# Towards Independence: an autonomous assistive device for safe at-home rehabilitation

**Authors:** James Bardet, Samuele Caneschi, Chiara Galletti, Lorenzo Germini, Camil Hamdane, Milo Imbeni

**Supervisors:** Nicolò Macellari, Sergio Hernandez

**Professors:** Grégoire Courtine, Silvestro Micera

## I. Introduction

Spinal cord injury (SCI) interrupts the normal flow of the nervous electrical signals through the spinal cord from the brain to the peripheral nervous system (PNS) and viceversa. SCI often leads to temporary or permanent loss of important motor, sensory and autonomic functions involving the body structures that are innervated by fibers descending from or ascending to areas below the injury.

Nowadays, epidural electrical stimulation (EES) to the lumbrosacral spinal cord allows for at least partial recovery, by reactivating spinal circuits. In particular, EES of lumbar segments has shown to facilitate standing and walking in both animal models and humans patients affected by SCI [1, 2], therefore it could provide an effective and promising treatment against locomotion deficits. But before SCI patients can voluntarily modulate their gait and support their body weight while standing, a long and intense rehabilitation training is needed. In the clinical trial study taking place at CHUV [3], SCI patients took part in a 5-6 months period of neurorehabilitation sessions 4 days/week at the hospital, followed by an optional extension up to 3 years of additional training at home, with continuous and regular check-ups. The rehabilitation sessions can only take place in the presence of another person (e.g. nurses, therapists, friends and family) who can monitor the patient for instability or falls and can keep the wheelchair close to him/her in case he/she feels the need to sit back on it. This necessary condition constitutes an important limiting factor. Thanks to the at-home system, with an easy-to-use user interface and a set of portable technologies, patients could indeed perform these sessions at their will without being surrounded by clinicians and without having to schedule these sessions in advance, increasing the frequency of the training and/or adapting it based on their fatigue state and improving the final results. Even at home SCI patients still need human support, and for this reason they still lack flexibility of performing the training sessions whenever and wherever they would like to.

Therefore, the goal of our project is to make patients more autonomous and independent, by proposing a strategy to ultimately build a human-following wheelchair. Through sensors, it is in fact possible to implement an algorithm that is able to detect the distance between the wheelchair and the SCI patient performing rehabilitation, and to then drive the motors of the wheelchair, so that it is maintained close enough to the subject. Thanks to this solution, it will be possible to optimize the efficiency and outcome of the neurorehabilitative training period. Through our strategy, we want to reduce the costs and optimize the training, while ensuring safety for the patients, that will always have their wheelchair within reach.

### A. *State of the art*

Robotised wheelchairs can restore autonomy to patients and usually cost a few thousand dollars [4]. Smart wheelchairs, in addition to this, integrate functions useful to users with sophisticated processing systems, such as navigational assistance and obstacle avoidance [5]. An application which is very similar to our use-case, outlined in the introduction, is the companion following functionality, which has already been implemented in commercial devices [6].

In order for the wheelchair to maintain a relative constant position to the person it is following (in our case, the patient) it needs to be governed by an algorithm able to translate information about the patient relative position to commands given to the wheels. That information can come from a variety of sensors of differing complexity and cost, influencing the final availability of the device to patients.

The rehabilitation protocol already includes Inertial Measurement Units (IMUs), which could in theory be used to deduce the distance and direction of the patients' movement. A more precise 2D circular light-based rangefinder (LiDAR) has been successfully used by Kobayashi et al. [7] to track the position and facing direction of the caregiver walking alongside the wheelchair. The processing is carried out by a laptop in the back pocket of the chair. The same system was then enhanced with bus boarding capabilities, with the use of a computer vision system running on a remote server [8].

Ultrasound rangefinders, capable of estimating the distance based on the time of flight of a sound wave, have the advantage of being viable even with low visibility or infrared reflective surfaces, which might compromise LiDAR readings. They have been employed to avoid obstacles in a walker-like assistive device [9] and in autonomous navigating wheelchairs [10], which integrate LiDAR mapping with ultrasound obstacle detection. An additional approach is to ensure a stable connection and identification to the person to be followed by the use of markers. One such example is ultrasound tethering [11], in which an ultrasonic beacon was attached to the companion's waist and localised through an array of receivers.

Camera-based depth sensing has been the subject of intense research due to its many applications in manufacturing and virtual reality [12]. The video from two cameras spaced apart (stereo vision) contains information about the depth in the pixel-wise shift between the two streams. However, this method (used by the Intel RealSense camera [13]) proves inaccurate in homogeneous scenes. Other approaches include the measurement of the deformation due to the reflection of an array of infrared light lasers, exploited by the Kinect camera [14], or more complex structured light based cameras, which however need expensive projectors.

Lastly, in recent years, deep-learning based computer vision has emerged as the gold standard in object recognition, pose estimation and even depth estimation [15]. The intense processing needed requires Graphics Processing Units (GPUs), capable of running these algorithms in real time, but only need a traditional RGB camera as input, without additional expensive sensors or projectors.

Our aim was to build a modular and inexpensive system that could be eventually integrated into existing robotic wheelchairs. In order to do so, we employed a small form-factor system-on-chip with integrated GPU, the Jetson Nano, which can be bought for less than 70$.

We excluded LiDAR sensors (over 250 $) and commercial depth cameras (400 $) from further testing given their cost, and focused instead on cheaper sensors such as IMUs, a single camera with deep learning and ultrasound rangefinders.

## II. Methods

We first attempted to limit our system to the data provided by IMUs in order to locate the patient, as they are already routinely used. Afterwards, we exploited deep learning algorithms to detect the patient and estimate the distance by using a single camera. In addition, an ultrasound rangefinder was tested. Finally, we demonstrated the feasibility of our approach by building a functioning prototype.

### A. Inertial Measurement Units

An inertial navigation system was explored as a first solution to determine the position of the patient. This system is based on Inertial Measurement Units (IMUs). They are electronic devices that measure and report linear acceleration and rotational rate, by combining measurements from multi-axis accelerometers, gyroscopes and magnetometers. Given the angular velocity from the gyroscope measurement, the device orientation is firstly obtained via Kalman filtering and integration. Based on the orientation, gravity is subtracted from the acceleration and a double integration on the residual acceleration is performed to get an estimation of the position [16]. However, all inertial navigation systems suffer from integration drift. Small sensor errors or biases accumulate in the double integration process, leading to quick explosion of the error in the position estimation [17, 18]. To implement the IMUs tracking, 16 sensors were placed on the legs of a healthy subject and measurements were performed during walking. As shown in Fig. 1, the angular position on all axes diverges, preventing an accurate estimation of orientation and thus position.

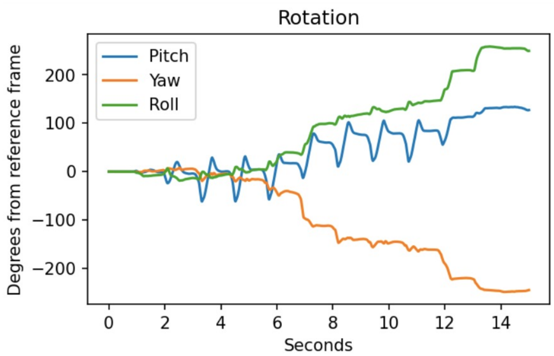


Figure 1: Evolution of angular position along the three axes. It is determined through double integration of IMUs placed on the subject's leg.

To ensure more accurate estimations, the position must be periodically corrected by implementing an Extended Kalman filtering. However, this requires inputs from other types of sensors and prevents the double integration of the IMUs from being a standalone solution for the tackled problem [17] [19].

Furthermore, deep learning algorithms have been successfully used to estimate the trajectory of a walking subject with IMUs. For example, the RoNIN network has been trained on over 40 hours of recordings from IMU sensors on smartphones [20]. Although well-grounded and easy to implement, this method does not ensure a sufficient precision for our goal, showing a prediction error with respect to the ground truth in the order of one meter (see Fig. 2).

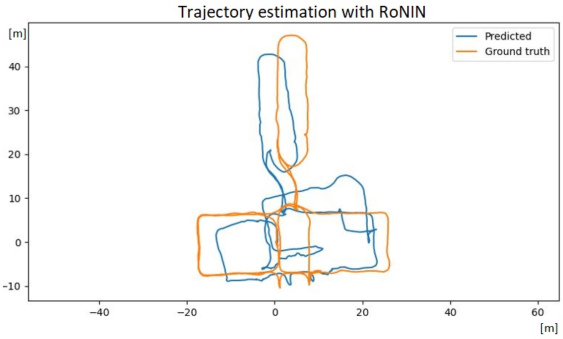


Figure 2: Trajectory estimation compared to the ground truth with the RoNIN algorithm [20].

### B. Computer Vision

We first tested a powerful human pose estimation algorithm on a laptop. The input video from the webcam is down-sampled to 640x360 pixels to achieve a frame rate of 5 frames/second. The live video is then processed using a custom-made Python algorithm based on MediaPipe [21] running on a CPU. MediaPipe is a framework developed by Google Research for building pipelines to perform inference over different sensory data. We have used MediaPipe Pose framework which allows a high-fidelity human pose estimation from video data, based on 33 3D body landmarks. It uses a background segmentation mask on the whole body from RGB video frames, based on BlazePose [22]. BlazePose is a lightweight Convolutional Neural Network (CNN) that is tailored for real-time inference on mobile devices. The model's architecture is based on a lightweight body pose detector followed by a pose tracker network that predicts the joints' coordinates in space (see Fig. 3 and 4).

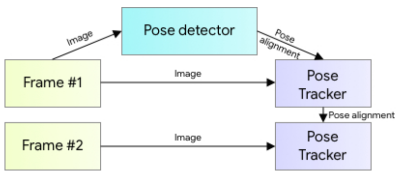


Figure 3: MediaPipe architecture [22].

With the 3D landmarks, we can then find the patient's relative distance from the camera by averaging the shoulder marker positions. The inference speed, even on a powerful laptop, is not enough for a live application: we therefore opted for a neural network accelerator for our prototype.

### C. Prototype

In order to demonstrate the feasibility of our approach we built a small prototype. It is controlled by an edge system-on-chip with an integrated 128 core GPU and 2 Gb RAM, the NVIDIA Jetson Nano. The video input comes from a Raspberry Pi Camera Module 2 with 77° field of view. We acquired a motor kit complete of L298N motor driver, 2 TT DC motors, 2 wheels and a caster ball as

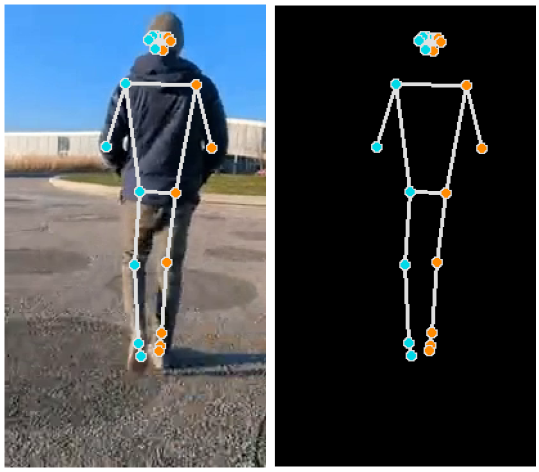

Figure 4: Video and pose estimation from MediaPipe.

third contact point. The chassis and camera holder designs were created in Blender and 3D printed with ABS on a Zortrax M200. All the components were soldered and screwed onto the chassis. The robot is powered with a 10000 mAh portable battery. The prototype independently connects to a predefined WiFi network and is controlled remotely with Jupyter notebooks.
Various attempts were made using at running MediaPipe (the pose estimation algorithm we tested on laptops) on the Jetson Nano. Due to incompatibility issues we had however to switch to YoloNet v5s [23], an already trained low-resource object-detection algorithm.

If a human is identified in the acquired frame, the bounding box is given as output. The motor control algorithm updates every frame, at more than 10 fps. It firstly checks whether the human fills the field of view: in this case, the robot is close enough (around 1 m) and stops. If the human moves away, the bounding box gets smaller and the robot moves forward. The algorithm aims at keeping the identified human centered. Depending on the deviation of the center of the bounding box from the mid-line of the image, the robot proportionally adjusts the control of the two motors in order to turn until the human is centered again and forward motion can be resumed.

### D. Ultrasound probe

As an integration to the above described algorithm, we also planned to add an HC-SR04 ultrasonic distance sensor (3$ on Amazon) to ensure redundancy and increased accuracy in the distance estimation. It was paired with a ESP32 low power micro controller with an on-chip screen, and was able to consistently estimate the distance of an object in its 15° field of view up to 4 m away (Fig. 5).

## III. Results

After assembling and software setup, the prototype was ready to go. We achieved successful

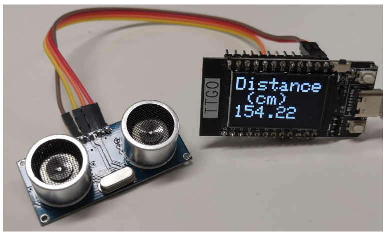


Figure 5: Ultrasound rangefinder and ESP32 micro-controller

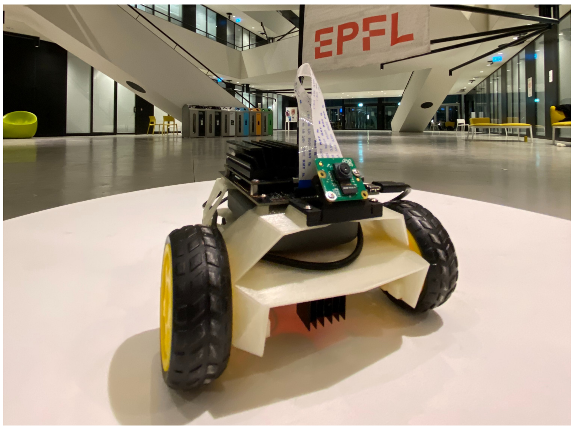


Figure 6: Our prototype on display.

test runs displaying an adequate inference speed (more than 10 fps despite the small size) and accuracy of the model. The robot is able to follow a human subject accurately, with only the visual input provided by the camera, stopping at the predefined distance.

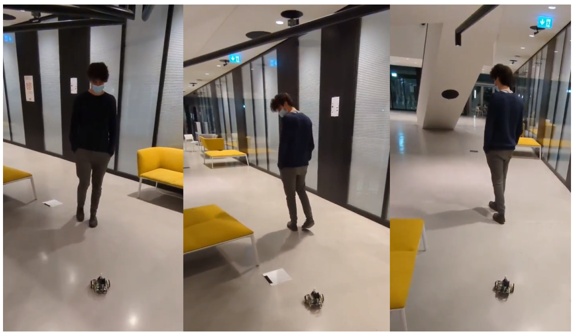

Figure 7: Our prototype following a subject.

Videos of the robot in action were recorded and uploaded on a Google Drive. We also worked on a small UI for presentation purposes that indicates the chosen input provided to the motor driver according to the video feed. An important aspect of our research was the focus on accessibility and ease of use. We believe that relying solely on computer vision makes the solution cheaper and more accessible by requiring less sensors: the total cost of the device was around 120 $, including the hardware which represents the robotic wheelchair.

### A. *Future outlook*

As previously outlined at its core our approach consists in a computer vision system for object recognition with hard coded heuristic for the path planning and navigation. Here we outline limitations and advantages of our prototype.

*1) Limitations:*

- The computer-vision-based subject's distance estimation may be affected by jittering in particularly challenging situations, and it may be less accurate in terms of resolution and robustness than other sensors, such as ultrasonic, depth camera or LiDAR. Measurements from such physical sensors could be integrated, for example in situations when the estimation has an instantaneous first derivative higher than a certain threshold. In the end, however, the system has to be field tested in real-life conditions.
- In its current form, the system cannot handle a multi-subject environment. This would need further improvements in the tracking algorithm, in order for it to remember features of the target person being followed.

*2) Advantages:*

- Despite a computer-vision-only approach, our system has shown remarkably high accuracy in the subject following task, as well as quick reaction time to the target's changes of direction, thanks to the fast GPU inference of the Jetson Nano Board. The following distance can also be changed according to preference.
- Subject detection and distance estimation remained very robust despite the presence of multiple objects in the environment, such as chairs and a sofa, mimicking a typical at-home rehabilitation setting (as our video shows).
- The computer vision pipeline for subject recognition and distance estimation could be easily fine-tuned with a transfer learning process [24], by training a part of the neural network specifically on videos of patients performing SCI rehabilitation. This should enable to reach even better performance and increase the robustness in the predictions. As of now, our machine vision neural net might encounter some trouble in recognizing subjects from the prototype low viewpoint or in detecting people with crutches.
- Our system represents a solution that is both cheaper and easier to implement than using other sensors, by just relying on a camera. We developed our prototype with an easy-to-service modular design in mind, that could be potentially integrated with different wheelchairs with minimal effort.

## References


[1] Eduardo Martin Moraud et al. "Mechanisms Underlying the Neuromodulation of Spinal Circuits for Correcting Gait and Balance Deficits after Spinal Cord Injury". en. In: *Neuron* 89.4 (Feb. 2016), pp. 814–828. ISSN: 08966273. DOI: 10.1016/j.neuron.2016.01.009. URL: https://linkinghub.elsevier.com/retrieve/pii/S0896627316000106 (visited on 01/03/2022).

[2] Claudia A. Angeli et al. "Altering spinal cord excitability enables voluntary movements after chronic complete paralysis in humans". en. In: *Brain* 137.5 (May 2014), pp. 1394–1409. ISSN: 1460-2156, 0006-8950. DOI: 10.1093/brain/awu038. URL: https://academic.oup.com/brain/article-lookup/doi/10.1093/brain/awu038 (visited on 01/03/2022).

[3] Jocelyne Bloch. *STIMO: Epidural Electrical Stimulation (EES) with robot-assisted rehabilitation in patients with Spinal Cord Injury*. en. [Online] Available: "https://clinicaltrials.gov/ct2/show/study/NCT02936453" [Accessed: 03/01/2022].

[4] Kim Acosta. *Best Electric Wheelchairs Of 2022*. 2022. URL: https://www.forbes.com/health/healthy-aging/best-electric-wheelchairs/.

[5] Jesse Leaman and Hung Manh La. "A Comprehensive Review of Smart Wheelchairs: Past, Present, and Future". In: *IEEE Transactions on Human-Machine Systems* 47.4 (2017), pp. 486–499. DOI: 10.1109/thms.2017.2706727.

[6] 2022. URL: https://doog-inc.com/en/type-ride/.

[7] Yoshinori Kobayashi et al. "Robotic wheelchair moving with caregiver collaboratively depending on circumstances". In: *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems - CHI EA '11* (2011). DOI: 10.1145/1979742.1979894.

[8] Shamim Al Mamun et al. "Companion Following Robotic Wheelchair with Bus Boarding Capabilities". In: *2018 Joint 7th International Conference on Informatics, Electronics Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision Pattern Recognition (icIVPR)* (2018). DOI: 10.1109/iciev.2018.8641059.

[9] Daniel Leite, Karla Figueiredo, and Marley Vellasco. "Prototype of Robotic Device for Mobility Assistance for the Elderly in Urban Environments". In: *Sensors* 20.11 (2020). DOI: 10.3390/s20113056.

[10] Matthew Aquilina, Marvin Bugeja, and Simon Fabri. "Smart Wheelchairs: Using Robotics to Bridge the Gap between Prototypes and Cost-effective Set-ups". In: *Proceedings of the 16th International Conference on Informatics in Control, Automation and Robotics* (2019). DOI: 10.5220/0007796801810189.

[11] Theja Pingali, Edward Lemaire, and Natalie Baddour. "Ultrasonic Tethering to Enable Side-by-Side Following for Powered Wheelchairs". In: *Sensors* 19.1 (2018), p. 109. DOI: 10.3390/s19010109.

[12] Zhiwei Xiong et al. "Computational Depth Sensing : Toward high-performance commodity depth cameras". In: *IEEE Signal Processing Magazine* 34.3 (2017), pp. 55–68. DOI: 10.1109/msp.2017.2669347.

[13] *D455 depth camera by INTEL*. 2022. URL: https://www.intelrealsense.com/depth-camera-d455/.

[14] Syamsiar Kautsar et al. "A Simple Algorithm for Person-Following Robot Control with Differential Wheeled based on Depth Camera". In: *2019 International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE)* (2019). DOI: 10.1109/icomitee.2019.8921165.

[15] Yue Ming et al. "Deep learning for monocular depth estimation: A review". In: *Neurocomputing* 438 (2021), pp. 14–33. DOI: 10.1016/j.neucom.2020.12.089.

[16] Hang Yan, Qi Shan, and Yasutaka Furukawa. "RIDI: Robust IMU Double Integration". In: *CoRR* abs/1712.09004 (2017). arXiv: 1712.09004. URL: http://arxiv.org/abs/1712.09004.

[17] Manon Kok, Jeroen D. Hol, and Thomas B. Schön. "Using Inertial Sensors for Position and Orientation Estimation". In: *Foundations and Trends® in Signal Processing* 11.1-2 (2017), pp. 1–153. ISSN: 1932-8346. DOI: 10.1561/2000000094. URL: http://dx.doi.org/10.1561/2000000094.

[18] Wikipedia site. *Inertial Navigation System*. https://en.wikipedia.org/wiki/Inertial_navigation_system.

[19] Daquan Feng et al. "Kalman-Filter-Based Integration of IMU and UWB for High-Accuracy Indoor Positioning and Navigation". In: *IEEE Internet of Things Journal* 7.4 (2020), pp. 3133–3146. DOI: 10.1109/JIOT.2020.2965115.

[20] Sachini Herath, Hang Yan, and Yasutaka Furukawa. "RoNIN: Robust Neural Inertial Navigation in the Wild: Benchmark, Evaluations, New Methods". In: *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, May 2020, pp. 3146–3152. ISBN: 9781728173955. DOI: 10.1109/ICRA40945.2020.9196860. URL: https://ieeexplore.ieee.org/document/9196860/.

[21] Camillo Lugaresi et al. *MediaPipe: A Framework for Building Perception Pipelines*. 2019. arXiv: 1906.08172 [cs.DC].

[22] Valentin Bazarevsky et al. *BlazePose: On-device Real-time Body Pose tracking*. 2020. arXiv: 2006.10204 [cs.CV].

[23] *Yolo v5 by Ultralytics*. 2022. URL: https://github.com/ultralytics/yolov5.

[24] Karl R. Weiss, Taghi M. Khoshgoftaar, and Dingding Wang. "A survey of transfer learning". In: *Journal of Big Data* 3 (2016), pp. 1–40.