Attention-driven Egocentric Computer Vision for Robotic Wheelchair Navigation

Haoxiang Li[†], Philippos Mordohai[†], Gang Hua[‡]
[†]Stevens Institute of Technology [‡]Microsoft Research
[†]{hli18, Philippos.Mordohai}@stevens.edu [‡]ganghua@microsoft.com

1. Introduction

According to the National Institute of Child Health and Human Development (NICHD), 2.2 million people in the United States depend on a wheelchair for day-to-day tasks and mobility [1]. Most of them are elderly or disabled individuals, to whom independent mobility is very important. However, operating the existing manual or power wheelchairs could be difficult or impossible for many individuals [6]. Even with power wheelchairs, people with severe upper body motor impairment may not have enough hand functionality to use the joystick. To accommodate these severely disabled individuals and support their independent mobility, researchers developed a number of alternative wheelchair controls [2, 3], which can improve the quality of life of these individuals.

Although users are able to navigate the wheelchair using the above methods, these hands-free driving controls require the users' full attention. For elderly or disabled individuals, it would be desirable to increase the level of autonomy during navigation, when possible, to reduce the effort required by the users. *The long term objective of our work is to provide wheelchair users with a range of navigation controls at varying levels of autonomy according to the needs of the situation they are in.* In this context, given a well-specified navigation target, the wheelchair system should be expected to navigate to it without involving the user. Existing control methods enable selection among a set of pre-defined locations or objects for autonomous navigation. However, it is inconvenient, if not impossible, for the existing methods to specify a navigation target that is a priori unknown. Even in an indoor scenario, it is difficult to label every potential navigation target and keep the labeling up-to-date over time.

In this work, we enable the user to select any object as the navigation target based on attention modeling using an egocentric camera as the primary interface between the user and the robotic wheelchair. This natural interface provides hands-free navigation with minimal effort for disabled individuals, who have limited or even no hand functionality at all. In addition to the egocentric camera, which is mounted on a baseball cap worn by the user, our robotic wheelchair is equipped with a consumer RGB-D camera (a Kinect V2) and a tablet for displaying information to the user. A microphone (the built-in mic of the egocentric camera) is used for simple speech recognition tasks. The Kinect is fixed with respect to the wheelchair and serves as the primary navigation sensor. A key property of our sensor suite is that its cost is low compared to many of the existing, competing technologies for hands-free navigation and also compared to the cost of the power wheelchair.

This robot enables a novel attention-driven navigation mode initiated using images captured by the egocentric camera, which makes target selection without prior knowledge about the targets possible. The user triggers this navigation mode with a voice command and moves his/her head to draw attention to the target. The robotic wheelchair shows the captured target in the frontal display (the tablet) and asks the user for confirmation by voice. Then the robotic wheelchair autonomously navigates to the selected target. We consider the approach presented here as proof of concept that egocentric vision can be an invaluable technology for people with disabilities. We believe that a number of other applications in assistive computer vision and robotics will be enabled by the core technologies presented in the remainder of the paper.

2. Egocentric Attention-driven Navigation

For users who might have limited or no hand functionality at all, navigating the wheelchair using alternative control methods requires their full attention on driving. On the other hand, autonomous navigation is only applicable when the destination is given to the motion planner.

Our method provides a natural interface to specify the destination for autonomous navigation. The egocentric vision naturally follows the user's attention. When the user stares at an object for a short period of time, it is regarded as an object of interest. Our system then autonomously navigates to it upon the user's request.

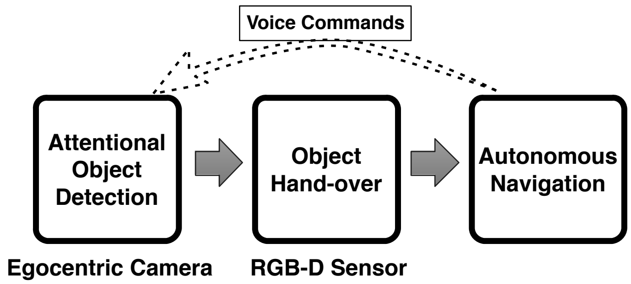

Figure 1. Work-flow of the proposed attention-driven navigation.

Figure 2. The scenario of our experiment.

As shown in Figure 1, in attention-driven navigation, the user initiates the attentional object detection with voice commands. For example, the voice command can be "robot, attention" or simply "attention". In the attentional object detection stage, the object of interest is detected from the egocentric camera view. The user sees the detected object in the frontal display. After the user confirms, the object localization process begins, in which the RGB-D sensor is used to obtain the location of the object. The robotic wheelchair then autonomously navigates to the object. During the autonomous navigation, the user is free to look around. Hence the user can initiate another attention-driven navigation to any other object without waiting for the current navigation to finish.

Our application scenario requires an efficient solution to detect the object of interest. Here we present an intuitive and practical technique. Given the image from the egocentric camera, we apply contour detection [7] to detect all the closed contours. Then for each closed contour, we estimate a bounding rectangle as a detection candidate. For each candidate window, its confidence score is defined as the Intersection-Over-Union (IoU) overlap ratio to a predefined attentional bounding box. Once the IoU overlap ratio exceeds a pre-defined threshold (0.2 in this work), the current frame is selected as the anchor frame and the object is marked as the candidate object of interest.

Then, we track the candidate object in the following frames. We estimate the optical flow from the anchor frame to the current frame and then apply RANSAC to estimate a homography. Once the number of well tracked frames (inlier ratio larger than 50%) exceeds a threshold (50 in this work), the candidate object is regarded as the attentional object. We then proceed to localize it in the RGB-D camera, which is fixed with respect to the wheelchair. After the user confirms via a voice command, the object is handed over by the egocentric to the RGB-D camera.

Because the user's gaze may be turned to the left or right with respect to the forward direction of the wheelchair when the attentional object is detected, it is possible that the object is out of the field of view of the Kinect. Hence, in order to localize the object, the robotic wheelchair needs to rotate to the right direction so that the Kinect sensor sees the object. Based on the head pose when the object is detected, we can infer the right direction to rotate the robotic wheelchair to complete the object hand-over.

While the robotic wheelchair is rotating (if necessary), we apply feature based matching to the RGB image from the Kinect sensor given the object from the egocentric camera. In this work, we use OpenCV's implementation of FAST [4] feature detector with ORB [5] feature descriptor. Then we use RANSAC to search for the best homography from matched points. We set the center of all inlier points on the image from the Kinect as the center of the object. From the corresponding depth image, we average the depth values in a 10×10 window around the object center and calculate the relative location of the object from the robotic wheelchair from the depth.

3. Experiment

We demonstrate the proposed attention-driven navigation in an indoor scenario (Figure 2). The video of the experiment is shared at https://youtu.be/2lg8GgYqfaY.

References

[1] How many people use assistive devices? nichd.nih.gov. 1
[2] T. Carlson, R. Leeb, R. Chavarriaga, and J. d. R. Millán. The birth of the brain-controlled wheelchair. In *IROS*, 2012. 1
[3] E. Perez, C. Soria, N. M. López, O. Nasisi, and V. Mut. Vision based interface applied to assistive robots. *International Journal of Advanced Robotic Systems*, 10, 2013. 1
[4] E. Rosten and T. Drummond. Machine learning for high-speed corner detection. In *ECCV*. 2006. 2
[5] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. Orb: an efficient alternative to sift or surf. In *ICCV*, 2011. 2
[6] R. C. Simpson, E. F. LoPresti, and R. A. Cooper. How many people would benefit from a smart wheelchair? *Journal of rehabilitation research and development*, 2008. 1
[7] S. Suzuki and K. Abe. Topological structural analysis of digitized binary images by border following. *Computer Vision, Graphics, and Image Processing*, 30(1):32–46, 1985. 2