Exploring with a Foveated Robot Eye System*

Thierry Baron† and Martin D. Levine‡
Centre for Intelligent Machines
McGill University
Canada

Yehezkel Yeshurun§
Department of Computer Science
Tel Aviv University
Israel

Abstract

Computer vision systems are becoming more and more anthropomorphic. In addition to taking an active vision approach, researchers are now investigating robot heads with biologically motivated space-variant sensors. We have designed such an exploring eye system for robot gaze control when observing an arbitrary scene. Our approach uses "retinal" images constructed from overlapping receptive fields that are mapped to form log-polar "cortical" images. We have implemented this system on a parallel SIMD machine using an efficient parallelization algorithm. In order to demonstrate our approach, we have formulated various types of attentional operators that are used to select interest points to drive the gaze control mechanism. We show that our system efficiently integrates images from space-variant snapshots in an arbitrary natural environment.

Introduction

Biological vision is foveated, highly goal-oriented and task-dependent. This observation, which is rather clear if we contemplate the behavior of practically every vertebrate, is now being seriously considered by the computer vision community, mainly in recent research on active vision. A foveated retina offers several important advantages, such as data reduction, as well as object size and rotation invariance.

We use the term foveated image for the combination of *periphery* and *fovea*. The periphery possesses an irregular pixel geometry and is postulated to detect movement and points of interest in the scene. The fovea is a small uniformly sampled disk with high resolution, thereby permitting a detailed analysis of objects lying at the center of the visual field.

The use of such sensors requires efficient mechanisms for gaze control, that in turn, are directed by specific attentional processes. In psychophysical terms, these algorithms are either *overt*, analyzing in detail the central foveated area, or *covert*, analyzing various regions within the field-of-view that are not necessarily in the central foveated area.

In this work, we illustrate the performance of gaze control, based on a parallel implementation of a foveated retina, to explore and construct a comprehensive image of an unknown environment. We use a winner-take-all algorithm to integrate the individual space-variant snapshots into a unified high resolution image mapped onto a geodesic dome.

1 The *FOVIA* system

1.1 Receptive Field Distribution

Conventional video cameras perform image sampling using a regular rectangular grid. However, in the primate retina, sampling is achieved by receptive fields (RF's) which take their inputs from a contiguous cluster of cones. Thus they combine information from all photoreceptors within a "circular" area using a particular weighting function. With such a structure, it becomes possible to *overlap* neighboring RF's as in the human visual system.

Wilson[6] has proposed a model to compute a foveated retina with overlapping RF's. We have extended this model (see [7] for details) to yield n concentric rings containing overlapping RF's in the periphery, which surrounds a uniform fovea. The radii of these rings and the RF's vary logarithmically and linearly, respectively, with eccentricity.

For each RF in the periphery we use a Gaussian weighting function :

$$f_I(i,j) = \sum_i \sum_j Gauss(i,j)\ \ I(i,j) \qquad (1)$$

where *Gauss* is a 2D Gaussian function centered on the RF.

*This research was partially supported by the National Centre of Excellence IRIS program, The Natural Sciences and Engineering Research Council of Canada, and the FCAR Program of the Province of Quebec.

†Thierry Baron was supported in part by a grant from the French Lavoisier Foundation

‡Martin D. Levine would like to thank the Canadian Institute for Advanced Research and PRECARN Associates for its support.

§Yehezkel Yeshurun was partially supported by a grant from the US-Israel BSF.

1.2 The Current *FOVIA* Design

FOVIA, an active vision system being developed in our laboratory [1, 7], is a robot hand-eye configuration. The computed foveated image, which is based on overlapping receptive field theory, is obtained using a SIMD computer[1]. The hand-eye system uses the foveated image to determine various saliency maps. It then produces specific foveations or gaze trajectories from the associated cortical image.

Our Maspar contains 2048 processors, structured in an 32×64 array, which are connected to a front-end machine. The images are digitized, blocked into 32×64 image tiles, and each pixel in each tile is read into and processed on a single processor. Tiles are analyzed sequentially.

The fixed array structure of the Maspar system permits eight directional links, each with long distance communications. All non-masked processors execute the same set of instructions on their own data. The algorithms must be entirely dedicated to this specific architecture in order to provide fast execution times. The minimization of processor inter-dependence, the distances of inter-processor communications, and the number of conditional statements plays an active role in the efficiency of the implementation.

1.3 Computation of the Foveated Image

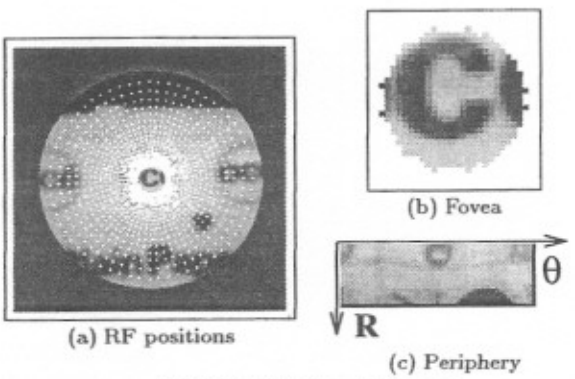
Rather than using equation (1) at each RF position in the periphery we benefit from our parallel architecture by using an iterative kernel convolution process. Hence, a particular size of Gaussian template can be obtained at each point in the periphery from the initial image by iteratively applying a fixed number of 3×3 convolutions [2]. Thus the required local averages within the RF's on different rings, which clearly have different radii, can be obtained by iterative convolution with a small weighting function or kernel.

The peripheral image f_l (where l denotes the half-size of the kernel K_l) can be obtained directly by parallel convolution with the full resolution input image f_0 using the recursive expression :

$$f_l(i,j) = \sum_{m=-1}^{1} \sum_{n=-1}^{1} K(m,n) f_{l-1}(i+m, j+n)$$

where K is the basic 3×3 kernel and f_{l-1} is the previous weighted-image result. Hence the value at any image pixel in the periphery is a weighted sum of its neighborhood pixels. By making the ultimate size of the Gaussian weighting function correspond to the radius of the RF, we can incrementally create the peripheral image. The foveal image is taken directly from the input on a one-to-one basis.

[1] A Maspar, which is a massively parallel computer system.



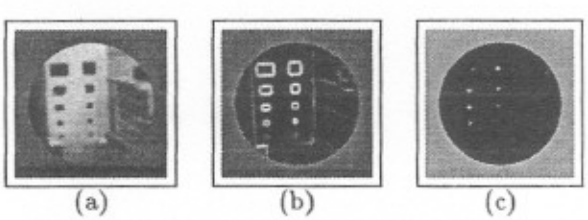
(a) RF positions (b) Fovea (c) Periphery

Figure 1: A foveated image.

Figure 1-(a) illustrates the position (R, θ) of the receptive fields according to our retinal model. The cortical representation of the foveal image is given in (b) and that of the peripheral image in (c). It can be shown that the latter is a log-polar mapping of the retinal image [1, 7].

2 The Saliency Map

Potential interest points are computed using the same operator on the whole foveated image, which we saw combines both the fovea and the periphery. This greatly simplifies the implementation of our isotropic point saliency operator. In comparison, the usual approach is to compute the interest points on the foveal and peripheral cortical maps separately [5] *after* the log-polar mapping. In this case, an isotropic operator applied to a cortical map corresponds to a *nonisotropic* Cartesian operator in the input image, since the size of RF's varies according to eccentricity.



(a) (b) (c)

Figure 2: Interest points detected in a foveated image (a) The foveated image. (b) The edge image. (d) Radial symmetry.

To compute the saliency maps, we used edge, axial and radial symmetry operators which are also implemented in parallel on the foveated image. The edge detector is standard. For symmetry detection, we employed an elegant solution based on the gradient magnitude and direction [4]. Note that this operator does not require prior object segmentation of the image and is very robust with respect to changing lighting conditions, perspective, and other noise degradations [3].

3 Foveational Control

Using the cortical saliency maps discussed in the previous section, we can control the camera viewpoint to conduct a context-free exploration of a scene. In the illustrated experiments, we use only the saliency map computed with the radial symmetry operator, chosen because of its superior stability (see Figure 2). However, in future, we plan to investigate the exploration of scenes using a combination of the different saliency maps.

3.1 The Foveational Mechanism

Different modes for controlling attention are possible with our system. In the current implementation, we use two sequential attentional mechanisms : an exploration and a data updating mode. These are necessary to acquire and maintain the necessary scene information as time progresses.

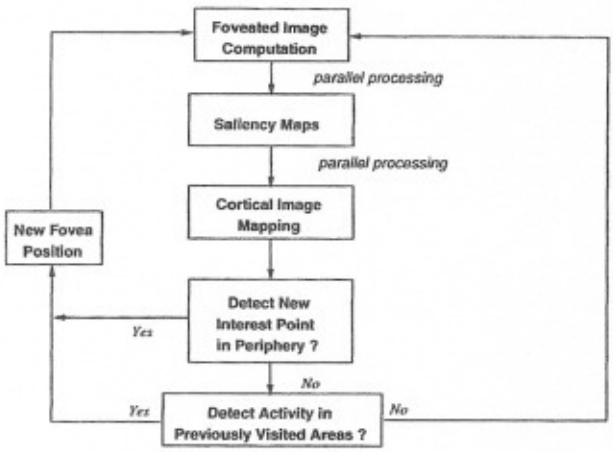


Figure 3: Foveational control.

Figure (3) illustrates the foveational control mechanism. At each cycle, we first compute the foveated image and the saliency maps. Secondly, cortical mapping is performed. We then select the point with the greatest saliency in the cortical image and use it as the next interest point. The cortical position of this point provides the next gaze point and thus a new region to explore. This process is repeated until no new interest points are detected.

3.2 Exploration

During exploration, the system uses the saliency maps to plan a trajectory which covers the interesting regions in the scene. A sequence of interest points is selected and visited in turn. Each visited position generates a restricted access area (80% of the foveal radius) around it. The next gaze point must lie outside this region.

At each foveation, we visit the point with the maximum value in the symmetry map. The ensuing behavior exhibits the following characteristics :

- Small displacements of the gaze point are encouraged, since symmetry detection is more pronounced near the fovea (related to progressively coarser edge detection with eccentricity).
- Long range displacements within the field-of-view are tolerated when the strongest response is in the periphery.
- Because of the restricted access area around the present gaze point, the local foveal region cannot participate in the selection of the next point Thus the system cannot get stuck at the same point.

After each foveation, the system provides image and other pertinent information (such as the associated pan and tilt angles, the value of the saliency, ...) about the gaze point to a long term memory resident on a separate workstation. These space-variant image snapshots of the environment are integrated into a geodesic dome representation (taking the pan and tilt angles as as the axes) using a winner-take-all strategy based on image resolution [7] .

After a finite number of foveations, the interesting parts of the image, as indicated by the symmetry operator, have been explored. If the scene is static, the system will eventually stop since no new information can be gathered.

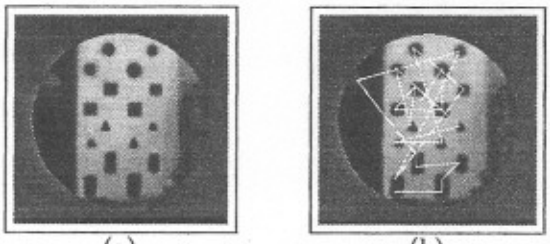
3.3 Data Updating

For a scene in which the objects may move or change, it is necessary to take into account the evolution of the already-visited gaze control positions. To keep track of ongoing activity in the scene, we continuously evaluate an activity indicator at each previously visited position. This indicator is based on the change in the saliency map value at that point. The information in the long term memory is updated should a change be detected at any of the visited positions. Furthermore, a new gaze point will automatically be selected since there would now be a previously *unvisited* gaze point with maximum saliency. Clearly this permits tracking of objects and we are now investigating how to link the gaze points using Kalman filtering.

3.4 Experimental Exploration Results

We present some results which illustrate the behavior of our system when viewing a simple image. In these experiments the object is a 2D image composed of simple geometrical patterns. First we use a

static camera ("covert" attention). Figure 4 illustrates the single foveated image (4.a) and the resulting scanpath (4.b) determined from the fixation points. From this experiment, we observe that the *FOVIA* system visits all of the salient positions in the scene once and then stops since no further change is detected. If we were now to mask one of the objects, then the system would return to the masked position in order to update the associated information at that position.



(a) (b)

Figure 4: Gaze control on a stationary image ("covert" attention)

We may repeat the same experiment, but now using the robot hand-eye system. In this dynamic case, the gaze control path and long term memory are both different from the static case.

The *FOVIA* system permits us to build an image representation of an environment. At each foveation, we merge the foveated image into the geodesic dome, thereby incrementally recreating the scene as shown in Figure (5). This global map is maintained over time by the updating activity.

Conclusions

The main goal of our project is to build a specific tool, based on analogies with biological systems, to perform foveated gaze control in an active vision system. Our implemented solution incorporates scene analysis to compute the foveated image and gaze control to move the sensor to the salient parts of the scene.

This paper presents the exploration activity of our foveated robot eye system : *FOVIA*. Analogies with human visual systems are integrated into the design of the sensing model and in the way we perform exploration of a scene without using contextual information. We use retina-like images to compute interest points and produce the corresponding cortical images from a log-polar mapping. The computations are done on the MASPAR, a massively parallel SIMD machine. The cortical images are then used to drive the attention of the system to the salient points of the scene. Foveational control permits us to build and maintain a global and integrated long term representation of the scene from a set of individual image acquisitions. In future experiments, we intend to employ this approach with a miniature sensor mounted on a mobile robot. The sensor will permit excellent camera agility and mobility, thereby producing very fast and accurate foveations.

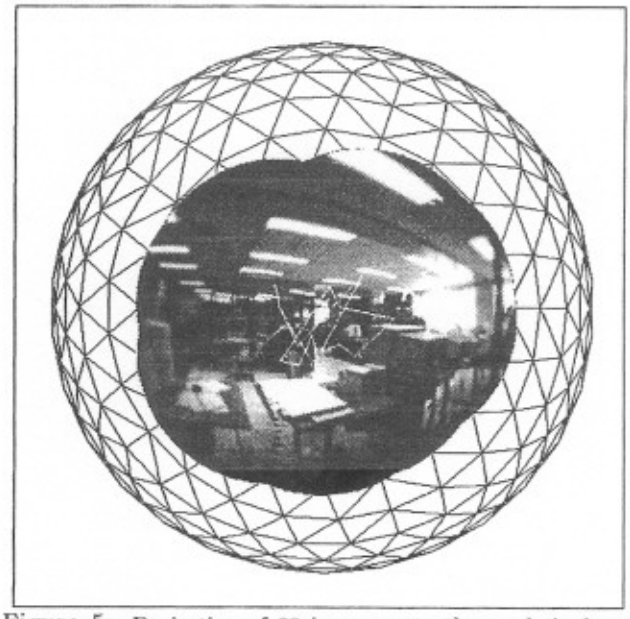


Figure 5: Projection of 23 images onto the geodesic dome which constitutes our long term memory map for room exploration.

References

[1] T. Baron, H. Yamamoto, M. D. Levine, and Y. Yeshurun. An integrated autonomous active vision system : *Xfovia*. Technical report, Centre for Intelligent Machines, McGill University, in preparation.

[2] J. J. Crowley and R. M. Stern. Fast computation of the difference of low-pass transform. *Pattern Analysis and Machine Intelligence*, 6(2):212–222, 1984.

[3] M. F. Kelly and M. D. Levine. Analysis of shape using annular symmetry operators. In *SPIE International Symposium on Optical Engineering and Photonics in Aerospace Sensing: Visual Information Processing III*, volume 2239, pages 2–13, Orlando, FL, April 1994.

[4] D. Reisfeld, H. Wolfson, and Y. Yeshurun. Generalized symmetry: A context free attentional operator. *International Journal of Computer Vision*, 1994. In press.

[5] G. Sandini and V. Tagliasco. An anthropomorphic retina-like structure for scene analysis. *Computer Graphics and Image Processing*, 14:365–372, 1980.

[6] S. W. Wilson. On the retino-cortical mapping. *International Journal on Man-Machine Studies*, 18:361–389, 1983.

[7] H. Yamamoto, Y. Yeshurun, and M.D. Levine. A vision system with a foveated image sensor. (title translated from Japanese). Transaction of the Institute of Electronics, Information and Communication Engineers. Accepted for publication, Sept., 1993 (In Japanese).