Hands-Free Wheelchair Control Using Vision-Based Intention Estimation

Proceedings of the 8th Jc-IFToMM International Symposium (2025)
The 31st Jc-IFToMM Symposium on Theory of Machines and Mechanisms
21-22 June 2025

Hands-Free Wheelchair Control Using Vision-Based Intention Estimation

Lorenzo Baglieri
Department of Mechanical and Aerospace Engineering, Politecnico di Torino, Italy,
lorenzo.baglieri@polito.it

Daisuke Matsuura
Department of Mechanical Engineering, Institute of Science Tokyo, Japan,
matsuura.d.aa@m.titech.ac.jp

Tsune Kobayashi
JTEKT Corporation and Dept. of Mech. Eng. Tokyo Institute of Technology, Japan,
kobayashi.t.cs@m.titech.ac.jp

Giuseppe Quaglia
Department of Mechanical and Aerospace Engineering, Politecnico di Torino, Italy,
giuseppe.quaglia@polito.it

Abstract
This paper presents a hands-free interface for controlling an omnidirectional electric wheelchair, designed to enhance mobility for individuals with motor impairments. The system utilizes multiple cameras to track face and torso movements, enabling natural and non-invasive driving system. A vision-based Intention Estimator integrates input from both the rider and caregiver, allowing for shared control. The proposed method ensures smooth navigation in confined indoor spaces, improving user experience. Experimental results demonstrate the feasibility of the approach, though challenges remain, particularly in addressing lighting conditions that affect facial landmark detection. Future work will focus on enhancing drivability and solving illumination issues to ensure reliable operation.
Keywords: Hands-free wheelchair, HMI, Vision-based control, Intention estimation, Face tracking, Service Robots, Disability.

1. Introduction

The rise of an aging society presents complex welfare challenges, including an increasing number of people with mobility issues and a growing demand for caregivers. By 2020, over 120 million people in Europe had mobility limitations[1], with 10% struggling to use traditional power wheelchairs, another 10% having problems steering without help, and 40% struggling with steering [2]. To address this, researchers have explored alternative propulsion systems [3], mobility mechanisms, and haptic interfaces for improved wheelchair operability [4].
However, most existing solutions still require manual input from users or caregivers. This research aims to develop an innovative human-machine interface (HMI) that enables hands-free wheelchair control, enhancing users' daily activities. The proposed HMI will be integrated into an electric, omnidirectional wheelchair designed for better navigation in confined indoor spaces [5].
Various HMI approaches, such as EEG [6], eye-tracking [7], tactile sensors [8], IMU [9], and tongue control [10], have been explored to improve wheelchair usability. The ultimate goal is to enhance mobility and quality of life, particularly for tetraplegic users, by adapting the HMI to individual needs while maintaining assisted navigation.

The proposed HMI estimates the intentions of both the rider and caregiver, enabling shared control of the wheelchair. A rule-based system prioritizes user actions, allowing one person to drive while the other makes adjustments. A vision system measuring face and torso movements will generate control commands, ensuring a natural and non-invasive experience. The project aligns with the UN Sustainable Development Goal 3: Good Health and Well-Being.

1.1 Wheelchair's components and working principles

The proposed system is an omnidirectional electric wheelchair with a hands-free interface, enabling natural side-by-side movement with a caregiver. Figure 1 illustrates the wheelchair's subsystems. The system receives input from multiple cameras (Rider CamR, Rider CamL, and the Caregiver Camera), which capture the face and torso movements of both individuals. These movements are represented within the rider's head (rh) and rider torso (rt) reference frames. A host computer handles the data acquisition and processing in three main steps: Face and Torso Detection, Intention Estimation, and Motion Planning. First, the system identifies facial and torso movements from the captured images. Then, the "Intention Estimator" determines the wheelchair's overall movement direction by integrating input from both the rider and caregiver. Lastly, the Motion Planner generates the wheelchair motion commands. A microcontroller then translates the planned trajectory into motor commands. The paper focuses on face and torso detection and intention estimation, particularly in scenarios where only the rider's movements are considered.

2. Face and torso detection

2.1 Explanation of process variables

Table 1 explains all the References Systems (R.F.s), the positions, the Rotation Matrix (R.M) and the Transformation Matrix (T.M.) that are taken into consideration for the calibration and for measuring the face pose.

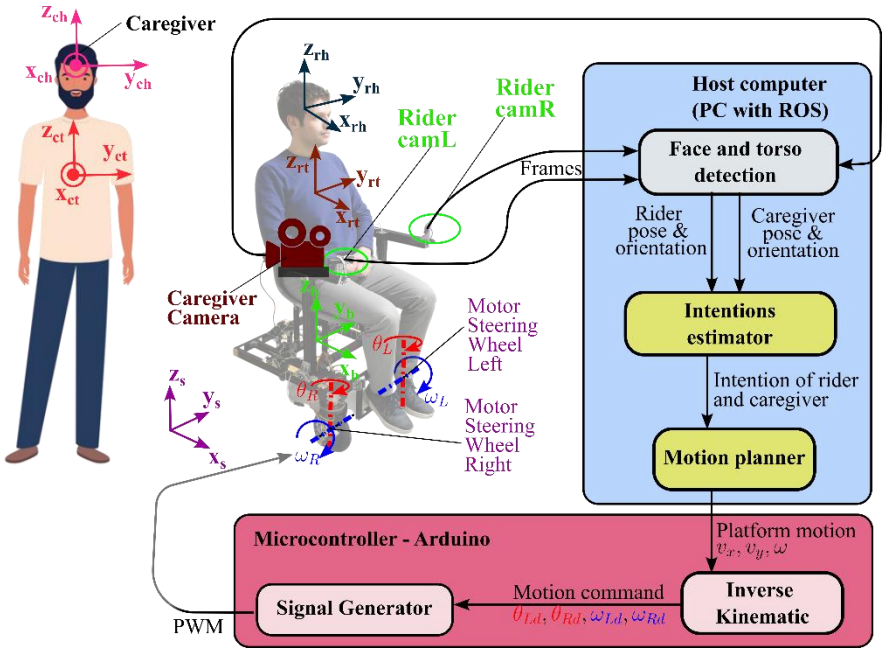

Figure 1 hands-free side-by-side wheelchair working principle with subsystems

Table 1 symbols used during the calibration and the face detection

General variables	
$camL$	Reference System of the left camera
$camR$	Reference System of the right camera
T_{camL}^{camR}	Transformation matrix from $camR$ to $camL$
Calibration process variables	
M	Reference system of the calibration markers
p_{camL}^{M}	Position of the calibration marker from camL
p_{camR}^{M}	Position of the calibration marker from camR
Stereovision process variables	
$p_{camL}^{l,n}$	Position of a nth landmark from camL
$p_{camR}^{l,n}$	Position of a nth landmark from camR
A_{camL}	Projection matrix of rh in camL RF
A_{camR}	Projection matrix of rh in camR RF
$x_{camL}^{l,n}$	Pixel coordinates of a nth landmark in camL
$x_{camR}^{l,n}$	Pixel coordinates of a nth landmark in camR
T_{camL}^{rh}	Transformation matrix from rh to $camL$

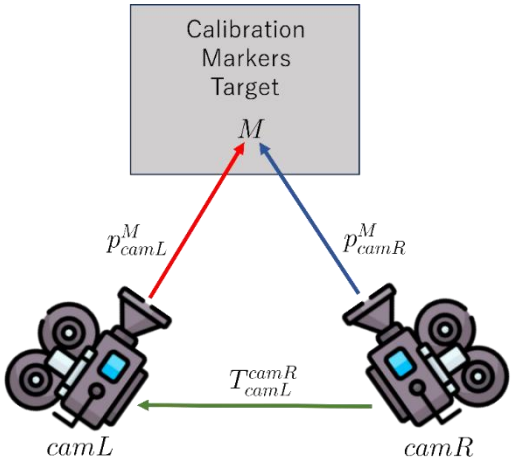

Figure 2 the two cameras are calibrated using the common marker target

With the Reference System, it is intended that three axes are orthogonal respecting the right-hand rules for the crossing product among each other. Regarding the position it is intended that a point in 3D-space in a certain R.F. is described with a 3x1 vector. With the Rotation Matrix, it is intended that the rotation around three coordinates to orient the origin R.F. like the measured R.F. and it is described with a 3x3 matrix.
The Transformation Matrix is intended as a 4x4 matrix which includes a rotation matrix and a position matrix,

$$T = \begin{bmatrix} R & \boldsymbol{p} \\ \boldsymbol{0} & 1 \end{bmatrix} \quad (1).$$

Where R is the 3x3 R.M., p is the 3x1 position vector, 0 is a 1x3 zero vector and 1 is the scale factor.

2.2 Calibration process

Figure 2 illustrates the calibration setup, which consists of two cameras (camL and camR) and a calibration target marker (M). The target features a chessboard structure embedded with 17 AruCo markers, creating a distinctive and easily recognizable pattern for comparison. The chessboard's actual square dimensions are uploaded into a software program to facilitate the calibration process.
Throughout multiple calibration trials—typically at least 20—the coordinates of the target markers ($\boldsymbol{p_{camL}^{M}}$ and $\boldsymbol{p_{camR}^{M}}$) are recorded from both the left and right cameras. By capturing these markers from the cameras' perspectives, the acquired images can be compared to the preloaded reference in

the software. This comparison enables the evaluation of intrinsic camera parameters, including focal length f_x, f_y, principal point c_x, c_y, and skew factor s. These parameters collectively form the intrinsic matrix for each camera.

$$K = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (2).$$

Additionally, by combining the two viewpoints from both cameras and knowing the actual size of a common object, triangulation can be performed to estimate the calibration matrix—specifically, the transformation matrix from camR to camL, denoted as T_{camL}^{camR}. This matrix is derived by minimizing the error across all triangulated measurements captured by both cameras. It plays a crucial role in computing the extrinsic matrix of each camera.

The extrinsic matrix is a 3x4 matrix that consists of the rotation matrix R, which defines the relationship between the world reference frame and the camera reference frame. The last column represents the position vector $\boldsymbol{p}$ of the camera reference frame relative to the world origin:

$$E = [R \quad \boldsymbol{p}] \quad (3).$$

For simplicity, the world origin is set at the camL reference frame. As a result, for the left camera, the rotation matrix R is the identity matrix, and the position vector $\boldsymbol{p}$ is a zero vector. For the right camera, both R and $\boldsymbol{p}$ are derived from the transformation matrix T_{camL}^{camR}, obtained during calibration. Finally, the two projection matrices, A_{camR} and A_{camL}, , are computed by multiplying the intrinsic and extrinsic matrices:

$$A_{camR} = K_{camR}E_{camR}, \; A_{camL} = K_{camL}E_{camL} \quad (4).$$

2.3 Stereovision: face pose detection

Figure 3 illustrates the configuration of the cameras and the rider when capturing the rider's movements. The five purple crosses on the rider's face indicate the key landmarks to be detected: both eyes, the nose, and the two mouth corners. These landmarks are essential for determining the face's pose relative to the camera's reference frame. This section focuses on solving the triangulation problem to determine the position of the landmarks using one of the two cameras.

The process begins with the pixel coordinates of the 5-zllandmarks determined above as captured by each camera, denoted as $x_{camL}^{l,n}$ and $x_{camR}^{l,n}$. As depicted in Figure 3, these pixel coordinates represent the projection of the 3D landmark positions onto the camera's image plane. This projection is obtained by multiplying the projection matrix by the position vector:

$$\begin{aligned} x_{canL}^{l,n} &= A_{camL}p_{camL}^{l,n}, \\ x_{canR}^{l,n} &= A_{camR}p_{camR}^{l,n} \end{aligned} \quad (5).$$

In equation 5, the unknowns are the position vectors $p_{camR}^{l,n}$ and $p_{camL}^{l,n}$. By utilizing the transformation matrix between the two cameras, the system can be simplified to express the unknowns in terms of a single variable, making it solvable:

$$\begin{bmatrix} A_{camL} \\ A_{camR}T_{camR}^{camL} \end{bmatrix} p_{camL}^{l,n} = \begin{bmatrix} x_{canL}^{l,n} \\ x_{canR}^{l,n} \end{bmatrix} \quad (6).$$

This system involves a 4x3 matrix, meaning the position vector is represented as a four-element vector, while the pixel coordinates form a three-element vector. The additional element in the pixel coordinates corresponds to perspective division, which acts as a scaling factor based on the object's

distance from the camera. Similarly, the fourth element in the position vector serves as a scale factor but is always set to 1. From equation 6, the landmark positions in the camL reference frame can be determined. These positions are then used in the next section to compute the orientation and position of the rider's head, which is crucial for estimating their intended movement.

3. Intention estimator's logic

3.1 Face Reference frame estimation

This section explains how the rider's head reference frame is determined using landmark positions obtained by both cameras in the previous section. As shown in Figure 3, the reference frame is aligned with key facial landmarks. The z-axis extends from the nose marker through the midpoint between the eyes. The y-axis originates from the nose marker and points toward the midpoint between the left eye and the left mouth corner. It is adjusted to ensure perpendicularity to the z-axis:

$$\overrightarrow{y_{rh}} = \overrightarrow{y_{rh}} - (\overrightarrow{y_{rh}} \cdot \overrightarrow{z_{rh}})\overrightarrow{z_{rh}} \quad (7).$$

The x-axis is obtained as the cross-product of the y- and z-axes.

3.2 Velocity Command

The rider's head reference frame serves as the basis for estimating motion intention. A transformation matrix, T^{rh}_{camL}, describes the head's position and orientation relative to camL. However, the primary focus is tracking head movements from a defined rest position (aligned with the spinal column). This requires selecting a transformation matrix $T^{r,0}_{camL}$ for the default head position.

Figure 4 illustrates how head movements are converted into velocity commands. The rider's head moves from the rest position *rh,0* to a new position *rh,a*, where the distance d_r from the reference center and the angle ϕ_r w.r.t the $x_{rh,0}$ are computed:

$$d_r = \sqrt{{x_{rh,a}}^2 + {y_{rh,a}}^2},$$
$$\phi_r = \arcsin\left({y_{rh,a}}/{d_r}\right) \text{ if } d_r \neq 0, \quad \phi_r = 0 \text{ if } d_r = 0 \quad (8).$$

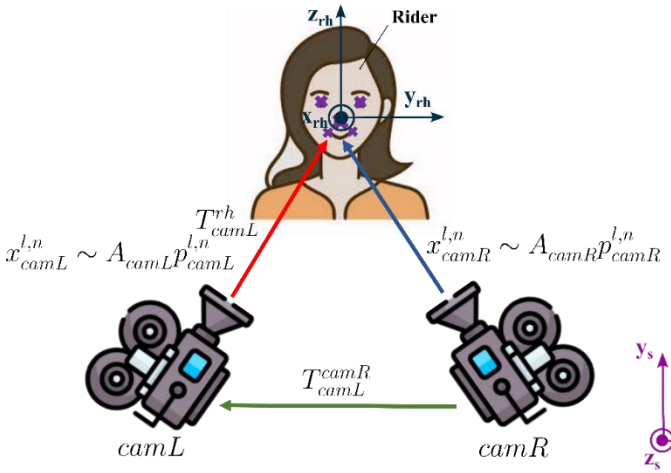

Figure 3 Cameras perform stereovision and catch head pose

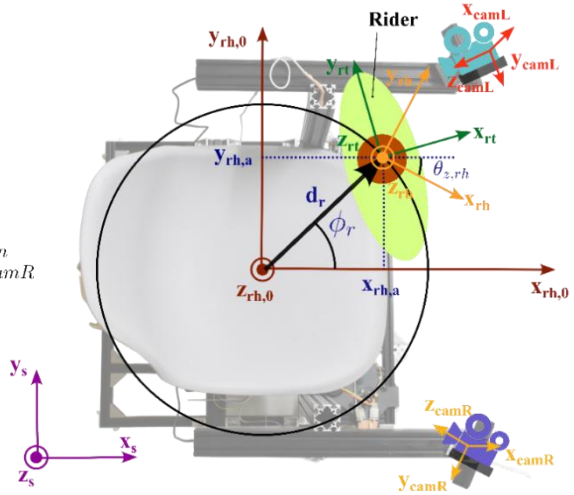

Figure 4 how the rider's movements from rest position generate reference signals that moves the wheelchair

A 5th-order polynomial function (as described by Tagliavini et al.) [11] is used to transform d_r and $\theta_{z,rh}$ into linear velocity v and angular velocity ω_z. This ensures smooth wheelchair movement by linking the deadband region to the linear control response. Table 2 summarizes the variables used to convert input signals into velocity commands. The velocity components in the Cartesian frame are calculated as:

$$v_x = v\cos(\Phi_r), \qquad v_y = v\sin(\Phi_r) \tag{9}$$

Before sending velocity commands to the inverse kinematics routine, they are processed by a Motion Planner, which smooths them using an exponential filter:

$$\begin{aligned} s_t &= \alpha x_t + (1-\alpha)x_{t-1}, \quad \text{with } t > 0 \\ s_t &= x_t, \quad \text{with } t = 0 \end{aligned} \tag{10}$$

Where s_t is the smoothed variable, x_t is the actual variable, x_{t-1} is the previous variable and α the smoothing value (in this case 0.6)

4. Experimental Tests

Figure 5a illustrates the experimental setup, consisting of an empty room with cones marking the test area. A monitor provided real-time feedback from the wheelchair's PC. A motion capture system, using four symmetrically placed cameras (two on a tripod and two on the opposite side), tracked the wheelchair's movement. The experiment evaluated the system's ability to move straight and rotate. The rider, seated in the wheelchair, used hands-free control to navigate along the perimeter of the cones. Due to some back-light problem it is also showed the shield putted in backside position to solve the issue.

Table 2 Values and parameter utilized in the velocity input generation

Variable	Value	Description
d_r	Measured data	Distance of the rider reference frame from the wheelchair reference frame in polar coordinates. Input of the polynomial function for linear velocity
Φ_r	Measured data	Direction of the rider reference frame from the wheelchair reference frame in polar coordinates.
$\theta_{z,rh}$	Measured data	Rotation of the head of the rider. Input of the polynomial function for angular velocity
v_{max}	0.3 m/s	Saturation for linear velocity.
$\omega_{z,max}$	0.5 rad/s	Saturation for angular velocity
d_{min}	4 cm	Dead band for the head motion to obtain a linear translation
θ_{min}	20°	Dead band for the head motion to obtain a rotation
d_{change}	6 cm	Value for the head motion switch from polynomial to linear behaviour
θ_{change}	40°	Value for the head rotation switch from polynomial to linear behaviour
v_{change}	0.1 m/s	Linear velocity value switch from polynomial to linear behaviour.
$\omega_{z,change}$	0.1 rad/s	Angular velocity value switch from polynomial to linear behaviour.
d_{max}	10 cm	Maximum linear movements before entering in the saturation zone
θ_{max}	90°	Maximum angular movements before entering in the saturation zone

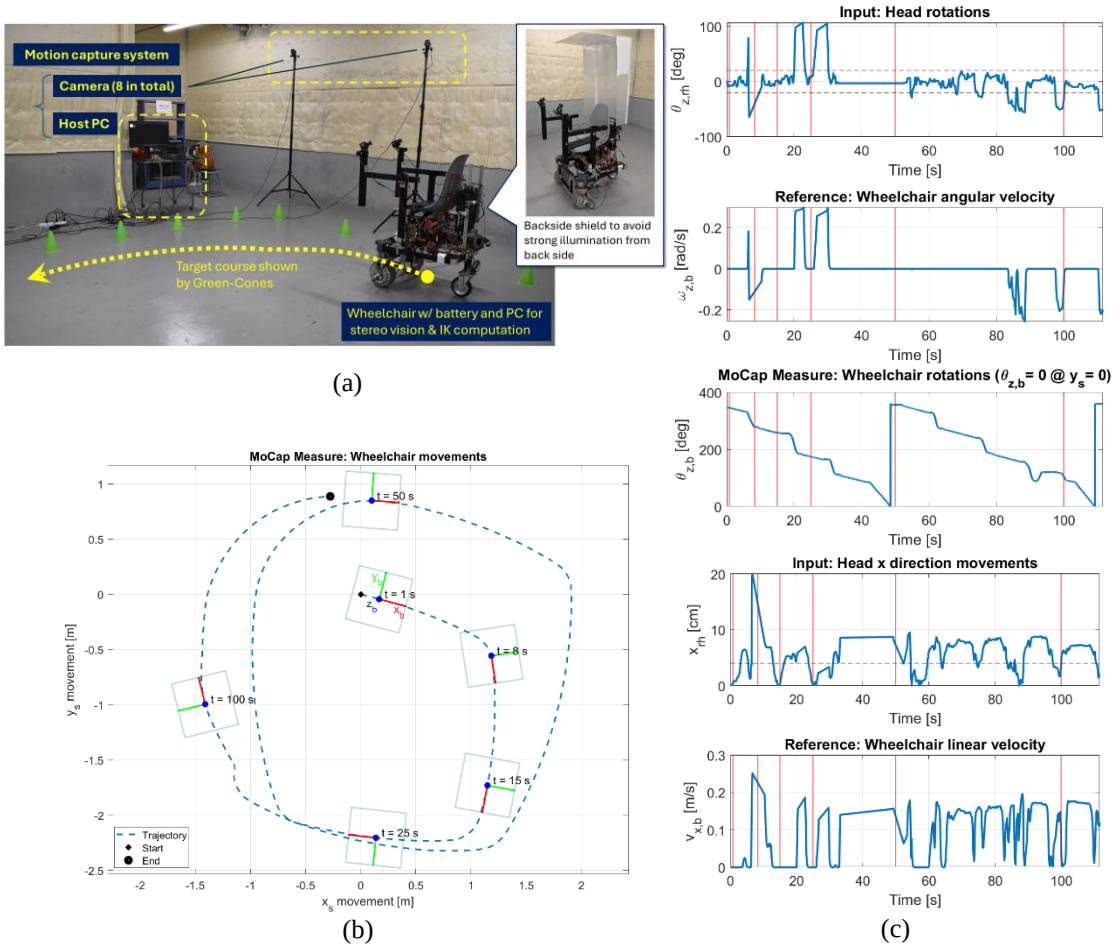

Figure 5: a - experimental setup used during the experiments. b - Trajectory and position of the wheelchair during the experiment. c – Head movements and reference velocity generated, angular position of the wheelchair.

Figure 5b presents the trajectory followed by the wheelchair during the experiment. The movements of the wheelchairs include both linear and angular velocities even at the same moment. There are highlighted some particular moments that can be found also in the head movements and reference velocity, in Figure 5c.

5. Conclusion

The papers describe a hands-free interface for wheelchair control based on the rider's head movements, which can also be adapted for shoulder movements and caregiver commands. Ongoing research focuses on shared control between the rider and caregiver, with results to be presented in future studies. While experimental findings are promising, challenges remain, particularly regarding lighting conditions affecting rider landmark detection, potentially halting wheelchair motion. Future work will aim to enhance drivability and address face illumination issues.

Acknowledgment

This publication is part of the project PNRR-NGEU which has received funding from the MUR – DM 351/2022 and has been supported by a joint research contract between JTEKT Corporation and Institute of Science Tokyo.

References

[1] 'Persons with disabilities - Employment, Social Affairs & Inclusion - European Commission'. Accessed: Feb. 26, 2024. [Online]. Available: https://ec.europa.eu/social/main.jsp?catId=1137

[2] L. Fehr, W. E. Langbein, and S. B. Skaar, 'Adequacy of power wheelchair control i terfaces for persons with severe disabilities : A clinical survey', vol. Fehr, Linda, W. Edwin Langbein, and Steven B. Skaar. 'Adequacy of power wheelchair control interfaces for persons with severe disabilities: A clinical survey.' Journal of rehabilitation research and development, 2000.

[3] G. Quaglia, E. Bonisoli, and P. Cavallone, 'A proposal of alternative propulsion system for manual wheelchair', *International Journal of Mechanics and Control*, vol. 19, pp. 33–38, Jan. 2018.

[4] L. Tagliavini and G. Quaglia, 'On the Design of MoviWE.Q: An Omnidirectional Electric-Powered Wheelchair for Indoor Mobility', in *Advances in Mechanism and Machine Science*, M. Okada, Ed., Cham: Springer Nature Switzerland, 2023, pp. 311–321. doi: 10.1007/978-3-031-45770-8_31.

[5] L. Baglieri, D. Matsuura, T. Kobayashi, and G. Quaglia, 'Development of a Mobility Platform Aiming to Achieve User-Friendly Functionality', in *Proceedings of Jc-IFToMM International Symposium Vol. 7 (2024)*, Japanese Council of IFToMM, 2024, pp. 103–110. Accessed: Sep. 23, 2024. [Online]. Available: https://www.jstage.jst.go.jp/article/jciftomm/7/0/7_103/_article/-char/ja/

[6] I. Iturrate, J. Antelis, and J. Minguez, 'Synchronous EEG brain-actuated wheelchair with automated navigation', in *2009 IEEE International Conference on Robotics and Automation*, May 2009, pp. 2318–2325. doi: 10.1109/ROBOT.2009.5152580.

[7] Y. Matsumotot, T. Ino, and T. Ogsawara, 'Development of intelligent wheelchair system with face and gaze based interface', in *Proceedings 10th IEEE International Workshop on Robot and Human Interactive Communication. ROMAN 2001 (Cat. No.01TH8591)*, Sep. 2001, pp. 262–267. doi: 10.1109/ROMAN.2001.981912.

[8] D. Matsuura, K. Kiyomoto, and T. Kobayashi, 'Development of a simple human-cooperative transportation system based on tactile information', *ICPE2022*, vol. Proceedings of the 19th International Conference on Precision Engineering.

[9] L. Baglieri, D. Matsuura, T. Kobayashi, and G. Quaglia, 'Vision Systems and IMU Signals to Design a Hand-Free Driving HMI', in *Advances in Italian Mechanism Science*, vol. 163, G. Quaglia, G. Boschetti, and G. Carbone, Eds., in Mechanisms and Machine Science, vol. 163. , Cham: Springer Nature Switzerland, 2024, pp. 283–291. doi: 10.1007/978-3-031-64553-2_33.

[10] M. E. Lund, H. V. Christiensen, H. A. Caltenco, E. R. Lontis, B. Bentsen, and L. N. S. Andreasen Struijk, 'Inductive tongue control of powered wheelchairs', in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, Aug. 2010, pp. 3361–3364. doi: 10.1109/IEMBS.2010.5627923.

[11] L. Tagliavini, L. Baglieri, G. Colucci, A. Botta, C. Visconte, and G. Quaglia, 'D.O.T. PAQUITOP, an Autonomous Mobile Manipulator for Hospital Assistance', *Electronics*, vol. 12, no. 2, Art. no. 2, Jan. 2023, doi: 10.3390/electronics12020268.