UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

A neural network model for learning to represent 3D objects via tactile exploration

Permalink

https://escholarship.org/uc/item/3515m810

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 40(0)

Authors

Yan, Xiaogang

Knott, Alistair

Mills, Steven

Publication Date

2018

eScholarship.org Powered by the California Digital Library
University of California

A neural network model for learning to represent 3D objects via tactile exploration

Xiaogang Yan (yanxg@cs.otago.ac.nz)
Department of Computer Science, University of Otago, North Dunedin 9016, New Zealand

Alistair Knott (alik@cs.otago.ac.nz)
Department of Computer Science, University of Otago, North Dunedin 9016, New Zealand

Steven Mills (steven@cs.otago.ac.nz)
Department of Computer Science, University of Otago, North Dunedin 9016, New Zealand

Abstract

This paper aims to answer the fundamental but still unanswered question: *how can brains represent 3D objects?* Rather than building a model of visual processing, we focus on modeling the haptic sensorimotor processes through which objects are explored by touch. This idea is inspired from two main facts: 1) in developmental terms, tactile exploration is the primary means by which infants learn to represent object shapes; 2) blind people can also represent and distinguish objects just by haptic exploration. Therefore, in this paper, we firstly establish the relationship between the geometric properties of an object and constrained navigation action sequences for tactile exploration. Then, a neural network model is proposed to represent 3D objects from these experiences, using a mechanism that is computationally similar to that used by hippocampal place cells. Simulation results based on a $2 \times 2 \times 2$ cube and a $3 \times 2 \times 1$ cuboid show that the proposed model is effective for representing 3D objects via tactile exploration and comparative results suggest that the model is more efficient and accurate when learning a representation of the $3 \times 2 \times 1$ cuboid with an asymmetrical geometrical structure than the $2 \times 2 \times 2$ cube with a symmetrical geometrical structure.

Keywords: tactile exploration; 3D object representations; constrained navigation action sequences; neural network

Introduction

How is the geometry of 3D objects represented in the mammalian brain? This question has attracted much attention in cognitive science (e.g. Biederman, 1985; Bülthoff et al., 1995; Logothetis & Pauls, 1995; Murata et al., 1997). Most researchers have focussed on deriving 3D object representations from vision (e.g. Georgieva et al., 2009; Logothetis & Pauls, 1995; Murray et al., 2003). However, retinal representations of 3D objects mean very little in themselves: they only acquire meaning by being correlated with motor movements, and in particular motor affordances (Gibson, 1950).

In this paper, we develop a model of how 3D object representations can be learned through tactile exploration. There are several recent models of this process (see e.g. Gemici & Saxena, 2014; Natale et al., 2004). In Gemici & Saxena (2014) and Natale et al. (2004), robots are trained to learn haptic representation of objects through tactile exploration and then manipulate objects. Our model derives from one particular intuition—namely that the process of exploring an object using a hand is computationally very analogous to the process whereby a freely moving agent explores its two-dimensional environment. This latter process has been intensively studied: it is known to involve the hippocampal region, in particular the system of place cells and grid cells in the hippocampus and entorhinal cortex (Etienne & Jeffery, 2004; Frank et al., 2000; Fyhn et al., 2004; Moser et al., 2008).

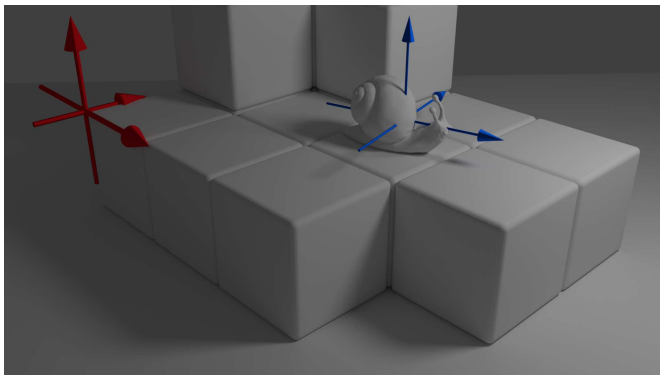


Figure 1: An agent (here a snail) exploring with only tactile feedback can use its egocentric (blue axes) observations to build up a model of its environment in an allocentric (external, red axes) frame.

A place cell's circuit receives information about the agent's locomotion movements ('dead reckoning'), and also perceptual information about the current environment, derived from vision (and olfaction in rats; see e.g. McNaughton et al., 2006). Through these *egocentric* inputs, it learns an *allocentric* representation of the local environment (Gramann et al., 2010; Holdstock et al., 2000; Klatzky, 1998; Vidal et al., 2004). For example, as shown in Fig. 1, the exploring agent (i.e. a snail in that figure) can use its egocentric inputs to build up an allocentric representation of its environment. In our model, we assume the hand is a 'navigating agent', which can execute various kinds of movements in its own coordinate system and explore the spatial information of the navigated environment. We will abstract away from the arm actions that produce these movements, so the hand is construed as an autonomous navigator. We will focus on the situation where this navigating hand is exploring the 'environment' of a 3D object. In this case, it receives various perceptual inputs about the form of the object, through the sense of touch: it can sense edges, and other surface features of the object. We then envisage a circuit, analogous to the place cells circuit, that takes hand-centered representations of hand movements and object features, and learns an 'allocentric' (i.e. object-centered) representation of the object being explored by the navigator.

In this paper, we first introduce a network for navigating 2D environments, which learns allocentric representations of places similar to those in the hippocampus. We then describe a modification of this network to model the parietal circuit involved in haptic exploration of 3D objects, that learns object-centered representations of places on 3D objects. Finally, we present an evaluation of this 3D model. The main contributions of this paper are as follows.

- Instead of representing 3D objects via vision like most researchers, the proposed model in this paper aims to represent 3D objects via tactile exploration.
- The relationship between navigation action sequences allowed for execution and the navigated object is established. By assuming the hand as a 'navigating agent', the model learns an 'allocentric' representation of a 3D object through navigation action sequences and perceptual inputs.
- Simulative results of the model exploring two 3D objects are presented and compared, which verifies that the model is effective on object representations and is more efficient and accurate for representing a cuboid, owing to its asymmetrical geometrical structure, than a similar cube.

A Model of 2D Navigation Using a Recurrent Self-Organizing Map

In this section, we describe a neural network, which represents the navigated environment through the navigating agent's egocentric dead reckoning information and perceptual information about environment landmarks.

The neural network of navigation derives from a model by Takac and Knott (see Dar & Knott, 2018 for an introduction). It is based on a recurrent self-organizing map, named modified self-organizing map (MSOM; Strickert & Hammer, 2005). In a regular self-organizing map (SOM), the units are distributed in a two-dimensional plane and each of the units are fully connected to every input unit via adjustable weights (Kohonen, 1982). During training, the units compete to be a winner for a certain input pattern and then the winner's as well as its neighborhood's weights are adapted to be more responsive for this input pattern. Finally, a non-recurrent SOM comes to represent frequently occurring *patterns* in its input units. In an MSOM, apart from the normal input units like SOM, its input contains a recurrent representation of its own state at previous time instance. After training, MSOM units come to represent frequently occurring *sequences* in its input units.

The MSOM we describe in this paper is designed to learn sequences of navigation actions executed when the navigating agent is exploring some environment. Its representations of these sequences implicitly encode 'allocentric' places in the environment, because the possible sequences are constrained by the geometry of the environment. For example, consider the 2D environment shown in Fig. 2. Say the navigating agent starts at location L1 facing North, and executes the navigation

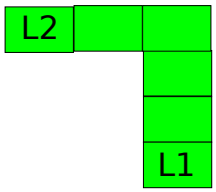


Figure 2: Geometrical description of a 2D environment, where L1 and L2 denote two different locations.

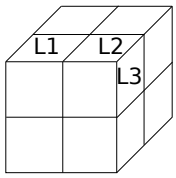


Figure 3: Geometrical description of a cube, where L1, L2 and L3 denote three different locations.

action sequence 'F→F→F→L→F→F' to reach location L2 (where 'F' and 'L' denote the movements of moving forward and turning left respectively). L2 is the only possible location that can be reached with this navigation sequence: the geometry of the environment effectively maps this sequence to a particular place in the environment. Therefore, by learning the object constrained navigation action sequences as well as the perceptual information about the object, the MSOM come to represent the navigated object.

Adapting the Model for Haptic Exploration of 3D Objects

In this section, we explain how to configure a general MSOM to represent 3D objects based on hand navigation action sequences. We intend the MSOM as a high-level model of the parietal circuit that controls haptic exploration of objects by the hand. There is good evidence that parietal cortex computes object-centered spatial representations. For instance, in humans, parietal lesions often lead to object-centered neglect (see e.g. Behrmann & Tipper, 1994), and fMRI has recently shown object-relative spatial activity (see e.g. Uchimura et al., 2015). In macaque, cells in parietal area 7a encode target location relative to an object, rather than in retina- or head-centric coordinates (Chafee et al., 2007).

When a navigating hand explores a 3D object, it can move directly, which includes moving directly forward, moving directly left, moving directly right and moving directly backward, and it can also move over an edge, which includes moving forward over an edge, moving left over an edge, moving right over an edge and moving backward over an edge. For example, with regard to a cube shown in Fig. 3, when the navigating agent is in location L1 facing East, it can move directly forward to reach location L2 while it cannot move forward over the edge. In contrast, when the agent reaches location L2, the agent cannot move directly forward whereas

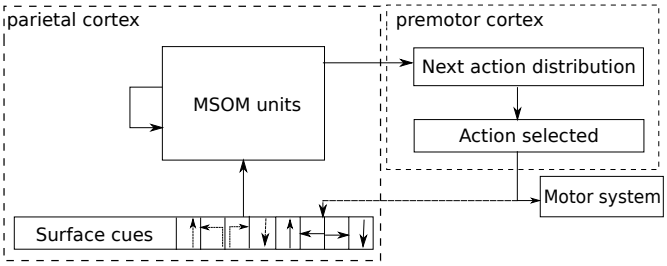


Figure 4: The architecture of the model for learning to represent 3D objects.

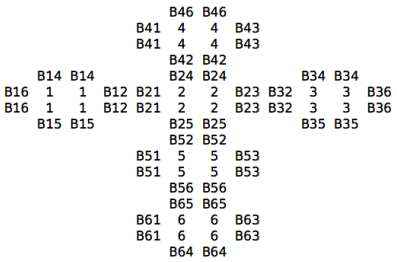


Figure 5: Geometrical description of a $2 \times 2 \times 2$ cube, where the number denotes available exploration surface of the cube, and the boundary between two surfaces is denoted by 'B' appended with such two surface numbers.

it can move forward over the edge to reach Location L3.

The architecture of the whole model for learning to represent 3D objects is illustrated in Fig. 4. The model aims to achieve the function of a navigator (i.e., the hand) when exploring a cuboidal structure. The hand is assumed to be able to execute the hand-centered navigation actions, which include movements to move directly (i.e., move directly forward, move directly left, move directly right and move directly back, denoted by $\uparrow$, $\leftarrow$, $\rightarrow$ and $\downarrow$ in solid lines respectively in Fig. 4) and movements to move over the edge (i.e., move forward over the edge, move left over the edge, move right over the edge and move backward over the edge, similarly denoted in dashed lines respectively in Fig. 4), and receive the tactile sensory inputs, like edge in front, edge on left, edge on right and edge behind. Note that the model is inspired from the circuit in brains which gleans haptic sensorimotor information of an object from peripheral sensors, then transfers the inputted information to the somatosensory cortex and finally obtains the object representation in the parietal cortex as well as the premotor cortex. More details about the model are shown in Yan et al. (2018).

Simulative Verification and Comparison

To test the performance of the model for learning to represent 3D objects, two typical 3D objects, a $2 \times 2 \times 2$ cube and a $3 \times 2 \times 1$ cuboid, are assigned to the model for exploration and representation. The length, width and height of the cube are 2 units and those of the cuboid are 3, 2 and 1 units respectively. Corresponding exploration results are illustrated and compared in this section.

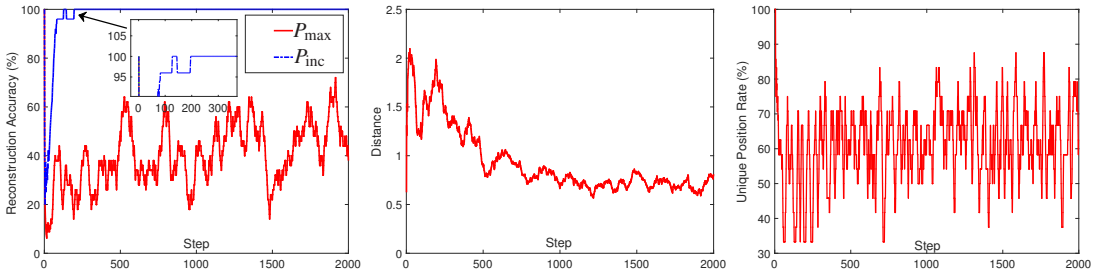
Performance Indicators

To evaluate the learning ability of the model for representing 3D objects, we first develop three performance indicators: *reconstruction accuracy*, *geodesic distance between reconstructed positions and actual position* and *uniqueness rate*. Note that to better evaluate the performance of the model, instead of obtaining average values of designed performance indicators during the whole training process, we utilize average values of performance indicators within a sliding window.

Reconstruction Accuracy With the agent exploring an object, the MSOM is trained to represent the input patterns. Based on the representation learnt via tactile exploration, for a certain MSOM activity pattern, the agent can reconstruct its position and orientation. To learn about how much information included in MSOM activities, the reconstruction accuracy is used to evaluate the learning ability of the model. At a given time instance, each position on the object is assumed as the agent's actual position with a corresponding probability, where higher probabilities denote more likely guessed positions. Thus, we design two criteria $P_{\max}$ and P_{inc}. We let ϕ denote the size of the sliding window, and then we define $P_{\max} = \frac{T(\alpha=\beta)}{\phi}$ and $P_{\text{inc}} = \frac{T(\alpha>0)}{\phi}$, where $T(\cdot)$ denotes how many times the given event happened in a given window in the sliding window series, α denotes the probability of the actual agent position in the reconstructed probability distribution and β denotes the maximal value in the reconstructed probability distribution.

Geodesic Distance Between Reconstructed Positions and Actual Position Another useful performance indicator is to measure the geodesic distance between reconstructed positions and actual position, which is defined as D_{geodesic}. Specifically, the criterion D_{geodesic} is designed to denote the sum of geodesic distances between each reconstructed position and the agent's actual position weighted by the reconstruction probability of the corresponding position. We assume $g(i,\delta)$ denotes the geodesic distance between position i and the actual agent position δ, and then we define $D_{\text{geodesic}} = \sum_{i=1}^{m}\sum_{j=1}^{4} g(i,\delta)p_{ij}$, where m denotes the number of available exploration positions on the object; $j = 1,2,3,4$ is used to respectively denote North South East West orientations; p_{ij} denotes the probability of the agent being in location i and with the particular j orientation in the reconstruction distribution.

Uniqueness Rate The above criteria say nothing about how thoroughly the hand explores the whole of an object. (In machine learning terms, they emphasize 'precision' over 'recall'.) Our exploration algorithm includes a 'boredom' routine, which encourages exploration of unknown places and avoids exploring objects in a loop, and such a routine is described in more details in Yan et al. (2018). As a measure of the algorithm's recall, we introduce another criterion called



(a) (b) (c)

Figure 6: Results of the model for exploring and representing a $2 \times 2 \times 2$ cube. (a) Reconstruction accuracy of the model for estimating the agent's position. (b) Geodesic distance between the reconstructed agent's position and the actual agent's position. (c) Uniqueness rate of the model during the exploration.

uniqueness rate U, which is the percentage of unique positions within a sliding window. Specifically, $U = \tau/\varepsilon$, where τ is the number of unique positions within a sliding window and ε is the size of the sliding window, which is the number of positions allowing tactile exploration of an object (e.g. 24 for a $2 \times 2 \times 2$ cube).

Example 1: A $2 \times 2 \times 2$ Cube

The geometrical description of the $2 \times 2 \times 2$ cube to be explored and represented is shown in an unfolded view perspective in Fig. 5. Note that in the figure, the numbers are used to denote different surfaces in the cube and 'B' with numbers appended stands for boundaries of the cube. Additionally, the number of one surface number in Fig. 5 stands for how many available exploration positions are in such a certain surface. As illustrated in Fig. 5, there are 4 available positions to be explored in each surface and in total, such a cube has 24 available positions. As stated before, the object constrains action sequences which can be performed to explore the object.

At the beginning, the agent chooses a random initial exploration position and orientation. Then, via executing constrained actions, the model is trained to represent such an object being explored. This training process is divided into 20 epochs and each epoch consists of 100 steps of successfully performed actions. Note that the unsuccessful action attempt is not counted and does not train the proposed model, since it is not allowed by the object which means it does not include any geometric information about the object being explored.

Results of the three performance indicators when the model is trained to represent the cube are illustrated in Figs. 6(a), (b) and (c) respectively. As shown in Fig. 6(a), after a short phrase of training (specifically, near 200 steps), the model can 100% predict the actual agent position with a nonzero probability, which substantiates the effectiveness of the model for learning to represent the cube. Additionally, as we can see from Fig. 6(a), the model becomes progressively better at predicting the actual agent's position, which benefits from the representation learning via tactile exploration. As depicted in Fig. 6(b), the geodesic distance between reconstructed positions and the actual agent's position becomes progressively smaller as the exploration proceeds, which further verifies the effectiveness of the model for representing the cube as well as other 3D objects. As shown in Fig. 6(c), the average uniqueness rate is about 60%, which means during 24 steps, the agent explores on average almost 15 unique positions; that is, it does not simply travel in a loop, which to some extent indicates the effectiveness of the model. Note that the model can also accurately predict the actual agent's position and orientation at the same time. Since the result of the model predicting the agent's actual position and orientation is similar with that shown Fig. 6(a) and (b), we omit it here for saving space. Therefore, the effectiveness of the model for representing 3D objects is justified.

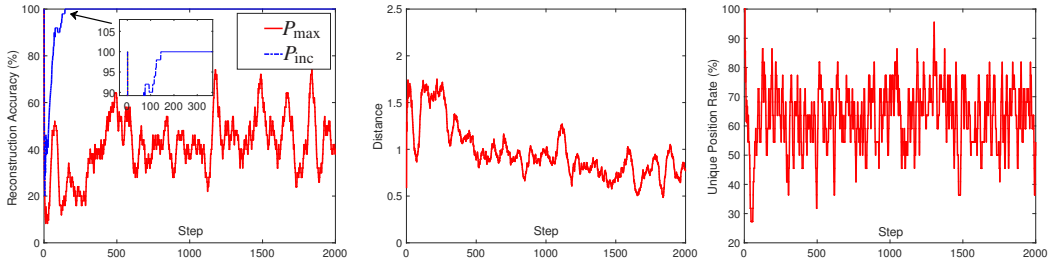
Example 2: A $3 \times 2 \times 1$ Cuboid

For further investigating its effectiveness for representing 3D objects as well as for comparison, the model also explores a $3 \times 2 \times 1$ cuboid. To save space, we omit its geometrical description, which is similar to the $2 \times 2 \times 2$ cube.

Results of this exploration, based on a random starting position and orientation, are shown in Fig. 7. As shown in Fig. 7(a), after about 150 training steps, the model can 100% predict the actual agent's position with a nonzero probability, which substantiates the effectiveness of the model for representing the cuboid. Similarly, the model becomes progressively better at predicting the agent's actual position as exploration proceeds. Moreover, as we can see from Fig. 7(b), the geodesic distance between reconstructed positions and the actual agent's position becomes progressively smaller as exploration proceeds, which suggests that the model's representation of the cuboid's geometry improves over this time. The average uniqueness rate 63% shown in Fig. 7(c) indicates during 22 steps, the agent explores 14 unique positions of the cuboid possessing 22 available positions, which means that the agent does not explore the object in a loop. Therefore, the model's effectiveness is demonstrated again.

Comparison

To further investigate the model's performance in learning to represent 3D objects, we quantitively compare simulative re-



(a) (b) (c)

Figure 7: Results of the model for exploring and representing a $3 \times 2 \times 1$ cuboid. (a) Reconstruction accuracy of the model for estimating the agent's position. (b) Geodesic distance between the reconstructed agent's position and the actual agent's position. (c) Uniqueness rate of the model during the exploration.

Table 1: Comparison of average values of criteria from step 1 to step 2000 and from step 200 to step 2000, respectively denoted as avg and avg_{200}, when exploring the $2 \times 2 \times 2$ cube and $3 \times 2 \times 1$ cuboid

		$P_{\max}$	P_{inc}	D_{geodesic}	U
# cube	avg	40.24%	98.09%	0.93	61.26%
	avg_{200}	41.84%	100.00%	0.85	61.95%
# cuboid	avg	41.72%	98.26%	0.97	63.08%
	avg_{200}	43.32%	100.00%	0.91	63.37%

Table 2: Comparison of maximal values of criteria from step 200 to step 2000 when exploring the $2 \times 2 \times 2$ cube and $3 \times 2 \times 1$ cuboid

	$P_{\max}$	P_{inc}	U
# cube	72.00%	100.00%	87.50%
# cuboid	**74.00**%	100.00%	**95.50**%

sults of the model obtained when representing the $2 \times 2 \times 2$ cube and $3 \times 2 \times 1$ cuboid via tactile exploration from different perspectives.

Firstly, by studying the time course of evaluation criteria, the learning ability of the model can be examined. As indicated in Fig. 6(a), the cube needs about 200 steps for the preliminary training, to the point when the model begins to have success in predicting the agent's actual position. By contrast, the cuboid needs only about 150 steps for the preliminary training, as shown in Fig. 7(a), which suggests that the model is more efficient in learning to represent the cuboid compared to the cube. Secondly, we compare average values of the criteria from step 1 to step 2000 with those from step 200 to step 2000 to study the representation learning ability of the model, which are shown in Table 1. As we can see from Table 1, all average values of all reconstruction accuracy criteria for the cube and cuboid from step 200 to step 2000 are greater than those from step 1 to step 2000, which means that after a procedure of exploration of objects, the model comes to predict the agent's position more accurately. In addition, all average values of geodesic distance criteria for the cube and cuboid from step 200 to step 2000 are smaller than those

Table 3: Comparison of minimum values of criterion from step 200 to step 2000 when exploring the $2 \times 2 \times 2$ cube and $3 \times 2 \times 1$ cuboid

	D_{geodesic}
# cube	0.56
# cuboid	**0.48**

from step 1 to step 2000, which further substantiates the effectiveness of the model for representing 3D objects.

When it comes to the maximal reconstruction accuracy values reached during learning, from Table 2, we can see that the geometry of the cuboid is learned better than that of the cube. As illustrated in Table 3, the minimum geodesic distance value of the cuboid reached during learning is smaller than that of the cube, which verifies again that the model is more successful in representing the cuboid than the cube.

Based on the above analysis, we can draw the conclusion that the model is effective for learning to represent 3D objects via tactile exploration. Moreover, multiple criteria substantiate that the model is more effective and efficient to represent a cuboid than a cube, which is due to that the extra asymmetrical space information involved in the cuboid contributes to its representation learning of the model.

Conclusion

This paper has developed a neural network model for learning to represent 3D objects through a navigating agent's tactile navigation action sequences. Simulative results based on two 3D objects, i.e., a $2 \times 2 \times 2$ cube and a $3 \times 2 \times 1$ cuboid have verified the efficacy and accuracy of the proposed model to learn representing 3D objects via tactile exploration. Finally, having shown a proof of concept for the model, we intend to examine in more detail whether the parietal cortex has circuitry which could implement it. SOMs are a reasonable high-level model of cortex (see e.g. Adesnik et al., 2012; Kohonen, 1982, 1993; Ritter et al., 1992); and there are various recurrent loops involving parietal cortex which could implement the recurrent component of an MSOM - for instance, the corticostriatal loops of Alexander et al. (1986).

Acknowledgment

The authors would like to thank Martin Takac for providing his code, his help and his earlier work of the SOM-based navigation model, which lays the foundation of this paper.

References

Adesnik, H., Bruns, W., Taniguchi, H., Huang, Z. J., & Scanziani, M. (2012). A neural circuit for spatial summation in visual cortex. *Nature*, *490*(7419), 226–231.

Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual review of neuroscience*, *9*(1), 357–381.

Behrmann, M., & Tipper, S. P. (1994). Object-based attentional mechanisms: Evidence from patients with unilateral neglect. In *Conscious and nonconscious information processing* (pp. 351–375). Cambridge: The MIT Press.

Biederman, I. (1985). Human image understanding: Recent research and a theory. *Computer vision, graphics, and image processing*, *32*(1), 29–73.

Bülthoff, H. H., Edelman, S. Y., & Tarr, M. J. (1995). How are three-dimensional objects represented in the brain? *Cerebral Cortex*, *5*(3), 247–260.

Chafee, M. V., Averbeck, B. B., & Crowe, D. A. (2007). Representing spatial relationships in posterior parietal cortex: single neurons code object-referenced position. *Cerebral Cortex*, *17*(12), 2914–2932.

Dar, H., & Knott, A. (2018). Learning and representing the spatial properties of objects via tactile exploration. *Technical Report OUCS-2018-01*, Department of Computer Science, University of Otago.

Etienne, A. S., & Jeffery, K. J. (2004). Path integration in mammals. *Hippocampus*, *14*(2), 180–192.

Frank, L. M., Brown, E. N., & Wilson, M. (2000). Trajectory encoding in the hippocampus and entorhinal cortex. *Neuron*, *27*(1), 169–178.

Fyhn, M., Molden, S., Witter, M. P., Moser, E. I., & Moser, M.-B. (2004). Spatial representation in the entorhinal cortex. *Science*, *305*(5688), 1258–1264.

Gemici, M. C., & Saxena, A. (2014). Learning haptic representation for manipulating deformable food objects. In *Intelligent robots and systems (iros 2014), 2014 ieee/rsj international conference on* (pp. 638–645).

Georgieva, S., Peeters, R., Kolster, H., Todd, J. T., & Orban, G. A. (2009). The processing of three-dimensional shape from disparity in the human brain. *Journal of Neuroscience*, *29*(3), 727–742.

Gibson, J. J. (1950). *The perception of the visual world.* Riverside Press, Cambridge, England.

Gramann, K., Onton, J., Riccobon, D., Mueller, H. J., Bardins, S., & Makeig, S. (2010). Human brain dynamics accompanying use of egocentric and allocentric reference frames during navigation. *Journal of cognitive neuroscience*, *22*(12), 2836–2849.

Holdstock, J., Mayes, A., Cezayirli, E., Isaac, C., Aggleton, J., & Roberts, N. (2000). A comparison of egocentric and allocentric spatial memory in a patient with selective hippocampal damage. *Neuropsychologia*, *38*(4), 410–425.

Klatzky, R. L. (1998). Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. In *Spatial cognition* (pp. 1–17).

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological cybernetics*, *43*(1), 59–69.

Kohonen, T. (1993). Physiological interpretation of the self-organizing map algorithm. *Neural Networks*, *6*(7), 895–905.

Logothetis, N. K., & Pauls, J. (1995). Psychophysical and physiological evidence for viewer-centered object representations in the primate. *Cerebral Cortex*, *5*(3), 270–288.

McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I., & Moser, M.-B. (2006). Path integration and the neural basis of the'cognitive map'. *Nature Reviews Neuroscience*, *7*(8), 663–678.

Moser, E. I., Kropff, E., & Moser, M.-B. (2008). Place cells, grid cells, and the brain's spatial representation system. *Annual review of neuroscience*, *31*, 69–89.

Murata, A., Fadiga, L., Fogassi, L., Gallese, V., Raos, V., & Rizzolatti, G. (1997). Object representation in the ventral premotor cortex (area f5) of the monkey. *Journal of neurophysiology*, *78*(4), 2226–2230.

Murray, S. O., Olshausen, B. A., & Woods, D. L. (2003). Processing shape, motion and three-dimensional shape-from-motion in the human cortex. *Cerebral cortex*, *13*(5), 508–516.

Natale, L., Metta, G., & Sandini, G. (2004). Learning haptic representation of objects. In *International conference on intelligent manipulation and grasping*.

Ritter, H., Martinetz, T., Schulten, K., Barsky, D., Tesch, M., & Kates, R. (1992). *Neural computation and self-organizing maps: an introduction*. Addison-Wesley Reading, MA.

Strickert, M., & Hammer, B. (2005). Merge som for temporal data. *Neurocomputing*, *64*, 39–71.

Uchimura, M., Nakano, T., Morito, Y., Ando, H., & Kitazawa, S. (2015). Automatic representation of a visual stimulus relative to a background in the right precuneus. *European Journal of Neuroscience*, *42*(1), 1651–1659.

Vidal, M., Amorim, M.-A., & Berthoz, A. (2004). Navigating in a virtual three-dimensional maze: how do egocentric and allocentric reference frames interact? *Cognitive Brain Research*, *19*(3), 244–258.

Yan, X., Knott, A., & Mills, S. (2018). A neural network model for learning to represent 3d objects via tactile exploration: technical appendix. *Technical Report OUCS-2018-05*, Department of Computer Science, University of Otago.