Max–Planck–Institut
für biologische Kybernetik
Spemannstraße 38 • 72076 Tübingen • Germany

Technical Report No. 62

Differences between Active-Explorers and Passive-Observers in Virtual Scene Recognition

Chris Christou[1] & Heinrich H. Bülthoff[2]

March 1998

Chris Christou is supported by a scholarship from the Max-Planck Gesellchaft.

---

[1] Max-Planck-Institute für biologisch Kybernetik, E–mail: chris.christou@tuebingen.mpg.de; to whom all correspondence should be addressed
[2] Max-Planck-Institute für biologisch Kybernetik, E–mail: heinrich.buelthoff@tuebingen.mpg.de

---

Differences between Active-Explorers and Passive-Observers in Virtual Scene Recognition

Christou C.G. & Bülthoff H.H.

Abstract. Recognition of a newly learned environment from both familiar and novel perspectives was investigated using a 3D-computer model in which observers made simulated translational and rotational head movements. They were encouraged to move around the environment to find and acknowledge spatially localized coded markers. During each acknowledgement the observers' viewing parameters were stored and later used for stimulus generation. The observers' simulated movements were restricted to a small region of the environment and complete head rotations were not allowed. To test the importance of making volitional movements during familiarization, two groups of observers were tested: *active-explorers* initiated their own movement through the environment while the *passive-observers* watched a playback of these movements. In the recognition tests, all participants were shown both familiar and novel views of both acknowledged and unacknowledged locations. Testing took place immediately and was repeated after 7 days. Results indicate that observers always found novel perspective views more difficult to recognize than familiar views, and surprise locations more difficult than acknowledged locations. The principle difference between active and passive was an advantage in recognizing novel direction views by active-explorers. This difference became more pronounced over the course of seven days. The results provide evidence for egocentric encoding and suggest that this can be reduced, if only marginally, by facilitating observer self-locomotion during learning.

1 Introduction

The nature of human spatial encoding, or representation, of spatial layout is under intense scientific investigation because mental representations play an essential role in many cognitive functions such as planning actions, navigating through the world and indeed all abilities requiring decisions based on information beyond what is immediately available through sensory stimulation. A behavioural demonstration of the ecological benefits of having spatial representation was provided by the experiments of Menzel (1978) who showed that chimpanzees have a superior advantage in localizing items of food whose spatial location was seen previously than control animals who had to perform a random search. The ability of the former to perform such localization implies that they are able to represent the spatial layout of the environment and ascribe degrees of importance to particular regions of it. However, the utility of such mental representation depends entirely on its generality or applicability to a variety of situations. That is, representation should not be specific only to the (often transient) conditions under which it was formed. For instance, the animals in the Menzel experiment did not start collecting food items in the order in which they were shown to them but appeared to make optimized decisions based on their instantaneous location. This implies they were able to reason based on their location and their representation of their environment which in turn implies that their representation was quite flexible.

The mechanisms and processes which facilitate such generalization during learning have not been established. Alternate models have been proposed regarding representational forms that could do the job. Generalization could for instance be the result of continual integration of various sources of information which lead to an abstract or schematic representation, no longer tied to specific sensory experience (see Neisser, 1967). An abstract representation of spatial layout within an environment consisting of various objects is provided, for instance, by a set of propositions specifying the identity and relationship between these objects (Winston, 1975). One useful feature of such a representational model would be the ability to recognize an environment from entirely novel directions because the spatial relations described by the representation do not depend on the position of the viewer. Such a view-independent theory of the encoding and recognition of objects has for instance been proposed by Biederman (1987) and Marr and Nishihara (1978).

In contrast, view generalization can be achieved using transformation procedures which operate on 2D image-based (sensory specific) features. Two distinct and ecologically plausible systems for overcoming differences in views of familiar objects have been described by Poggio & Edelman (1990) and Ullman & Basri (1991). These models predict that recognition performance should drop with increasing viewing distance from familiar (stored) views and this has been corroborated by experimental studies both on humans (Bülthoff & Edelman, 1992) and monkeys (Logothetis, Pauls & Bülthoff, 1994). There is also increasing evidence for such view-specificity in the recognition of scenes or large-scale environments (Shelton & McNamara, 1997; Rieser, 1989).

The bulk of previous studies in scene recognition has found evidence for decreased specificity and increased abstraction in the representation of scenes, especially over long periods of time (Rowland et.al., 1978; Hock & Schmelzkopf, 1980). However, a major difficulty encountered by these studies has been that it is extremely difficult either to keep variables constant or eliminate unwanted cues. Examples of problematic variables in real environments include changes in illumination (sunny, cloudy days), cast shadow visibility, visibility of surface textures and visual landmarks. In studying ones' ability to recognize a scene based on the spatial relations between objects, precise control of such variables is very important. To counteract such 'cues' to recognition some researchers have used line-drawn scenes and other artificial stimuli. Such methods lack realism not only in the content of the scene depicted but more importantly in the means of familiarization. Under natural circumstances familiarization with an environment usually takes place during observer locomotion and interactivity. The importance of interactivity in spatial learning is paramount, for instance, in the developmental theory of Piaget & Inhelder (1967) which considers that adequate spatial representation is only formed through a process of interaction with the environment. Thus, during spatial learning people perform motor actions or movements which either guide locomotion and orientation within the environment or which provide additional sensory feedback when more spatial information is desired. The way people form spatial concepts is through such action-perception-action cycles.

A means of facilitating natural locomotion while controlling sensory cues is to use the tools of Virtual Reality (VR) and Virtual Environments (VE). A VE can be defined as any portrayal of three-dimensional space that has no physical basis. This definition is also applicable to paintings as well as static computer images although the essential difference is that real-time rendering allows one to simulate movement through a VE. Current computer graphics models allow highly realistic environments to be simulated. High speed interfaces and dedicated graphics hardware allows for the simulation of movement and interaction in these environments. At all times the level of control (of what the environment contains and what the observer sees) is dictated by the experimenter. Perhaps one objection to the use of VEs in studying mental encoding is that the representation formed within a virtual environment is somehow different from that derived from the real world and that recognition of a VE does not necessarily reflect recognition abilities under natural circumstances. This point is valid only if the usual means of familiarisation are missing in the simulation. As stated above, the provision of interactive control of movement may be very important in providing depth cues and also in facilitating natural learning strategies (e.g., see Siegel & White, 1975). Furthermore, the availability of rich sensory information even when the observer is static may also be very important in conveying an impression of depth. These cues to depth may in turn reflect on the elaborateness or completeness of ones' spatial representation.

In the present experiment we wished to investigate the ability to recognize novel views of familiar locations within a newly learned VE when familiarization was carefully controlled but where interactive movement was faciliated. Because the normal process by which we become familiar with an environment is usually unrestricted (except, for instance, those viewing restrictions imposed by our height and accessibility of the environment) there seems to be little opportunity to test the difficulty of such generalization to novel views. However, one common situation in which generalization is tested is when we need to walk through a new environment and then find our way back to our original starting point. Unless specific navigational landmarks are remembered people often have difficulty in these situations. This is probably because they are unfamiliar with the newly learned environment as perceived from a new direction. If we reduce the potential of distinctive features or objects becoming used as visual landmarks we may ask whether we are able to generalize visual recognition to novel perspectives using only information regarding the geometrical layout. Generalization in this case would have important implications both for the kinds of information stored in long-term memory and for the processes necessary for recognition.

We compared two modes of familiarization to guage the influence of interactivity on the speed and accuracy of scene recognition. In the training stage of this experiment participants were familiarized with a VE while searching for, and acknowledging spatially localized coded markers. We employed two groups of observers: active-explorers and passive-observers. The training of these two groups differed only in the ability of the 'active-explorers' to make self-initiated movements. To guage the completeness of encoding in our experiment, we tested recognition ability both for spatial locations which had to be acknowledged by observers and also other unacknowledged or 'surprise' locations lying in between these acknowledged locations. Previous studies on childrens' spatial awareness have found that structural detail is grouped or clustered around specific locations but the relations between these clusters were themselves inconsistent with the environment (Schadler & Siegel, 1973). We were therefore keen to look for differences between acknowledged and surprise views and compare these differences across active and passive observers. Finally, we wished to determine the influence of delay in testing on recognition ability. Many of the experiments that have provided evidence for schema-based representations (e.g., Rowland et. al., 1978) would predict a decrease in performance for exemplar images after a delay in testing together with an increased ability in recognizing unfamiliar views. This is compatible with the transition to a more schematic representation. We investigated this by testing observers in their recognition of a VE both immediately and then after a seven day interval.

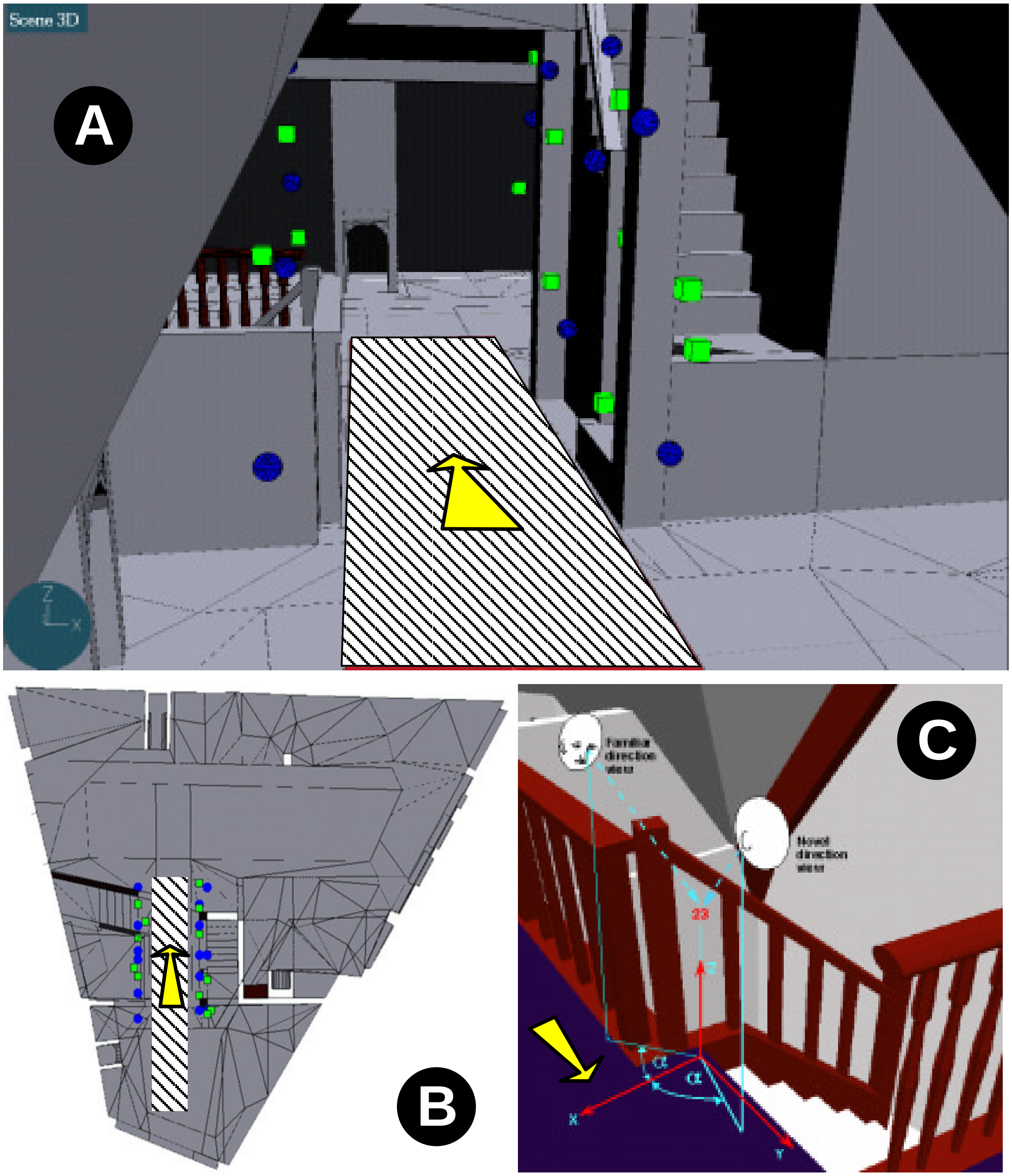

Figure 1: Schematic depiction of the VE used in the experiment. In all cases the *familiar direction* is indicated by an arrow. (A) Perspective view showing the position of the two sets of markers (represented either by spheres or cubes). Observers were allowed to move across this walk-way and to turn their heads by up to ±60° to either side of the indicated direction. (B) Plan view showing the walk-way (rectangular region with diagonal lines) and the locations of the markers to its left and right. (C) Figure indicating the method used to calculate novel views. The observers vantage point during acknowledgement of each marker was rotated by an angle of 2α in a clockwise or anti-clockwise direction about the Z axis. The value of α was calculated as the angle between the observers vantage point projected onto the floor and the X axis direction as indicated.

2 Methods

2.1 Subjects

Subjects were 32 paid individuals between 17 and 34 years of age who had not participated previously in such experiments and had not seen the test VE. Previous results in a similar task (Christou & Bülthoff, 1997) revealed no effect of gender on performance in this task but gender pairing was carried out where possible. All subjects received written instructions for each stage of the experiment. Subjects who made themselves available for retest did so knowing that this involved a recognition test of the same environment.

2.2 Materials

The test VE consisted of a 3D computer generated model of an attic, shown in figure 1, created from approximately six thousand polygons using the Medit Modelling program running on a Silicon Graphics High Impact computer. The layout of the VE was derived from the architects drawings of a real house located in Tübingen, Germany. This house has a very irregular structure which we thought would provide an interesting setting, especially as we wished to eliminate most visual cues such as textures and distinctive objects.

The IRIS Performer applications programming interface was used to render the VE in real-time. This rendering reflected the active-explorers movements which were input via a Spacemouse (Spacetec IMC Co., Massachusetts, USA). The Spacemouse is a 6 degrees-of-freedom motion input device used primarily in Computer Aided Design because it allows intuitive rotation and translation to be input. Movements of the simulated head were initiated by turning the Spaceball in a clockwise or anti-clockwise direction. Changes in pitch were initiated by turning the ball about a horizontal axis and translational movements were initiated by pushing the ball in any direction. We ensured active-explorers' competence with the Spaceball prior to actual scene familiarization by allowing them to perform the exploration task in an unrelated VE. Both training and testing of observers was facilitated by a Silicon Graphics Octane computer with dual R10000 processors. A high resolution (1280x1024) monitor portrayed the rendered images.

2.3 Procedure

2.3.1 Training

The experiment consisted of two stages: an initial familiarization stage followed by an immediate recognition test, which was repeated after 7 days. Observers were told that training would be followed by a recognition test but were not informed that the test involved recognizing novel perspective views.

The active-observers searched for coded markers whose locations were always easy to find and arranged on either side of the 'walk-way' along which subjects were allowed to move (see figure 1). These movements where recorded at 60Hz in the form of simulated view coordinates (heading and pitch) and relative position in the environment. For passive-observers, these coordinates were later used by an IRIS Performer program which rendered the appropriate views at the same frame-rate as the data was generated. The passive-observers therefore experienced the actual movements of the active-explorers and the play-back was stopped when the markers became visible. At this stage passive-observers acknowledged the code and play-back was continued. This meant that each active/passive pair saw the same views of the attic and the essential difference was the ability to make self-initiated movement.

2.3.2 Testing

Both groups were tested immediately after training using an old/new recognition task where images from both familiar/novel directions and acknowledged/unacknowledged locations were presented to them in a single block. They were informed that pictures could be taken from any perspective and that some of the images could be from a similar but structurally different 'distractor' environment. Their task was to classify each trial as old or new by pressing keys on a keyboard. No performance feedback was given. Although there was no time limit imposed on responses, observers were instructed to respond as quickly as possible and responses made after 10 seconds were discarded from the

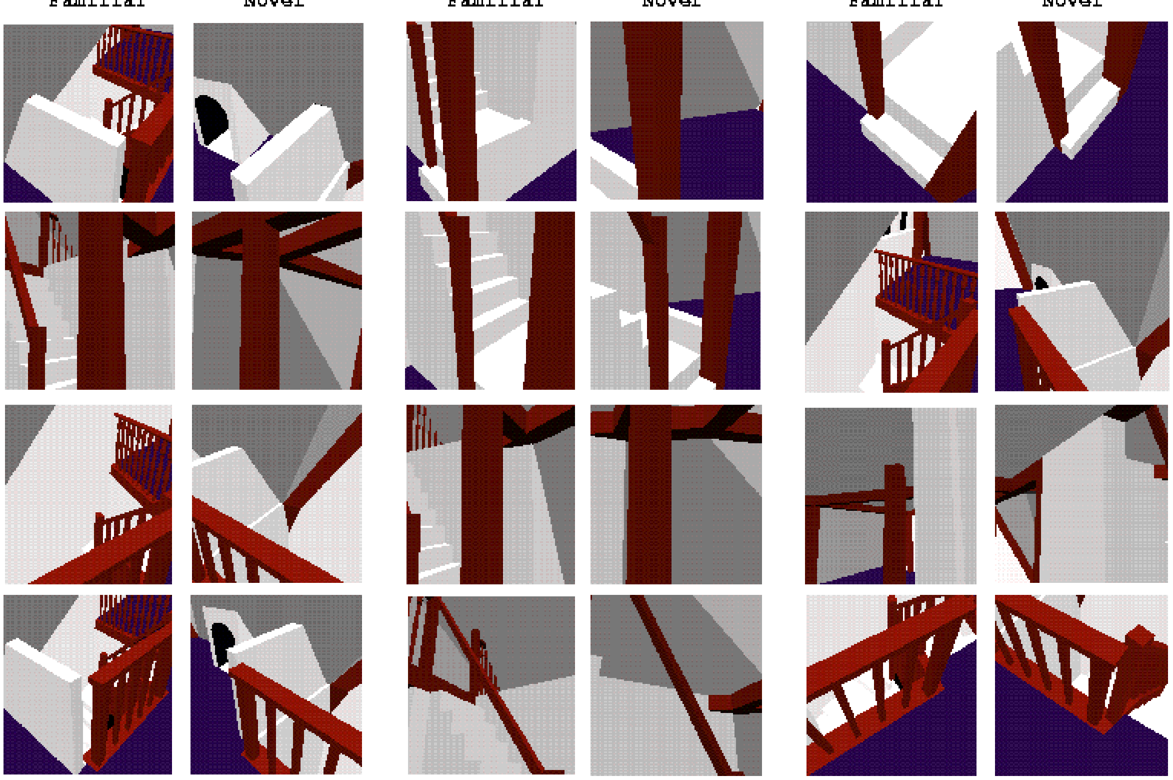

Figure 2: Some example (familiar and novel direction) views of the VE that served as stimuli in the experiment. Images corresponding to both location sets (A and B, see text) are depicted.

analysis.

The stimuli consisted of 600x600 pixel images depicting familiar and novel direction views of the learned environment. They were generated individually for each observer according to their spatial location and viewing coordinates (in terms of heading and pitch) at the time they acknowledged each marker. Familiar views were generated simply by rendering the view seen at the time of acknowledgement. The novel direction views were generated by rotating the familiar direction about a vertical axis (see figure 1C). The magnitude of this rotation was twice the magnitude of the angle between the instantaneous heading and the X axis direction (perpendicular to the direction of the walk-way). Given the restrictions imposed on heading this meant that the angle of rotation for novel views ranged from 60 to 120 degrees. The stimuli corresponding to unacknowledged locations were calculated from the experimenters settings while they performed the exploration procedure under the same conditions as subjects. These stimuli therefore remained constant for all subjects.

In order to test the recognition of surprise locations two sets of marker locations (A and B) were used (see figure 1A). Because of the possibility that some locations were easier to recognize than others, observers in each of the passive and active groups were further divided into two groups; one group would acknowledge marker set A while the second group acknowledge marker set B. Equal numbers of observers acknowledged both sets and results were averaged. All observers were tested on all views (familiar and novel) corresponding to all marker locations (whether acknowledged or otherwise). No feedback was given during test. In total, the stimulus set consisted of familiar and novel direction views of the 15 locations indicated by markers and a further 15 locations not indicated by markers (see figure 2). This made a total of 30 target images.

The observers task consisted of discriminating between images of the familiar target environment and a similar distractor environment. The latter was in fact derived from the target with the following adjustments: First, all polygons constituting the attic where mirror-

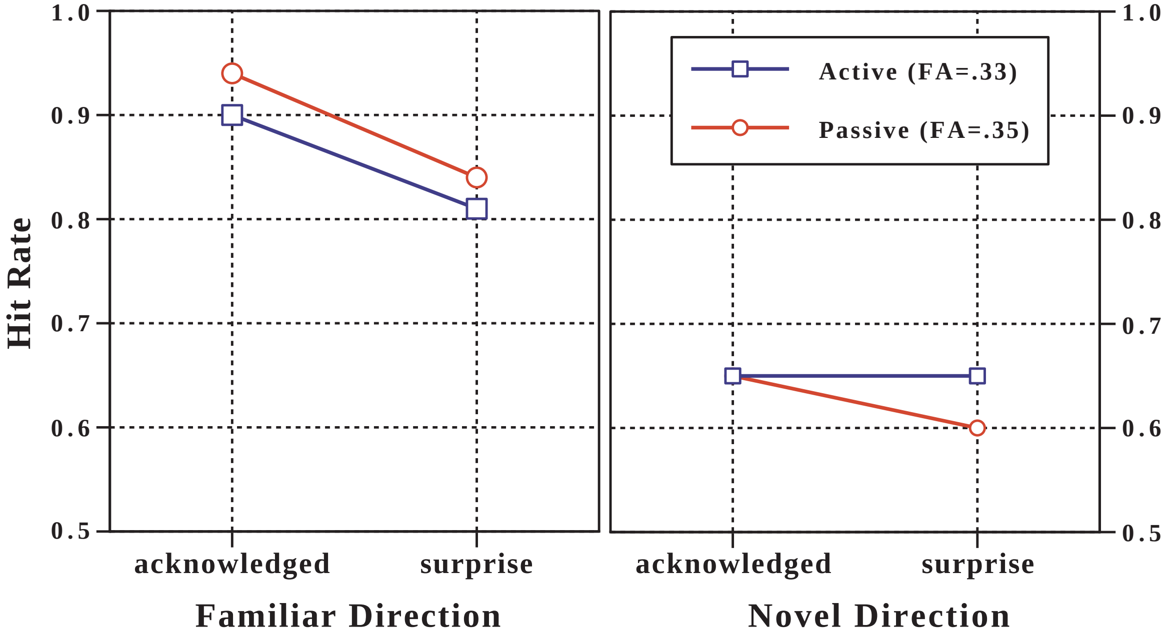

Figure 3: Mean proportion of correct responses according to Mode of Training (Active or Passive), the type of spatial location corresponding to test images (either Acknowledged or Surprise) and the viewing direction of each image (either Familiar or Novel).

reflected about the YZ plane, perpendicular to the 'walk-way'. Second, structural adjustments were made to stairs and components such as chimnies and wooden beams were displaced. This produced a new environment having similar components yet with distinctive differences from the target. Images were presented in random order and were preceeded by the presentation of a fixation cross. Observers were told to expect images taken from all perspectives within these environments.

3 Results

3.1 Immediate Testing

The observers responses were stored as *hit rates* (that is, proportion of correct responses) and the *response time* (RT) for each correct response was also noted. These two measures provided a gauge of the level of difficult inherent in the recognition of each group of stimuli. Because all conditions were presented within the same block the false alarm rates for active and passive observers were constant across conditions. Mean false-alarm rates for active-explorers were 0.33 and for passive-observers they were 0.35. An ANOVA with (2 X 2 X 2) mixed factorial design was used to analyze the performance data. The ANOVA design consisted of a between-subjects factors of Mode of Learning with two levels (Active and Passive), a within subjects factor of Location with two levels (Acknowledged and Surprise), and a within subjects factor of View Direction also with two levels (Familiar and Novel). The ANOVA revealed that the mode of learning made no significant overall difference to performance in terms of hit rate [$F_{1,30} = .049, p < 0.6$] or RT [$F_{1,30} = 0.74, p < 0.39$]. Mean hit rates for all 32 subjects are shown in figure 3. There was a significant main effect of View on hit rate [$F_{1,30} = 201.15, p < .0001$] and RT [$F_{1,30} = 36.06, p < .0001$] with novel direction views requiring more time to process and producing more errors, in agreement with previous experiments (Christou & Bülthoff, 1997). There was also a significant main effect of Location on hit rate [$F_{1,30} = 6.74, p < .02$] but with no corresponding influence on RT [$F_{1,30} = .22, p < 0.25$]. A planned comparison revealed that the hit rates observed for familiar direction views of acknowledged locations (mean = .92) were significantly higher than those for familiar direction views of surprise locations (mean = .83; [$F_{1,30} = 13.11, p < 0.005$]. There was no such difference for novel direction views [$F_{1,30} = 0.72, p = 0.4$]. The interaction between Location and View was significant [$F_{1,30} = 4.81, p < 0.05$].

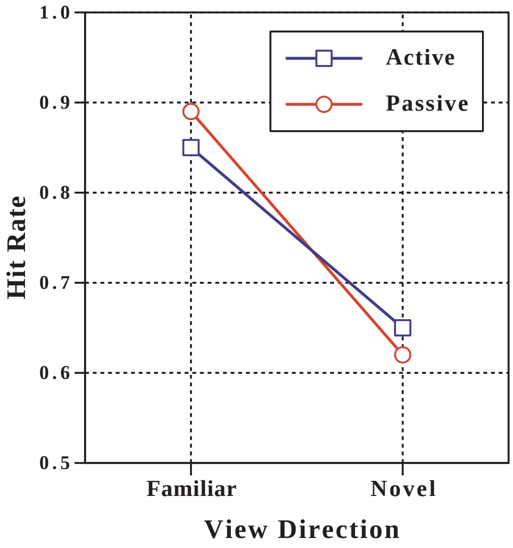

Figure 4: Plot showing the interaction between Mode of Training (active, passive) and View (familiar, novel).

The ANOVA also revealed a nearly significant interaction between Training and View [$F_{1,30} = 3.2, p < 0.08$] (see figure 4). The passive-observers appeared to do better than active-explorers for familiar views (mean hit rates collapsed across Location were 0.89 and 0.85 respectively). However, this relationship is reversed for novel direction views (mean hit rates were 0.62 and 0.65). Although still inconclusive such a trend would suggest that the performance of the active observers is more stable across familiar and novel views than that of the passive observers. We were therefore keen to determine whether this relationship was extended over the seven-day delay period in testing.

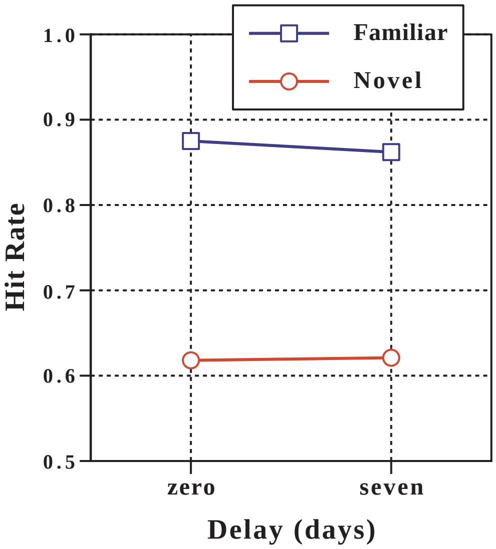

Figure 5: Plot showing the influence of the seven day period between testing on hit rate (proportion correct). There is a slight increase for novel views over this period although this difference was not significant.

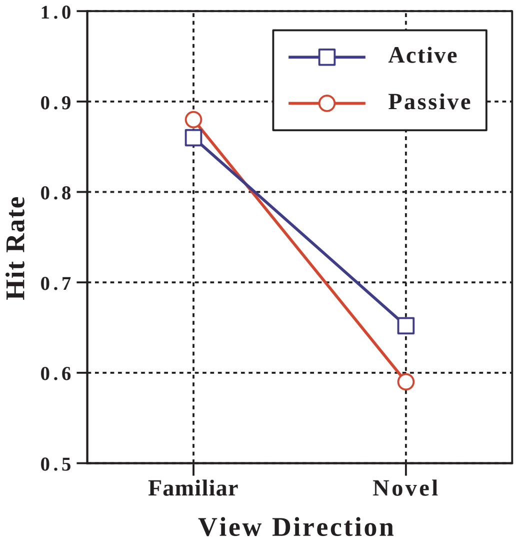

Figure 6: Plot showing the interaction between Mode of Training (active,passive) and View (familiar, novel) after a delay of seven days

4 Results:-after 7 Days

In total, 17 of the original 32 observers were able to come back exactly seven days after training for a retest. Of these, nine were participants of the active group and eight were participants of the passive group during training. An ANOVA with (2 X 2 X 2 X 2) mixed factorial design with one between-subjects factor (Training) and three within-subjects factors (Location,View and Delay) was used to analyze the hit rates and RTs. The between subjects factor of Training again consisted of two levels (active, passive). The within-subjects factor of Location consisted of two levels (acknowledged, surprise), as did that of View (familiar, novel) and Delay period (0 days, 7 days). We were initially interested in seeing whether any differences occurred in the familiar and novel direction view performance over the seven day period. This relationship is depicted in figure 5 which shows a very small improvement in novel view performance and a small decrease in performance for familiar views. Such a change is predicted after a transition from exemplar-based (or image-based) representation to schema-based representation (e.g., Rowland et. al., 1978) However, our analysis revealed there was no significant overall effect of Delay Period on hit rates or RT and again no overall effect of Mode of Training.

The significant effect of View persisted over the seven day interval both in terms of hit rate [$F_{1,15} = 265.0, p < 0.001$] and RT [$F_{1,15} = 11.6, p < 0.005$]. There was also a significant effect of Location on hit rate [$F_{1,15} = 11.1, p < 0.005$] but not on RT [$F_{1,15} = .05, p < 0.82$], as after immediate testing. Overall, acknowledged locations resulted in significantly higher hit rates (mean = 0.77) than surprise locations (mean = 0.72). Interestingly, the ANOVA also revealed that the marginal interaction between Training and View shown after immediate testing had now reached significance [$F_{1,15} = 7.6, p < 0.02$]. This relationship is depicted in figure 6 which shows mean data for all subjects averaged across Location and Delay. A *post hoc* comparison (Newman-Keuls) showed that the differences in hit rates between active and passive observers for novel direction views were significantly different ($p < 0.01$) whereas those for familiar views were not. It therefore appears that the novel direction views, which required considerably more processing to recognize than familiar direction views, were more easily recognized by the active explorers than by passive observers. There was no corresponding interaction with respect to RT.

5 Discussion

We have used a desktop VE to study the acquisition of spatial information and to assess what kind of limitations occur in the recognition of 3D spatial environments after simulated locomotion. Our previous experiments have shown that performance in this task is strongly dependent on the familiar direction views which are experienced during training. The recognition of novel direction views is always more difficult, as reflected by reduced accuracy and increased response times. The current experiments were intended to test the influence of three additional factors on this ability to generalize to novel direction views. First, we wanted to assess the advantages offered to active explorers who are allowed to initiate their own movements as compared to passive observers. This distinction is analogous to comparing the accumulation of spatial information by drivers of an automobile to that of the automobiles' passengers. Second, we wanted to test the degree of generality of the representation formed of the VE by including views of unexpected or surprise locations. Thirdly, we wanted to determine if any changes occur in spatial representation over an intervening period after initial testing.

Clearly, the mode of training could be an important factor in how well observers cope with surprise views and how much of the spatial information they acquire is retained over a seven day period. One may envisage for instance that the active explorers are more attentive during their search through the environment (e.g., see Appleyard, 1970) and thereby are able to recognize surprise views more accurately than the passive observers. Also, the strength of a 'memory trace' of an environment may be determined by the degree of importance the environment has for the observer. It is reasonable to assume that the attic would be more meaningful to the active-explorers who were allowed to navigate through it to find markers (which was in itself, not a trivial task).

Our results showed no overall difference between active-explorers and passive-observers with respect to acknowledged and surprise locations indicating that both groups were equally attentive of the attics' spatial structure. There was however, an enhanced ability in the recognition of novel direction views by active-explorers, which was magnified after the delay in testing. There was no such difference for familiar directions which again indicates that this difference is not due to lack of attention on behalf of the passive-observers. In fact, immediately after training the passive-observers displayed a superior ability in recognizing familiar direction views. Hence it seems that the ability to initiate ones' own movements through the scene result in a more elaborate mental representation which can facilitate enhanced recognition accuracy. This effect is made more significant over a delay period. Why should this be the case? One possibility is that the coupling of self-initiated movement and optical feedback as a result of this movement enhances the pick-up of geometrical information which results in the elaboration of ones' mental model. The importance of such coupling is captured by theories of 'active-vision' (e.g., Aloimonos, 1993) which posit that a major advantage of an active or mobile visual system is the ability to alter ones view of the scene to achieve some particular goal. In our case, although passive-observers watched the movements of the active-explorers they lacked the knowledge of why these actions were performed. This dissociation between perception and action may have reflected on their ability to integrate structural information and eventually affected their recognition performance. Clearly, closer theoretical and empirical analysis is required to determine how vision and action are coupled during spatial learning.

This difference only affected novel views and became more apparent after the delay period. Perhaps the differences between active and passive observers would have been greater if the familiarisation process involved a more cognitive task relating to the VE rather than the random search for markers. Further differences may be revealed when locomotion is initiated, for instance, by actual walking rather than by using the Spacemouse. The Spacemouse is an indirect means of control and subjects needed prior training in its use. As such, its use may not fully utilise the information derived from the natural coupling between limbs and visual perception. In general, the current experiment involved a particular desk-top VR implementation and many more differences may be revealed between active and passive observers in a fully immersive VE.

Additionally in this experiment, we introduced both familiar and novel direction views of *surprise* locations. We found a similiar benefit of familiar direction even for surprise locations. Though we observed a significant difference in hit rates for acknowledged and surprise location familiar direction views a similiar difference was not observed for novel directions. These results suggest two things. First there does appear to be an enhancement of recognition performance for acknowledged locations but this is only for familiar directions not for novel directions. This indicates that novel view sensitivity, which was quite substantial in some obervers, is not just an artifact of drawing subjects attention to these specific locations. The fact that there was no difference between acknowledged and surprise locations for novel directions suggests that performance results reflect a natural ability to generalize recognition. Secondly, even surprise familiar direction performance was significantly better than novel direction performance. This indicates that the difficulty in generalization of novel direction views is not simply a result of having to recognize apparently unfamiliar images but suggests that the apparent rotation of observer vantage point introduced the major difficulty.

There is therefore a considerable effect of view dependency revealed by this experiment which would speak in favour of the various theories of object recognition which posit that novel view recognition would be difficult when interpolation between familiar views is not possible (e.g., Ullman & Basri, 1991). However, the rotations in vantage point observed here not only changed the appearance of familiar structure but, to a small extent, introduced new detail which was not seen from the familiar direction. We reduced this possibility somewhat by placing markers close to particular structures like stairs and bannisters. However, we cannot therefore offer firm evidence against the various theories of object recognition that are based on the use of three-dimensional or schema-based mental representation (e.g., Lowe, 1986; Biederman, 1987) because such theories also predict reduced performance when key components of objects are not visible.

Finally, we turn to the general influences of the delay period on recognition. Subjects received exactly the same stimuli during both tests although they were not informed of this. Previous studies have suggested that over time subjects representations of an environment evolve from exemplars, used in training or study, to more abstract schemas (e.g., Hock et. al., 1980). If this is the case, we would have expected to see both a diminution of familiar view performance, coupled with an increase in performance for novel direction views. Although there was such a trend in our data, this was not significant. The differences between acknowledged and surprise locations for familiar directions were also still significant after the delay period. This further suggests that no such transition occurred as one would expect that the effects of the special status of images corresponding to acknowledged locations diminishes over time. It could be that over extended periods of time more concrete differences between the various conditions will be revealled and current efforts are directed at probing the performance of subjects after delay periods greater than seven days. However, there is little reason why tasks requiring more abstract representations should not be performed immediately after learning. Our previous studies (Christou & Bülthoff, 1979) revealled that subjects could correctly pick-out the corresponding floor-plan of the target environment from a series of distractors immediately after training, even though they exhibited a similar view-dependency as reported here. It seems more likely that some tasks requiring apparently abstract knowledge can be performed by reconciling more sensory specific egocentric representations and that abstract schemas are only formed as part of a process in which conflicts occur in existing representations and current sensory input. In the present studies no such conflicts occurred and therefore no changes occurred in subjects representations of the scene.

In summary, the current experiment has shown that restrictions imposed on observers viewing of a large-scale VE are reflected in their ability to generalize to the recognition of novel, unexperienced, views. It appears that the relative ease with which we ordinarily recognize our everyday environments must be the result of the unrestricted accummulated awareness of appearance from many perspectives. Furthermore, even along the familiar direction of our VE acknowledged locations were more easily recognized than in-between, surprise, locations showing that the mental representation for scenes reflects the attentional significance assigned to some locations and this significance persists over time. Finally, although allowing subjects to control their own movements did not affect familiar direction performance it did enhance their ability to recognize novel direction views as compared to the passive-observers. This suggests that active control of movement through a scene might result in the formation of a more flexible mental representation.

References

Biederman I. (1987)
Recognition-by-Components: A theory of human image understanding, *Psychological Review*, **94**, 115-147.

Bülthoff, H. H., & Edelman, S. (1992). Psychophysical support for a 2-D view interpolation theory of object recognition, *Proceedings of the National Academy of Science*, **89**, 60-64

Christou, C.G. & Bülthoff, H.H. (1997) View-direction specificity in scene recognition after active and passive learning, Technical Report No. 53, *Max-Planck-Institute for Biological Cybernetics*, Tübingen, Germany.

Gibson, J.J. (1979). *The Ecological Approach to Visual Perception*, Boston: Houghton Mifflin.

Hock, H. S., & Schmelzkopf, K. F. (1980) The abstraction of schematic representations from photographs of real-world scenes, *Memory and Cognition*, **8 (6)**, 543-554.

Logothetis, N. K., Pauls, J., & Bülthoff, H. H. (1994) View-dependent object recognition by monkeys, *Current Biology*, **4**, 401-414.

Lowe, D. G. (1986) *Perceptual Organization and Visual Recognition*, Boston:Kluwer.

Marr, D. & Nishihara H. K. (1978) Representation and recognition of the spatial organization of three-dimensional shapes, *Proceedings of the Royal Society of London, Series B*, 200, 269-294.

Menzel, E. W. (1978) Cognitive mapping in chimpanzees, In S. H. Hulse, F. Fowler & W. K. Honig (eds.) *Cognitive Process in Animal Behaviour*, Hillsadle, N.J.: Lawrence Erlbaum Associates Inc.

Neisser, U. (1967) *Cognitive Psychology*, New York:Appleton-Century-Crofts.

Piaget, J. & Inhelder, B. (1967) *The Childs Conception of Space*, New York:Norton.

Poggio, T. & Edelman, S. (1990) A network that learns to recognize three-dimensional objects, *Nature (London)*, **343**, 263-266.

Rieser, J. J. (1989). Access to knowledge of spatial structure at novel points of observation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **15**, 6, 1157-1165.

Rowland, G. L., Franken, R. E., Bouchard L. M., & Sookochoff, M. B. (1978) Recognition of familiar scenes from new perspectives, *Perceptual and Motor Skills*, **46 (3,2)**, 1287-1292.

Schadler, M. & Siegel, A. W. (1973) Young childrens knowledge of a familiar spatial environment, Paper presented at the meeting of the *Psychonomic Society*, St Louis.

Siegel, A. W. & White, S. H. (1975). The development of spatial representations of large-scale environments, In H. W. Reese, *Advances in Child Development and Behavior*, **10**, New York: Academic Press.

Shelton, A. L., & McNamara, T. P. (1997). Multiple views of spatial memory, *Psychonomic Bulletin & Review*, **4**,1, 102-106.

Ullman, S. & Basri, R. (1991) Recognition by linear combinations of models IEEE *Trans. Patt. Anal. Mach. Intell*, **13**, 992-1005.

Winston, P. H. (1975) Learning structural descriptions from examples, in P. H. Winston (ed.), *The Psychology of Computer Vision*, New York:McGraw-Hill.