process. Here we extend that work in the following ways:

- The candidate junction points are detected in regions and at scale levels determined by the local image structure. This forms the bottom-up attentional mechanism.

* 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. We would also like to thank Kourosh Pahlavan, Akihiro Horii and Thomas Uhlin for valuable help when using the robot head.

[1] By foveation we mean active acquisition of image data with a locally highly increased resolution. Lacking a foveated sensor, we simulate this process on our camera head.

- The analysis is integrated with a head-eye system allowing the algorithm to actually take a closer look by zooming in to interesting structures.
- The loop is further closed, including an automatic classification. In fact, by using the active visual capabilities of our head we can acquire additional cues to decide about the physical nature of the junction.

In this way we obtain a three-step procedure consisting of (i) selection of areas of interest, (ii) foveation and (iii) determination of the local image structure.

1 Background: Classifying Junctions by Active Focusing

The basic principle of the junction classification method [1] is to accumulate local histograms over the grey-level values and the directional information around candidate junction points, which are assumed to be given, e.g. by an interest point operator. Then, the numbers of peaks in the histograms can be related to the type of junction according to the following table:

Intensity	Edge direction	Classification hypothesis
unimodal	any	noise spike
bimodal	unimodal	edge
bimodal	bimodal	L-junction
trimodal	bimodal	T-junction
trimodal	trimodal	3-junction

The motivation for this scheme is that for example, in the neighbourhood of a point where three edges join, there will generically be three dominant intensity peaks corresponding to the three surfaces. If that point is a 3-junction (an arrow-junction or a Y-junction) then the edge direction histogram will (generically) contain three main peaks, while for a T-junction the number of directional peaks will be two etc. Of course, the result from this type of histogram analysis cannot be regarded as a final classification (since the spatial information is lost in the histogram accumulation), but must be treated as a hypothesis to be verified in some way, e.g. by backprojection into the original data. Therefore, this algorithm is embedded in a classification cycle. More information about the procedure is given in [1].

1.1 Context Information Required for the Focusing Procedure

Taking such local histogram properties as the basis for a classification scheme leads to two obvious questions: Where should the window be located and how large should it be[2]? We believe that the output from a representation called the scale-space primal sketch [11, 12] can provide valuable clues for both these tasks. Here we will use it for two main purposes. The first is to coarsely determine regions of interest constituting hypotheses about the existence of objects or parts of objects in the scene and to select scale levels for further analysis. The second is for detecting candidate junction points in curvature data and to provide information about window sizes for the focusing procedure.

In order to estimate the number of peaks in the histogram, some minimum number of samples will be required. With a precise model for the imaging process as well as the

[2] This is a special case of the more general problem concerning how a visual system should be able to determine where to start the analysis and at what scales the analysis should be carried out, see also [13].

noise characteristics, one could conceive deriving bounds on the resolution, at least in some simple cases. Of course, direct setting of a single window size immediately valid for correct classification seems to be a very difficult or even an impossible task, since 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 modalities would be affected. Conversely, if it is too small then the histograms, in particular 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[3].

Therefore, what we make use of instead is the process of *focusing*. Focusing means that the resolution is increased locally in a *continuous* manner (even though we still have to sample at discrete resolutions). The method is based on the assumption that stable responses will occur for the models that best fit the data. This relates closely to the systematic parameter variation principle described in [11] comprising three steps

- vary the parameters systematically
- detect locally stable states (intervals) in which the type of situation is qualitatively the same
- select a representative as an abstraction of each stable interval

2 Detecting Candidate Junctions

Several different types of corner detectors have been proposed in the literature. A problem, that, however, has not been very much treated, is that of at what scale(s) the junctions should be detected. Corners are usually treated as pointwise properties and are thereby regarded as very fine scale features.

In this treatment we will take a somewhat unusual approach and detect corners at a coarse scale using blob detection on curvature data as described in [11, 13]. Realistic corners from man-made environments are usually rounded. This means that small size operators will have problems in detecting those from the original image.

Another motivation to this approach is that we would like to detect the interest points at a coarser scale in order to simplify the detection and matching problems.

2.1 Curvature of Level Curves

Since we are to detect corners at a coarse scale, it is desirable to have an interest point operator with a good behaviour in scale-space. A quantity with reasonable such properties is the *rescaled level curve curvature* given by

$$\tilde{\kappa} = |L_{xx}L_y^2 + L_{yy}L_x^2 - 2L_{xy}L_xL_y| \tag{1}$$

This expression is basically equal to the curvature of a level curve multiplied by the gradient magnitude[4] as to give a stronger response where the gradient is high. The motivation behind this approach is that corners basically can be characterized by two properties: (i) high curvature in the grey-level landscape and (ii) high intensity gradient. Different versions of this operator have been used by several authors, see e.g. Kitchen, Rosenfeld [9], Koenderink, Richards [10], Noble [16], Deriche, Giraudon [3] and Florack, ter Haar et al [5, 7].

[3] This effect does not occur for an ideal (sharp) corner, for which the inner scale is zero.

[4] Raised to the power of 3 (to avoid the division operation).

Figure 1(c) shows an example of applying this operation to a toy block image at a scale given by a significant blob from the scale-space primal sketch. We observe that the operator gives strong response in the neighbourhood of corner points.

2.2 Regions of Interest — Curvature Blobs

The curvature information is, however, still implicit in the data. Simple thresholding on magnitude will in general not be sufficient for detecting candidate junctions. Therefore, in order to *extract* interest points from this output we perform blob detection on the curvature information using the scale-space primal sketch. Figure 1(d) shows the result of applying this operation to the data in Figure 1(c). Note that a set of regions is extracted corresponding to the major corners of the toy block. Do also note that the support regions of the blobs serve as natural descriptors for a characteristic size of a region around the candidate junction. This information is used for setting (coarse) upper and lower bounds on the range of window sizes for the focusing procedure.

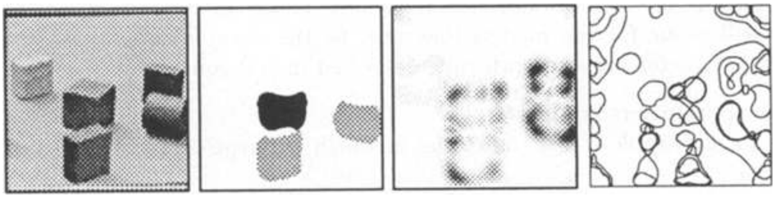

Fig. 1. Illustration of the result of applying the (rescaled) level curve curvature operator at a coarse scale. (a) Original grey-level image. (b) A significant dark scale-space blob extracted from the scale-space primal sketch (marked with black). (c) The absolute value of the rescaled level curve curvature computed at a scale given by the previous scale-space blob (this curvature data is intended to be valid only in a region around the scale-space blob invoking the analysis). (d) Boundaries of the 50 most significant curvature blobs (detected by applying the scale-space primal sketch to the curvature data). (From Lindeberg [11, 13]).

A trade-off with this approach is that the estimate of the location of the corner will in general be affected by the smoothing operation. Let us therefore point out that we are here mainly interested in *detecting* candidate junctions at the possible cost of poor localization. A coarse estimate of the position of the candidate corner can be obtained from the (unique) local maximum associated with the blob. Then, if improved localization is needed, it can be obtained from a separate process using, for example, information from the focusing procedure combined with finer scale curvature and edge information.

The discrete implementation of the level curve curvature is based on the scale-space for discrete signals and the discrete N-jet representation developed in [11, 14]. The smoothing is implemented by convolution with the discrete analogue of the Gaussian kernel. From this data low order difference operators are applied *directly* to the smoothed grey-level data implying that only nearest neighbour processing is necessary when computing the derivative approximations. Finally, the (rescaled) level curve curvature is computed as a polynomial expression in these derivative approximations.

3 Focusing and Verification

The algorithm behind the focusing procedure has been described in [1] and will not be considered further, except that we point out the major difference that classification procedure has been integrated with a head-eye system (see Figure 2 and Pahlavan, Eklundh [17]) allowing for algorithmic control of the image aquisition.

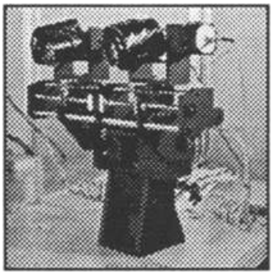

Fig. 2. The KTH Head used for acquiring the image data for the experiments. The head-eye system consists of two cameras mounted on a neck and has a total of 13 degrees of freedom. It allows for computer-controlled positioning, zoom and focus of both the cameras independently of each other.

The method we currently use for verifying the classification hypothesis (generated from the generic cases in the table in Section 1, given that a certain number of peaks, stable to variations in window size, have been found in the grey-level and directional histogram respectively) is by partitioning a window (chosen as representative for the focusing procedure [1, 2]) around the interest point in two different ways: (i) by back-projecting the peaks from the grey-level histogram into the original image (as displayed in the middle left column of Figure 5) and (ii) by using the directional information from the most prominent peaks in the edge directional histograms for forming a simple idealized model of the junction, which is then fitted to the data (see the right column of Figure 5). From these two partitionings first and second order statistics of the image data are estimated. Then, a statistical hypothesis test is used for determining whether the data from the two partitionings are consistent (see [2] for further details).

4 Experiments: Fixation and Foveation

We will now describe some experimental results of applying the suggested methodology to a scene with a set of toy blocks. An overview of the setup is shown in Figure 3(a). The toy blocks are made out of wood with textured surfaces and rounded corners.

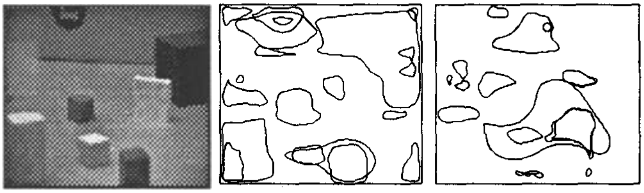

Fig. 3. (a) Overview image of the scene under study. (b) Boundaries of the 20 most significant dark blobs extracted by the scale-space primal sketch. (c) The 20 most significant bright blobs.

Figures 3(b)-(c) illustrate the result of extracting dark and bright blobs from the overview image using the scale-space primal sketch. The boundaries of the 20 most significant blobs have been displayed. This generates a set of regions of interest corresponding to objects in the scene, faces of objects and illumination phenomena.

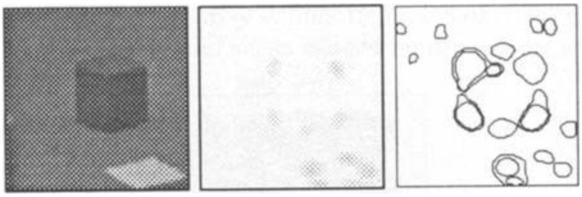

Fig. 4. Zooming in to a region of interest obtained from a dark blob extracted by the scale-space primal sketch. (a) A window around the region of interest, set from the location and the size of the blob. (b) The rescaled level curve curvature computed at the scale given by the scale-space blob (inverted). (c) The boundaries of the 20 most significant curvature blobs obtained by extracting dark blobs from the previous curvature data.

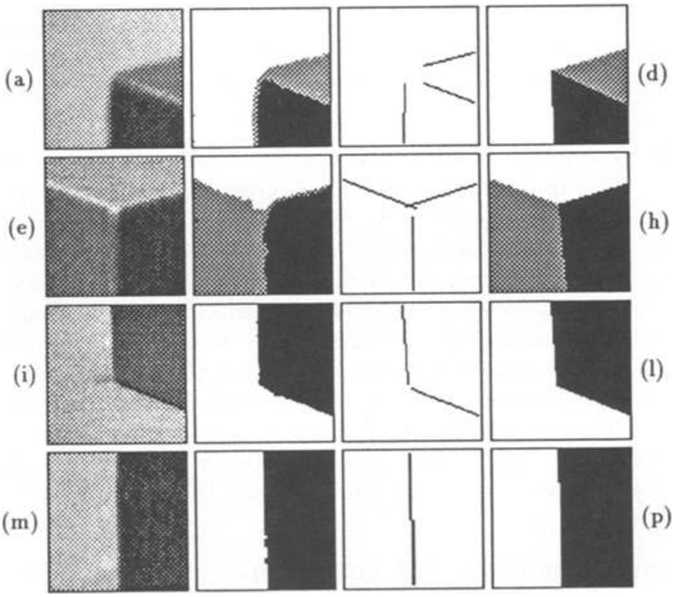

Fig. 5. Classification results for different junction candidates corresponding to the upper left, the central and the lower left corner of the toy block in Figure 4 as well as a point along the left edge. The left column shows the maximum window size for the focusing procedure, the middle left column displays back projected peaks from the grey-level histogram for the window size selected as representative for the focusing process, the middle right column presents line segments computed from the directional histograms and the right column gives a schematic illustration of the classification result, the abstraction, in which a simple (ideal) corner model has been adjusted to data. (The grey-level images have been stretched to increase the contrast).

In Figure 4 we have zoomed in to one of the dark blobs from the scale-space primal sketch corresponding to the central dark toy block. Figure 4(a) displays a window around that blob indicating the current region of interest. The size of this window has been set from the size of the blob. Figure 4(b) shows the rescaled level curve curvature computed at the scale given by the blob and and Figure 4(c) the boundaries of the 20 most significant curvature blobs extracted from the curvature data.

In Figure 5(a) we have zoomed in further to one of the curvature blobs (corresponding to the upper left corner of the dark toy block in Figure 4(c)) and initiated a classification procedure. Figures 5(b)-(d) illustrate a few output results from that procedure, which classified the point as being a 3-junction. Figures 5(e)-(l) show similar examples for two other junction candidates (the central and the lower left corners) from the same toy block. The interest point in Figure 5(e) was classified as a 3-junction, while the point in Figure 5(i) was classified as an L-junction. Note the weak contrast between the two front faces of the central corner in the original image. Finally, Figures 5(m)-(p) in the bottom row indicate the ability to suppress "false alarms" by showing the results of applying the classification procedure to a point along the left edge.

5 Additional Cues: Accomodation Distance and Vergence

The ability to control gaze and focus does also facilitate further feature classification, since the camera parameters, such as the focal distance and the zoom rate, can be controlled by the algorithm. This can for instance be applied to the task of investigating whether a grey-level T-junction in the image is due to a depth discontinuity or a surface marking. We will demonstrate how such a classification task can be solved monocularly, using focus, and binocularly, using disparity or vergence angles.

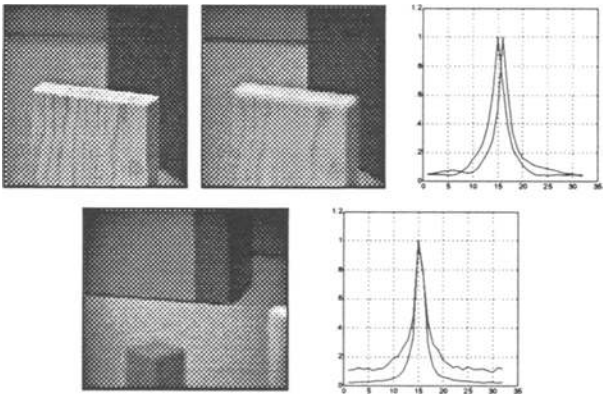

Fig. 6. Illustration of the effect of varying the focal distance at two T-junctions corresponding to a depth discontinuity and a surface marking respectively. In the upper left image the camera was focused on the left part of the approximately horizontal edge while in the upper middle image the camera was focused on the lower part of the vertical edge. In both cases the accomodation distance was determined from an auto-focusing procedure, developed by Horii [8], maximizing a simple measure on image sharpness. The graphs on the upper right display how this measure varies as function of the focal distance. The lower row shows corresponding results for a T-junction due to a surface marking. We observe that in the first case the two curves attain their maxima at clearly distinct positions (indicating the presence of a depth discontinuity), while in the second case the two curves attain their maxima at approximately the same position (indicating that the T-junction is due to a surface marking).

In Figure 6(a)-(b) we have zoomed in to a curvature blob associated with a scale-space blob corresponding to the bright toy block. We demonstrate the effect of varying the focal distance by showing how a simple measure on image sharpness (the sum of the squares of the gradient magnitudes in a small window, see Horii [8]) varies with the focal distance. Two curves are displayed in Figure 6(c); one with the window positioned at the left part of the approximately horizontal edge and one with the window positioned at the lower part of the vertical edge. Clearly, the two curves attain their maxima for different accomodation distances. The distance between the peaks gives a measure of the relative depth between the two edges, which in turn can be related to absolute depth values by a calibration of the camera system. For completeness, we give corresponding results for a T-junction due to surface markings, see Figure 6(d)-(e). In this case the two graphs attain their maxima at approximately the same position, indicating that there is no depth discontinuity at this point. (Note that this depth discrimination effect is more distinct at a small depth-of-focus, as obtained at high zoom rates).

In Figure 7 we demonstrate how the vergence capabilities of the head-eye system can provide similar clues for depth discrimination. As could be expected, the discrimination task can be simplified by letting the cameras verge towards the point of interest. The vergence algorithm, described in Pahlavan et al [18], matches the central window of one camera with an epipolar band of the other camera by minimizing the sum of the squares of the differences between the grey-level data from two (central) windows.

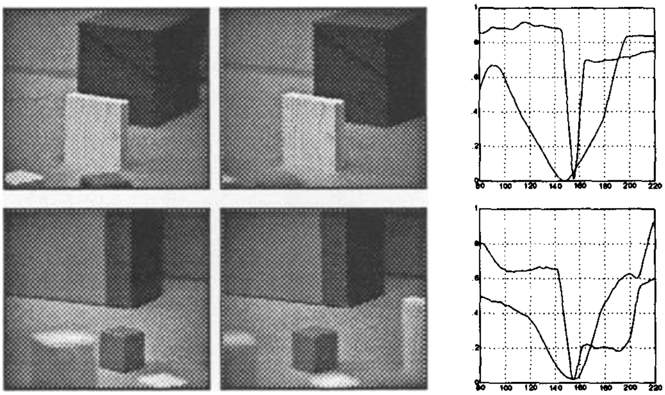

Fig. 7. (a)-(b) Stereo pair for a T-junction corresponding to a depth discontinuity. (c) Graph showing the matching error as function of the baseline coordinate for two different epipolar planes; one along the approximately horizontal line of the T-junction and one perpendicular to the vertical line. (d)-(e) Stereo pair for a T-junction corresponding to a surface marking. (f) Similar graph showing the matching error for the stereo pair in (d)-(e). Note that in the first case the curves attain their minima at different positions indicating the presence of a depth discontinuity (the distance between these points is related to the disparity), while in the second case the curves attain their minima at approximately the same positions indicating that there is no depth discontinuity at this point.

Let us finally emphasize that a necessary prerequisite for these classification methods is the ability of the visual system to foveate. The system must have a mechanism for focusing the attention, including means of taking a closer look if needed, that is acquiring new images.

6 Summary and Discussion

The main theme in this paper has been to demonstrate that feature detection and classification can be performed robustly and by simple algorithms in an *active* vision system. Traditional methods based on prerecorded overview pictures may provide theoretical foundations for the limits of what can be detected, but applied to real imagery they will generally give far too many responses to be useful for further processing. We argue that it is more natural to include attention mechanisms for finding regions of interest and follow up by a step taking "a closer look" similar to foveation. Moreover, by looking at *the world* rather than at prerecorded images we avoid a loss of information, which is rather artificial if the aim is to develop "seeing systems".

The particular visual task we have considered to demonstrate these principles on is junction detection and junction classification. Concerning this specific problem some of the technical contributions are:

- Candidate junction points are detected at adaptively determined scales.
- Corners are detected based on blobs instead of points.
- The classification procedure is integrated with a head-eye system allowing the algorithm to take a closer look at interesting structures.
- We have demonstrated how algorithmic control of camera parameters can provide additional cues for deciding about the physical nature of junctions.

In addition, the classification procedure automatically verifies the hypotheses it generates.

References

1. Brunnström K., Eklundh J.-O., Lindeberg T.P. (1990) "Scale and Resolution in Active Analysis of Local Image Structure", *Image & Vision Comp.*, 8:4, 289-296.
2. Brunnström K., Eklundh J.-O., Lindeberg T.P. (1991) "Active Detection and Classification of Junctions by Foveation with a Head-Eye System Guided by the Scale-Space Primal Sketch", *Tech. Rep.*, ISRN KTH/NA/P–91/31–SE, Royal Inst. Tech., S-100 44 Stockholm.
3. Deriche R., Giraudon G. (1990) "Accurate Corner Detection: An Analytical Study", *3rd ICCV*, Osaka, 66-70.
4. Dreschler L., Nagel H.-H. (1982) "Volumetric Model and 3D-Trajectory of a Moving Car Derived from Monocular TV-Frame Sequences of a Street Scene", *CVGIP*, 20:3, 199-228.
5. Florack L.M.J., ter Haar Romeny B.M., Koenderink J.J., Viergever M.A. (1991) "General Intensity Transformations and Second Order Invariants", *7th SCIA*, Aalborg, 338-345.
6. Förstner M.A., Gülch (1987) "A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centers of Circular Features", *ISPRS Intercommission Workshop.*
7. ter Haar Romeny B.M., Florack L.M.J., Koenderink J.J., Viergever M.A. (1991) "Invariant Third Order Detection of Isophotes: T-junction Detection", *7th SCIA*, Aalborg, 346-353.
8. Horii A. (1992) "Focusing Mechanism in the KTH Head-Eye System", *In preparation.*
9. Kitchen, L., Rosenfeld, R., (1982), "Gray-Level Corner Detection", *PRL*, 1:2, 95–102.
10. Koenderink J.J., Richards W. (1988) "Two-Dimensional Curvature Operators", *J. Opt. Soc. Am.*, 5:7, 1136-1141.
11. Lindeberg T.P. (1991) *Discrete Scale-Space Theory and the Scale-Space Primal Sketch*, Ph.D. thesis, ISRN KTH/NA/P–91/8–SE, Royal Inst. Tech., S-100 44 Stockholm.
12. Lindeberg T.P., Eklundh J.-O. (1991) "On the Computation of a Scale-Space Primal Sketch", *J. Visual Comm. Image Repr.*, 2:1, 55-78.
13. Lindeberg T.P. (1991) "Guiding Early Visual Processing with Qualitative Scale and Region Information", *Submitted.*
14. Lindeberg T.P. (1992) "Discrete Derivative Approximations with Scale-Space Properties", *In preparation.*
15. Moravec, H.P. (1977) "Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover", *Stanford AIM-340.*
16. Noble J.A. (1988) "Finding Corners", *Image & Vision Computing*, 6:2, 121-128.
17. Pahlavan K., Eklundh J.-O. (1992) "A Head-Eye System for Active, Purposive Computer Vision", *To appear in CVGIP-IU.*
18. Pahlavan K., Eklundh J.-O., Uhlin T. (1992) "Integrating Primary Occular Processes", *2nd ECCV*, Santa Margherita Ligure.
19. Witkin A.P. (1983) "Scale-Space Filtering", *8th IJCAI*, Karlsruhe, 1019-1022.