Automated Creation of Image-Based Virtual Reality*

Eric Bourque and Gregory Dudek
Centre for Intelligent Machines
McGill University
3480 University Street, Montreal, Canada H3A 2A7
ericb,dudek@cim.mcgill.ca

Proceedings of SPIE, “Sensor Fusion and Decentralized Control in Autonomous Robotic Systems” Volume 3209, 1997, pages 292-301

ABSTRACT

While virtual reality is a powerful tool for a range of applications, it has the following two associated overheads that fundamentally limit its usefulness:

(1) The creation of realistic synthetic virtual environment models is difficult and labour intensive;

(2) The computing resources needed to render realistic complex environments in real time are substantial.

In this paper, we describe an approach to the fully automated creation of *image based* virtual reality (VR) models: collections of panoramic images (cylindrical or spherical images) that illustrate an environment. Traditionally, a key bottleneck for this kind of modelling is the selection and acquisition of sample data. Our approach is based on using a small mobile robot to navigate in the environment and collect the image data of interest. A critical issue is selecting the appropriate sample locations of the modelling process: this is addressed using a computational mechanism that resembles human attention. Our objective is to select regions that differ from the surrounding environment. We do this using statistical properties of the output of an edge operator.

Specifically, we guide a camera-carrying mobile robot through an environment and have it acquire data with which we construct a VR model. We then demonstrate the effectiveness of our approach using real data.

Keywords: virtual reality, image registration, visual attention, autonomous mobile robotics, attention, interest operator, environment modelling

1. INTRODUCTION

1.1. Motivation

Graphical models of an environment can be used for a range of applications from architectural studies to environmental inspection to telerobotic control. Graphical models that provide a realistic visual experience are frequently referred to as virtual reality (VR) models. The standard approach to VR modelling consists of using an *a priori* manually-constructed 3D model of the environment for real-time graphic rendering from a desired viewpoint. One factor limiting the utility of VR modelling is that the construction of a realistic synthetic environmental model can be extremely labour intensive. For example, the modelling and texturing of a single object can take months.[†] In addition, the computational burden involved in rendering scenes for model-based VR can be substantial. Finally, obtaining a truly realistic result for an arbitrary environment remains exceedingly challenging.

Image-based virtual reality refers to the use of real image data (i.e. photographs) of an existing environment or model to create a VR environment. The image-based VR interface allows a user to look in arbitrary directions from a given viewpoint, and to jump between pre-computed viewing locations. Although observer motion is constrained, image-based VR permits extremely realistic scenes to be displayed and manipulated in real time using commonplace computing hardware. The commercial product QuickTime VR (a trademark of Apple Computer) exemplifies this technology.

[*] This work is supported by a grant to G. Dudek from the National Sciences and Engineering Research Council of Canada.

[†] The construction of a single complete textured model of an airplane in the film “Con Air” took two months,[1] and the rendering alone in the Disney Production, “Toy Story” took 800,000 machine hours.[2]

The use of image-based VR addresses the shortcomings of limited realism and/or high computational load imposed by conventional model-based VR. Unfortunately, it only partially alleviates the intensive effort needed to create a VR world model: the acquisition of the requisite images to construct an image-based VR model still entails effort and expertise. In addition, if a scene must be modelled under different conditions, for example during the night and the day, or during weekdays and weekends, the image acquisition processes will have to be carried out repeatedly. Furthermore, selecting suitable vantage points to produce an evocative and complete VR model is in itself an important issue. This paper deals with the automated acquisition and construction of image-based VR models by having a robotic system select and acquire images from different vantage points. The objective is to provide a fully or partially automatic system for both the selection and acquisition of the needed image data. In principle, this can be augmented by additional cues provided by a human operator.

1.2. Applications

Image-based VR modelling appears promising in several contexts in addition to entertainment. In particular, these include task domains where the scene to be examined is either too remote, too dangerous, or inconvenient for a human operator to visit directly. As such, the potential application contexts overlap those for teleoperated robotics. This range of applications includes museum previews by computer (for example on the world wide web), security or surveillance applications where an environment must be inspected at some periodic interval, or the exploration of remote locations such as undersea or on another planet.

1.3. Overview

A primary bottleneck in the use of image-based VR is that the creation of the models is time consuming and requires specialised expertise. The key issues are the selection of suitable vantage points to cover the interesting aspects of the environment and the acquisition of suitably calibrated images. This is followed by post-processing of the image data to provide the VR model. When an image-based VR model consists of a collection of viewpoints between which the user can move, it is referred to as a “multi-node model”. The selected viewing locations form the nodes of a graph that determines possible (discontinuous) motions that a user may experience when using the model. In this paper, we describe an approach to the fully automated creation of image based VR models of a finite environment with essentially no human intervention.

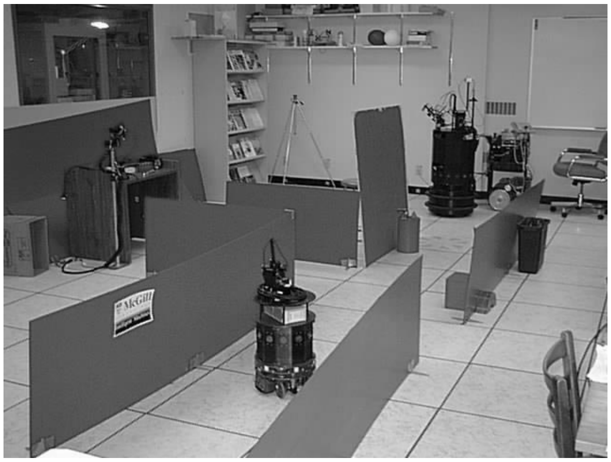


Figure 1. Laboratory set up for the environmental VR experiment. The RWI mobile robot is shown in the bottom portion of the picture, while the Nomad (not running in this particular experiment) is shown above.

Our approach is based on using a small mobile robot (shown in Fig. 1) to navigate about in an environment and collect the image data of interest. The main problem is how to select the appropriate sample locations for the modelling process: these will form the nodes of the multi-node model. We deal with this using a computational mechanism inspired by models of human attention. Our approach assumes that an autonomous exploration algorithm is available. While several such algorithms have been developed and implemented in our lab and elsewhere,[3–6] their details are outside the scope of this paper. The current work simply presupposes that the robot travels along some trajectory through the environment, and that it can estimate its current position at any time. In fact, the exploration could even be manually controlled.

Since our objective is to construct a virtual environment that appears subjectively realistic to human observers, our approach is inspired by models of human visual environment exploration. In particular, human exploration of either an environment or an image is driven by a shifting attentional "spotlight".[7] In building models of human attention, substantial research has been devoted to the computational mechanism involved.[8,9] We concentrate here, rather, on the *locations* to which attention is driven. One class of attentional processing is characterised by visual saccades to areas of high curvature, or sharp angles.[10] More generally, things which are "different" or inconsistent with their surroundings tend to attract visual attention. Thus, our approach is to compute a map over an image (perhaps a 3-dimensional image) of how much each point attracts attention. Selecting the extrema of this map provide a set of attentional features.

Our computational procedure for defining features is dependent on finding edges in the image. Edge structure has been used extensively in computational vision for several reasons including the apparent psychophysical relevance of edge structures, and because there is evidence that that visual edges tend to be highly correlated with the projection onto the image of real physical events in the world (eg. object boundaries, markings).[11] Higher level features such as large-scale edges, curves, circles, or corners, are difficult to detect robustly, and organising edge elements produced by an edge operator (*edgels*) into semantic tokens is a notoriously difficult problem. Edgels, however, have the advantage of being robust to variations in illumination, and are strongly suggestive of actual geometric structures in the environment.

It is with this in mind that we have formulated a metric for visual attention based on edgel density. For example, areas with rich geometric content will have a high edgel density. To focus attention at locales that are notable, our attention mechanism is driven to locations where the local edgel density differs substantially from the mean edgel density.

2. APPROACH

The set of all possible views or images obtainable from a fixed location in the environment can be described as a *viewing sphere* or spherical image. More specifically, for every ray projected from a location in R^3, in a direction along the unit sphere S^2 we get an associated intensity from the environment. This transformation can be expressed as:

$$M_{3D} : R^3 \oplus S^2 \longrightarrow R \tag{1}$$

or

$$M_{3D}(x, y, z, \phi, \theta) = i \tag{2}$$

where (x, y, z) are spatial coordinates, (ϕ, θ) refer to the orientation of a light ray, and i is the intensity observed. This parameterisation of light rays is related to the *light ray manifold* defined by Langer and Zucker[12] and the Lumigraph.[13]

In our particular case, we have a camera mounted on a pan and tilt unit at a fixed location on a mobile robot. For the purposes of this paper, let us assume that the robot is constrained to a flat floor, and thus we restrict the camera to a 2-dimensional plane. This constrains the origin of the ray to R^2, and we have the idealised 2-D observer in a 3-D world:

$$M_{2D} : R^2 \oplus S^2 \longrightarrow i \tag{3}$$

or

$$M_{2D}(x, y, \phi, \theta) = i. \tag{4}$$

A minor variation is the case of an idealised camera which only pans, which is the case for much image-based VR. Since we are now dealing with a camera, as opposed to a single ray, the result of the transformation is an image or a set intensities given by a cone about the camera direction:

$$M_C : R^2 \oplus S \longrightarrow R^2 \tag{5}$$

where $M_C(x, y, \phi) = \mathbf{I}$ now denotes a pixel-indexed image $I(a, b)$ implicitly dependent on the field of view of the camera. Each pixel is, of course, also specified by Eq. 4. An entire *spherical* panoramic image $I(x, y)$ where each pixel is a ray corresponding to Eq. 4 is given by

$$M_S : R^2 \longrightarrow R^n \tag{6}$$

where n is the number of pixels in the image, thus leading to a parameterisation of a *set* of images $\mathbf{I}_{x,y}(a, b)$ whose individual pixels implicitly depend on the viewing parameters of the camera.

It will prove convenient to consider the problem of directing attention to a sub-image of a large image. In this case, we specify its pixels as $I_{x,y}(\phi, \theta)$ so that it notationally resembles the specification of a pixel from a set of panoramas in Eq. 6:

$$I_{x,y}(\phi, \theta) = I(x + \phi, y + \theta) \tag{7}$$

An image-based VR model is founded on approximating a continuous set of spherical images given by Eq. 6 from some (discrete) set of representative points in the environment. In practice, image based VR allows a user to move between specific locations and look in (almost) any viewing direction from any of these locations.

To construct an image-based model, we must first gather a set of images from each point $\mathbf{P}_i = (x_i, y_i)$ in the environment we wish to model. These images are then tiled into a mosaic which can be subsequently mapped onto a viewing volume.[14,15] In practice, the mosaic is produced by "stitching" or fusing all of the images from one sample location into one composite image.[14] This involves registering consecutive images with one another using methods analogous to those used in stereo correspondence. In practice, this implies that camera rotation should be about the nodal point of the camera, that the scene should be static (or the sequence should be acquired as quickly as possible), that lighting should remain constant, and that camera motion and parameters must be minimised. These types of constraints, while conceptually trivial, substantially complicate the manual acquisition of image data for VR mosaics.

The shape of the panoramic image that is used can vary: both spherical and cylindrical projections have attractive properties, with cylindrical projections being predominant in existing applications. This gives the viewer a limited viewing hemisphere, in that information is lacking at the vertical extremes. For any viewing vector $\mathbf{v} = (r, \phi, \theta)$ where r represents the zoom factor, and ϕ, θ are the Euler angles, we can then map the appropriate field of view onto a planar surface for display.[15] The sampling location $\mathbf{P}_i$ in the environment now encompasses all possible viewing directions, within the constraints of the cylindrical map, and is defined as a *node*. To construct a navigable environment, several such nodes must be created, as well as a method defined for inter-nodal movement. In practice, one can define *hot-spots* within the images to create such links in the nodal graph. The desired result is to obtain a graph composed of such nodes which encompass all the distinctive regions in the environment, as well as a means of navigating smoothly between them. That is, if two nodes are chosen which have no overlapping visual information, it would be desirable to have a node in between which would allow a smooth transition. It is the automated selection of the nodal positions $\mathbf{P}_i$ which we will now develop further.

2.1. The 2-D Analogue

The logistics of acquiring sufficient data to evaluate alternative viewpoint selection algorithms in real environments is substantial. In order to readily test alternative methodologies in a controlled and pragmatic setting, we have used a analogue of the full problem based on selecting regions of interest within a two dimensional image. This also has the advantage of being closely related to existing models of human attentional processing.[8]

As noted above, the problem of selecting a sub-image from within a larger image can be expressed in the form of Eq. 7. If we consider spherical panoramas under orthogonal projection (i.e. without zooming due to perspective) the geometry and semantics of windowing within a 2-D image and real panorama selection from a set of spherical images are very similar. Thus, we can evaluate panoramic viewpoint selection methods by looking at their performance as region-of-interest selectors in 2-D images.

3. METHODOLOGY

To encode an environment using image-based VR nodes, we must first determine which locations in the environment, that is, which viewing cylinders from the set of all those possible, are most worthy of retention. We accomplish this by establishing which viewing spheres are most *interesting*, where interest is measured by the extent to which a location attracts visual attention.

To construct our attention operator, we will consider the case of simple two dimensional images. In principle, we would like to exploit image geometry and semantics. Work in human psychophysics suggests that various types of geometric structure such as line endings, oriented line segments, or curves "pop out" of an image when they are different from the rest of the scene.[8] Since edge linking and segmentation remain open problems in a generic context, we settle, instead, on exploiting the variations in the *distribution* of edge *elements* as cues to attentional fixation. Since both psychophysics and intuition suggest that we wish to concentrate on regions that are unusual or distinctive, we can evaluate the extent to which regions of an image differ from the mean. We begin with an image $I_{x,y}(a,b)$ and compute the binary edge map $E(I_{x,y}(a,b))$. A generic and computable metric for image content is local edge-element density. We compute this by convolving the image with a windowing operator to obtain the local edgel density $D(i,j)$. While in principle a Gaussian windowing function is suitable, in the interest of real-time performance we use a square-wave kernel of size AB:

$$D(i,j) = \frac{1}{4AB} \int_{j-B}^{j+B} \int_{i-A}^{i+A} E(I_{x,y}(a,b))\ da\ db \tag{8}$$

Note that this can be computed in the context of images from either Eq. 6 or 7.

The interest value of a point is then given by the absolute deviation from the mean local edgel density $\hat{D}$:

$$\mathcal{D}(i,j) = |\hat{D} - D(i,j)| \tag{9}$$

We then sort these points $\mathbf{P_i}$ based on their absolute deviation $\mathcal{D}$ from the mean to provide a list of the K most-interesting locations for which nodes are created. In practice, additional constraints such as assuring no two points are too close together are desirable. For the purposes of the present synopsis, we will simply assure that no two regions on the list of interesting places are permitted to overlap. If they do, we evaluate pairwise combinations and delete the less interesting of the two.

In order to illustrate and experimentally exemplify our approach in this paper, we first model the environment using a single image as in Eq. 7.

4. IMPLEMENTATION

Our approach to attention described above assumes that the statistics of the edge distribution of the environment are fully available when decisions are made. Such a paradigm is sometimes referred to as an *off-line algorithm.* In this context, it involves an analysis of image data from every point in the environment, followed by a selection of the best few locations for which panoramic image nodes are subsequently created and interconnected. Examples of the performance of this approach are presented in Section 5.

4.1. On-line Viewpoint Selection with α-backtracking

In practice, as the robot moves through the environment it would be highly advantageous to make decisions when locations are encountered so that there is no need to either acquire and store immense amounts of data, or backtrack to re-visit selected locations to obtain the panoramic images. To do this, nodes must be selected based only on partial information of the statistical distribution of image content over the environment giving rise to an *on-line algorithm.* Assuming that the off-line algorithm performs well, we seek an on-line algorithm whose performance is a good approximation of that obtained with the off-line method.

We can assure that the on-line algorithm exhibits arbitrarily good performance, as compared to the ideal of the off-line algorithm, by permitting the robot to backtrack. We can define the *forward interest* of a point from partial information as

$$\mathcal{D}_t(i,j) = |\hat{D}_t - D(i,j)| \tag{10}$$

where the subscript t denotes statistics computed from the initial fraction $t \in (0,1]$ of the entire data set. We define *on-line viewpoint selection with α-backtracking* as a variant of the off-line algorithm such that the best K non-overlapping points are selected as the exploration proceeds. As each point is selected, a corresponding panoramic node is constructed. Density values are also stored for all other points visited. As the exploration proceeds t increases and the forward interest values of previously visited locations may evolve. If a prior unselected point *which is no further back than a fraction α of the current trajectory length* becomes more interesting than one of the K selected points, the robot backtracks and uses it instead of the point it replaces. Clearly, the performance (in terms of the points selected) of this algorithm approaches the ideal as α approaches one.

4.2. Single Image

Due to the complexity of testing a system involving a mobile robot, we have implemented a system capable of analysing a single image representing the entire environment as described in section 2. For the off-line version of the problem, we consider moving a box of constant size over the entire image and calculating the edgel densities, as discussed in section 3. This enables us to easily compute the required statistics to choose the regions of the image which have edge densities which are the largest absolute deviations. One key point of using the absolute variance is exemplified when considering a textured image. In this case, it is regions which are large deviations *below* the mean which will be of interest.

For the on-line version of the problem, we move a rectangular window over a sub-part of the image and calculate the approximations incrementally. This can provide a compromise that achieves good performance with limited storage requirements. Unfortunately, the characteristics and performance of the on-line algorithm are outside the scope of this paper.

4.3. Environmental VR

Our approach to environmental VR is based on having a mobile robot traverse the 2-dimensional environment to be mapped[‡]. It is independent of the traversal strategy, although it assumes that the topology of the trajectory is known so that the multi-node model can be constructed. In addition, in order to avoid closely-spaced sample nodes, an approximate *local* estimate of distance is desirable (eg. from odometry). Our technique has been implemented to function with robots from both RWI and Nomadic Technologies, both of which provide odometry data whose accuracy is far in excess of our requirements.

The robot we used has a digital camera mounted on a pan and tilt unit, and is capable of translating and rotating, thus giving a the final camera position a total of 4 degrees of freedom (DOF). In practice, we have fixed the tilt of the camera, and therefore define an (x, y, θ) location for each image taken in the environment. During the exploration, the robot periodically stops and gathers a set of images in one or more orientations, so that they may be evaluated for candidacy within the set of salient points. The most complete sampling of the environment demands a set of images which collectively cover the view in every direction; that is the data presented in this paper.

The difference between the off-line and on-line implementations is apparent at this point. In the case of the on-line algorithm the robot must immediately decide whether or not to create a VR node and obtain additional data if necessary. If the current location is not selected, the image data acquired can be discarded. In the off-line case, on the other hand, the robot must retain all of the data. The final decision regarding node selection is made after the exploration is complete. Figure 2 shows a flow chart of the two approaches. In the off-line case, the robot must acquire all the additional data for a panorama at each location, or return to the selected locations after selection is complete. These nodes are then incorporated together to form a complete image-based VR representation of the previously unknown environment.

The need for additional images for panorama creation, above and beyond those needed for viewpoint selection, has several sources. Under-sampling a location may be sufficient for interest determination as a result of task constraints. In addition, several existing algorithms for stitching images to generate a panorama require substantial overlap (on the order of 50 *per cent*) between source images.[14,16] Since an overlap of this size simply multiplies the number

[‡] While there is no reason the methods we have discussed could not be extended to R^3, it is simply outside of the scope of this paper.

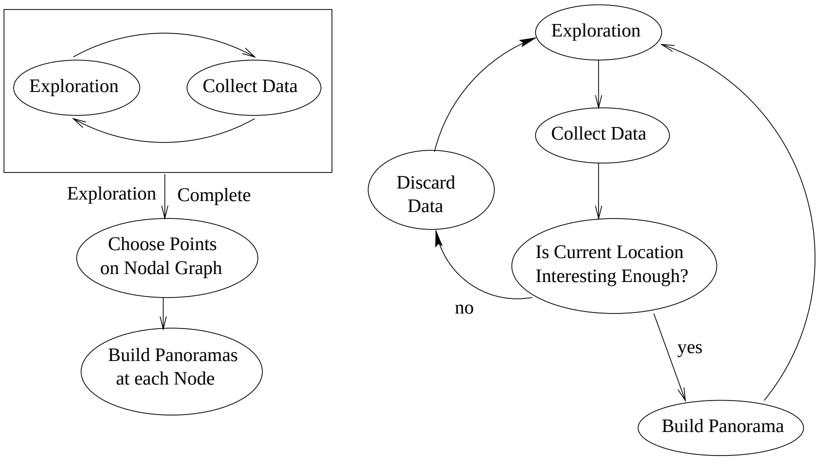


Figure 2. Flow charts for the off-line (left) and on-line (right) implementations of the algorithm.

of times we have seen artifacts in the environment, it does not have an effect on the final outcome of the chosen locations. Therefore taking images which barely overlap allows the amount of data stored to be reduced.

5. EXPERIMENTS

5.1. Single Image

For the purposes of our single image experiments, we acquired several photographs of natural scenes using a digital camera. The goal of these experiments was to choose points in the images which most closely matched the points which would attract the focus of a human observer. The software we developed for this analysis convolved the image using the kernel described in Eq. 8, and subsequently sorted the resulting interest points $\mathbf{P}_i$ by their descending absolute deviation. A simple heuristic was used to guarantee that there was no overlap in the resultant selections. For the images shown in Fig. 3, a kernel of size 90×90 pixels was used.

In order to demonstrate the advantages of our interest operator, we chose photographs which had interest points which were well defined semantically, but which contained substantial texture, thus allowing them to be potentially confusing to other types of operators.

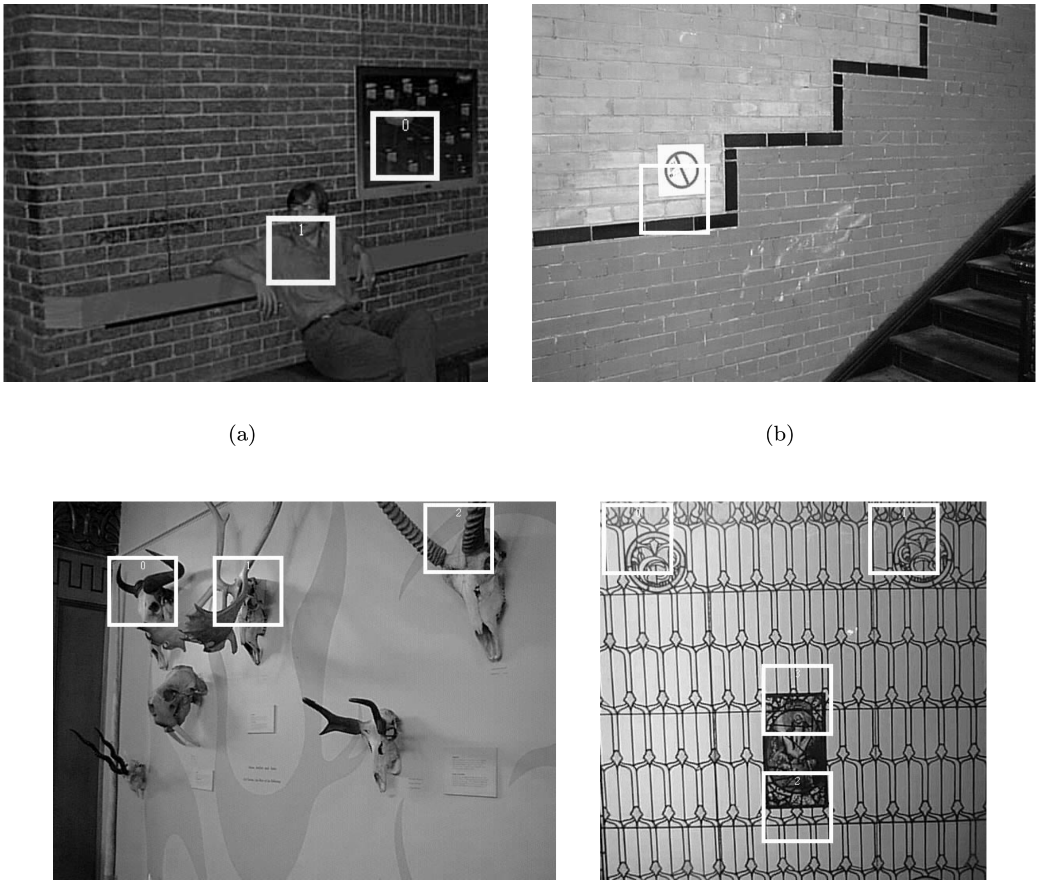
We believe that the locations chosen by our interest operator were in fact the correct ones, that is, that they match most closely what would attract the attention of a human observer. Although the formal substantiation of these results is beyond the scope of this paper, we are pursuing this objective.

5.2. Environmental VR

We have examined the performance of our environmental sampling technique in the context of a small fully-controlled test environment that can be manipulated at will (see Fig. 1). In the configuration shown here, the environment resembled an office-like world or a section of a maze. We used a small mobile robot with a top-mounted camera to navigate this environment[§]. The field of view of the camera mounted on the pan and tilt unit of the robot was such that the tops of the images were just slightly below the tops of the walls of the test maze. This was to ensure that no spurious image data was brought into the experiment. Two pictures were mounted on the inside of the walls at different locations, and a few small objects were dispersed throughout the environment. At each of the junctions there were views into the “open world” which were considerably different from the somewhat constrained internal environment.

The robot followed the path shown in Fig. 4, and collected 35 images from 12 vantage points for a total of 420 images gathered within the environment. In this experiment the robot used a simple path-planner to guide it

[§] An alternative, larger, robot is used for conventional environments.



(a)

(b)

(c)

(d)

Figure 3. Results of interest operator on 2-D images.

through the entire short trajectory. The actual image data was collected with a control system that automatically directed the camera at the candidate interest points.[17] The robot is equipped with sonar sensors that can be used for navigation or collision avoidance. Raw data from these was used to illustrate the layout of the environment in approximate form. The sonar map resulting from the exploration is shown in Fig. 5.

The image data was then analysed and the images were sorted by their descending absolute deviation as described in section 3. Figure 6 shows various images and their rank in the resultant sorted collection of images.

As was expected, the selected regions were those which encompassed the edge information at the extrema of the edgel density distribution. All of the junctions in the environment revealed information which was quite different from that contained within, and were therefore suitable candidates for selection. Other potential candidates from the environment were the objects, and pictures placed on the inside of the walls. The selection algorithm performed very well in this regard – the top selection was one which contained two objects: a gold samovar[¶], and a picture on the wall. Given the intended applications of this research, these results are very promising.

Also of interest are the image samples which placed in other areas of the order. Since the order is a direct function of image distinctiveness, we would hope that the samples become decreasingly interesting as their sorted list

[¶] A Russian urn used for boiling water for tea.

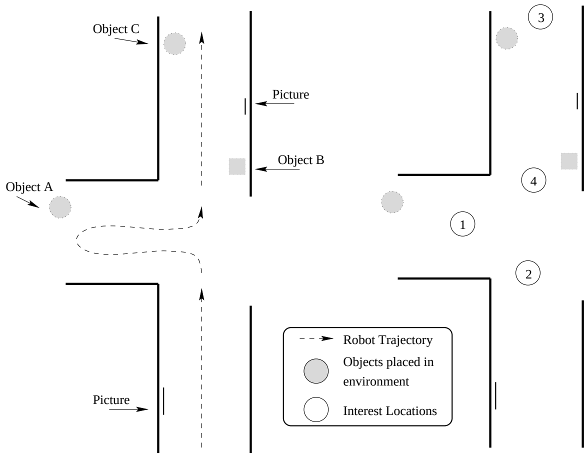
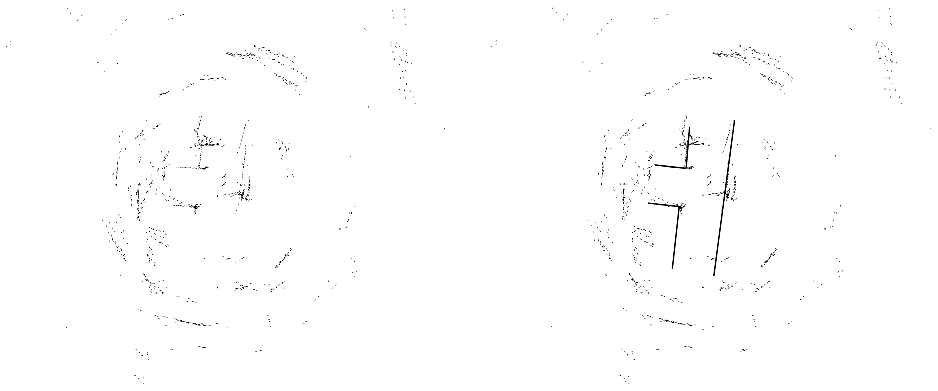


Figure 4. Left: Trajectory taken by the RWI through the unknown environment. Right: Node locations chosen by the multiple image, off-line implementation. The numbers inside the circles denote order of distinctiveness.



(a) Sonar map generated during the exploration of the unknown environment.

(b) Same map with walls drawn over the sonar data.

Figure 5.

is traversed. In our experiments, the samples from the middle and bottom of the order have revealed a diminishing amount of disparity.

In particular, our direct goal is to provide a complete nodal graph through the distinctive parts of the environment. The right-most portion of Fig. 4 exemplifies this accomplishment.

Since it is difficult to provide a virtual reality experience in a publication, we have decided to include the panoramic photographs resulting from the top four selected locations. By comparing Fig. 1 which portrays the environment from above, and the four panoramas in Fig. 7, it is clear that the environment has been captured.

6. SUMMARY AND DISCUSSION

In this paper, we have outlined an approach to the selection of representative views to convey the appearance of an otherwise unknown environment. The method assumes that a set of spherical or cylindrical images are the medium used to store and convey the appearance; these are presented using image-based virtual reality.

Our secondary objectives are to limit the number of spherical images acquired and used, and to minimise the trajectory length needed to acquire the images. In the experiments presented here, the image acquisition is accomplished using a mobile robot. By using an attention mechanism, the robot acquires images only at "interesting" locations.

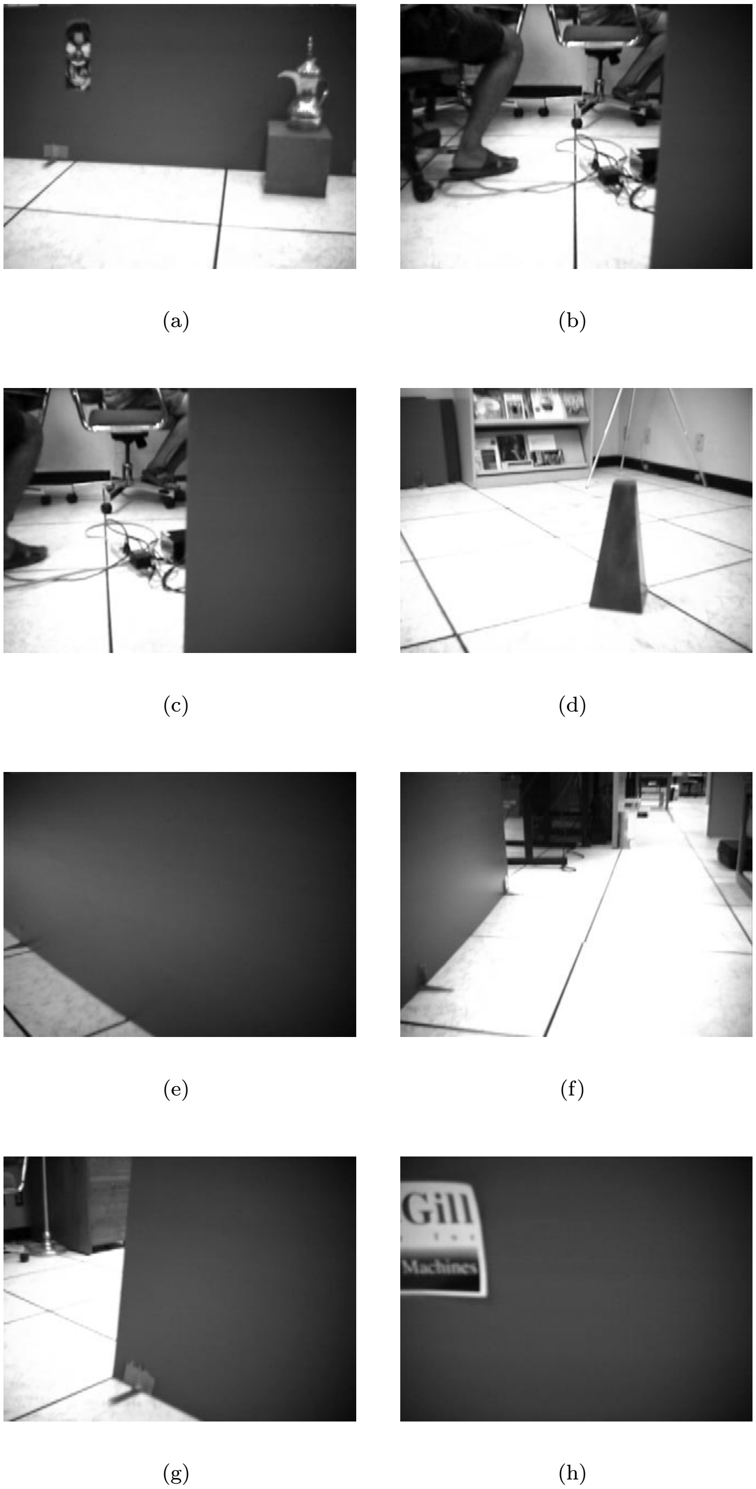
The image samples we obtain seem to effectively capture many important aspects of the scene being observed. For some applications, it may be desirable to explicitly specify certain views of interest *a priori*, for example for an art gallery it may be important to have frontal-parallel views of the pictures. Such views could be specified either manually (using map information) or procedurally (using a task-specific set of criteria). Exactly how to facilitate the specification of domain specific viewing constraints remains a subject for future investigation.

Even when the desired views are chosen manually, the generic attention mechanism is useful for selecting additional views of the environment to provide continuity and completeness in the VR viewing experience. In this case, the objective is to complement the manually specified views. In addition to supplementing the manually selected views, this work provides a mechanism for regularly and effortlessly updating the VR model; this update ability is an advantage even if *all* the views are selected manually.

In ongoing work, we are developing alternative attentional functions to specify where the views should be selected. In addition, we are developing a quantitative framework to measure the quality of the VR model generated, using both psychophysical criteria and image reconstruction paradigms.

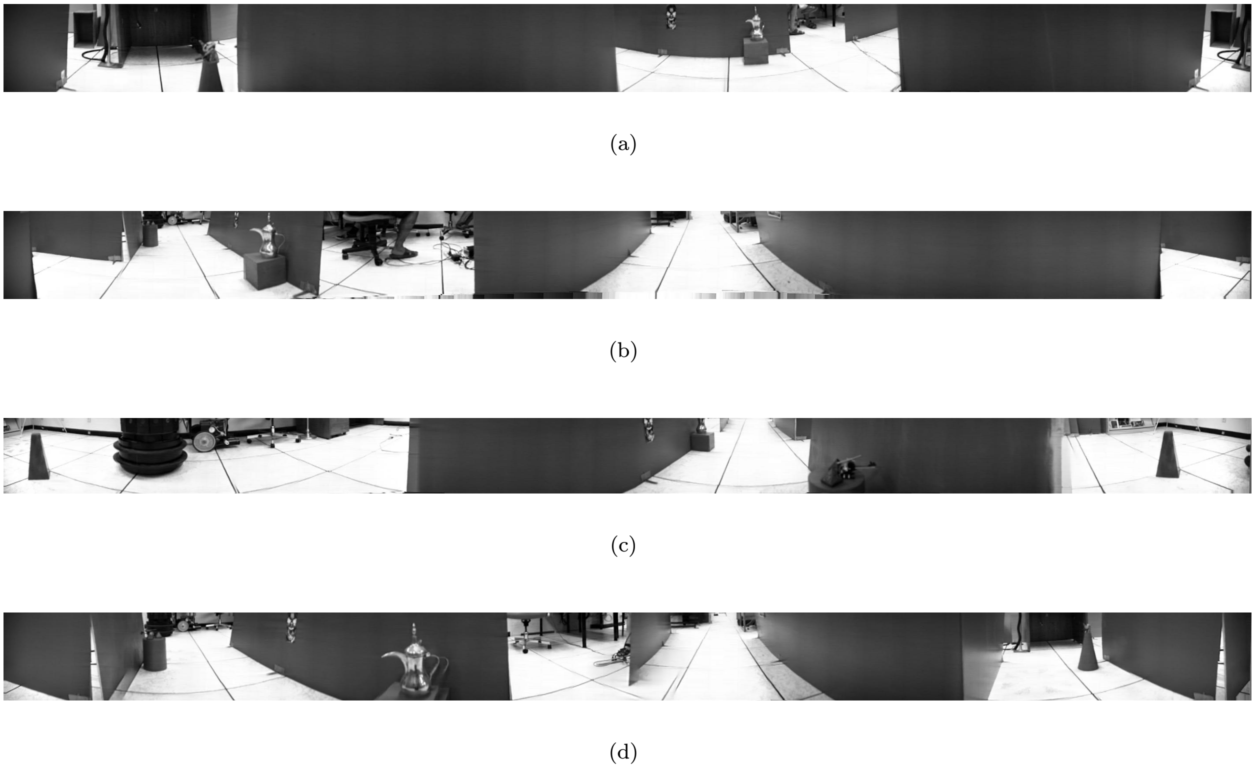
REFERENCES

1. J. Rygiel, "Digital effects in motion pictures," *Invited talk, Computer Vision and Pattern Recognition* , June 1997.
2. A. Thompson, "Toy wonder," *Entertainment Weekly* **304**, December 1995.
3. N. Roy, "Multi-agent exploration and rendezvous," Master's thesis, McGill University, 1997.
4. G. Dudek, M. Jenkin, E. Milios, and D. Wilkes, "Robotic exploration as graph construction," *IEEE Transactions on Robotics and Automation* **7**(6), pp. 859–865, 1991.
5. G. Dudek, P. Freedman, and I. M. Rekleitis, "Just-in-time sensing: efficiently combining sonar and laser range data for exploring unknown worlds," in *Proceedings of the IEEE International Conference on Robotics and Automation*, vol. 1, pp. 667–671, (Minneapolis, MN), April 1996.
6. B. Kuipers and T. Levitt, "Navigation and mapping in large-scale space," *AI Magazine* , pp. 25–43, summer 1988.
7. R. M. Shiffrin and W. Schneider, "Controlled and automatic human information processing: II. General theory," *Psychological Review* **84**, pp. 127–190, March 1977.
8. A. M. Triesman, "Perceptual grouping and attention in visual search for features and objects," *Journal of Experimental Psychology: Human Perception and Performance* **8**(2), pp. 194–214, 1982.
9. J. K. Tsotsos, "Analysing vision at the complexity level," *Behavioral and Brain Sciences* **13**(3), pp. 423–496, 1990.
10. D. Noton and L. Stark, "Eye movements and visual perception," *Scientific American* **224**, pp. 35–43, June 1971.
11. D. Marr, *Vision*, W.H. Freeman, San Francisco, 1981.
12. M. Langer, "Diffuse shading, visibility fields, and the geometry of ambient light," in *Proceedings of the Fourth International Conference on Computer Vision*, pp. 138–147, May 1993.
13. S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, "The lumigraph," in *Proceedings of the ACM SIGGRAPH*, pp. 43–54, August 1996.
14. R. Szeliski, "Video mosaics for virtual environments," *IEEE Computer Graphics and Applications* **13**, pp. 22–30, March-April 1996.
15. S. E. Chen, "QuickTime VR – An image based approach to virtual environment navigation," in *Proceedings of the ACM SIGGRAPH*, pp. 29–38, ACM, (New York), 1995.
16. A. Computer, *QuickTime VR 2.0 Authoring Tools Suite*, Apple Computer, 1997.
17. P. Ciaravola, "An automated robotic system for synthesis of image-based virtual reality," Tech. Rep. CIM-TR-97-12, Centre for Intelligent Machines, McGill University, 1997.



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 6. Selected viewpoints from environmental VR experiment. (a)-(d) were chosen as most interesting, (e)-(h) were evaluated to be substantially less interesting.



(a)

(b)

(c)

(d)

Figure 7. Fully stitched panoramic photographs. (a)-(d) correspond to locations 1-4 illustrated in figure 4 Note that there is a small overlap at the endpoints: this aids in the creation of a cylindrical panorama.