An MFIS for Computing a Raw Cognitive Map

W.K. Yeap, M.E Jefferies and *PS.* Naylor
AI Laboratory,
Computer and Information Science
University of Otago, New Zealand

Ph.+64 03 4798586
COSCWKY@OTAGO.AC.NZ
COSCMEJ@OTAGO.ACNZ

Abstract

This paper extends Yeap's [1988] computational theory *of* cognitive maps, focusing on the problem of computing a raw cognitive map by an autonomous agent. In addition to having a 150° view of the environment as input, the agent also maintains a representation of her immediate surroundings. This representation is referred to as an MFIS, a Memory For one's Immediate Surroundings. Arguments for the use of the MFIS are *presented.* The main questions that we ask in implementing our ideas are: (i) what frame of reference is appropriate for the MFIS? and (ii) how does the MFIS change as the agent moves through the environment? A program has been implemented successfully and the main algorithms used and the results of running the program are presented.

1 Introduction

Psychologists have long suggested that humans and animals compute a representation of the environment (i.e. a cognitive map) which allows them to find their way about in it and return to places of interest ITolman, 1948]. More recently, a computational theory of cognitive maps was developed which further suggested that the cognitive mapping process (CMP) computes a raw map and a full map [Yeap, 1988]. This paper describes further extensions of this theory dealing specifically with the raw map. In particular, we focus on the problem of computing a *raw map of a* large environment, based on *one's* memory of ones immediate surroundings (a structure which is referred to as an MFIS) rather than on ones immediate perception. Results from our simulation studies are presented and briefly discussed.

A raw map is a representation of the physical environment computed from one's sensory perception of the world. This representation forms the basic input to the later cognitive mapping processes, Hence, the input for computing a raw map is information from the senses and the output is a representation containing explicit information about our physical world which is relevant for later tasks. We adhere to a modular structure for the CMP similar to that suggested by Man- [1982] in his study of the visual process since we view the CMP as a natural extension of Marr's theory of vision at the level of the 2D Sketch. Hence, the next module to compute after the Sketch is the raw map, not the 3D Sketch as suggested by Marr and Nishihara [1978]. We shall digress briefly to discuss this point in a little more depth because, as Marr pointed out, one key problem in formulating a computational study is to identify clearly the likely modularisation.

Marr argued convincingly why we need a Sketch and abandoned trying to segment an image using specialised knowledge about the nature of the scenes. His argument was that "since most early visual processes extract information about the visible surface, it is these surfaces, their shape and disposition relative to the viewer, that need to be made explicit at this point in the processing". Following his argument, we suggest that the next module should compute the spatial layout of these surfaces as the viewer moves among them rather than forming 3D descriptions of the individual surfaces. The latter task should come much later in the process since not all information necessary for computing a 3D description is available at the initial encounter. This observation is further supported by some recent psychological experiments which demonstrated that a viewer-centered description and not an object-centered description of objects is remembered by subjects in an object recognition task [Tarr and Pinker, 1989; Rock *et al* 1989]. Like Marr's argument for computing the 2½D Sketch which does not preclude the fact that higher-level knowledge is useful, we are also not suggesting that we do not compute 3D objects. It is, however, unlikely that they should be computed *immediately* from the 2½D Sketch. Doing so would be like perceiving the objects without a spatial context and this would be contrary to the way we normally perceive our environment. That objects appear to us with a spatiality that we should not ignore has been raised sporadically in several areas of research |Veap and Handley, 1990]. Furthermore, computing a 3D Sketch is only one aspect of a more general problem of concept formation and we must therefore begin by asking why objects are formed and how, before suggesting specific algorithms to compute any aspect of it.

In the theory, it is suggested that the raw map is a network of Absolute Space Representations <ASRs), i.e. representations of individual local environments which have been visited (or entered) by the viewer. In this paper, we describe a program which simulates an agent moving with a 150° view through the environment. Information from each view is used to update the MFIS from which ASRs are computed. The questions that need to be asked are:

(i) how to compute an ASR from the MFIS?,
(ii) what frame of reference would be appropriate for the MFIS?, and
(iii) how does the MFIS change as the agent moves through the environment?

Computing ASRs from the MFIS implies that they are no longer computed directly from what could be perceived in a single view (either 360° or smaller) but from one's memory of one's immediate surroundings. This was not considered before in the earlier work and in robotics research concerned with building autonomous mobile robots [Chatila, 1982; Iyengar *et al* 1985; Rueb and Wong, 1987). A new algorithm was designed to compute an ASR from the MFIS and is described in detail in |Yeap *et al,* 1990] Section 2 briefly describes the idea of computing an ASR from the MFIS. The second and third questions are the main concern of this paper. Section 3 discusses these problems and related work. Section 4 describes the algorithm and presents the results of our implementation. Section 5 concludes the paper with some comments on the usefulness of our simulation study.

2 Background on Computing an ASR from Memory

Consider an agent who has just entered a new space and remains stationary at a point. To compute an ASR, one is faced with the problem of identifying the surfaces (unfamiliar or otherwise) in this new environment which, to the agent, appear to enclose her and separate this part of the environment from the rest of it. As well as the surfaces which can be seen from the current viewing position, there are surfaces which are hidden from the agent's view because they are occluded or behind her. If these surfaces were a part of the agent's recent experience of the environment then they, along with the surfaces which can actually be seen, are used to compute an ASR. As Attneave and Farrar [1977] pointed out, "our internal representation of the world around us is based in part on current sensory input, but in much greater part on past sensory inputs, i.e., upon memory".

The key to successfully computing an ASR is to identify which surfaces (in memory) are best combined at each step of the construction process. This is usually straightforward where two surfaces are already connected but in a complex, cluttered environment there are gaps in walls and obstacles of various sizes which occlude walls and other obstacles so that it is not immediately clear which surface should form the next part of the boundary. Our algorithm tackles this problem by characterising six possible connections which may occur between any two adjacent surfaces. Depending on the type of connection, imaginary surfaces are used to fill in gaps between real surfaces and real surfaces may be clipped forming sub-surfaces. A different weight is applied to each surface so that, for example, real surfaces are normally selected in preference to imaginary ones. At each stage of the construction process a preferred subset of all the possible connections is "grown" until the initial surface is reached. The set of surfaces which form the minimum weighted boundary defines the ASR for the current local space. For details of the algorithm see earlier work (Yeap *et al,* 1990].

3 The MFIS

Early AI models on cognitive maps [Kuipers, 1978; Davis, 1984] used a kind of "You Are Here" pointer to identify a person's current position in the environment. Such a pointer obviously under-represents the amount of information available at one's current position but it nonetheless indicates the usefulness of such information for computing a cognitive map. Furthermore, based upon the observations of our own visual system, such as a limited view and limited range vision, and our visual behaviour, such as the need for active scanning [Norton and Stark, 1971 J, psychologists have also suggested that we need to integrate the different views to form a more complete picture of what we perceive of our surroundings (e.g. [Gibson, 1979]). Robotics researchers have shown that it is necessary for their robots to compute a global representation of the whole environment in addition to partitioning the environment into smaller spaces [Yeap *et al,* 19901. Although computing a global representation of the *whole* environment is not necessary, computing a global representation of the immediate surroundings seems to be desirable [Yeap and Handley, 1990]. There are also advantages for doing so. For example, using it, one can detect that one is re-visiting a local space which has just been visited [Yeap, 1988] (and see below) and it provides a richer description of the environment for computing an ASR [Yeap *et al,* 1990].

Information in the immediate surroundings can be represented using either an egocentric (defining spatial positions in relation to the self) or an allocentric (defining spatial positions external to the self) frame of reference. It is clear that from an implementation viewpoint, it is inefficient to use an egocentric reference frame. To use an allocentric reference frame, one has to specify where the reference frame should be centered. The choice of this external point need not be chosen arbitrarily if one uses the current ASR as the frame of reference. An ASR is defined by a boundary (see section 2) and any part of it would be suitable. In the implementation we choose to use the entrance to the current ASR as the center of the reference frame. When the agent moves out of the current ASR, the MFIS is shifted to center on the entrance to the next ASR. Figure 1 shows how the MFIS defined as such, would follow the viewer as she moves through the environment.

The extent of the MFIS can be defined arbitrarily but, more importantly, need not be defined exactly, say, in metrical terms. Varying its size is a tradeoff between how much information is remembered (and hence how useful the MFIS is) and how much effort is required to compute it. Given an MFIS with a fixed size, part of an ASR will often be excluded as it lies outside the area covered by the MFIS (Figure 1). Figure 2 demonstrates that one advantage of having the MFIS is to help one recognise that nearby local spaces were visited before. We therefore do not want to remove a part of an ASR just because it falls outside the area covered by the MFIS. Since its extent is defined arbitrarily, it is better to include the whole ASR if a part of it lies within the area covered by the MFIS. Figure 3 shows how the MFIS defined as such would grow and shrink as the viewer moves through the environment.

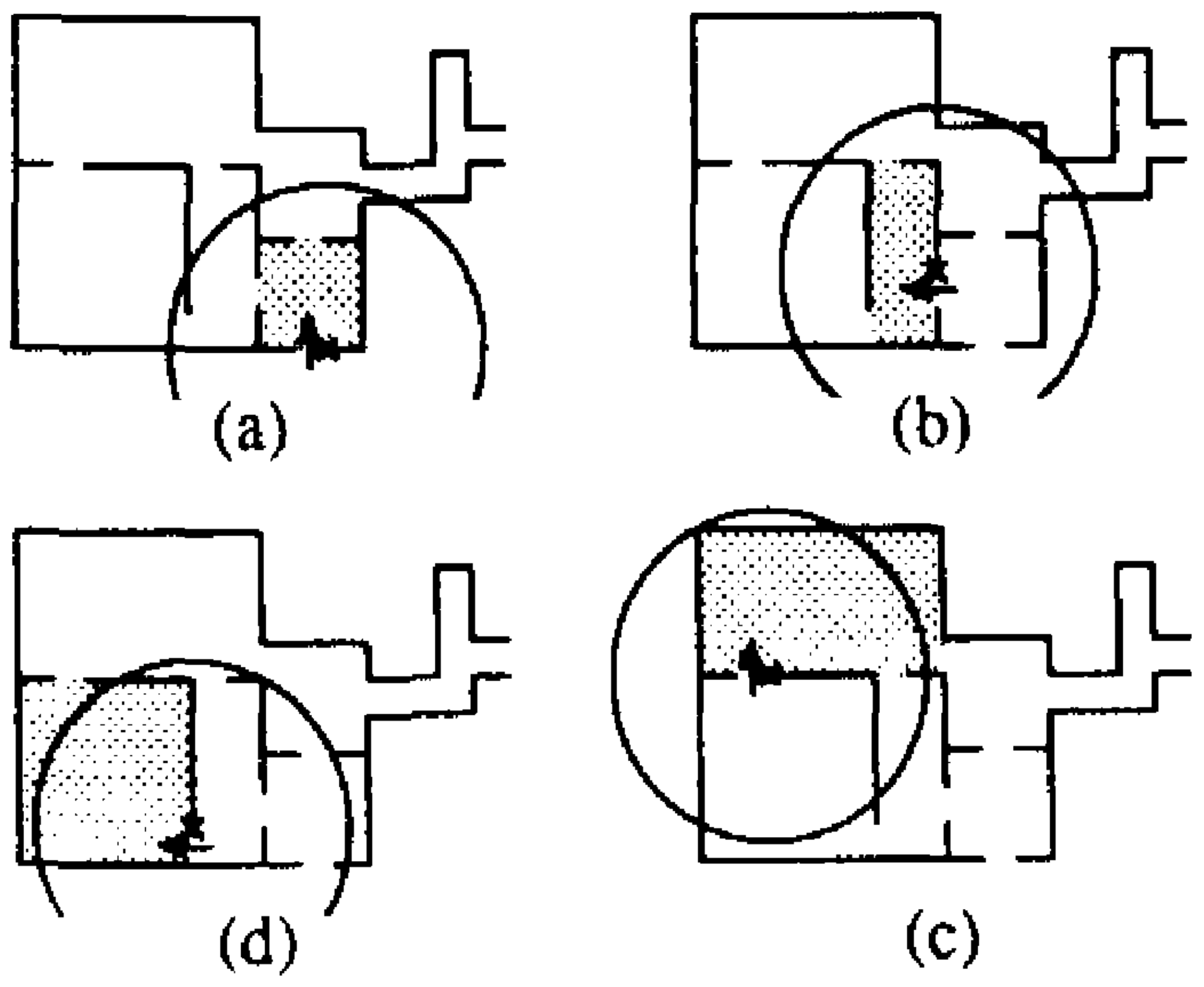

Figure 1 Defining an MFIS using the current ASR at the point marked X as shown from (a) to (d)- When the viewer moves out of the ASR, the MFIS is shifted to centre on the new ASR. The arrow indicates the position of the viewer.

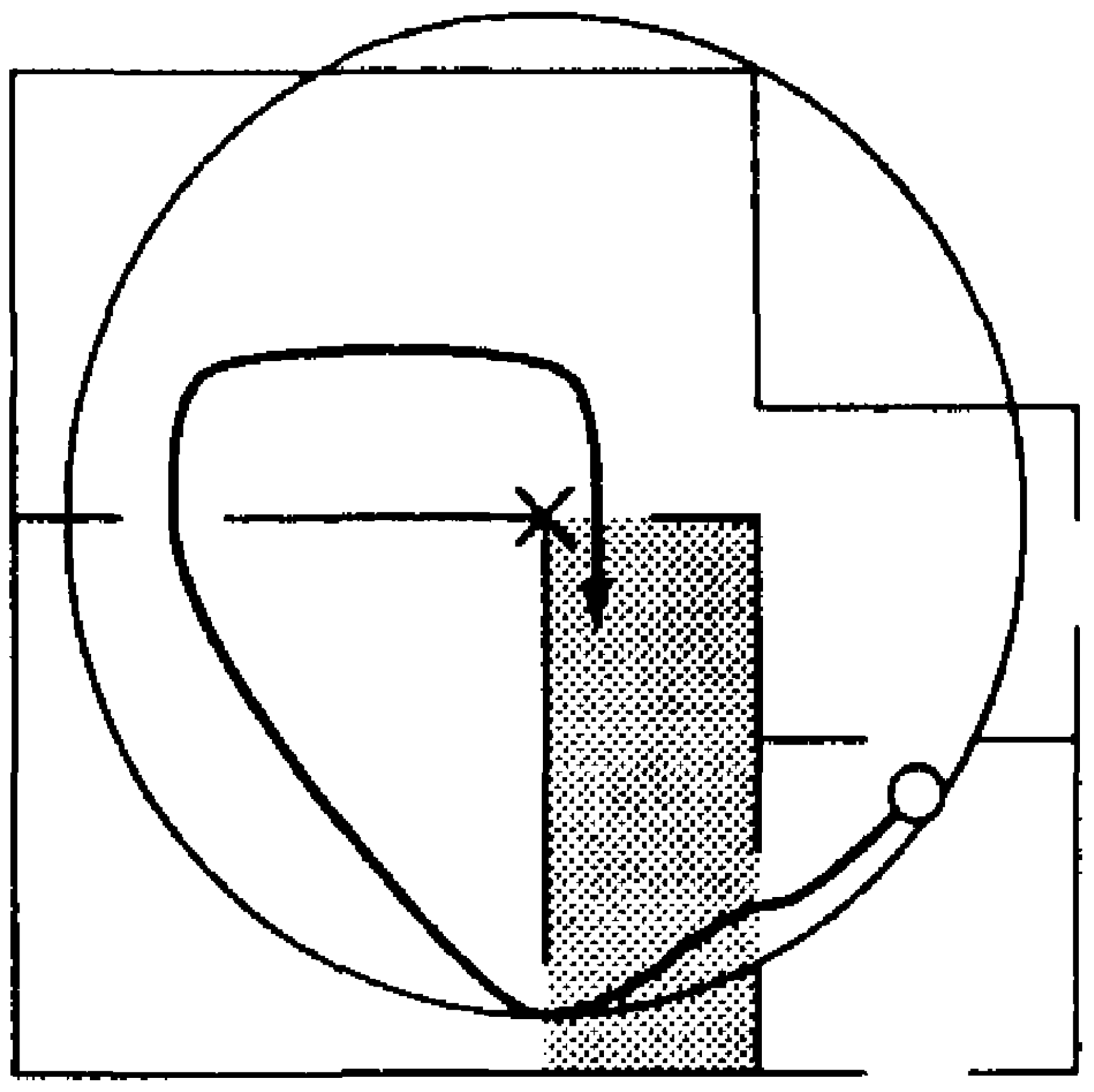

Figure 2. Recognising a local space using the viewer's position in the MFIS. The solid line shows the viewer's path through the environment. The viewer is currently re-visiting the shaded area. One can recognise that one has been to this place by noting the global position of where things are in the MFIS

In addition to those surfaces defined as part of an ASR, two other types of surfaces are also included in the MFIS. The first are those surfaces which are currently in view and the second are those surfaces which were previously in view but do not belong to an ASR. The former are included in the MFIS every time the MFIS is updated with information from the current view. As such, it does not matter where these surfaces lie. The latter must fall within the boundarv of the MFIS and if not, they are removed from the MFIS. Note that in the current implementation, surfaces are removed from the MFIS only when the MFIS itself is shifted.

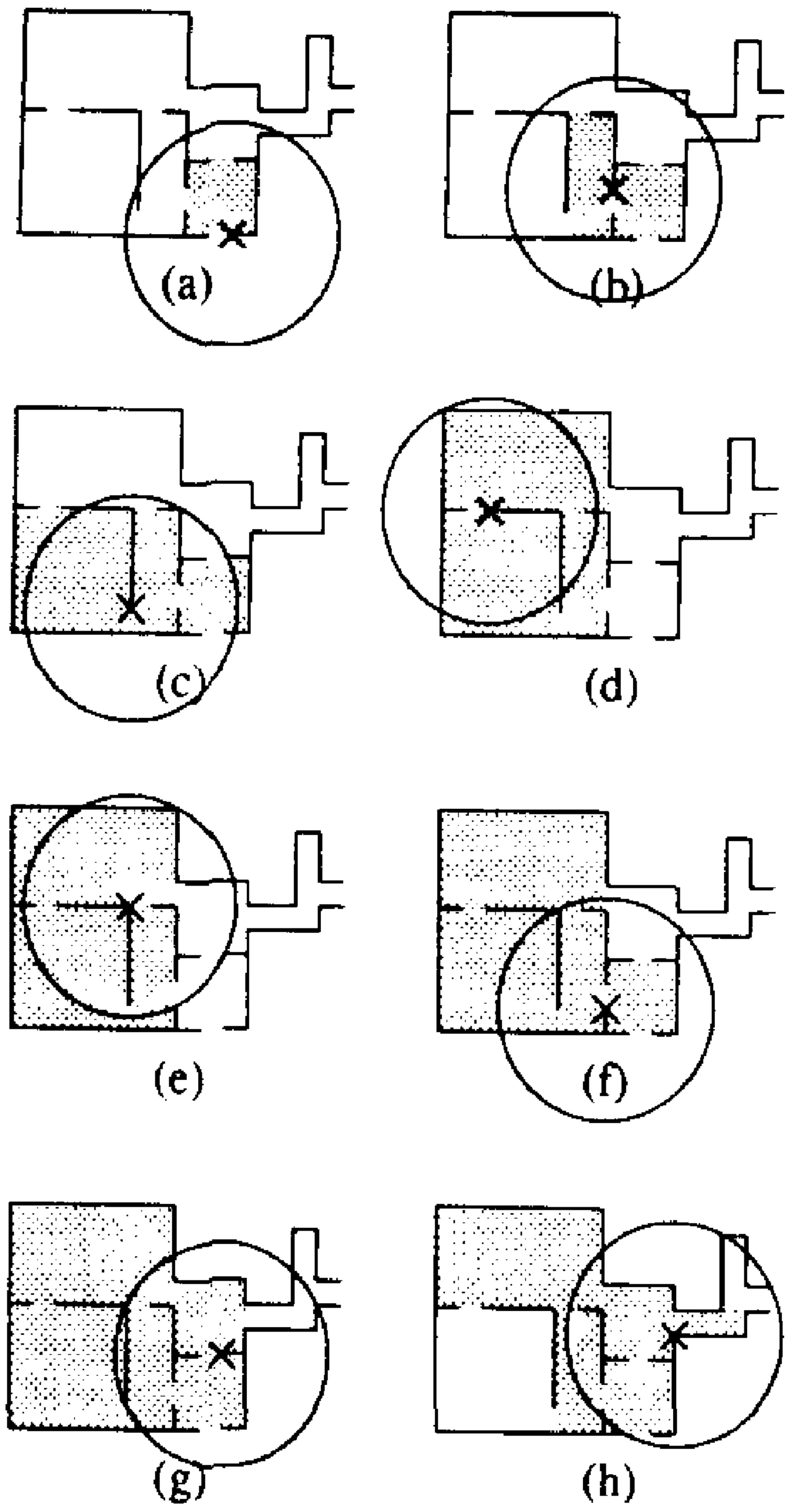

Figure 3. Defining the extent of an MFIS whereby an ASR is either included in it or excluded. As is shown from (a) to (d), the shaded areas indicate the actual size of the MFIS (only expressed in terms of the number of ASRs) as the viewer moves through the environment.

Thus, the MFIS contains the following three kinds of surfaces:

(i) those which are part of an ASR - the whole ASR must be included in the MFIS,

(ii) those which are currently in view - by current, we mean those surfaces which are perceived when in the current ASR. These surfaces are added to the MFIS irrespective of whether they fall within or outside the area covered by the MFIS, and

(iii) those which were perceived when outside the current ASR and are not part of any ASR - these are the least important surfaces and one's ability to remember them is influenced by a wide variety of other factors. They must lie within the area covered by the MFIS.

4 Implementation and Results

Figure 4 shows the steps taken to maintain the MFIS for computing a raw map. As in earlier work (Yeap, 1988], we detect the re-visiting of a local environment using three kinds of information, namely: the exit information, the viewer's position in the MFIS, and the shape of the local environment. However, the difference in this implementation is that the MFIS now "moves" with the viewer as she moves through the environment and ASRs are computed at each appropriate point in the environment rather than being given.

In the implementation, we continue to use a cartesian co-ordinate system, both in describing the surfaces in each view and in the MFIS. In the former, the co-ordinate system is centered on the viewer. Transforming the MFIS is straightforward and the algorithm need not be presented here. The algorithm, RETAIN_SURFACES, which decides which surfaces to remove from the MFIS once it is transformed is given below.

```
RETAIN_SURFACES

;; MFIS_L is a list of all the surfaces which lie
;; within the area covered by MFIS before it was
;; shifted.
;; MFIS_ASRs is the list of ASRs in the MFIS and
;; CV is the list of surfaces in the current view.
;; MRS.ASRs, MFIS,L and CV are all global data
;; structures.

1. Remove all surfaces from MFIS J- belonging to
   any ASR which is no longer in the current MFIS.

2. With the remaining surfaces in MFIS_L Do

    2.1 Retain surfaces in it which belong to the
        current view (CV).

    2.2 Retain surfaces in it which belong to an ASR
        in MFIS_ASRs.

    2.3 Retain surfaces which do not meet the
        criteria of 2.1 and 2.2 only if they lie within
        the boundary of the shifted MFIS.
```

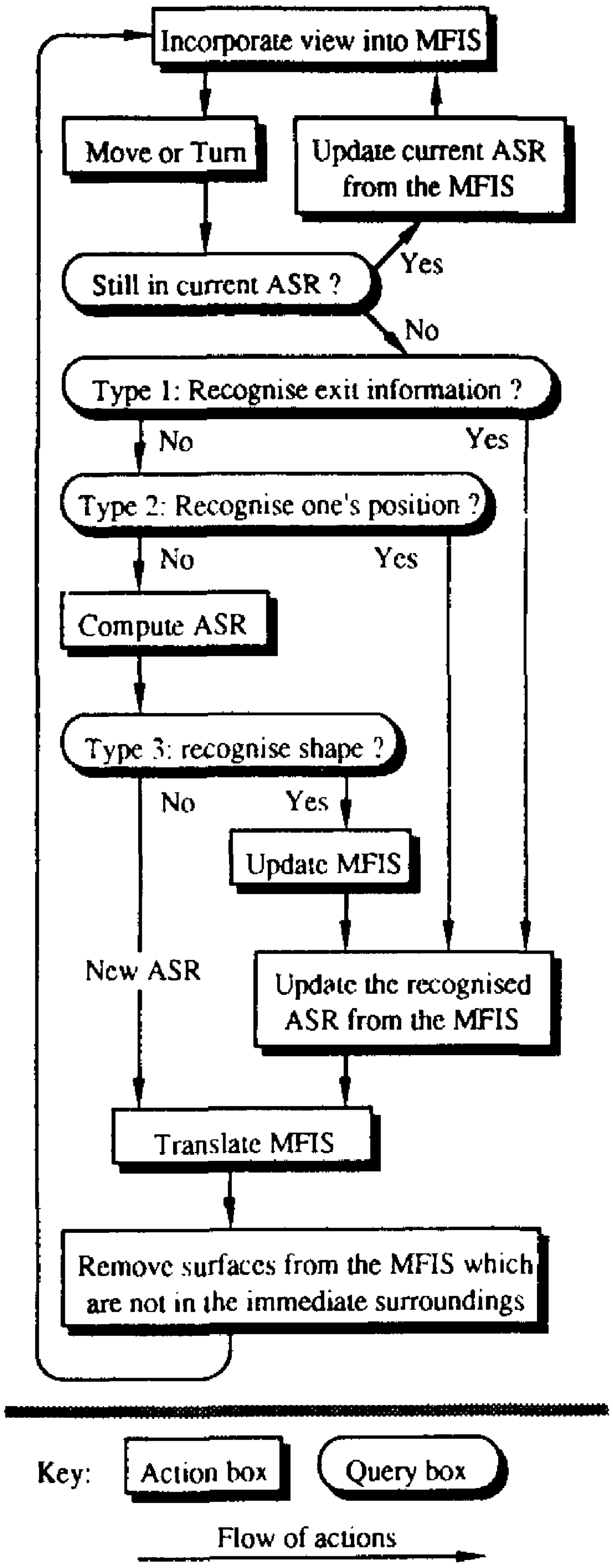

Figure 4. A Flowchart for maintaining the MFIS, from which a Raw Cognitive Map is computed.

Note that in the implementation we use two different boundaries to indicate the *size* of the MFIS. The reason is as follows. Since we have argued that an ASR must either be retained completely or removed from the MFIS, the designated area for their inclusion need not be large. But, we find that this area would be too small for those surfaces which are not part of any ASR. Hence, one solution is to have two extents for the MFIS.

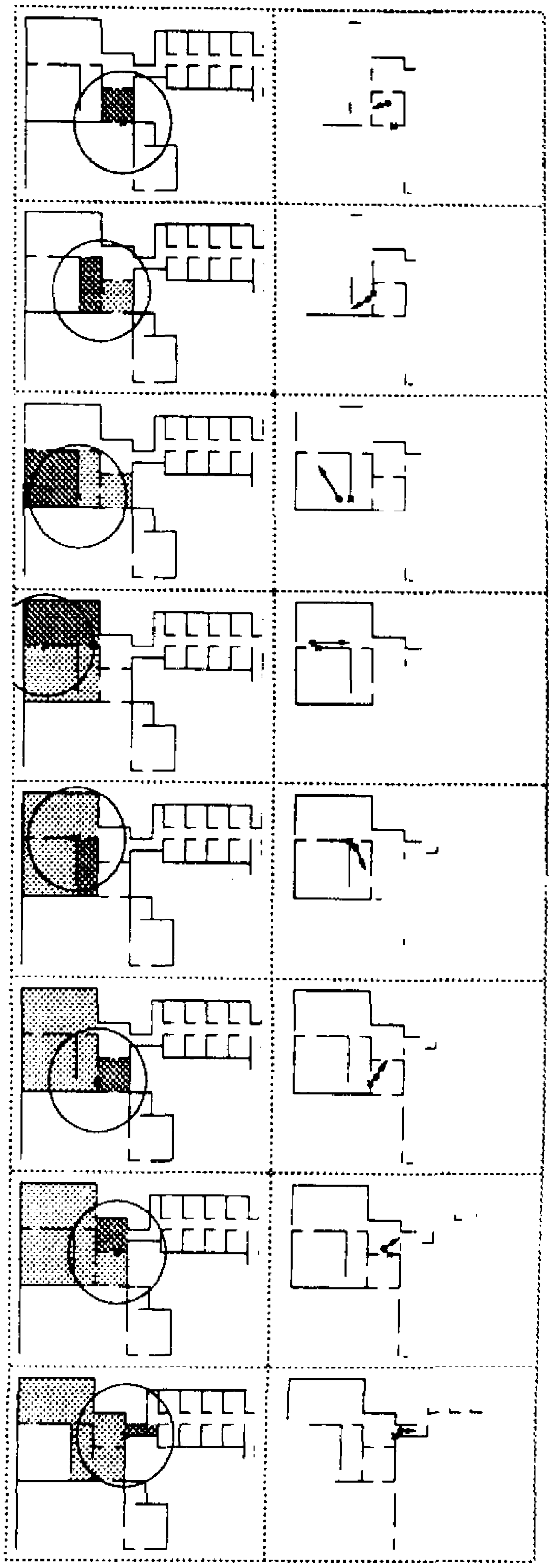

Figure 5. shows how the MFIS *(right)* changes as the viewer moves through the environment *(left)*.

Examples of how the MFIS changes as the viewer moves through this environment are shown in Figure 5, For a complete result, see [Naylor, 1991],

5 Conclusion

The current study shows that an ASR is not only used as an indication of the viewer's *current* local environment but also as a locus for describing the viewer's immediate surroundings. The current implementaiton reveals several problems in computing the individual ASR for building a raw map of the environment. For example, the shape of the ASR obtained does not always correspond to the shape of the room (especially the corridor) and updating for both the partial ASR and the MFIS becomes a complex problem. These problems are currently being investigated using our program.

Acknowledgements

Research in the AI Laboratory is supported by grants from the Otago University and the University Grant Committee. We thank Professor Iyengar of Louisiana State University for sending us reprints of his group's work. We thank our colleague, Chris Handley, and Charles Newhook for reading earlier drafts.

References

[Attneave and Farrar, 1977] F. Attneave and P. Farrar. The Visual World Behind the Head. *American Journal of Psychology,* 90, 4, 549-563.

[Chatila, 19821 R. Chatila. Path Planning and Environment Learning in a Mobile Robot System, *Proceedings of the European Conference on Artificial Intelligence,* France, 211-215.

[Davis, 1984] E. Davis. *Representing and Acquiring Geographic Knowledge.* PhD Thesis and Book. Department of Computer Science, Yale University.

[Gibson, 1979] J.J. Gibson. *The Ecological Approach to Visual Perception.* Boston: Houghton Mifflin Company.

[Iyengar *et al,* 1985] S.S. Iyengar, CC Jorgensen, S.V.N. Rao and C.R. Weisbin. Learned Navigation Paths in Unexplored Terrain. *AI Applications: The Engineering of Knowledge-Based Systems,* Edited by C.R.Weisbin, 148-155.

[Kuipers, 1978] B. Kuipers. Modelling Spatial Knowledge. *Cognitive Science,* 2, 129-153.

[Marr, 1982] D. Marr, *Vision.* New York: W.H. Freeman and Company.

[Marr and Nishihara, 1978] D. Marr and H.K. Nishihara. Representation and Recognition of the Spatial Organisation of Three-Dimensional shapes. *Proceedings of the Royal Society of London,* B 200, 269-294.

[Naylor, 1991] P.S. Naylor. *An Implementation of the Raw Cognitive Map.* MSc. Thesis (to be submitted), Department of Computer Science, University of Otago.

[Norton and Stark, 1971] D. Norton and L. Stark. Eye Movements and Visual Perception. *Scientific American,* 224, 34-43.

[Rock *et al* 1989] I. Rock, D. Wheeler and L. Tudor. Can We Imagine How Objects Look from Other Viewpoints? *Cognitive Psychology,* 21, 185-210.

[Rueb and Wong, 1987] K.D. Rueb and A.K.C Wong. Structuring Free Space as a Hypergraph for Roving Robot Path Planning and Navigation. *IEEE Transactions on PAM1,* 9, 2, 263-273.

[Tarr and Pinker, 1989] M]. Tarr and S. Pinker. Mental Rotation and Orientation-Dependence in Shape Recognition. *Cognitive Psychology,* 21, 233-282.

[Tolman, 19481 E. C Tolman, Cognitive Maps in Rats and Men. *Psychological Review,* 55, 4, 189-208.

[Yeap, 1988] W-K Yeap. Towards a Computational Theory of Cognitive Maps. *Artificial Intelligence,* 34, 297-360.

[Yeap and Handley, 1990] W.K. Yeap and CC. Handley. Four Important Issues in Cognitive Mapping. *Proceedings of the International Conference on Simulation and Adaptive Behaviour: From Animals to Animats.* 176-183. Cambridge: MIT Press.

[Yeap *et al,* 1990] W.K. Yeap, P.S. Naylor and M.E. Jefferies- Computing a Representation of the Physical Environment - A Memory-Based Approach. *Proceedings of the Pacific Rim International Conference on Artificial Intelligence,* Nagoya, 847-852.