Department of Mechatronic Engineering

MXEN4000 - Mechatronic Engineering Research Project 1

Progress Report

Navigation Assistance for a Semi-Autonomous Smart Wheelchair

Semester 1, 2022

Jakob Wyatt
19477143

Supervisor(s): Yifei Ren & Siavash Khaksar

Abstract

Table of Contents

List of Figures

List of Tables

1 Introduction

Many people with motor disabilities rely on wheelchairs for movement, and powered wheelchairs have enabled greater independance for people with disability. Despite the huge benefit powered wheelchairs have granted, the use of this technology can be inaccessible or unsafe for people with amyotrophic lateral sclerosis (ALS) or vision impairment, who may be unable to use a joystick or see their environment clearly.

1.1 Aims

The aim of this research is to develop a semi-autonomous smart wheelchair system. This research is done in collaboration with Glide, a WA wheelchair manufacturer, who have provided an existing powered wheelchair (CentroGlide) to use as a base for this functionality. By developing assistive technology for the wheelchair, the user is granted greater mobility, confidence, and independence.

1.2 Problem Definition

There are multiple engineering research project students who are part of this team, working on elements such as controller design, navigation assistance, and object detection. This work specifically focuses on pathway assistance, which identifies suitable paths for the wheelchair to drive on. If a user unintentionally drives off their desired path, this can lead to uneven terrain and possibly falling from the wheelchair. By guiding the user along a path, these safety issues can be mitigated.

Emphasis is placed on the ’semi-autonomous’ aspect of the wheelchair. An important requirement of this project is that the user still has control over their wheelchair, and can override any semi-autonomous functionality if required. When false positives occur within the smart wheelchair system, the users mobility should not be compromised.

Another requirement of the system is that any sensors mounted to the wheelchair should not impede the users comfort or the wheelchairs manouverability. Many wheelchair users have specific requirements for wheelchair seat adjustments, to avoid pressure sores and

Figure 1: CentroGlide in Reclined Configuration

discomfort. Figure 1 shows the wheelchair configuration when fully reclined, demonstrating that some sensor mounting locations are infeasible.

The smart wheelchair system should also be commercially viable - high-cost components and sensors are infeasible. Internet connectivity should not be a requirement for the system to operate either - the round trip time required to communicate with a server would compromise the safety of a user. Because of this, all processing is performed locally on the wheelchair.

Sensor	Advantages	Disadvantages
Stereo Camera	High Resolution	
MMWave Radar		
3D Lidar		High Cost
2D Lidar		
Ultrasonic Radar		
Inertial Measurement Unit (IMU)		
Servo Motor Encoder		

Table 1: Sensor Comparisons

2 Literature Review

Initial research on semi-autonomous wheelchairs

2.1 Sensors and Hardware

To percieve the environment, the wheelchair should be fitted with various sensors and a compute element to process the sensors output.

Table 1 shows some sensors that were considered for use in the smart wheelchair. Selecting a sensor to use is not necessarily an either-or decision. Sensor fusion algorithms such as the Extended Kalman Filter (EKF) or Unscented Kalman Filter (UKF) [1] allow outputs from multiple sensors to be used together to improve their accuracy. Additionally, some sensors may be used for different applications on the smart wheelchair.

2.2 Scene Understanding

Scene understanding is a broad field, and involves using computer vision methods on visual or spatial data to gain better knowledge about the surrounding environment. Convolutional Neural Networks (CNNs) are commonly used for this application, as they are able to exploit the local nature of image features to reduce the number of required computations.

Image classification is a core problem within this field, and involves identifying the subject of an image (such as an animal or object). AlexNet [2], based on the earlier digit-recognition CNN LeNet-5 [3], was one of the first deep CNNs applied to this problem. Alexnet was trained on the large ImageNet dataset [4], which consists of 15M images and 22K categories, and achieved an error of only 15.3% on a 1000 class subset. The underlying architecture uses a series of 5 convolutional layers and 3 fully connected layers.

Neural network architectures have become deeper and more accurate over time, enabled by both growth in computational power and dataset size. VGG-16 [5] and GoogLeNet [6] are 16 and 22 layers deep respectively, and approached human performance on the ImageNet dataset. ResNet [7] is up to 156 layers deep, and exceeds human performance at image classification with an error of 3.57%. ResNet uses a 'skipping' architecture to improve network training, where the output of a layer relies on the input of a previous layer.

Object localization is another core problem within this field, and involves identifying the location of objects within an image, as well as classifying them. This is important for our wheelchair, as we want to be able to identify the location of a pedestrian or obstacle within the environment. R-CNN [8] was one of the first object classification models which utilized convolutional networks, by identifying potential bounding boxes and running an image classifier on these bounding boxes. Fast and Faster R-CNN [9][10] improved the speed of this model by running an image classifier backbone once on the entire image, and using a CNN to improve identification of bounding boxes. Pascal VOC [11] and MS COCO [12] are datasets which are commonly used to evaluate object classification models.

YOLO (You Only Look Once) [13][14][15][16] is another object classification model which focuses on improving performance. In particular, YOLOv4 [16] reaches over 60 fps on the Tesla V100, which enables its use in real time applications such as autonomous driving and security camera footage. YOLO divides an image into an $S \times S$ grid, and uses a single convolutional network to output both bounding box predictions and image classification for each grid square. Low-probability and overlapping bounding boxes are then removed before the final output.

2.3 Assistive Control

3 Methodology

3.1 Hardware

The smart wheelchair should have the ability to sense, process, and manouver within the surrounding environment. To do this requires some necessary hardware, including a sensor system, compute element, and motor controller. Due to the 2021-2022 chip shortage, hardware selection was identified as a process that should occur relatively quickly.

The literature review provides a comparison between different sensor types and models. For outdoor navigation assistance, a forward facing stereo camera was selected as the best option for this project - specifically, the Zed 2 Mini. For the compute element, a Nvidia Jetson Xavier NX will be used, due to compatability with existing deep learning frameworks and low power usage.

3.2 Dataset Collection

To train and evaluate machine learning models, a dataset was collected.

4 Current Work

The first stages of smart wheelchair development involved:

1. Identifying desired sensors and hardware for the wheelchair.
2. Choosing an appropriate mounting point for these sensors.
3. Researching the field of machine perception and computer vision (both applied to wheelchairs and more generally).
4. Collecting an initial video dataset, enabling work to begin on labelling and algorithm evaluation.

4.1 Hardware

After consideration of the available options, it was decided to use a stereo camera as the main forward facing sensor, with 2D LIDAR used for the side and rear of the wheelchair.

The front of the joystick control unit was selected as the best mounting point for the stereo camera, due to several reasons:

1. Clear view of the environment in front of the wheelchair.
2. Not obstructed by the user in any wheelchair configuration.
3. When needed, the user can move the joystick control unit out of the way, which also moves the camera out of the way.

However, there are some challenges faced when using this mounting point, which must be addressed.

1. Shaky video footage due to low rigidity in joystick mount.
2. Close to the front of the wheelchair, which reduces visibility of the sides of the wheelchair.
3. Maximum camera width of 150 mm before doorway manouverability is affected.

To evaluate the effectiveness of this mounting point, a GoPro (Hero 4) was attached using a temporary mount and a 34 minute driving dataset was collected around Curtin University. It was found that camera shakiness could be reduced by using a stiffer mounting solution, however some shakiness would always remain due to the unstable mounting surface. It was also found that alternative sensors, such as 2D LIDAR, would be required for features such as doorway navigation and docking, due to the low field of view (FOV).

5 Future Work

6 References

[1] E. Wan and R. Van Der Merwe, "The unscented Kalman filter for nonlinear estimation," in *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium*, IEEE, 2000, pp. 153–158. DOI: 10.1109/ASSPCC.2000.882463.

[2] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, pp. 1097–1105, 2012, ISSN: 0001-0782. DOI: 10.1145/3065386.

[3] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998, ISSN: 0018-9219. DOI: 10.1109/5.726791.

[4] Jia Deng, Wei Dong, R. Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2009, pp. 248–255, ISBN: 1063-6919. DOI: 10.1109/CVPR.2009.5206848.

[5] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," 2014.

[6] C. Szegedy, W. Liu, Y. Jia, *et al.*, "Going Deeper with Convolutions," 2014.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2016, IEEE, 2016, pp. 770–778, ISBN: 1063-6919. DOI: 10.1109/CVPR.2016.90.

[8] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," 2013.

[9] R. Girshick, "Fast R-CNN," in *2015 IEEE International Conference on Computer Vision*, vol. 2015, IEEE, 2015, pp. 1440–1448, ISBN: 1550-5499. DOI: 10.1109/ICCV.2015.169.

[10] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” 2015.

[11] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, “The Pascal Visual Object Classes (VOC) Challenge,” *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2009, ISSN: 0920-5691. DOI: 10.1007/s11263-009-0275-4.

[12] T.-Y. Lin, M. Maire, S. Belongie, *et al.*, “Microsoft COCO: Common Objects in Context,” 2014. DOI: doi.org/10.48550/arXiv.1405.0312.

[13] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” 2015. DOI: doi.org/10.48550/arXiv.1506.02640.

[14] J. Redmon and A. Farhadi, “YOLO9000: Better, Faster, Stronger,” 2016. DOI: doi.org/10.48550/arXiv.1612.08242.

[15] J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement,” 2018. DOI: doi.org/10.48550/arXiv.1804.02767.

[16] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” 2020. DOI: doi.org/10.48550/arXiv.2004.10934.