Tracking in a space variant active vision system

Tracking in a Space Variant Active Vision System

F. L. Lim, G. A. W. West and S. Venkatesh
Department of Computer Science,
Curtin University of Technology,
GPO Box U1987,
Perth 6001,
Western Australia.
email: {limfl, svetha, geoff} @cs.curtin.edu.au

Abstract

Without the ability to foveate on and maintain foveation, active vision for applications such as surveillance, object recognition and object tracking are difficult to build. Although foveation in cartesian coordinates is being actively pursued by many, multi-resolution high accuracy foveation in log polar space has not been given much attention. This paper addresses the use of foveation to track a single object as well as multiple objects for a simulated space variant active vision system. Complex Logarithmic Mapping is chosen firstly because it provides high resolution and wide angle viewing. Secondly, the spatially variant structure of log polar space leads to an object increasing in size as it moves towards the fovea. This is important as we know which object is closer to the fovea at any instant in time.

Keywords: Space Variant Active Vision, Complex Logarithmic Mapping, Tracking, Raytracing.

1. Motivation

Active vision had emerged as a new paradigm [1, 9] which integrates vision with behaviour by having control over the camera parameters. In order to solve a vision task, an active vision system has the ability to actively scan the changing environment to extract necessary information. Only a portion of an image termed the region of interest is processed to extract information. The process of locating the region of interest for high resolution processing is called foveation. Most of the current active vision systems have a uniform low resolution image and have control over the camera parameters, in this case zooming, to obtain high resolution data. Unfortunately, zooming results in a small viewing angle and requires more mechanical operations. It is important to have a wide viewing angle as it covers a larger field of view allowing the search for regions of interest. Most of the current active vision systems either i) have one window which switches between a wide viewing angle looking for regions of interest and zoom in for high resolution small viewing angle processing; or ii) have two windows active at all times, one window for small viewing angle high resolution processing, and another window for wide viewing angle looking for regions of interest (motion detection). Alternatively, the implementation of space variant active vision systems in log polar space allows high resolution foveal viewing and an exponential decrease in resolution for simultaneous periphery viewing. In this paper, we address the use of the log polar map in tracking single moving objects in complicated backgrounds and multiple objects tracking in simple binary images. The overall objectives of this paper are:

1. Apply the foveation method to develop a tracking system to actively track a moving object from frame to frame continuously in a complicated background.
2. Apply the method to track multiple objects.

We have generated a simulated tracking system that uses the open loop stage developed in our foveation algorithm [6]. The tracking system tracks a moving object from frame to frame in an indoor scene. The object is tracked by considering the distance of the object with respect to the current foveation point or camera position. Gaussian noise is also added to the image to test the robustness of the algorithm.

For the multiple object tracking system, we divide the log polar map into four regions: guarded region, active region, passive region and semi-visible region. At any foveation, a search is carried out in the active region to decide which object to track in the next foveation. The competition among objects to be foveated next is based on i) size of objects; ii) number of pixels each object moved; and iii) distance of objects from the fovea. The winner is the object

Proceedings of ICPR '96

that scores the highest and exceeds a threshold value. In situations where there is no object that qualifies as the winner, the currently foveated object in the guarded region will be tracked.

2. Single Object Tracking

Single object tracking has been discussed in [4] for $3.9 \times 10^4 \text{mm}^2/\text{pixel}$ resolution. In this paper, we extend the tracking system by reducing the resolution to $1.4 \times 10^4 \text{mm}^2/\text{pixel}$ and adding Gaussian noise into the system to test its robustness. Resolution is defined as the ratio of $\frac{\text{area in mm}^2}{\text{area in pixels}}$ where pixel is an area. For more information on resolution for spatially variant sensors, refer to [3].

The application we consider is proximity tracking [5] which is concerned with tracking the closest object. Initially, the moving object is found in the periphery of the image and the camera is moved to foveate on the object. To bootstrap, the moving object is detected by computing the difference between the images of the scene with and without the object. Then the camera is moved to foveate on the moving object. As the object moves, the camera position is updated to keep foveating by considering the closest interesting feature. This process of object motion and tracking iterates for as long as required.

We use the open loop stage in our foveation algorithm to track the moving object. Consider the example of a moving ball in a living room as shown in figure 1a. For the trivial example of a clean background, the tracking process is successful. However, problems occur when the ball passes near other features (such as the fire place) because the tracking system gets confused. We solved this problem by predicting the object motion in the normal fashion, i.e. it is assumed the ball moves in the same direction and has the same velocity as for the previous move.

In the experiment, initially, the camera is foveating on an arbitrary point in the living room scene as shown in figure 1a. Figure 1b is the log polar mapping of figure 1a. Figure 1c shows the bootstrapping condition where the camera is foveated on the ball. Figure 1d is the corresponding log polar image of figure 1c. The stripped regions represent the ball. Once the ball is foveated, the ball is moved to another location. Figure 2 shows some frames of a sequence of 160 frames generated by the tracking system. Three images are shown per frame. The left image in each row indicates the living room scene in cartesian coordinates. The centre log polar image indicates the object after motion. The right image in each row shows the ball after tracking using the open loop foveation in log polar space. We only allow one iteration of open loop tracking since tracking does not required accurate foveation. Instead speed is more important so as to get near the object being tracked. Most of the pairs of lines in the closed loop image have a small foveation error. This

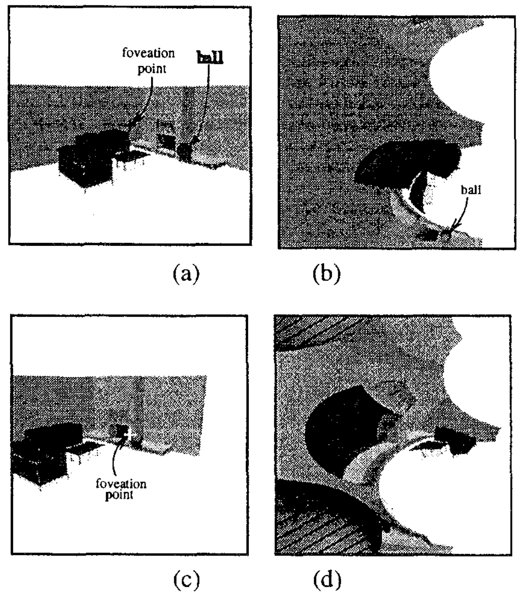


Figure 1. (a) Bootstrap condition in cartesian space; (b) image of figure 1a in log polar space; (c) first step in the tracking; (d) log polar image of figure 1c.

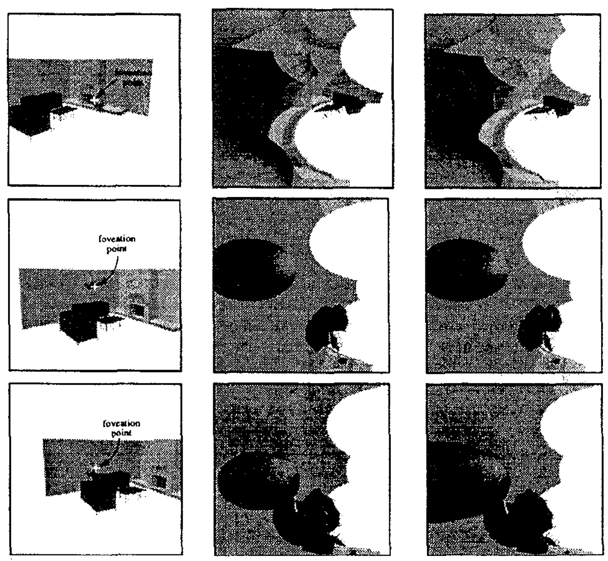


Figure 2. Tracking in log polar space. Column1: image in cartesian space; column 2: image in log polar space before tracking; column 3: image in log polar space after tracking.

is apparent as accurate foveation results in two horizontal lines at the fovea.

To test the robustness of the tracking algorithm, Gaussian noise of $\sigma = 5$ is added to the log polar images in figure 3. For comparison, figure 3 contains the same iteration as in figure 2. It is apparent that the tracking system follows a different trajectory when noise is added.

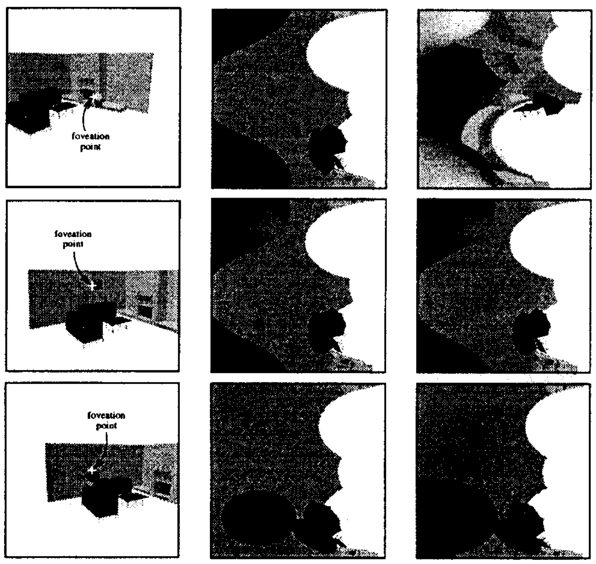
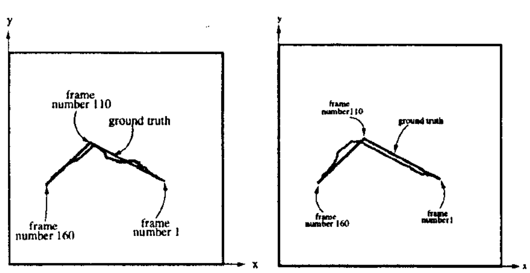


Figure 3. Tracking in log polar space with noise. Column1: image in cartesian space; column 2: image in log polar space before tracking; column 3: image in log polar space after tracking.

Figure 4 shows the ball trajectory in 2D for the tracking system without and with added Gaussian noise of $\sigma = 5$. In both cases, the ground truth is shown for comparison. The tracking system produces good results for both cases.



(a) (b)

Figure 4. Moving ball trajectories of (a) without and (b) with added Gaussian noise of $\sigma = 5$.

3. Multiple Objects Tracking

The tracking system that we have described above has been extended to multiple object tracking.

3.1. System Description

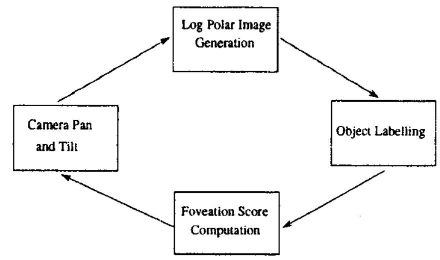


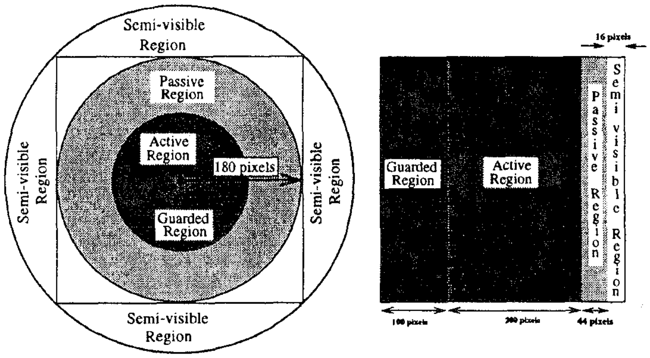
Figure 5. Schematic diagram of image acquisition and processing.

Figure 5 shows the schematic diagram of the multiple object tracking system. Initially, the log polar image is generated by using raytracing allowing very high resolution at the fovea and exponential decreasing resolution towards the periphery. This image is then analysed where every object in the image is labelled to distinguish one object from another. The decision on which object to foveate on is then carried out by considering the size, distance and number of pixels moved by objects in two consecutive images in order to select the next object to foveate on (which can be the currently foveated object). After the decision has been made, the camera pan and tilt are computed and the camera moved in order to foveate on the object. This sequence of information acquisition and image processing is executed as long as required.

3.1.1. Log Polar Image Generation

Log polar images are generated in the experiment by using raytracing. By simulation, very high resolution images can be generated, i.e. that match the human foveal resolution which is not achievable using currently available log polar sensors [2, 7, 8]. With this ability, we are able to analyse retinal images. For example, the feature invariance characteristic of the log polar image requires high resolution [3]. The log polar image is divided into four regions: guarded region, active region, passive region and semi-visible region as shown in figure 6a in cartesian space and figure 6b in log polar space.

The guarded region is of 4 pixels in radius (in cartesian space). This is a small region at/near the fovea. It is a safe



(a) (b)

Figure 6. (a) Different regions in cartesian space; (b) corresponding regions in log polar space.

region because at any instant, often the currently foveated object falls in this region.

The second region is the active region of 97 pixels in radius (in cartesian space) which is located in the periphery of the image. Computation is done at all times to check for any interesting event i.e. motion. For example, if there is a fast moving object, then that object may be foveated on next.

The passive region is a region that does not receive any attention. Objects that fall in this region are not significant (too small to be analysed in log polar space). For example, a small movement in this region is not apparent and hence will not be considered until the object has moved into the active region. The principal is that a big object, nearer to the fovea in log polar space is more significant than a small far away object. Log polar space in this case is suitable as it is of multi-resolution in that an object that is closer to the fovea looks bigger than an object further from the fovea.

The semi-visible region is so named because this region is partially shown up in cartesian space and no processing is done in this region.

The idea of dividing the log polar image into four regions is that we only extract information in guarded and active regions. Since the motion is not apparent in the other two regions, it is not necessary to process those regions. In this way, we process less data in less time compared to processing the whole image.

3.1.2. Foveation Score Computation

In order to decide which object to foveate on, a scheme has been developed which considers the size, distance and number of pixels moved by objects in two consecutive images to decide which object to foveate on either from the active region or the guarded region. The foveation function, fov_{obj} is defined as:

$$fov_{obj} = \frac{size}{distance} + Movement \tag{1}$$

where Movement is defined as the number of pixels travelled by an object in two consecutive frames.

The object to foveate on next has the highest score which is greater than a threshold value. In case there is no object that scored higher than the threshold value, and thus the currently foveated object will be continually foveated on.

To evaluate equation 1, we require two images, I_i and I_{i+1}. From equation 1, the size and distance of the object in log polar space are obtained from image I_{i+1}. The number of pixels moved by each object is then computed in log polar space using:

$$\text{Movement} = \exp(D_{i+1}) - \exp(D_i), \tag{2}$$

where D_i and D_{i+1} are the distances from the fovea in log polar space for image I_i and image I_{i+1} respectively.

Before the evaluation of the number of pixels moved by objects, we need to know which object in image I_i corresponds to which object in image I_{i+1}. In this case, we calculate the actual size of the object in cartesian space. We could not determine the size of the object easily in log polar space as log polar space is multi-resolution, which means an object that is closer to the fovea becomes larger than a larger object that is far away from the fovea.

4. Experimental Results

In the experiments, there are four fast moving objects in the scene with different speed and sizes. The number and position of objects in the scene varies depending on the current camera position and field of view. Consider figure 7, The foveation point is indicated by a " + " in all cartesian space images. Three images are shown per frame. The left image in each row indicates the scene in cartesian space. The centre log polar image indicates the scene before object tracking. The right image is obtained after tracking in log polar space. Initially, in the bootstrap condition, the closest object from the fovea is foveated as shown in figure 7b before foveation and figure 7c after foveation in log polar space. Computation is carried out to decide which object to foveate next based on equation 1. From frames 1 to 4, there is no object in the periphery that qualifies to be foveated on next and hence the camera keeps foveating on the currently foveated object. However, the foveation score for object 1 in frame 5 has exceeded the threshold value. By considering the number of pixels moved by the currently foveated object and the number of pixels moved by object 1, it is decided to foveate on object 1 from the periphery as shown in figure 7p in cartesian space.

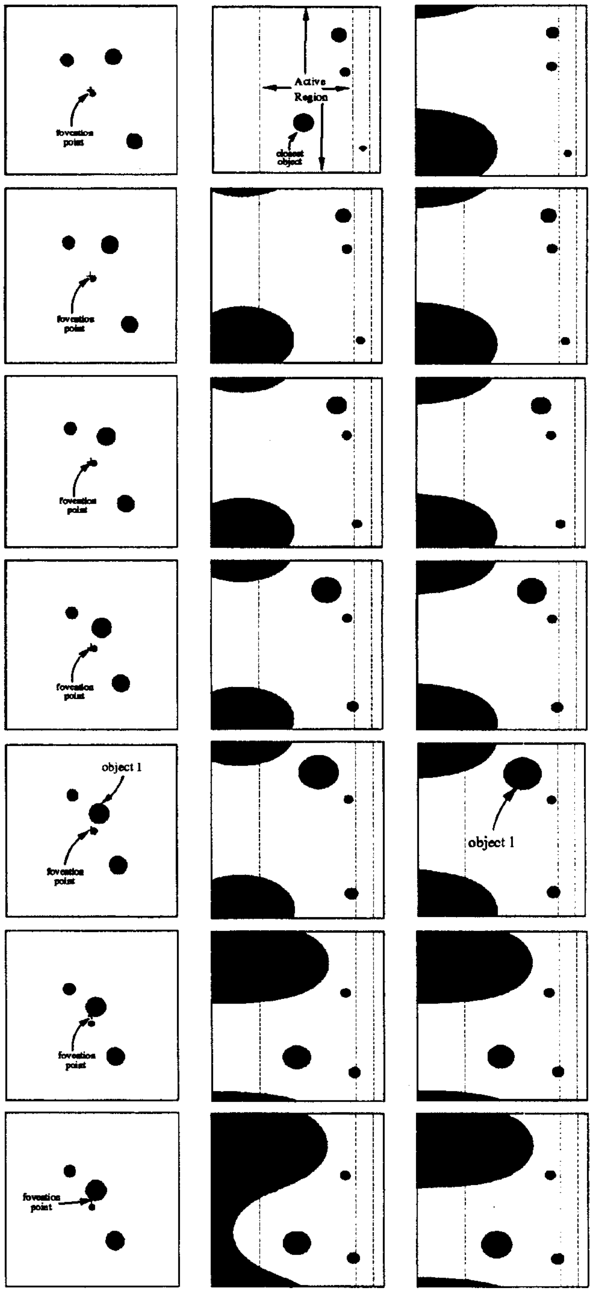


Figure 7. Multiple object tracking. Column1: Image in cartesian space; column 2: Image in log polar space before tracking; column 3: Image in log polar space after tracking.

5. Conclusions

We have shown that it is possible to implement a space variant active vision system in log polar space. The advantages of such a system is that we are able to obtain wide angle viewing and high resolution simultaneously which removes the necessity of camera zooming. Wide viewing angle is important for motion detection in the periphery. This enables the establishment of interesting features and hence the camera can be made to foveate on interesting objects. Finally, we have presented a way of deciding which objects to foveate on based on distance, size and the number of pixels travelled by objects in two consecutive images.

References

[1] D. H. Ballard. Animate vision. In *Artificial Intelligence*, volume 48, pages 57–86, 1990.

[2] B. B. Bederson. *A Miniature Space Variant Active Vision System*. PhD thesis, New York University, 1992.

[3] F. L. Lim, S. Venkatesh, and G. West. Resolution consideration in spatially variant sensor. *13th ICPR '96, Vienna, Austria*, 1996.

[4] F. L. Lim, S. Venkatesh, and G. A. W. West. Human saccadic eye movements and tracking by active foveation in log polar space. *SPIE Symposium on Electronic Imaging: Science and Technology*, 2657, 1996.

[5] F. L. Lim, G. A. W. West, and S. Venkatesh. An investigation into the use of log polar space for foveation, feature recognition and tracking. Technical Report 4, Curtin University of Technology, P.O. Box U 1980, July 1995.

[6] F. L. Lim, G. A. W. West, and S. Venkatesh. Features detection and recognition in active vision utilizing foveation in log polar space. In *Second Asian Conference on Computer Vision, Singapore*, pages 375–379, Dec, 1995.

[7] A. S. Rojer and E. L. Schwartz. Design considerations for a space variant visual sensor with complex logarithmic geometry. 2:278–285, 1990.

[8] V. D. Spiegel, G. Kreider, C. Claeys, I. Debusschere, G. Sandini, P. Dario, F. Fantini, P. Bellutti, and S. Soncini. A foveatted retina-like sensor using ccd technology. In C. Mead and M. Ismail, editors, *Analog VLSI Implementation of Neural Systems*, pages 189–211, 1989.

[9] M. J. Swain and M. Stricker. Promising direction in active vision. Technical Report CS 9-27, University of Chicago, 1991.