A computational theory of human perceptual mapping

W. K. Yeap (wai.yeap@aut.ac.nz)
Centre for Artificial Intelligence Research
Auckland University of Technology, New Zealand

Abstract

This paper presents a new computational theory of how humans integrate successive views to form a perceptual map. Traditionally, this problem has been thought of as a straightforward integration problem whereby position of objects in one view is transformed to the next and combined. However, this step creates a paradoxical situation in human perceptual mapping. On the one hand, the method requires errors to be corrected and the map to be constantly updated, and yet, on the other hand, human perception and memory show a high tolerance for errors and little integration of successive views. A new theory is presented which argues that our perceptual map is computed by combining views only at their limiting points. To do so, one must be able to recognize and track familiar objects across views. The theory has been tested successfully on mobile robots and the lessons learned are discussed.

Keywords: perceptual map; cognitive map; spatial layout; spatial cognition.

Introduction

How do humans integrate successive views to form a perceptual map? The latter is a representation of the spatial layout of surfaces/objects perceived in one's immediate surroundings. That we have such a map is evident in that we do not immediately forget what is out of sight when we turn or move forward (see Glennerster, Hansard & Fitzgibbon (p.205, 2009) for a similar argument). However, researchers studying this problem from four different perspectives, namely how we represent our environmental knowledge (i.e. a cognitive map (Tolman, 1948; O'Keefe & Nadel, 1978)), what frame of references we use, how we see our world, and how robots create a map of their own world, have offered solutions which when taken together create a paradoxical situation. It is noted that because the problem lends itself to a straightforward mathematical solution whereby information in one view is transformed to their respective positions in the next view, much of the current studies implicitly or explicitly assume that a solution to this problem would involve such a step. This step is problematic when used to explain how humans integrate their views and the lack of an alternative method has hampered progress.

In this paper, a new computational theory of human perceptual mapping is presented. It abandons the idea of integrating successive views to form a perceptual map. Instead, it argues that what is afforded in a view is an adequate description of the current spatial local environment and hence it does not need to be updated until one moves out of it. Only then, another view is added to the map. As a result, the map is composed of views selected at different times during one's exploration of the environment. However, these views need to be organized into a coherent global map and a method has been suggested. It requires recognizing objects found in the selected views in all the in-between views that have not been selected. These objects allow one to triangulate one's position in the map and add new views to the map in their appropriate position. The theory has been tested successfully with different implementations on mobile robots and the resulting maps produced were found to exhibit several interesting characteristics of a human perceptual map.

A Perceptual Paradox?

Researchers who investigated how spatial memories are organised often suggest the existence of a two-system model: an egocentric model and an allocentric model (Mou, McNamara, Valiquette & Rump, 2004; Burgess, 2006; Rump & McNamara, 2007). These two models are very different implementations of the same basic mathematical model described above and therefore have different costs associated with their use. In particular, the former keeps track of the relationship between the self and all objects perceived. As one moves, one needs to constantly update all objects position in memory with respect to the viewer's new position. The latter creates a global map of all objects perceived using a frame of reference independent of the viewer's position. These researchers claimed that the former is best suited for organising information in a perceptual map while the latter is best for a cognitive map. However, little is said about how information encoded in an egocentric perceptual map is transferred into an allocentric cognitive map. If this is achieved via switching frame of reference, then the process is straightforward and from a mathematical standpoint, the two representations are considered equivalent. In this case, a perceptual map is a subset of a cognitive map and holds only the most recently perceived information.

Researchers who investigated the nature of cognitive maps from studying resident's memory of their environment (both adults and children) often emphasized that the map is fragmented, incomplete and imprecise (e.g. Lynch, 1960; Downs & Stea, 1973, Evans, 1980). This does not mean that the map is devoid of metric information but rather, one's memory of such information is often found to be distorted systematically as a result of applying cognitive organizing principles (Tversky, 1992). Some well-known examples of these distortions include the regularization of turns and angles (Byrne, 1979), and over- and under- estimation of distances due to factors such as direction of travel (Lee, 1970), presence of barriers (Cohen & Weatherford, 1981), and others. More recent studies have also shown that metric

knowledge could be learned very early in one's exposure to a new environment (Ishikawa & Montello, 2006; Buchner & Jansen-Osmann, 2008). In Ishikawa and Montello's (2006) study, they also found large individual differences. Most participants either manifested accurate metric knowledge from the first session or they didn't, and knowledge of both groups did not show much improvement in some subsequent trials. Note that by accurate, it is meant that participants could "estimate directions and distances, and draw sketch maps more accurately after first exposure to the routes than would be expected by pure chance alone" (p. 118). All these observations on the nature of cognitive maps suggest that one's perceptual map should also be fragmented, incomplete and imprecise.

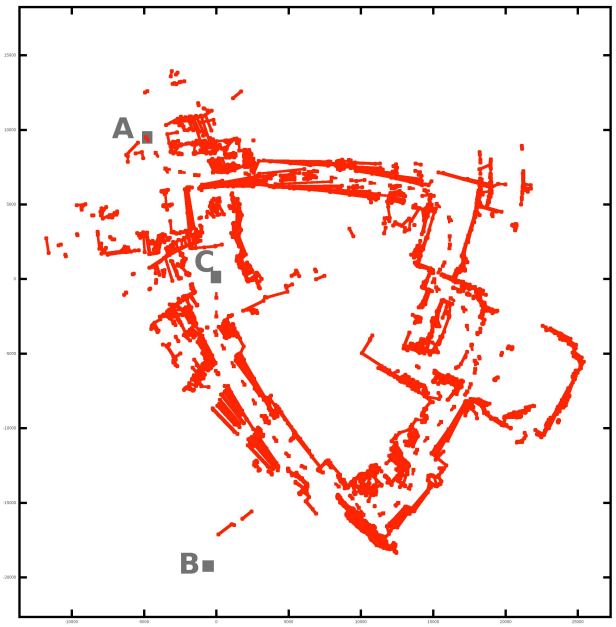

Figure 1. A distorted map

Yet, robotics researchers (e.g Thrun, 2008) who have been developing mapping algorithms using the transformation approach have shown that the map produced must be as accurate as possible. Errors found in robot sensor readings are known to seriously affect the map created. Figure 1 shows a typically distorted map computed by a mobile robot equipped with a laser sensor and without using any error correction procedure. The robot's path is as shown in Figure 2 and a rectangular shaped map should have been produced instead of the triangular one shown in Figure 1. With the map computed, one would have difficulties orienting oneself and there is also a danger that one could easily mistaken that one is returning to a familiar part of the environment. For example, at point C, the robot should be at point B in the physical space and the robot could thus be mistaken that it is re-entering a familiar part of the environment. Robotics research thus tells us that errors cannot be left unchecked when using such a procedure to compute a map. In short, the map computed needs to be precise. With hindsight, this is not surprising since the mathematical process used is aimed at producing an accurate map.

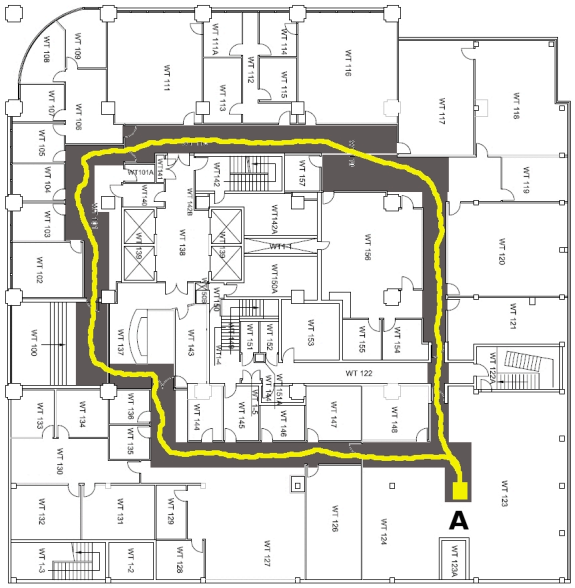

Figure 2. The test environment and the robot's path

If the perceptual map is necessary to be precise, it is surprising that our perceptual system has not evolved to support such computations. Take vision, for example. Our visual perception of the world is highly illusory (Hurlbert, 1994; Snowden, 1999) and thus, unlike computer vision, what we get is not a true geometrical description of what is out there (Fermuller, Cheong & Aloimonos, 1997; Bridgeman & Hoover, 2008; Glennerster, Hansard & Fitzgibbon, 2009). We have high visual acuity only in the small foveal region of the retina and thus a large part of our input lacks clarity and detail. Our eyes need to make rapid movements (known as saccades) to bring different regions into the foveal. Experiments on whether humans integrate successive views at the saccade level reveal that we fail to notice many kinds of changes occurring between saccades. This phenomenon is known as "change blindness" (see reviews of such work in Irwin, 1996; Intraub, 1997; Irwin & Zelinsky, 2002; Simos & Rensink, 2005) and it argues against the idea that successive views are integrated to form a single unified representation.

The above studies, when taken together, raise serious doubts as to the appropriateness of a transformational approach to human perceptual mapping.

A Theory of Human Perceptual Mapping

Logically, a perceptual map is a representation of the environment, as it is perceived. Thus, its input is a sequence of views, each being an integrated representation of information delivered by all its sensors. For simplicity, one could consider information from a single sensor and especially if it is the most important sensor. For humans, this is vision. With vision, Yeap (1998) argued that the input

should be at the level of Marr's (1982) 2½D sketch - a representation describing the shape and disposition of surfaces relative to the viewer. Yeap and Jefferies (1999) further argued that one should made explicit representations of local environments in a perceptual map and that these representations are computed from integrating successive views. The latter idea is again reminiscent of what was discussed earlier and must now be discarded.

If a representation of one's local environment is not computed from integrating successive views, what could be the alternative? In finding an answer, we make two observations. First, observe that a view affords us more than a description of the surfaces in front of us. It tells us what and where things are, where we can move to next, what events are unfolding, where there might be dangers, and others (Gibson, 1950). In short, a view is in fact a significant representation of a local environment and it should be made explicit in the map as a description of a local environment rather than as some spatially organised surfaces. Second, observe that the world we live in is relatively stable. That is, it does not change much when we blink our eyes or take a few steps forward. As such, there is no immediate need to update the view in our perceptual map as we move. For example, consider your first view of a corridor when entering it and assume an exit can be seen at the other end. If you walk down this corridor to the exit, then the description of the corridor space afforded in the first view adequately describes the local environment you are going through. Updating this description to include, for example, a view of a room besides the corridor as you walk past it will enrich the description, but is unnecessary if the room is not entered.

The tricky part of the problem is: if one does not constantly update the view in the map as one moves, how does one know where one is in the map or that one is still in the current local environment? Also, when does one begin to update the map and how? One possible solution is to keep track of objects seen in the initial view in all subsequent views. If some could be found, one could triangulate one's position in the map and thus localising oneself. However, at some limiting points, one will not be able to do so and this is when one needs to expand the map to include a new view (albeit, a new local environment). If the new view to be added is selected at a point just before reaching a limiting point, it could be added to the map using the same method of triangulation. From a human perceptual mapping standpoint, this solution is attractive since humans have developed powerful mechanisms for recognising objects.

Two points regarding the application of this method are worth noting here. First, for this method to work, it is important that one is able to track objects across successive views and for human vision, the fact that there is significant overlap between views ensures that this could be done. Second, the accuracy of this method depends on how accurately one can identify the position of the tracked objects in the map (or more precisely, the position of those points needed for triangulation). For humans, it is unlikely that the position of these points is always identified accurately and thus the map produced will be rough and vary among different individuals. The latter is a point emphasized in Ishikawa and Montello's (2006) study mentioned earlier.

A general algorithm for implementing this new theory can now be specified. Let PM be the perceptual map, V_0 be one's initial view, and R be some reference objects identified in V_0. Initialise PM with V_0. For each move through the environment, do:

Move and update:

1. Execute move instruction and get new view, V_n.
2. Search for the reference objects in V_n and remove from R those that are not found.
3. If R still contains a sufficient number of reference objects, go to step 1.
4. Expand PM, create a new R and go to step 1

In summary, the theory specifies that what is made explicit in a perceptual map is an integrated global representation of views selected during a journey. This is because each of these views provides an adequate description of the spatial layout of the local environment experienced. The basic algorithm for implementing the theory involves recognising objects in the current view that were remembered in the perceptual map, and using them to triangulate position of unknown objects (including the self) in the map. Compared to the traditional approach, this approach offers a simpler and less computationally expensive method for computing a perceptual map.

On Implementation and Results

Does the theory work? Can it produce a reasonably accurate perceptual map? One way to test the theory is to implement it and as Marr (1982) argued, the significance of a computational theory is that its implementation can be done independently. Hence, the theory was tested on a different platform – a mobile robot equipped with a laser sensor[1]. The details of our implementations will be reported elsewhere. This section highlights some key aspects of the implementation and the lessons learned so that in the next section, the significance of the theory is discussed with a concrete example.

To begin with, the theory leaves open two key implementation issues, namely how and what objects are selected for tracking across views, and how and when a new view is added to the perceptual map. These issues would depend on the kind of perceptual apparatus one has and one's needs in dealing with the environment. For our robot, the following is implemented. Laser points in each view are turned into lines denoting surfaces perceived. Any reasonably sized surfaces with at least an occluding edge are

[1]In reality, the reverse is true. The perceptual mapping problem was first investigated by considering how a robot, although with a different sensor, could solve a similar perceptual mapping problem. I refer to such robots as "albots" (Yeap, 2011).

tracked across views. The latter condition is imposed to ensure a good reference point exists for calculating the relative position of other surfaces in the map. Using laser, one's ability to perform recognition is limited. Thus, to track these surfaces between views, we use the traditional transformation method to locate them. To decide when to add a new view, the robot first detects if it has exited the local environment (by detecting that its current view has less than two tracked surfaces). Then it adds its previous view to the map (since with less than two tracked surfaces in the current view, it cannot add the current view to the map). When adding a new view, no attempt is made to update overlapping surfaces between the two views. All information in the perceptual map that occupies the same area covered by the current view will be deleted and replaced by what is in the view. The rationale here is that details are unimportant as long the overall shape of the environment is maintained.

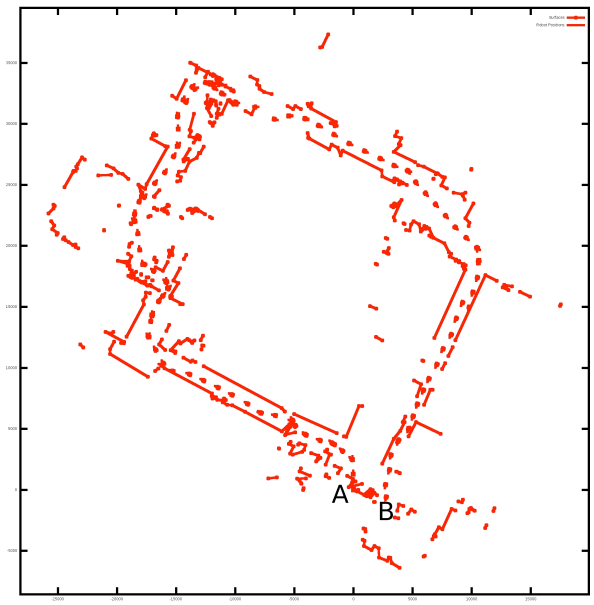

Figure 3. The perceptual map produced.

The robot algorithm used in this implementation is:

1. Execute move instruction and get a new view, V_n.
2. If it is a turn instruction, use V_n to expand PM and create a new R. Go to step 1.
3. Search for the reference objects in V_n by transforming previous view to the new view using the mathematical transformation approach.
4. If less than two objects are found, use V_{n-1} to expand PM and V_n to create a new R. To expand PM, one replaces what is in front of the robot in PM with what is seen in V_{n-1}. Go to step 1.
5. Remove reference objects in R that are no longer in view. Go to step 1.

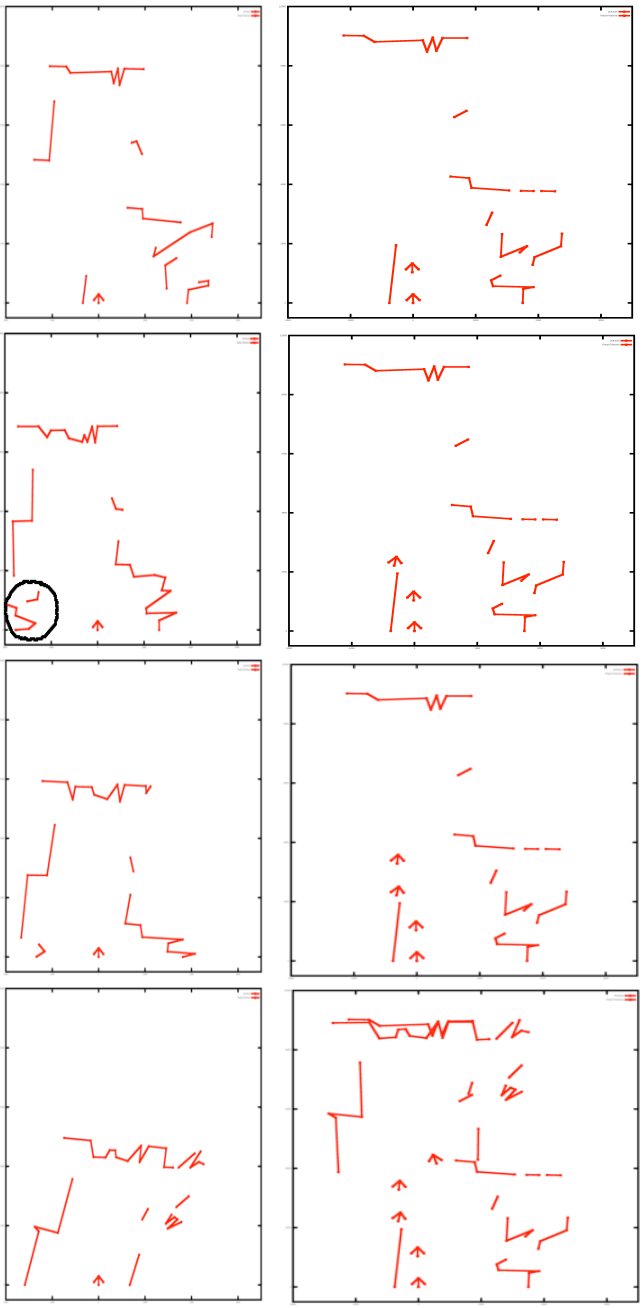

Figure 4. A trace of the mapping process

Figure 3 shows the perceptual map produced as the robot traversed the path through the environment in Figure 2. The dotted line indicates the approximate path of the robot. Points A (start) and B (end) should be the same points. Unlike the map as shown in Figure 1, this map preserves the overall shape of the environment visited. Figure 4 (left column) shows four consecutive steps of the robot. The right column shows the map expanded only ay the fourth step. The circle marks what information is missing in the map. Note that the position of the robot in the map (the little arrows) is estimated and it does not correspond to the exact position of the robot in the physical environment.

Discussion

The perceptual map shown in Figure 3 is imprecise and incomplete in the sense that it is not accurate in metric terms and has perceived surfaces missing. Yet, the overall shape of the environment experienced is maintained (as compared with the map in Figure 1). The theory thus works, at least on a mobile robot.

The present implementation, using a mobile robot with a perceptual system different from that of humans, shows that one can select different kinds of information as a reference object. This demonstrates the generality of this new approach. For humans, one expects a more complex method to select the reference objects in view. For the robot using a laser sensor, it is limited to selecting 2D line surfaces. Although the map computed by the robot is incomplete and imprecise, it is complete and precise in the sense that the overall shape of the environment is well preserved. This is partly due to the choice of information used as reference objects and partly due to the fact that the test environment is indoors. Both conditions enable the robot to detect several reference objects appearing directly in front of, and not far from, the robot. Consequently, the perceptual map is expanded more frequently, producing a more complete map. Furthermore, occluding edges of the reference targets provide good reference points for relative positioning of new information and the laser sensor provides accurate distance measurement of these points and especially if they are not too far away. Both conditions enable a fairly accurate map to be computed. From a robotics perspective, the map computed is considered surprisingly accurate since no error correction was done at the sensing level.

The perceptual map thus varies in details, both in terms of precision and completeness due to how often the map is expanded and the accuracy of the information used for expanding the map. This variability can explain the individual differences of human perceptual maps (Ishikawa and Montello, 2006). In an outdoor environment, it is likely that one selects reference objects consisting of large and easily visible distant objects. If so, one's perceptual map might not be expanded that often and consequently, one can experience not remembering much even through one has walked through a locally complex environment. In such cases, what is remembered can be the reference objects themselves. This might explain the emergence of landmarks in cognitive maps. The theory thus predicts some target features in one's perceptual map will become landmarks in one's cognitive map under the circumstances described.

The use of reference objects to expand a perceptual map has support in the literature on human vision. It has been reported that nearly all animals with good vision fixate on an object as they move, followed by some fast saccades that shift the direction of gaze (Carpenter, 1988; Land, 1999). These studies focused on why we have saccades but for the present study, it is the fixation of the eyes on an object that is more revealing. Such a mechanism allows humans (and animals) to locate and fixate on a reference object as they move and then uses saccades to improve the quality of the information perceived. One fixates using the high visual acuity region, which provides detailed reference object information. This aids later recognition and working out the position of other objects in the perceptual map.

Glennerster et al. (2009) provide an alternative explanation for the above observation to support their idea that humans do not continuously integrate successive views into a precise integrated 3D model. The reason for the latter is to explain an interesting finding from their experiments showing humans failure to notice the expansion of a room around them in an immersive virtual environment. To account for their findings, they proposed that humans compute a view graph of their environment rather than a precise 3D model. Each node in the graph is a stored snapshot of a scene and the link between them records the motor output required to move between nodes. The view graph idea is also popular for modeling animal spatial behavior and for robots (e.g Scholkopf & Mallot, 1995). However, they noted the view graph idea does not explain how different views are combined to form a consistent representation, i.e. a perceptual map. They claimed that this is an important and unsolved challenge. Interestingly, the theory proposed here could be considered as view-based since each local environment entered into the perceptual map is an individual view of the environment. However, each view is not captured as a node in a graph and there is no encoding of instructions to move from one node to the other. This theory provides a possible mechanism for integrating views to build a global map.

That the perceptual map is not updated from each successive view is strongly supported by the change blindness phenomenon. However, there is often a claim among these researchers that change blindness argues for a rethinking of how vision works and that no global map is computed. As O'Regan (1992) puts it succinctly: "the outside world is considered as a kind of external memory store which can be accessed instantaneously by casting one's eyes to some location." This theory provides an alternative way in which a global map can be computed without updating from each successive view and it is evidently clear that such a map is much needed in our interaction with the environment (Glennerster et al., 2009). The fact that the map is not constantly updated could also explain why our perception of the world is a stable one. If one were to use the transformation method, then the locations of all the points in the map are constantly adjusted to accommodate what is in the current view. If one were to trace the map computed at each step, one could see the shape of the map changes constantly as it adjusts the errors in the map. This is not the case here. The local environment once perceived in a given view will not change until much later. This gives the impression of having a very stable map (see Figure 4).

Conclusion

A computational theory of human perceptual mapping is presented which shows how a perceptual map is computed

without integrating successive views. The theory is supported by various accounts of how humans perceive their world and in particular our lack of attention to changes and the illusory nature of our perception. The theory has provided tentative account of various observations about human spatial cognition and in particular how a stable world is perceived and how landmarks might emerge. The implementation of the theory shows how the map computed is both imprecise and incomplete and yet still preserves a good shape of the environment. The implementation also shows how the theory could be implemented differently to produce map with different precisions and details and this was offered as an explanation as to why individual differences are observed.

Acknowledgments

I would like to thank my students, Zati Hakim, Md. Zulfikar Hossain, and Thomas Brunner, who have collaborated on this project, and to the reviewers who have given valuable comments.

References

Bridgeman, B. & Hoover, M. (2008). Processing spatial layout by perception and sensorimotor interaction. *Quarterly Journal of Experimental Psychology, 61*, 851-859

Buchner, S., & Jansen-Osmann, P. (2008). Is route learning more than serial learning? *Spatial Cognition and Computation, 8*, 289-305

Burgess, N. (2006). Spatial memory: How egocentric and allocentric combine. *Trends in Cognitive Sciences*,

Byrne, R.W. (1979). Memory for urban geography. *Quarterly Journal of Experimental Psychology, 31*, 147-154

Carpenter, R.H.S. (1988). *Movements of the eyes*. London: Pion Ltd.

Cohen, R. & Weatherford, D.L. (1981). The effect of barriers on spatial representations. *Child Development, 52*, 1087-1090

Downs, R.M., & Stea, D. (1973). *Image and environment: Cognitive mapping and spatial behaviour*. Chicago: Aldine.

Evans, G.W. (1980). Environmental cognition. *Psychological Bulletin, 88*, 259-287

Fermuller, C., Cheong, L.F., & Aloimonos, Y. (1997). Visual space distortion. *Biological Cybernetics, 77*, 323-337

Glennerster, A., Hansard, M.E., & Fitzgibbon, A.W. (2009). View-based approaches to spatial representation in human vision. In D. Cremers, B. Rosenhahn, A.L. Yuille, & F.R. Schmidt, (Eds.), *Statistical and Geometrical Approaches to Visual Motion Analysis*. Berlin: Springer-Verlag.

Hurlbert, A.C. (1994). Knowing is seeing. *Current Biology, 4*, 423-426

Intraub, H. (1997). The representation of visual scenes. *Trends in Cognitive Sciences, 1*, 217-222

Irwin, D.E. (1996). Integrating information across saccadic eye movements. *Current Directions in Psychological Science, 5,* 94-100

Irwin, D.E., & Zelinsky, G.J. (2002). Eye movements and scene perception: Memory for things observed. *Perception & Psychophysics, 64*, 882-895

Ishikawa, T., & Montello, D.R. (2006). Spatial knowledge acquisition from direct experience in the environment: Individual differences in the development of metric knowledge and the integration of separately learned places. *Cognitive Psychology, 52*, 93-129

Lee, T. (1970). Perceived distance as a function of direction in the city. *Environment and Behavior, 2*, 40-51

Lund, M.F. (1999). Motion and vision: Why animals move their eyes. *Journal of Comparative Physiology A, 185*, 341-352

Lynch, K. (1960). *The image of the city*. Cambridge, MA: MIT Press.

Marr, D. (1982). *Vision*. San Francisco, CA: Freeman.

Mou, W., McNamara, T.P., Valiquette, C.M., & Rump, B. (2004). Allocentric and egocentric updating of spatial memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 142-57

O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map.* Oxford: Clarendon Press

O'Regan, J.K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology, 46*, 461-488

Rump, B., & McNamara, T.P. (2007). Updating in models of spatial memory. In Barkowsky, T.; Knauff, M.; Ligozat, G.; and Montello, D.R. eds. *Spatial Cognition V Reasoning, Action, Interaction.* Berlin, Heidelberg.: Springer.

Scholkipf, B., & and Mallot, H.A. (1995). View-based cognitive mapping and path planning. *Adaptive Behavior, 3*, 311-348

Simons, D.J., & Rensink, R.A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences, 9*, 16-20

Snowden, R.J. (1999). Visual perception: Here's mud in your mind's eye. *Current Biology, 9*, R336-R337

Tolman, E.C. (1948) Cognitive maps in rats and men. *Psychological Review, 55*, 189-208

Thrun, S. (2008). Simultaneous localization and mapping. In M.E. Jefferies & W.K. Yeap (Eds.), *Robotics and Cognitive Approaches to Spatial Mapping*. Springer Tracts in Advanced Robotics.

Tversky, B. (1992). Distortions in cognitive maps. *Geoforum, 23*, 131-138

Yeap, W.K. (1988). Towards a computational theory of cognitive maps. *Artificial Intelligence 34*, 297-360

Yeap, W.K., & Jefferies, M.E (1999). Computing a representation of the local environment. *Artificial Intelligence 107,* 265-301

Yeap, W.K. (2011). How Albot_0 finds its way home: A novel approach to cognitive mapping using robots. *Topics in Cognitive Science, in press*