Denoising of human motion data for a 3D gesture recognition system for a two-wheeled inverted pendulum robot

Jorge Solis, Fernanda Amaral Melo
Department for Engineering and Physics, Karlstad University, Universitetsgan, 2, 65188, Karlstad, Sweden

Abstract
The development of a human-friendly robot vehicle for carrying-medical tools (iCAR) was presented iCAR is composed by a statically stable mobile robot vehicle with on board sensors. A time-delay neural network was designed and implemented for the 3D gesture recognition. A successful gesture recognition percentage of 91% was obtained for 8 gestures. Due to the fact the 3D gesture recognition system implemented into the iCAR has been designed for a statically stable mobile platform; in this paper, the authors focused in denoising the human motion data for a dynamically stable mobile platform instead. We have proposed a denoising algorithm which takes into consideration the physical constrains of the human motion and it adds a little time-delay in the filtered motion. In particular, a probabilistic analysis was taken into consideration, where the difference between the positions of two consecutive frames was removed if it is out of a trust interval. For this purpose, the mean average and standard deviation were calculated. Its values were used to define the trust interval at each frame, where the constant k that multiplies the standard deviation was defined empirically. The probabilistic analysis was applied to the original data and an increase of 2.9% for the success recognition ratio (from 92.29% to 95.20%) was obtained with a negligible raise on the processing time of the system

Keywords: Gesture recognition, mobile robot.

1 Introduction

In industrialized countries, regional disparities in healthcare and welfare services, increased medical expense caused by aging societies and shortages of medical staff have become serious problems. In Sweden, it is expected that about 25% of Swedes population will be older than 65 years old by 2060[1]. For this purpose**, robot** technology (RT) is expected to be an important key to find solutions to these problems. In particular, roboticists have been developing assistive robots for health care and welfare applications to improve the security, independence and quality of the elderly so they can stay in their homes longer [2], [3]. As a consequence, this opens the opportunity to free up time so that medical staff can provide care for the patients who really may need a human support [4], [5].

In particular, different walking-aid robots have been proposed during the last decades [6]-[12]. In particular, the walking-aid robots can be classified in two main groups according to the mobility factor [6]: active-type walkers driven by a servo motor [8]-[9] and passive-type walkers driven by a servo brake [10]-[11]. Spenko [7] proposed the PAMM system together with a smart cane robot with a relative small size but the maneuverability is compromised by the cost. Fukuda [6] introduced an intelligent cane robot consisting of a stick, a group of sensors for recognizing the user's intentions and an omnidirectional mobile platform. However, the physical support is provided by means of a fixed length and stiffness aluminum stick and cannot be customized depending on the needs of the specific user (required level of physical support depending on the undergoing daily activity) and environmental conditions (indoor or outdoor). From those researches; a special focus has been done in terms to increase the level of multimodal interaction, sensing and control to facilitate the perception of the environment for a better guidance and provide a static physical support to avoid falling down. However, dynamic physical support (e.g. by means of a variable stiffness mechanism), the adaptability to the user/task needs (e.g. human-in-the-loop control), and the multipurpose design concept (e.g. provide support to the elderly and/or care gives) have been scarcely studied.

For this purpose, at Karlstad University, the author [12] has proposed to incorporate and develop the concept of robotic human science introduced by Takanishi [13] and to enable its application in a multipurpose human-friendly robot for assisting elderly persons (e.g. walking-support) as well as assisting care givers (e.g. carrying-medical tools). On the one hand, models of human motor control and learning, as well as cognition should allow creating truly interactive human-friendly robots; on the other hand, modelling human-friendly robots allows the development for reverse engineering and scientific understanding of human motion, perception and cognition. The focus of the research is embodying perceptual (sensing the incoming stimuli), cognitive (processing the incoming stimuli) and bodily-kinaesthetic (response to the incoming stimuli as a result of combining perceptual and motor skills) capabilities.

Due to the complexity of the proposed research, two assistive robots vehicles are under development at Karlstad University: an intelligent carrying-medical tools robot vehicle [14] and a human-friendly assistive robot vehicle for supporting physically elderly [15]. The development of a human-friendly robot vehicle for carrying-medical tools (*iCAR*) was presented in (14). *iCAR* is composed by a mobile robot vehicle with on board sensors, and two-actuated and four-passive wheels (Figure 1). A simplified fuzzy logic controller has been implemented for the navigation control. The iCAR was able to correct its posture in order to follow the subject after a transitional period of time (about 4 seconds). On the other hand, a time-delay neural network (TDNN) was designed and implemented for the 3D gesture recognition (8 gestures were defined). A successful gesture recognition percentage of 91% was obtained [14].

Figure 1.Human-friendly robot vehicle for carrying-medical tools (iCAR) developed at Karlstad University.

Figure 2.Human-friendly walking assisting robot vehicle (*hWALK*) developed at Karlstad University.

The development of a human-friendly walking assistive robot vehicle (*hWALK*) was presented in [15]. The *hWALK* is composed by a two-wheeled inverted pendulum mobile robot, a 3-DOFs desktop haptic interface, a mobile computer and a wireless module for communication purposes (Figure 2). In order to improve the velocity control of the two-wheeled inverted pendulum of the *hWALK*, a LQR has been added as compensator for the wheel angular velocity to the existent PID controller (16). On the other hand, an algorithm based on computing the integral of the motor measured current was proposed in order to detect a ramp in order to cope with inclined surfaces. From the experimental results, the errors in the linear and velocity displacement were reduced based on the proposed control. On the other hand, the feasibility to detect the uphill was verified [16].

In this paper, both the *iCAR* and *hWALK* have been integrated into a multiupose human-friendly assistive robot vehicle (*KFriend*) for providing assisting care givers and supporting physically elderly. In particular, due to the fact the 3D gesture recognition system implemented into the *iCAR* has been designed for a statically stable mobile platform (the RGB-D camera was mounted into a mobile platform composed by two-actuated and four-passive wheels), the authors focused in denoising the human motion data of the 3D gesture recognition system for a dynamically stable mobile platform (the RGB-D camera was mounted into a two.-wheeled inverted pendulum). Experiments are proposed in order to verify the performance of the denoising algorithm.

The goal of human motion denoising is not only to remove noises and outliers from motion data but also to preserve both the embedded spatio-temporal patterns of human motion and the human body structural constraints [17]. The

Figure 3. The multipurpose human-friendly assistive robot vehicle (*KFriend*) developed at Karlstad University

human motion data denoising problem is still an open research problem. The classic signal denoising methods like Gaussian low-pass and wavelet transformation are adopted to filter motion data ([18], which are low computational costly. However, these methods do not take into consideration the relationship between the different human joints and the spatio-temporal patterns. Linear time-invariant filters to denoise noisy motion data have been also proposed [19]. Such kind of dynamic filter converts a physically inconsistent motion into a consistent one. However, the relationship between the different human joints and the spatio-temporal patterns are not considered as well.

On the other hand, dynamic system-based methods represented by Kalman filter and linear dynamic system (LDS) have been proposed to discover the hidden variables and learn their dynamics [19]. However, the prediction of the current state depends on the past data, so that the filtered motion data exhibit a time-delay. More recently Lai [21] reformulated the human motion completion and denoising problems into a low-rank matrix optimization framework. However, the low-rank matrix completion theory would not be applicable under dynamic conditions.In our application, the human motion data is prone to random variations with low and high frequency components due to the dynamically stable two-wheeled inverted pendulum of the *KFriend*. For this purpose, the above denoising methods do not take into consideration the physical constrains of the human motion and may not be suitable for filtering both low and high frequency components in our real-time application. In this paper, we have proposed a denoising algorithm which takes into consideration the physical constrains of the human motion and it adds a little time-delay in the filtered motion. The proposed method has been compared against the classical Butterworth filter commonly implemented in commercial human modeling software.

This paper is organized as follows: at first, a brief overview of the multipurpose human-friendly assistive robot vehicle is described. Then, the details of the proposed denoising algorithm described. Finally, the proposed experiments to verify the performance of the proposed denoising algorithm are detailed and discussed.

2 Multipurpose Assistive Robot Vehicle

2.1 Overview

The multipurpose human-friendly assistive robot vehicle (*KFriend*) has been designed to assist care givers for transporting medical tools as well as to provide physical support to the elderly while walking. The *kFRIEND* is composed by (Figure 3): a two-wheeled inverted pendulum robot with on board general purpose controller, a RGB-D camera, a 3-DOFs haptic interface and a ZigBee wireless module.

2.2 3D Gesture recognition system

In order to command the KFriend in order to carry items from one spot in the environment to another, 8 gestures were considered (Figure 4): *engage* (for initiating the interaction), f*ollow* (for tracking and tracking the user), s*top* (for indicating the robot to stop), *come-left* (for indicating the robot to move to the left), *come-right* (for indicating the robot to move to the right), d*isengage* (for ending the interaction), *emergency-stop-h* and *emergency-stop-v* (for stopping the robot's operation).

In order to recognize the proposed gestures, the Time-Delay Neural Network (TDNN) has been considered. TDNNs are characterized by two important characteristics: a hierarchy can be constructed from a three layer arrangement allowing the formation of arbitrary nonlinear decision surfaces by means of the backpropagation and the network is capable to discover acoustic-phonetic features and the temporal relationships between them independent of position in time.

The TDNN consists of three layers input, hidden, and output layers, where the basic unit is modified by introducing time delays (D_1 to D_N). A TDNN unit has the ability to relate and compare current input to the past history events. The response of TDNN in time *t* is based on all previous inputs. A mapping performed by the TDNN produces a *y(k)* output at time *k* as shown in Eq. (1), where *u(k)* is the input at time *k* and *N* the maximum adopted time delay. After been adequately trained, TDNN have been used successfully for prediction, because they are able to capture the dynamics of a system and to foresee the output in the current time.

$$y(k) = f(u(k), u(k-1), u(k-2), \cdots, u(k-N)) \quad (1)$$

In our case; the architecture of TDNN has been designed with an input layer with 480 nodes, 3 hidden layers with 220, 120 and 152 nodes respectively; and 1 output layer with 8 nodes. As for the input layer, a total of 12 features was considered obtained from a combination of normalized displacement and speed to represent the three joints describing the trajectory of an arm motion; right elbow, right wrist and right hand joints for the right arm as well as the left elbow, left wrist and left hand joint for the left arm.

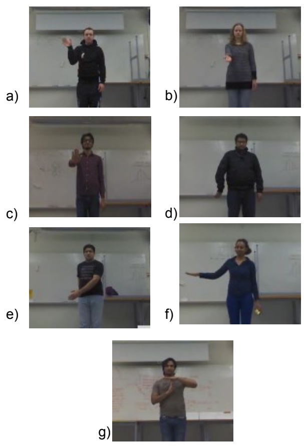

Figure 4.Definition of set of gestures: a) engage; b) follow; c) stop; d) come-left; e) come-right; f) disagange and g) emergency-stop.

As for the representation of the displacement features, the *spine* joint information (p_s^t) of the user is considered as the origin of the coordinate system of the displacement features. Thus, the displacement vector (s_{joint}^t) is defined as Eq. (2); where p_{joint}^t is defined as Eq. (3). On the other hand, in our work, the speed feature has been also taken into account. The speed feature of a joint (v_{joint}^t) is obtained as the rate of change of the position of a joint as given by the displacement vector as shown in Eq. (4).

$$s_{joint}^t = p_{joint}^t - p_s^t \quad (2)$$

$$p_{joint}^t = [x_{joint}^t, y_{joint}^t, z_{joint}^t]^T \quad (3)$$

$$v_{joint}^t = \left\| \frac{\Delta s_{joint}^t}{\Delta \tau^t} \right\| \quad (4)$$

2.3 Denoising of human motion data

At first, the authors considered to implement the moving average and Butterworth filter as are common filters used in digital signal processors and human modeling software.

The moving average is optimal for a common task: reducing random noise while retaining a sharp step response [22]. The noise generated in our application has high frequency components so the moving average could be suitable for this purpose. The moving average, as defined as Eq. (5), is a weighted mean of time series data from several consecutive periods that is continually recomputed as new data becomes available. The result gets more smoothen as the neighborhood ratio increases. The ratio size (*R*) and weight distribution (ω_i) that increased most the system performance is shown in Eq. (6). In Figure 5, the smoothen output of the moving average filter.

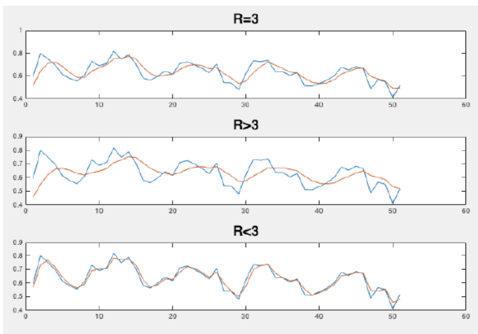

Figure 5. Moving average experimental results (blue color: original signal; red color: filtered signal)

$$x_i = \frac{1}{\sum w_i} \sum_{t-R}^{t+R} x_i \omega_i \quad (5)$$

$$x_i = \frac{4x_i + 3x_{i-1} + 2x_{i-2} + x_{i-3}}{10} \quad (6)$$

The Butterworth filter is a common solution for this type of high frequency denoising. The Butterworth filter is an optimal filter with maximally flat response in the passband [23]. The transfer function can be written as Eq. (7), where the scaled transfer function can be found by replacing *s* by s/ω_c (in our case, ω_c = 0.102 rad/sample). It consists on applying a Fast Fourier Transform on the input data to convert it to the frequency domain, then multiplying it for a function that attenuates high frequencies, and finally converting it again to time domain. The result is a considerably smoothen signal, however, this domain conversions takes time to be done, what is a disadvantage to real-time systems, such as iCAR's recognition module. Therefore, the Butterworth filter was applied here only to make a comparison with the moving average results.

$$H(s) = \frac{1}{(s+1)(s^2+s+1)} \quad (7)$$

In addition, the use of signal processing techniques for adapting a gesture recognition system made for static situations to work well in dynamic ambient as well as improving its performance when the system is receiving noisy data. Based on preliminary experiments with the moving average, it has been notice that is not able to remove the noise due to the oscillations on the mobile platform of the two-wheeled inverted pendulum robot. Therefore, a probabilistic analysis was taken into consideration, where the difference between the positions of two consecutive frames was removed if it is out of a trust interval. The main principle of the proposed method is that the distance of a joint position in two consecutive frames should be physically realizable. Since the user is not performing gestures with sudden movements, all the variations that are higher than the previous ones should be removed. The mean average and standard deviation are calculated as shown in Eq. (8) and Eq. (9) respectively. Its values are used to define the trust interval at each frame as shown in Eq. 10, where the constant *k* that multiplies the standard deviation was defined empirically (*k* = 20).

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (8)$$

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu)^2} \quad (9)$$

$$t_i = \mu + k\sigma \quad (10)$$

3 Experiments and Results

The probabilistic analysis was applied at the original data for cleaning the noise generated by the mobile platform. The final result is an increasing of 2.9% (from 92.29% to 95.20%) on the Success Percentage of the Recognition Module, and a negligible raise on the processing time of the system. The percentages per gesture are shown in Figure 9. The final average percentage (95.20%) is a satisfactory result, and even bigger than the one obtained with the TDNN training data set (94%) that doesn't have the movement of the mobile platform. The final average percentage of 95.20% is a satisfactory result under dynamically stable conditions, and even bigger than the one obtained with the TDNN training data set of 94% under statically stable conditions.

Table 1: Success ratio per gesture

Method	G1	G2	G3	G4	G5	G6	G7	G8
Original	0.96	0.90	0.68	0.98	0.99	0.81	0.95	0.97
Probabilistic Analysis	0.96	0.96	0.85	0.98	0.99	0.85	0.97	0.97

4 Conclusions

In this paper, the development of a multipurpose human-friendly assistive robot vehicle has been introduced. In particular, the first prototype composed by a two-wheeled inverted pendulum robot, a RGB-D camera and a 3-DOFs haptic interface has been described. In particular, the denoising algorithm for the 3D human motion data for the 3D gesture recognition has been described. A set of experiments were proposed to verify the performance of the proposed denoising algorithm. From the experimental results, a 95.20% of success ratio has been obtained.

References

(1) Wilén, A., Ljungberg, A. (2009). The future population of Sweden 2009–2060, Statistics Sweden, Forecasting Institute, pp. 20.

(2) Bertmar, H. (2012). "Hälsorobotar: Robotar som hjälper och vårdar", *Hjälpmedelsinstitutet*, 11318.

(3) Solis, J., Takanishi, A., (2011) "Robotic-Assisted Technology for Medical Training Purposes", Biomechatronics in Medicine and Health Care, Le Li and Kai-Yu Tong (Eds.), PanStanford, pp. 171-186.

(4) Ishii, H., et al. (2009). "Development and Experimental Evaluation of Oral Rehabilitation Robot That Provides Maxillofacial Massage to Patients with Oral Disorders," International Journal of Robotics Research, Vol. 28(9), pp. 1228–1239.

(5) Yohan, N., Solis, J., et al. (2012) "Development of Airway Management Training System WKA-4: Provide Useful Feedback of Trainee Performance to Trainee during Airway Management," in Proc. of the Int. Conf. on Complex Medical Engineering, pp. 423-428.

(6) Fukuda, T., Huang, J., D., Pei, Sekiyama, K. (2015) "Motion control and fall detection of intelligent cane robot, Intelligent Assistive Robots, Springer Tracts in Advanced Robotics, vol. 106, pp. 317-337.

(7) Yu, H., Spenko, M., Dubowsky, S. (2003) "An Adaptive Shared Control System for an Intelligent Mobility Aid for the Elderly. Auton. Robots, vol. 15(1), pp. 53–66.

(8) Dubowsky, S., Genot, F., Godding, S.K., Skwersky, H.A., et al. (2000) "PAMM-a robotic aid to the elderly for mobility assistance and monitoring: A helping-hand for the elderly," in Proc. of the IEEE International Conference on Robotics and Automation, pp. 570–576.

(9) Hirata, Y., Baba, T., Kosuge, K. (2003) "Motion control of omni-directional type walking support system walking helper," in Proc. of the 12th IEEE International Workshop on Robot and Human Interactive Communication, pp. 85–90.

(10) Rentschler, A.J., Cooper, R.A., Blaschm, B., Boninger, M.L. (2003) "Intelligent walkers for the elderly performance and safety testing of vapamaid robotic walker" Journal of Rehabilitation Research and Development, vol. 40, pp. 423–432.

(11) Hirata, Y., Hara, A., Kosuge, K.: Passive-type intelligent walking support system "RT Walker," in Proc. of the IEEE International Conference on Intelligent Robots and Systems, pp. 3871–3876.

(12) Solis, J.; et al. (2015) "Towards the development of multipurpose assistive vehicle robot for the ambient assisted living," International Conference on Automation Science and Engineering, pp. 1145-1150.

(13) Takanishi, A. (1995). "In Memoriam: Professor Ichiro Kato", Autonomous Robots, vol. 2, pp. 7-10.

(14) Solis, J. et al. (2014). Development of a human friendly robot vehicle for carrying-medical tools: Embodying perceptual capabilities, 14th Mechatronics Forum Int. Conf. Mechatronics, pp. 370-376

(15) Solis, J., De la Rosa, J.P. (2015). Development of a human-friendly walking assist robot vehicle: System Integration and Preliminary Stability Tests, 14th World Congress in Mechanism and Machine Science.

(16) Solis, J., Sansanayuth, T., Shojaei, E. (2016). Velocity control improvement for the Human-Friendly Assist Robot Vehicle, 2016 IEEE/SICE International Symposium on System Integration, pp. 331-336.

(17) Feng, Y. Ji, M., Xiao, J., Yang, X., Zhang, J.J., Zhuang, Y., Li, X. (2015). Mining Spatial-Temporal Patterns and Structural Sparsity for Human Motion Data Denoising, IEEE Transactions on Cybernetics, Vol 45(2), pp. 2693-2706.

(18) Lou, H. and Chai, J- (2010) Example-based human motion denoising, IEEE Transactions on Visualization and Computer Graphics, Vol. 16(5), pp. 870–879.

(19) Yamane, K. and Nakamura, Y. (2000) "Dynamics filter-concept and implementation of online motion generator for human figures," in Proc. IEEE International Conference on Robotics and Automation, pp. 688–694.

(20) Li, L., McCann, J., Pollard, N.S., Faloutsos, C. (2010). BoLeRO: A principled technique for including bone length constraints in motion capture occlusion filling, in Proc. Symposium Computer Animation, pp. 125–135.

(21) Lai, R., Yuen, P., Lee, K. (2011). Motion capture data completion and denoising by singular value thresholding," in Proc. Eurographics, pp. 45–48.

(22) Smith, S.W. (1997) The scientist and engineering's guide to digital signal processing, California Technical Pub, p. 277.

(23) Winder, S. (2002). Analog and digital filter design, Newnes, p. 54.