Primate-inspired vehicle navigation using optic flow and mental rotations

Ronald C. Arkin*, Frank Dellaert, Natesh Srinivasan and Ryan Kerwin

School of Interactive Computing, 85 5th ST NW, Georgia Institute of Technology, Atlanta, GA, USA 30332

ABSTRACT

Robot navigation already has many relatively efficient solutions: reactive control, simultaneous localization and mapping (SLAM), Rapidly-Exploring Random Trees (RRTs), etc. But many primates possess an additional inherent spatial reasoning capability: mental rotation. Our research addresses the question of what role, if any, mental rotations can play in enhancing existing robot navigational capabilities. To answer this question we explore the use of optical flow as a basis for extracting abstract representations of the world, comparing these representations with a goal state of similar format and then iteratively providing a control signal to a robot to allow it to move in a direction consistent with achieving that goal state. We study a range of transformation methods to implement the mental rotation component of the architecture, including correlation and matching based on cognitive studies. We also include a discussion of how mental rotations may play a key role in understanding spatial advice giving, particularly from other members of the species, whether in map-based format, gestures, or other means of communication. Results to date are presented on our robotic platform.

Keywords: Mobile robot navigation, optical flow, mental rotations

1. INTRODUCTION

In many respects, robot navigation is a solved problem. There already exist a large number of techniques for moving a mobile robot from one location to another: simultaneous localization and mapping (SLAM)[1]; reactive behavior-based control[2]; Rapidly Exploring Random Trees (RRTs)[3] to name but a few. Some of these methods are biologically plausible, others less so or not at all.

A question our research group has been pondering is the fact that many primates possess the ability to perform mental rotations (spatial transformations from one frame of reference to another of abstract representations of objects or space) and wondering what possible role this cognitive asset can play in getting from one point to another. Mental rotations refer to the ability to manipulate 2D and 3D cognitive object representations.

Specifically our research addresses the question of how mental rotations can enhance and supplement existing robot navigational capabilities. To answer this question we explore the use of optical flow as a basis for extracting abstract representations of the world, comparing these representations with a goal state of similar format and then iteratively providing a control signal to a robot to allow it to move in a direction consistent with achieving that goal state. This paper reports on our insights and progress to date while we study a range of transformation methods that can implement the mental rotation component of the architecture, some based on cognitive studies. Results to date are presented on our robotic platform.

*arkin@cc.gatech.edu; phone 1 770 977-4630; www.cc.gatech.edu/ai/robot-lab

2. MENTAL ROTATIONS

2.1 Biological Considerations

We have previously reported on the significance of mental rotations in biology[4,5]. We summarize these findings in this section. The cognitive ability of mental rotation has been observed in numerous animals (mostly primates): humans[6,7], rhesus monkeys[8], baboons[9,10], and sea lions[11,12] serve as exemplars. There exists another related transformational process in nature referred to as rotational invariance, which is more common in non-primate species such as pigeons[13] and is believed responsible in part for their homing ability. Mental rotation and rotational invariance differ in terms of their timing constants in their performance, among other factors. There also exists significant controversy regarding the underlying representations used in the mental rotation process: they have been posited to be either visual analogues[7] or propositional models[14], with little resolution on the matter to date.

While we are not particularly interested in reproducing the specific methods by which primates conduct mental rotations for use in robots, we believe based on the principle of biological economy[15] that this cognitive system serves a useful purpose to the animal. We further posit the research hypothesis that mental rotations may play a yet to be verified role in navigation, particularly with respect to the communication of information from one conspecific to another. This may occur through non-verbal graphical communication such as maps[16,17], non-verbal communication such as gestures, and perhaps even direct verbal communication.

2.2 Architectural Considerations

Towards gaining an understanding of this role of mental rotations for robotics applications, we are creating a test harness to explore a range of representation and correlation methods derived from depth maps of the environment in which the agent (in our case a robot) is immersed. Starting from an initial depth map representation of a goal state, iterative snapshots are acquired from the current position which the robot compares to the goal state, determining the movement required to reduce the difference between the states, and then moving in that direction, repeating that process until the goal state is achieved. This process is depicted in Figure 1. The test harness, currently using both Kinect depth data and visual optic flow as the basis for the abstract representations, is used to generate the navigational impetus as shown in Figure 2.

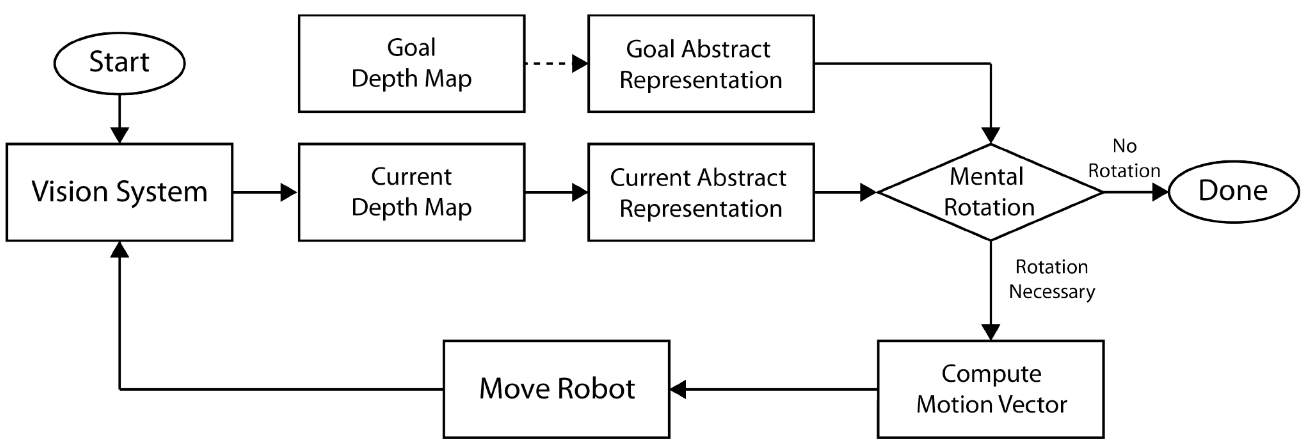


Figure 1: High-level view of algorithmic process of using mental rotation for navigation

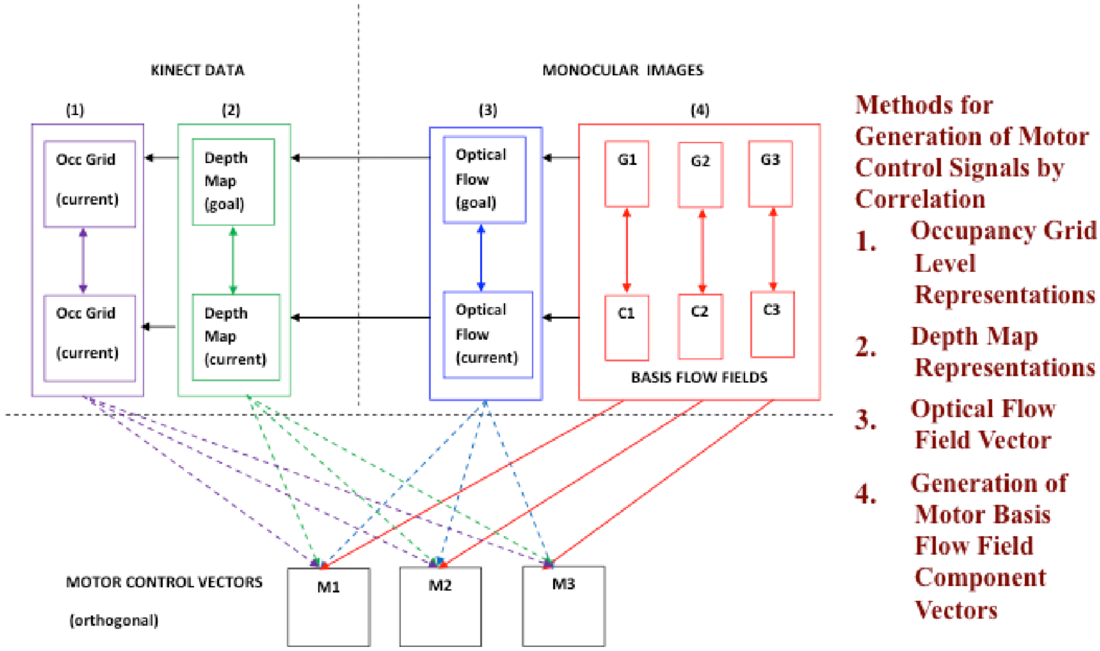


Figure 2: Multiple Representational Pathways to Motor Control

Overall there exists a superficial resemblance to some aspects of research in visual homing[18], but a key differentiating factor is that this process is conducted on an abstract representation derived from the image, not within the image itself.

The next section describes in more detail the process by which optical flow can serve the navigational needs of our mental rotation framework through the generation of a depth map of the environment.

3. OPTICAL FLOW FOR DEPTH MAP GENERATION

Optical flow is a measure of the movement of a 3D point in image space. This arises as a result of relative motion between the camera and the scene that it is imaging. Here, we consider the case when the camera is moving and the objects in the scene are stationary. As a result, the optical flow is dependent on camera egomotion and scene structure alone.

If we are given the optical flow and the camera egomotion, we can compute the depth of the scene. This is because camera egomotion, optical flow and scene depth are tightly coupled together in the following per-pixel differential constraint developed by Longuet et. al[19]

$$\boldsymbol{u} = \frac{1}{d} A\boldsymbol{v} + B\boldsymbol{\omega} \tag{1}$$

where, $\boldsymbol{v} = (v_x \quad v_y \quad v_z)^T$ is the translational velocity, $\boldsymbol{\omega} = (\omega_x \quad \omega_y \quad \omega_z)^T$ is the rotational velocity of the camera, d is the depth at pixel i and $\boldsymbol{u}$ is the optical flow at pixel i. The optical flow can be seen as a sum of two components, a depth-dependent translational component, $\frac{1}{d}A\boldsymbol{v}$ and a rotational component, $B\boldsymbol{\omega}$.

In this section we demonstrate a method that will enable us to decompose the optical flow into components that relate to camera egomotion. We achieve this by using a basis flow model developed by Roberts et. al[20]. We first obtain the camera motion using an image gradient-based approach. We then linearly combine the rotational bases by weighting them according to the estimated rotations and subtract the net rotational flow from the total dense flow to obtain the optical flow corresponding to translation. We further show that this can be achieved within a Graphics Processing Unit (GPU)[21] based framework that achieves real-time performance.

3.1 Estimation of camera motion

We estimate the camera motion by using a purely image gradient based method by exploiting the idea of basis flows[20]. The basis flows are a learned linear mapping from the camera motion to the optical flow that reduces Eq. (1) to the following simple form:

$$u = W \begin{pmatrix} \boldsymbol{v} \\ \boldsymbol{\omega} \end{pmatrix} \tag{2}$$

where W is the learned basis flow[20]. The advantage of the basis flow model is that it removes the dependency on the per-pixel depth term, d. However the trade-off of using this method is that, while rotational basis flows are scene independent, the translational basis flows depend on the large-scale scene structure and have to be re-learned for a new scene. The method described in Roberts et al.[20] provides the means to compute motion using the basis flow model. However while this paper talks about a sparse approach, we use a fully dense approach to estimate motion. This is advantageous because it makes the estimate of motion more accurate and also scalable to GPU architectures that enable us to do the motion estimation at high frame rates (120Hz) in real time. However, the details of the motion estimation itself are beyond the scope of this paper, but can be found here[22]. In this paper, we are interested in using this estimate to obtain the translational flow for performing the task of mental rotations.

3.2 Estimation of Translational Flow

The translational flow is the optical flow that is obtained if the camera were under pure translation. However, if the camera is undergoing complex motions, the optical flow is a combination of multiple components each of which corresponds to a specific motion component. We are interested in obtaining the optical flow corresponding to translation by decomposing the total optical flow.

We define the problem with the following example. Consider the case when a new object is introduced into a scene where the basis flows have already been learned. Our goal is to obtain an abstract spatial representation either in the form of a depth map or any other representation that captures the spatial details of the object. From Eq. (1), we know that the optical flow corresponding to translation carries *all* the information on the scene depth. Hence, we are interested in obtaining the translational flow on this object as a result of a complex camera motion.

In order to get the translational flow, we cannot directly use Eq. (2). This is because the translational basis flows carry no information on the new object, since the basis flows are learned without the objects in the environment. However, we exploit the fact that rotational bases are scene independent to obtain the net rotational flow by linearly combining the rotational bases after weighting them by the estimated rotational velocity.

We first obtain the total flow using a TV-L1 dense flow model[23]. This has the advantage that it is scalable to parallel architectures and can be computed quickly using commodity graphics processing hardware. We then subtract the net rotational flow from the total flow to obtain the flow corresponding to pure translation. Figure 3 shows the pipeline for obtaining the translational flow. We can see from the net estimated rotational flow that it has no information on the structure of the scene. Once the rotational flow is subtracted from the total flow, the structure become more evident in the obtained translational flow. We can further use the translational flow to get refined depth maps within the GPU by exploiting the linear constraint that exists between inverse depth, *d* and the translational flow.

4. COMPUTING MENTAL ROTATION

To compute mental rotations in the context of our system, we need to determine the rotational and translational offset of a target scene between a view from the robot's current position and a similar view from its goal location. In a high-level sense, this requires us to recognize the scene in both images, determine which parts of the scene are equivalent between images, and then compute the transformations that could be applied to the current image to align it with the goal image spatially (Fig. 4).

Rather than attempting to encode every possible element of the scene, we are currently using an abstract representation that is composed of lines found along the outlines of the objects visible in the scene. In previously reported research[4], we have used occupancy grids[24] derived from depth information as the underlying representation for rotation and correlation, but this required the projection of the depth imagery into the ground plane, which is not cognitively plausible. Working with lines within the image, as posited in some cognitive approaches to mental rotation[25], is more consistent with our stated goal of primate-inspired navigation.

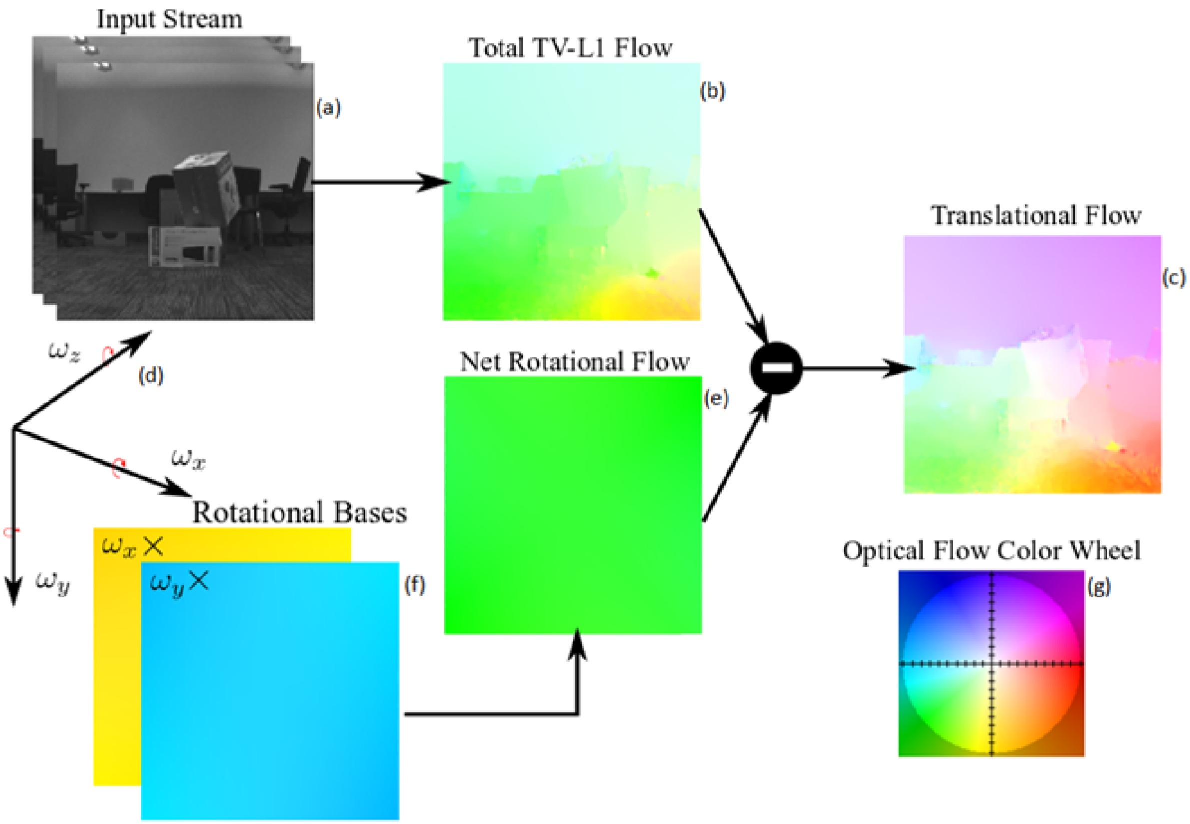
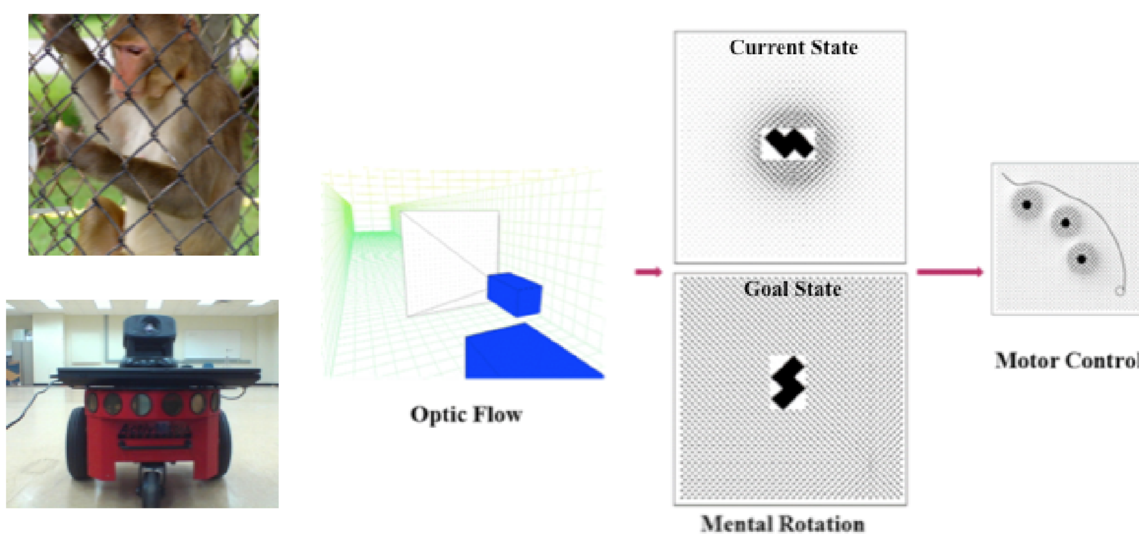


Figure 3. (a) Input image stream. (b) Computed total, TV-L1 flow. (c) Computed translational flow. (d) Estimated egomotion of camera (axes showing the components). (e) Computed rotational flow, obtained by a weighted linear combination of the rotational bases. (f) The rotational basis flows corresponding to pitch [yellow] and yaw [blue]. (g) Optical flow color wheel showing conversion from color space to vector space.



.

Figure 4. Abstract view of primate inspired mental rotations. Note the vectorial representations common to optic flow, mental rotation, and motor control.

Once the scene has been parsed into such an abstract representation and the equivalent components have been determined between current and goal imagery, recovering the rotational and translational transformations necessary to align the images simply entails finding the average spatial and rotational offset across lines between images. However, finding a useful set of lines and mapping them requires more complex methods.

To detect the image lines, we use the Fast Line Finder (FLF) algorithm[26]. This algorithm calculates the image gradient for each pixel, groups pixels with gradient direction aligned within a given threshold, and then fits a line to each of these 'line support regions'. We have previously used FLF for other navigational purposes in our research[27,28]. Since the algorithm can be tuned to favor lines aligned to orientations specified in the parameters, it is particularly useful in applications where some structure of the target is known. However, in applications where the algorithm is not biased towards certain orientations, this can lead to noise, where lines seen in the image that are not meaningful to the scene are included in the list of returned lines. We are investigating two methods for processing the lines returned from the FLF algorithm, depending on whether they are a close approximation of the scene, or contain false detections and missing lines. For the sake of simplicity, in this discussion we assume an ideal line representation of the scene from the goal image, only worrying about the quality of lines gathered from the current image (Figure 5). As this goal state may be provided from a non-visual source, such as a graphical map, this is still consistent with the hypothesis of advice giving stated earlier in Section 2.

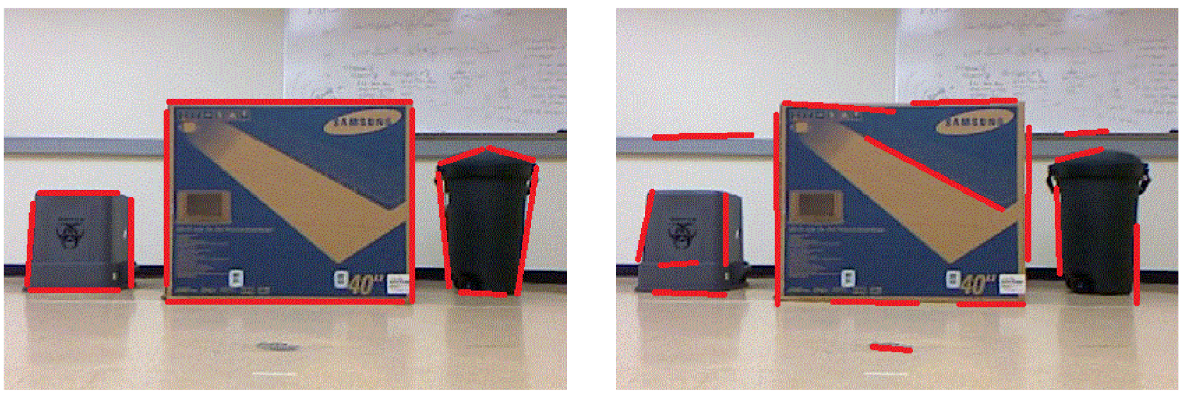


Figure 5. Possible line representations of the scene with (left) ideal detection and (right) noisy detection.

The first method investigated was developed by the Qualitative Reasoning Group at Northwestern University as the mental rotational component of a system for solving geometric analogy problems[25]. In the context of their original application, Lovett et al. performed their analysis in three major steps: parsing two input images into their basic lines, finding the correspondence of lines between images, and determining the rotation of one image to match it to the other. Because Lovett et al. deal with simple line drawings for input, and thus can easily parse their images into individual lines, we will assume that the lines we detect in the previous step represent a good depiction of the scene as viewed from the robot's current location. In this case, we need a method to map lines from the current image to the lines in the goal image.

In their system, Lovett at al. used a software tool developed in their laboratory called the structure-mapping engine (SME) to achieve this mapping[29]. The SME takes as input descriptions of both a base system and a target system, described in terms of entities, attributes and inter-entity relationships contained within the system. For example, a description of the solar system in this manner would include planets, relative sizes, orbits, etc. Given two inputs of this type, the SME calculates all possible mappings of entities from the base system to the target system and scores these mappings based on the similarity of attributes and relationships between entities. For instance, the representation of the solar system could be compared to a representation of an atom, with electrons mapped to planets and the nucleus mapped to the sun. The SME returns the mapping with the best score across all such mapped pairs.

Since the SME was developed to be a high-level system to handle all kinds of input, it is possible to prepare the information contained in a line drawing such that it fits the structure necessary for use in the SME. In the case of mental rotations, Lovett et al. treated their lines as the entities, with inter-line relationships as the attributes for each entity (such as 'leftOf', 'adjacentTo' etc.). When given such a description of both the base image and the target image, the SME returned an optimal mapping of lines between images. While this methodology works well in their application, the same effect can be achieved by using a similar method but without imposing the use of the SME and its very particular input structure onto the line relationships. Thus, in our own case, we use a mapping method based on the functioning of the SME, but tailored to our specific problem.

In the case where our abstract representation of the current lines is imperfect, we have developed a method based on target tracking to determine an optimal mapping onto the goal image. Each line in the goal image is treated as a target, while each detected line in the current image is treated as a measured approximation of the target line's position at the current time. Specifically, our method is based on Probabilistic Multi-Hypothesis Tracking (PMHT)[30]. This method was developed for target tracking in cases where many false detections occur due to noise in the data collection process. While most other tracking methods aim to choose one measurement from the noisy data set as the true position of the target, the PMHT method instead assigns a certain probability to each measurement that it truly represents the target's location. In our case, this can be thought of in terms of the probability that the lines observed in the current image correspond to a given target line from the goal image. We can represent this as

$$P(J \mid Z, G) \tag{3}$$

where Z is the set of measured lines from the current image, G is the set of target lines in the goal image, and J is a set of probabilities that each measured line $z_i \in Z$ maps to a target line $g_j \in G$. Thus the subset of lines $z_{gj} \in Z$ that should be associated with a particular line $g_j \in G$ will have the highest probabilities in J of being mapped to g_j. We can make a good guess of the true position of the desired line from the goal image based on these approximations by using the Expectation Maximization (EM) algorithm. In this case, the contributions of the different candidate lines are treated as a distribution similar to a mixture of Gaussians, from which the EM algorithm finds an optimal solution. Since the EM algorithm iteratively improves its predictions over time, the longer the robot runs, the better its predictions of the goal lines current position will become. Thus we expect we can determine a valid mapping of lines between the current and goal images. The results of this approach are in progress.

5. OCCUPANCY GRID RESULTS

In this section we review some other preliminary results, using the test harness based on the occupancy grid approach mentioned earlier[5] with a Kinect sensor providing the necessary depth information. Figure 6 illustrates the various elements in use in the algorithm. Figure 7 left provides a snapshot of the start position while Figure 7 right shows the end state for the robot. Note the similarity between the goal and end states in Figure 7 right when compared to Figure 7 left indicating the validity of the approach for this example of translational motion. A narrated video of this experiment is available at: http://www.cc.gatech.edu/ai/robot-lab/brc/movies/brc_video.mp4

6. SUMMARY

The role of mental rotations in primate navigation is not well understood and even less so in robot navigation. While a few previous studies have considered applied computational models[31,32] for describing mental rotation, our approach to this problem appears unique. The motivation is different than the norm as well, as we are not as concerned with optimal routes but rather understanding why this capability exists in primates and whether it can serve a similar, hopefully useful, role in robot navigation. We hypothesize that this role may be involved with receiving advice from other sources when navigation is stymied by traditional methods and the need to consult a conspecific or map arises.

To date we have explored methods involving occupancy grids and line-based representations with sensor depth map sources including visual optic flow and a Kinect sensor. The research continues and future results will test our hypotheses further.

ACKNOWLEDGMENTS

This research is supported by the Office of Naval Research under grant #00014-11-1-0593. The authors would like to acknowledge Ivan Walker and Joan Devassy for their valuable contributions to this project.

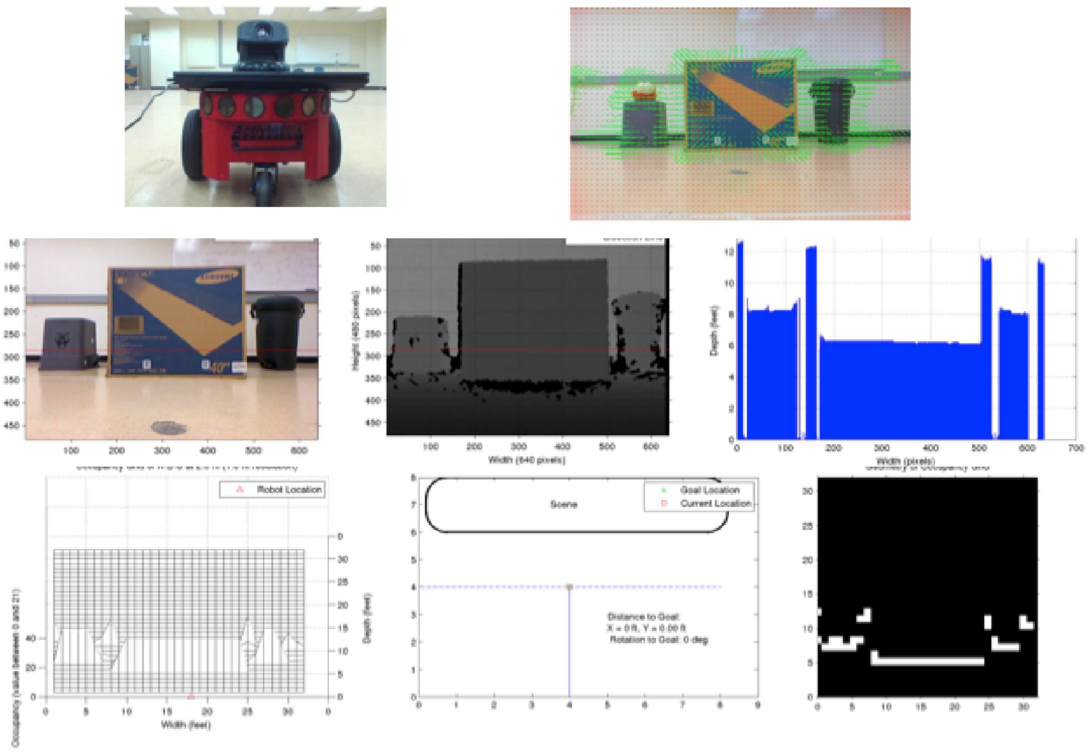


Figure 6: Experimental text components. (Top left) Pioneer robot used for experiment. (Top right) depth map generated from optical flow. (Middle left) last image of optical flow sequence. (Middle center) Kinect depth. (Middle right) depth of image row in middle of scene. (Bottom left) occupancy grid generated from depth. (Bottom center) experimental layout. (Bottom right) binarized occupancy grip for correlation matching.

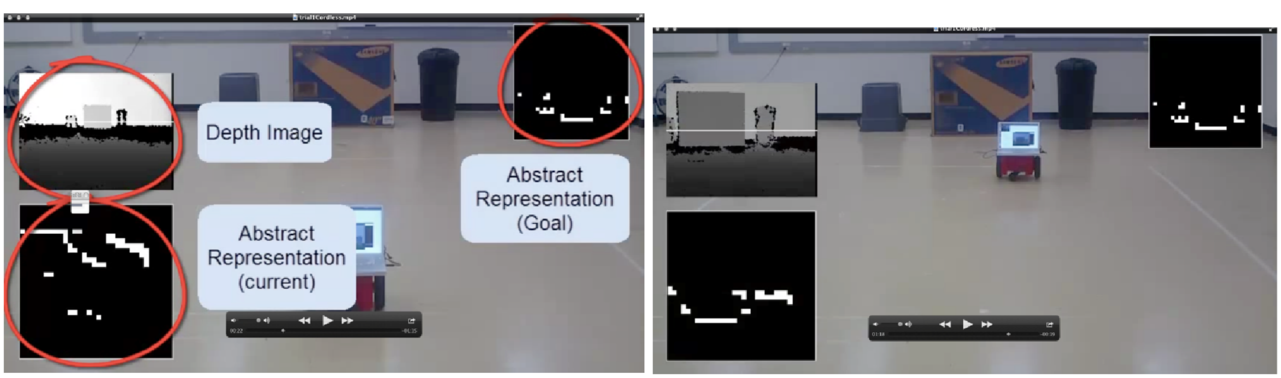


Figure 7: (Left) The start of an experimental run. 6 meters separate the start and goal state. (Right) End state of experimental run. Note the similarity between the goal and current representations when compared to the Left figure.

REFERENCES

[1] Thrun S., Burgard W., Fox D., [Probabilistic Robotics], MIT Press, Cambridge MA (2005).

[2] Arkin R.C., [Behavior-based Robotics]. MIT Press, Cambridge MA, (1998).

[3] Choset, H., Lynch, K., Hutchinson, S., Kantor, G., Burgard, W., Kavraki, L., and Thrun, S., [Principles of Robot Motion: Theory, Algorithms, and Implementations], MIT Press, (2005).

[4] Arkin, R.C., "The Role of Mental Rotations in Primate-inspired Robot Navigation", Cognitive Processing, 13(1), 83-87 (2012).

[5] Arkin, R.C., Dellaert, F., and Devassy, J., "Primate-inspired Mental Rotations: Implications for Robot Control", Proc. IEEE International Conference on Robotics and Biomimetics, (2012).

[6] Shepard R and Cooper L [Mental images and their transformations], MIT Press, Cambridge, MA, (1982).
[7] Khooshabeh P, and Hegarty M., "Representations of Shape during Mental Rotations" Proc. AAAI Spring Symposium Series, (2010).
[8] Kohler, C., Hoffman, K., Dehnhardt, G., and Mauck, B., "Mental Rotation and Rotational Invariance in the Rhesus Monkey (Macaca mulatta)", Brain Behav. Evol. 66, 258-166, (2005).
[9] Hopkins, W., Fagot, J., and Vauclair J., "Mirror-Image Matching and Mental Rotation Problem Solving by Baboons (Papio papio): Unilateral Input Enhances Performance," Journal of Experimental Psychology: General, 122(1), 61-72, (1993).
[10] Vauclair J, Fagot J, Hopkins, W. "Rotation of Mental Images in Baboons when the Visual Input is Directed to the Left Cerebral Hemisphere", Psychological Science 4(2):99-103, (1993).
[11] Stich, K. P., Dehnhardt G, & Mauck B., "Mental rotation of perspective stimuli in a California sea lion (Zalophus californianus)", Brain, Behav. Evol. 61:102-112, (2003).
[12] Mauck B, Dehnhardt G. "Mental Rotations in a California Sea-Lion (Zalophus Californianus)", Journal of Experimental Biology 200:1309-1316 (1997).
[13] Hollard, V. D., & Delius, J. D., " Rotational invariance in visual pattern recognition by pigeons and humans". Science, 218, 804-806, (1982).
[14] Pylyshyn Z., "What the mind's eye tells the mind's brain: a critique of mental imagery", Psychological Bulletin 80: 1–24, (1973).
[15] Smith, L., [Behaviorism and Logical Positivism], Stanford University Press, (1986).
[16] Shepard, R. and Hurwitz, S, "Upward Direction, Mental Rotation, and Discrimination of Left and Right Turns in Maps", Cognition, 18,161-193, (1984).
[17] Aretz, A. and Wickens, C., "The Mental Rotation of Map Displays", Human Performance, 5(4), 303-328, (1992).
[18] Franz, M., Schoekopf, B., Mallot, H., Bullthoff, H., "Where did I take that snapshot? Scene-based homing by image matching", Biol. Cybern. 79, 191-202 (1998).
[19] Longuet-Higgins, HC and Prazdny .K, "The Interpretation of a Moving Retinal Image" proceedings of the Royal Society of London. Series B, Biological Sciences (1934-1990), 208(1173):385–397, (1980).
[20] Roberts .R, Potthast .C, and Dellaert .F, "Learning general optical flow subspaces for egomotion estimation and detection of motion anomalies" IEEE Conf. on Computer Vision and Pattern Recognition, (2009).
[21] NVIDIA, "What is GPU Computing?", March 15, 2013, http://www.nvidia.com/object/what-is-gpu-computing.html
[22] Srinivasan, N., Roberts R., Dellaert F., "High Frame Rate Egomotion Estimation," 9th Int. Conference on Computer Vision Systems (2013) [submitted].
[23] Wedel. A, Pock .T, Zach .C, Bischof .C, and Cremers .C, "Statistical and geometrical approaches to visual motion analysis. Chapter: An Improved Algorithm for TV-L1 Optical Flow" pages 23–45. Springer-Verlag, Berlin, Heidelberg, (2009).
[24] Elfes A. "Occupancy Grids: A Stochastic Approach to Mobile Robot Perception and Navigation", Dissertation, Carnegie-Mellon University (1989).
[25] Lovett, A., Tomai, E., Forbus, K., & Usher, J., "Solving Geometric Analogy Problems Through Two-Stage Analogical Mapping," Cognitive science 33(7), 1192-1231 (2009).
[26] Kahn, P., Kitchen, L., and Riseman, E. M., "A fast line finder for vision-guided robot navigation," IEEE Transactions on Pattern Analysis and Machine Intelligence 12(11), 1098-1102 (1990).
[27] Arkin, R.C., Riseman, E. and Hanson, A.,"Visual Strategies for Mobile Robot Navigation", Proc. IEEE Computer Society Workshop on Computer Vision, Miami Beach FL, pp. 176-181 (1987).
[28] Arkin, R.C., Murphy, R.R., Pearson, M. and Vaughn, D., "Mobile Robot Docking Operations in a Manufacturing Environment: Progress in Visual Perceptual Strategies", Proc. IEEE International Workshop on Intelligent Robots and Systems (IROS '89), Tsukuba, Japan, pp. 147-154 (1989).
[29] Falkenhainer, B., Forbus, K. D., & Gentner, D., "The structure-mapping engine: Algorithm and examples," Artificial intelligence 41(1), 1-63 (1989).
[30] Streit, R. L., and Luginbuhl, T. E., "Maximum likelihood method for probabilistic multihypothesis tracking," Proceedings of SPIE, 394-405 (1994).
[31] Taylor H, Brunye T, Taylor S., Spatial Mental Representation: Implications for Navigation System Design. Reviews of Human Factors and Ergonomics, 4(1)1:40 (2008).
[32] Kozhevnikov, M., Motes, M., Rasch, B., and Blajenkova, O., "Perspective-Taking versus Mental Rotation Transformations and How they Predict Spatial Navigation Performance", Applied Cognitive Psychology, 20:397-417, (2006).