Context-Based Face Control of a Robotic Wheelchair

Arturo Escobedo, Jorge Rios-Martinez, Anne Spalanzani, Christian Laugier

Context-Based Face Control of a Robotic Wheelchair

Arturo Escobedo[1], Jorge Rios-Martinez[1], Anne Spalanzani[2], Christian Laugier[1] †
[1]INRIA Rhône-Alpes, Grenoble, France
[2]UPMF-Grenoble 2 - INRIA - LIG, France

Abstract— **In this article a method to perform semi-autonomous navigation on a wheelchair is presented, contextual information from the environment as user's habits and points of interest are employed to infere the user's desired destination in a global map. Illogical steering signals comming from the user-machine interface input are filtered out to improve the overall performance of the system. Examples using a face tracking system and voice recognition are presented. The estimation of the user's desired goal is performed employing a Bayesian network. An autonomous navigation system is used to control the wheelchair's low level navigation while the user is just concerned at pointing to the desired destination.**

Index Terms— **Shared control, Bayesian network, semi-autonomous navigation, wheelchair control, face pose.**

I. Introduction

The aging of world's population is bringing the need to provide robotic platforms capable to assist elder people to move [1], [2]. It is necessary that such transportation is reliable, safe (at least as much as a human) and comfortable. Patients and medical staff have a strong desire for the services that a smart wheelchair can offer [3]. Some users cannot use a normal power wheelchair because they lack the required motor skills, strength, or visual acuteness for some others the main reason to use such robotic systems would be the increased ease of use that they can provide. When using a robotic wheelchair, the occupant must feel that this mode of travel is tailored to its needs. The vehicle or wheelchair must meet specific needs: those of people with disabilities or reduced mobility or just those of people without disabilities but who want a service of comfort . No matter what the mobility assistance device is (car, wheelchair, walking aid...), navigation in human environments is a central problem. If one aims to develop a robotic device, many different technologies must be combine: perception, prediction, fusion, navigation, control. Even more the system must integrate a way to share the control with the user to guarantee safe navigation while avoiding frustration due to the disregarding of the user's requests by the autonomous navigation system. The operation of the platform discussed in this paper has been designed around the following requirements:

- Usability: People with motor disabilities or aging people often have problems using joysticks and other standard control devices. The system should account for this, for example by favoring the most reasonable actions when presented with an ambiguous command.
- Safety: The system should avoid collisions with both static and dynamic entities.
- Respect of social conventions: When moving, a robot may considerably disturb people around it, especially when its behavior is perceived as unsocial. Even worse, the wheelchairs passenger may be held responsible for that behavior. It is thus important to produce socially acceptable motion.

This article is structured as follows: Section II offers an overview of related works. A general description of the system architecture is presented in III while section IV focuses in our technique for estimating the user intended destination and control of the wheelchair. In section V examples of execution on our real platform are exhibited. Section VI presents conclusions about the work and perspectives.

II. Related Work

Many efforts have been made to develop robotic wheelchairs that operate similarly to an autonomous robot so that the user gives a final destination and supervises as the smart wheelchair moves to the goal [4], [5], [6]. Other smart wheelchairs limit their assistance level to collision avoidance where the user is in charge of most of the navigation task. These systems do not normally require prior knowledge of an area or any specific alterations to the environment. They require instead more planning and continuous effort on the part of the user [7], [8], [9] .

Shared control is presented in situations in which the assisting device combines the control input coming from the robot and the user in order to cooperate in the task [10], [9]

The estimation of the user's plan is a key point in many shared control tasks because it allows to the automatic controller/robot to adequate its actions to the desire of its user. Inferring the user plan is necessary whenever the interface with the user doesn't allow him to explicitly dictate this to the robot as with many popular electric wheelchair interfaces (Brain control interface, Face tracking, gaze tracking, sip and puff, joystick, etc). A robotic wheelchair can assist by taking over low-level control, requiring the user to use the input method only to give high-level directional commands like right, left, back, forward, etc.

Some methods to perform an implicit estimation of the user intention from simple inputs have been proposed in [10], [8], [9]. They model the user intent as possible trajectories to follow, then a probability distribution is maintained over the set of trajectories and finally the selection of the most

† {jesus.escobedo-cabello, anne.spalanzani, christian.laugier}@inria.fr
This work has been founded by the INRIA large scale initiative project PAL, Personally Assisted Living

probable one is done using the input from the user within a Bayesian framework.

In [11] a learned Partially Observable Markov Decision Process (POMDP) is used to estimate the intended destination into a predefined map of the environment in a high level topological manner. This means that the user focuses on driving the wheelchair from one spatial location to another without having to worry about all the low level control. The spatial representation used is based on a topological graph representation of the environment, where vertices are locations and edges represent a viable path connecting two locations as a result of performing an action. Places of interest are selected as spatial locations in the environment where the user spends significantly most of his time.

The method presented in this article to infer the user intended goal aims to build a model combining a Bayesian network approach with a topological goal based representation of the environment. In [11], [8] they both used a joystick as input device while in this work a more natural human-machine interface based on a face tracking system and voice recognition was used to command the wheelchair (V-C) while the navigation is performed using a human aware planning algorithm that avoids uncomfortable situations when the wheelchair is navigating among humans [12], [13].

III. System Architecture

When performing robotics research, often the scope of the investigation is limited to a well-defined area of the system, such as a software module which performs some type of planning, reasoning, perception, or control. However, to get a robotic system up and running for experiments, a much larger software ecosystem must exist.

Fig. 1 presents an overview of our systems architecture. It is divided into several subsystems, some of them are being developed by our team while others were taken from external sources to perform necessary tasks that are not crucial for our research domain.

- User Intentions Estimation: The user intention subsystem estimates the desired goal within the map of the environment among a list of possible predefined goals. Those locations can be previously selected by an expert caregiver, the user, or learned by the system using machine learning techniques. The user intention estimation computes the probability for each typical goal given the current position of the wheelchair and the user command and then selects the goal with the highest probability. The computation of probabilities is done using a Bayesian network approach.
- Tracking: The off-board tracker provides global information about moving obstacles which is the learning input for our motion prediction module. It is built as a conventional detect-then-track system. The tracking subsystem is also necessary to identify the interactions between people (e.g. two persons involved in a conversation).

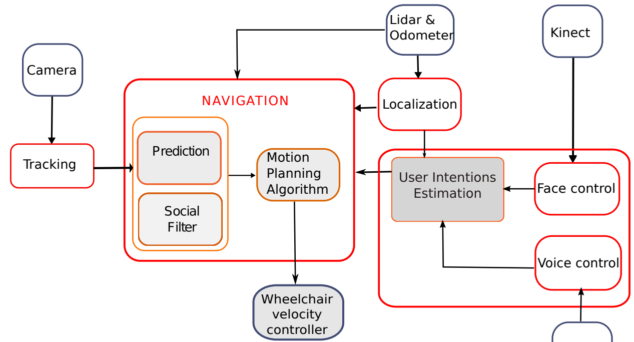


Fig. 1. System Architecture Overview

- Prediction: Processes data from the trackers and transforms it into probabilistic predictions about the configuration of the free space in the future environment. The motion prediction subsystem takes tracking data (i.e. position, orientation and velocity) and outputs grids, representing the posterior probability of the space being occupied at a given time step in the future. Prediction itself is accomplished with a Growing Hidden Markov Model (GHMM) [14] and an Extended Kalman Filter.
- Social Filter Detects social interactions and creates virtual obstacles corresponding to those interaction zones. In order to produce socially acceptable motion, it has been proposed the Social Filter, which integrates constraints inspired by social conventions in order to evaluate the risk of disturbance and take it into account when making the autonomous navigation planning. We focus on detecting and predicting conversations in the environment surrounding the wheelchair [13].
- Motion Planning: The navigation subsystem includes a laser-based localization module and a motion-planner which integrates predictions to compute safe trajectories that are fed to the execution module. The motion planner is based on RiskRRT [15], a partial motion planner which integrates motion predictions to provide safe trajectories. This algorithm was thought to operate in dynamic, uncertain environments, it supposes that the moving pedestrians detected in the environment follow typical motion patterns that are represented by Growing Hidden Markov Model (GHMM). This motion planner generates human friendly paths respecting people's personal and interaction spaces, as provided by the social filter.

IV. User Intentions Estimation System

The user intentions are modelled as topological poses into a predefined map. Those poses are defined by the user habits (places where the user spends most of his time during the day) and interesting points taken from the map of the environment as doors, desks and other facilities.

The reasoning method used is based on a Bayesian Network Fig. 2.

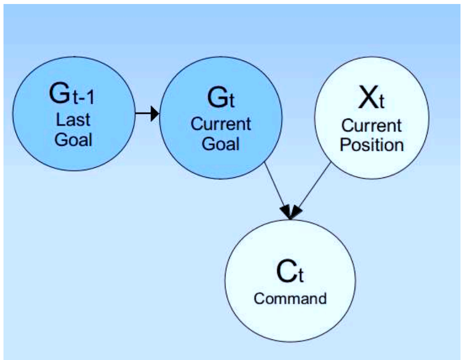


Fig. 2. The Bayesian network used to estimate the user's intended *goal* G from the current position X, the user command C and the prior knowledge of the environment

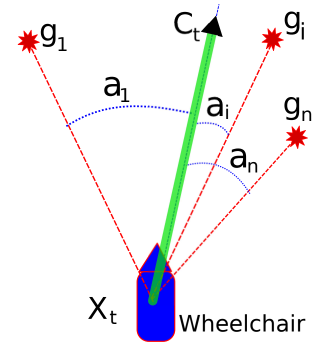


Fig. 3. The probability distribution for a given command C_t (big arrow) is proportional to the angle a_i formed with respect to each goal g_i in the environment.

The variable *command* C is dependent on the variables *goal* G and *position* X (because the user normally will point towards the desired goal). To estimate the status of the *goal* variable the command direction coming from the user-machine interface and the current user's position are applied as evidence. It is represented as $P(G_t^i|C_tX_t)$ which can be computed as a Bayesian filter:

$$P(G_t^i|C_tX_t) = P(C_t|X_t, G_t^i) \sum_j [P(G_t^i|G_{t-1}^j)* P(G_{t-1}^j|C_{t-1}X_{t-1})] \quad (1)$$

The notation G_t^i is used to express $G_t = g_i$ where g_i is one of the predefined goals in the environment as appear in Fig. 3.

$P(C_t|X_t, G_t^i)$ represents the probability of giving a command C_t when the user is located at position X_t and her goal is at position G_t^i at current time t.

Under the assumption that commands are directed straight to the goal position rather than anywhere else, as shown in Fig. 3, the non normalized probability $P'(C_t|X_t, G_t^i)$ can be computed as follows:

$$P'(C_t|X_t, G_t^i) = \frac{1-|a_i|}{\pi} \quad (2)$$

This probability is the normalized:

$$P(C_t|X_t, G_t^i) = \frac{P'(C_t|X_t, G_t^i)}{\sum_i P'(C_t|X_t, G_t^i)} \quad (3)$$

The a_i term is the angle between the command's direction and the goal's direction.

$P(G_t^i|G_{t-1}^j)$ expresses the probability that the current estimated goal is different to the last one. This works as an smoothness term which avoids abrupt changes in the estimated goal from one command to another. Caution should be taken when choosing the value of this term because large values lead to a slow response whenever the user changes of intended of goal. After practical experimentation it was defined a value of $P(G_t^i|G_{t-1}^j)$ that is 10 times bigger if the last estimated goal was the same.

$$P(G_t^i|G_{t-1}^j) = \begin{cases} 10/(n+9) & \text{if } G_t^i = G_{t-1}^j \\ 1/(n+9) & \text{else} \end{cases} \quad (4)$$

Where n is the number of possible goals in the environment.

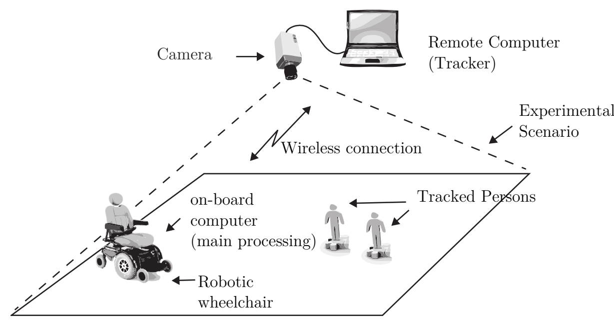


Fig. 4. Overview of the full experimental setting

V. Experimental Setup and Results

The proposed experimental setting is shown in Fig. 4. The main entity is the robotic wheelchair with all the on-board sensors and computer Fig. 5. The scenario proposed for the experiments is a human populated environment where people can be moving and interacting. Those persons can be tracked using the camera mounted on the top of the scenario. A remote computer is in charge to send the tracking information to the wheelchair.

A map of the environment is built using the LIDAR mounted on the wheelchair and some important goals are set up by hand within this map. A Kinect mounted in front of the user's face is used as input device to control the direction of the wheelchair while a microphone supplies the sound data for the voice recognition system.

A. Visual Tracking Subsystem

A camera mounted over the scenario is used to track the present people. A marker based visual tracking system is used to accurately track the position and orientation of

Fig. 5. Robotic wheelchair used for the described experiments.The mobile base includes all the electronic components and the computer in charge of the low level control of the wheelchair

special marked cards. In order to track the people in the experimental scene they wear markers as explained in [12].

B. Wheelchair

The equipment used is the robotic wheelchair shown in Fig. 5 that consists of a mobile base equipped with a seat, all the on-board electronics and different attached devices. Wheelchair's sensors consist of a LIDAR (Light Detection and Ranging) model SICK LMS-200, quadrature encoders for odometry measurements, emergency bumpers sensors (contact switches) and 1 Kinect sensor used as command input for the user-intentions subsystem presented in this work.

C. Face Control Subsystem

The user can control the robotic wheelchair by using the movements of his face, with or without using the user intention estimation system. This is accomplished by means of a face tracking system that estimates the direction of the sight of the user using data from a Microsoft's Kinect [16]. The images taken by the 2D RGB camera are used to set up a region of interest over the depth data coming from the infrared sensor. The face of the user is located using a Haar detector and a set of SWIFT features are selected over it, those features are used to perform the 2D tracking using the Lucas-Kanade method. The identification of the face pose is done by a random forest classifier which takes as input the 3D data from the Kinect sensor and gives the estimated position of the face [17], the results of the face tracking are shown in figure 6.

The wheelcair can be controlled in semi-autonomous mode employing the user intention prediction module or in manual mode in which the user is in charge of driving by her self.

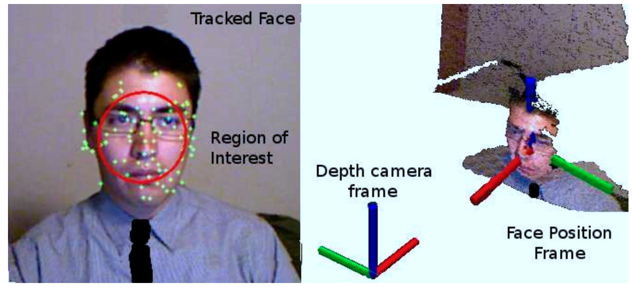


Fig. 6. In the presented approach the user drives the wheelchair by using his face. The face is tracked by processing the RGB Kinect's image (left) to set a region of interest used to estimate the face position from range data (right).

In manual mode the user controls the wheelchair's angular speed moving her head while the linear speed is controlled with voice commands as 'faster, full-speed slower, stop' explained in sec. V-D.

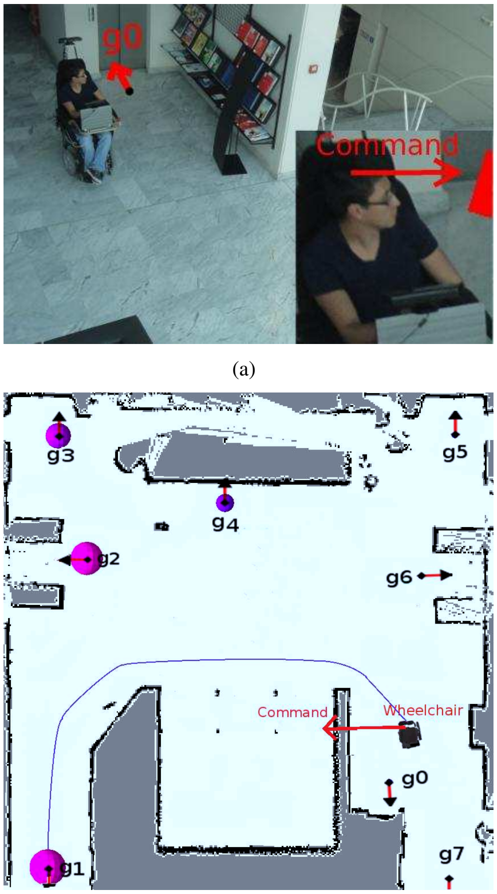
In semi-autonomous mode the user shows the direction to her desired destination facing towards it. Whenever a new command is read from the face pose estimation system. The user's intention module computes the goal with the highest posterior probability, depicted in Fig.7 (b) as the biggest sphere.The navigation module receives the map of the environment, the list of humans present in the scene and the currently estimated goal to compute the necessary trajectory to the goal as shown in Fig.7 (b). In the example the user is looking to the wall located to the left however the wheelchair instead of moving towards this wall goes to the goal that is in the other side of the hall which is the one with the highest probability in that direction. When moving the user does not have to worry about the necessary planning to avoid obstacles because the autonomous navigation system is in charge of that. In the example Fig.7, the user is seeing to the left therefore it is more probable that she is aiming to go to any of the goals in that direction.

D. Voice Control Subsystem

In the frame of the presented approach it was used the speaker-independent, continuous speech, recognition system *Pocketsphinx* [18] from Carnegie Mellon University.

This system was trained on a large speech corpus developed for acoustic phonetic research, such that appropriate acoustic models can be precomputed. Pocketsphinx allow the designer to specify a set of syntactic rules (or grammar) which specifies constraints on the ordering of words within a sentence. This grammar can be useful to enhance speech recognition quality by constraining the hypothesis space. In general a small vocabulary makes speech recognition more accurate, therefore a dictionary that focuses in a very small fixed set of tasks was considered (Go, Start, Move, Stop, Break, cancel, faster,. . .).

The voice interface is used to fulfill some lack in functionality from the head control system. The main function is to switch between manual and semi-autonomous mode by



(a)

(b)

Fig. 7. Experimental scenario. a) INRIA's hall. Possible goals are marked with small arrows and the current given command (direction of the head) is depicted with a big one. b)Results of the system in a real scenario. The size of the spheres in the environment represent the value of the computed posterior probability for each destination. The computed trajectory is also shown.

saying '*semi-autonomous*' or '*manual*' commands. In semi-autonomous mode the voice interface is used to '*stop*' the wheelchair whenever it is required. To improve the reliability of the system some other synonyms were considered to perform the same action ('*break*', '*cancel*', and a panic exclamation '*ah!*'). When manual mode is selected, voice control can be used to '*stop*' or '*start*' the movement of the wheelchair if the user says the '*stop*' command once, the wheelchair stops moving linearly but keeps turning in the direction of user's face. If a second '*stop*' command is then used the wheelchair stops completely until it receives a *move*, *start*, *forward*, *go* or *backward* command.

VI. Conclusions and Future Work

The proposed experimental platform is an ongoing effort and the results we have obtained should be considered preliminary. The most important results up to this point are:

The proposed autonomous navigation system (*RiskRRT* with *Social Filter* and *GHMM* predictor) has been extensively validated [12]. This approach consistently yields to more socially acceptable and safer navigation in dynamic environments.

The user's intention algorithm has proved to be useful to translate simple input commands (direction of the desired movement) into high level orders (the desired destination) that can be used to feed the autonomous navigation system developed by emotion team at INRIA Rhône-Alpes.

It is necessary to take into account cases where there exist ambiguity in the possible desired goals. As it can be seen in Fig. 7 even if g_1 was chosen over g_2 both of them have similar probability values (size of the sphere) In those cases when the system can not decide accurately which is the intended goal it should be requested some extra information from the user to make a better choice. Another open question is how to distinguish and omit normal head movements that does not aim to control the wheelchair.

In order to work in a non-supervised environment, the user intention algorithm must be extended combining machine learning techniques in order to add the capability to adapt autonomously to the user's specific disability.

Despite all the well known limitations of the Kinect (useful only indoors, limited field of view, etc.) using it as input device can be advantageous to assist the elderly because it provides a more natural way of interaction when giving directions (we usually look to where we want to go) so they can be more confident when using the wheelchair.

References

[1] R. C. Simpson, "How many people would benefit from a smart wheelchair? the journal of rehabilitation research and development," *Rehabilitation Research and Development*, vol. 45, pp. 53–72, 2008.

[2] E. G. S. M. Lachmann and A. Wrench, "Assessment of need for special seating and/or electronic control systems for wheelchairs among people with severe physical disabilities," *Clinical Rehabilitation*, vol. 7, pp. 151–156, 1993.

[3] L. Fehr, W. E. Langbein, and S. B. Skaar, "Adequacy of power wheelchair control interfaces for persons with severe disabilities: a clinical survey." *Journal of rehabilitation research and development*, vol. 37, no. 3, pp. 353–60, 2000.

[4] S. P. Levine, D. a Bell, L. a Jaros, R. C. Simpson, Y. Koren, and J. Borenstein, "The NavChair Assistive Wheelchair Navigation System." *IEEE transactions on rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, vol. 7, no. 4, pp. 443–51, Dec. 1999.

[5] T. Kollar, S. Tellex, D. Roy, and N. Roy, "Grounding Verbs of Motion in Natural Language Commands to Robots," in *International Symposium on Experimental Robotics*, 2010.

[6] s. . M. y. . . Hemachandra Madhawa, title = Narrated Guided Tour Following and Interpretation by an Autonomous Wheelchair, Ph.D. dissertation.

[7] G. Vanacker, J. del R Millán, E. Lew, P. W. Ferrez, F. G. Moles, J. Philips, H. Van Brussel, and M. Nuttin, "Context-based filtering for assisted brain-actuated wheelchair driving." *Computational intelligence and neuroscience*, vol. 2007, p. 25130, Jan. 2007.

[8] E. Demeester, A. Huntemann, D. Vanhooydonck, G. Vanacker, A. Degeest, H. Brussel, and M. Nuttin, "Bayesian Estimation of Wheelchair Driver Intents: Modeling Intents as Geometric Paths Tracked by the Driver," *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, no. Iros, pp. 5775–5780, Oct. 2006.

[9] E. Demeester, a. Huntemann, J. del R. Millan, and H. Van Brussel, "Bayesian plan recognition for Brain-Computer Interfaces," *2009 IEEE International Conference on Robotics and Automation*, pp. 653–658, May 2009.

[10] E. Demeester, M. Nuttin, D. Vanhooydonck, and H. Van Brussel, "Assessing the user's intent using bayes' rule: application to wheelchair control," in *Proceedings of the 1st International Workshop on Advanced in Service Robotics*, 2003, pp. 117–124.

[11] T. Taha, J. V. Miro, and G. Dissanayake, "POMDP-based long-term user intention prediction for wheelchair navigation," *2008 IEEE International Conference on Robotics and Automation*, pp. 3920–3925, May 2008.

[12] P. S. D. Vasquez, "Human aware navigation for assistive robotics," in *13th International Symposium on Experimental Robotics*, 2012.

[13] J. Rios-Martinez, A. Spalanzani, and C. Laugier, "Understanding human interaction for probabilistic autonomous navigation using Risk-RRT approach," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sept. 2011, pp. 2014–2019.

[14] D. Vasquez, T. Fraichard, and C. Laugier, "Growing Hidden Markov Models: a Tool for Incremental Learning and Prediction of Motion," *International Journal of Robotics Research*, vol. 28, no. 11–12, pp. 1486–1506, 2009.

[15] C. Fulgenzi, A. Spalanzani, C. Laugier, and C. Tay, "Risk based motion planning and navigation in uncertain dynamic environment," INRIA," Research Report, 2010. [Online]. Available: http://hal.inria.fr/inria-00526601/en/

[16] "Pi face tracker webpage," 2012. [Online]. Available: http://www.pirobot.org

[17] G. Fanelli, J. Gall, and L. Van Gool, "Real time head pose estimation with random regression forests," *Computer Vision and Pattern Recognition*, 2011.

[18] "Cmu sphinx webpage," 2012. [Online]. Available: http://cmusphinx.sourceforge.net