

Towards an Intelligent Robotic Walker for Assisted Living using Multimodal Sensorial Data

Georgia G. Chalvatzaki, Georgios Pavlakos, Kevis Maninis,
Xanthi S. Papageorgiou, Vassilis Pitsikalis, Costas S. Tzafestas and Petros Maragos
School of Electrical and Computer Engineering, National Technical Univ. of Athens, Greece

Abstract—We aim at developing an intelligent robotic platform that provides cognitive assistance and natural support in indoor environments to the elderly society and to individuals with moderate to mild walking impairment. Towards this end, we process data from audiovisual sensors and laser range scanners, acquired in experiments with patients in real life scenarios. We present the main concepts of an automatic system for user intent and action recognition that will integrate multiple modalities. We demonstrate promising preliminary results, firstly on action recognition based on the visual modality, i.e. color and depth cues, and secondly on the detection of gait cycle patterns that exploit the laser range data. For action recognition we are based on local interest points, 3D Gabor filters and dominant energy analysis, feeding a support vector machine. Then the recognized actions can trigger the gait cycle detection that detect walking patterns by exploiting the laser range data, modeled by hidden Markov models. In this way, we shall acquire the overall patient's state and the robot shall autonomously reason on how to provide support.

Keywords: assisted living for the elderly, visual action recognition, laser sensor gait analysis and tracking

I. INTRODUCTION

As the elderly population rises in our societies, mobility issues become a major problem. Recent research shows that, approximately 20% of people aged over 70 years, and 50% of people aged over 85, face mobility difficulties, [1]-[3]. In the EU, the rising life expectancy is estimated to bring about an increase of 40% of population aged over 80 during 1995-2015; thus mobility disabilities are expected to pose even greater significance. Walking deficiency hampers daily activities, while causing drawbacks both physically and psychologically, due to lack of exercise and independence. Moreover, the insufficiency of nursing staff, [4], leads to the need of robotic assistants that incorporate functionalities such as posture support and stability, walking assistance, navigation in indoor/outdoor environments and health monitoring. We envisage a robotic assistant walker that combines multimodal action recognition and behaviour analysis by integrating various sensing technologies that assists the elderly people with specific forms of moderate to mild impairment in a particular user-centered, context-adaptive and natural manner for an assisted living, [5].

There are several attempts towards developing intelligent robotic mobility aids, i.e. robotic wheelchairs and robotic walkers, especially rollators, [6], which constitute a mechanized

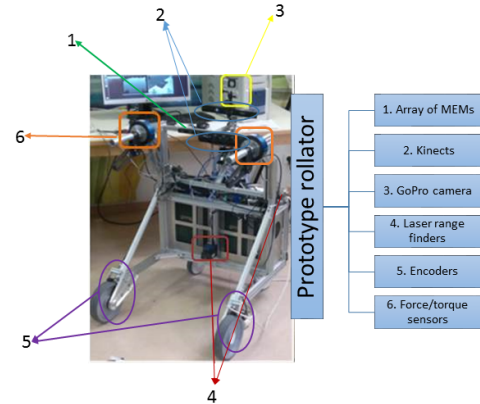


Fig. 1. Experimental platform: first sensorized prototype used for data acquisition and recordings.

version of the standard walking frame attached to the wheels, mainly used for balance problems. The robotic walkers, incorporate features such as physical support, sensorial and cognitive assistance, health monitoring and human - machine interaction, [7]. These walkers are either passive or active devices, [8]. There have been attempts of integrating various sensors onto robotic walkers, such as fusion techniques of sensorial data from a camera for face tracking and a laser sensor for user legs' detection, [9], and a framework for multimodal human-robot interaction using data from a laser scanner and two inertial measurement units, [10].

Our research aims at developing a robotic walker for indoor environments that localizes the user, estimates the body pose and recognizes human actions, gestures and intentions. A crucial feature of our approaches is the integration of multiple modalities, [11]. Herein, we develop a system that takes as input the multimodal data, first processes the visual data to perform action recognition, and then the recognized actions can trigger gait processing that detects walking patterns by exploiting the laser range data. Initial results of our work are presented recently in [12], [11], [13].

II. EXPERIMENTAL PLATFORM AND MULTIMODAL DATA

The experimental platform consists of a passive rollator prototype that moves forward only after applying force on it, equipped with various sensors, as shown in Fig. 1. The sensing technologies include: (i) two Laser Range Finder sensors (LRF), a Hokuyo UTM-30LX placed at the front of the platform facing the walking area for environment mapping and

This research work was supported by the European Union under the project "MOBOT" with grant FP7-ICT-2011-9 2.1 - 600796

obstacle detection, a Hokuyo UBG-04LX-F01 rapid LRF at the back of the platform mounted at a height of approximately 40cm from the ground for detecting lower limbs movement, (ii) two 6 DOF HR3 force/torque sensors placed at the handles, to measure the applied forces between the patient and the rollator, (iii) two types of visual sensors: a GoPro HD camera to record patient's upper body movements and two Kinect-for-Windows (KFW) sensors. The first horizontal KFW captures the torso, waist, hips and the upper part of the limbs and the second KFW faces downwards at the lower limbs, (iv) an array of 8-microphone MEMS is used for audio capturing. This platform is used for data acquisition in predefined scenarios of everyday life actions involving elderly people with mobility inabilities.

III. MULTIMODAL: VISUAL - HUMAN ACTION RECOGNITION

Towards multimodal action and gesture recognition, we develop robust computer vision techniques to process multiple cues, such as spatiotemporal RGB appearance and depth data from Kinect sensors. These data are employed to automatically recognize human actions such as stand-up, sitting intent, dynamic walking/stability behavior of the user and gestures recognition, as related to action intention.

For gesture recognition, we employ features extracted both on handshape and movement information, which are then modeled via hidden Markov models (HMM). We also explore multimodal fusion techniques to take advantage of audio information and enhance the robustness of our framework, [11]. Preliminary results are promising, despite the specificities of the task, as the not very active movements, the variance among users, and the inconsistency in gesture performance.

For action recognition, our front-end, [12] detects local spatio-temporal interest points, which are voxels in a video that maximize a saliency function. The latter is an energy map computed by multiband spatio-temporal filtering with 3D Gabor filters and dominant energy analysis. We apply the *Teager-Kaiser Energy Operator* (TKEO) to each filtered channel and the most dominant channel's magnitude forms the energy map of every pixel. TKEO has the ability to track energy oscillations and separates them into their amplitude and frequency components with excellent spatio-temporal resolution and small complexity. The detections are represented through local gradient based descriptors computed in a volume around each detected point. For all the above we also take advantage of depth information to localize and segment the user. Thus the spatio-temporal interest points are better localized. An example of user segmentation and the detected features are presented in Fig. 2.

The above described descriptors are processed to form *Bag-of-Features* (BOF) histograms. For this, we form groups of descriptors by clustering. The centroids are the visual features that represent a video through the frequency of appearance of each descriptor. We use a moving temporal window to extract features to construct a BOF histogram for each temporal frame. BOF representations are evaluated through support vector machine (SVM) classifiers with χ^2 -kernels. These result to the probabilities of a video frame to belong to a specific action. A background model (BM) is employed to represent sections not belonging in our vocabulary, such as silent sections and visual noise.

Finally, we end up with probabilities for each action segment. These probabilities are smoothed by a Gaussian/median window and are feeded forward to the Viterbi algorithm, [14], as observation likelihoods. Viterbi is combined with a penalty applied on the diagonal of its transition matrix to constrain frequent action transitions. This is applied in order to avoid action transitions in consequent frames. The same algorithm provides the final result of the recognized action sequence (see Sec.V). The recognition of actions provide information for the processing of gait cycles by predicting the intention of the subject to walk.

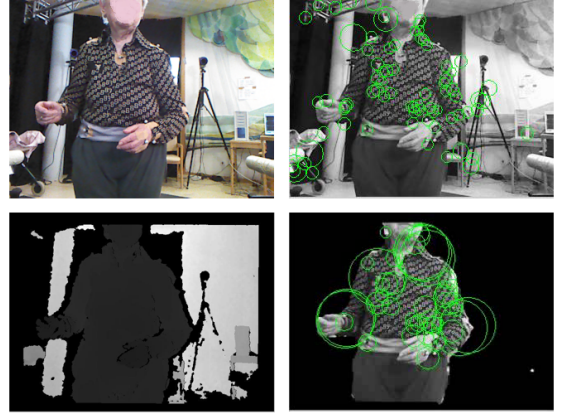


Fig. 2. Interest points detection. Left: RGB/depth images. Right: Interest points for the original frame (top), and after employing depth stream to localize the user (bottom). Elimination of false alarms in the background improves recognition performance.

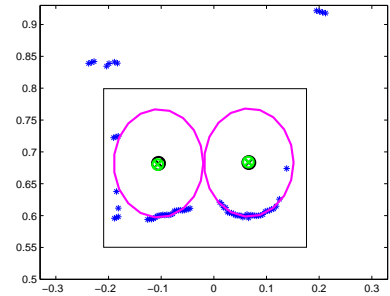


Fig. 3. Limbs' detection from range data. Blue stars correspond to the raw laser data; black 'X' correspond to the detected leg centers; green 'X' correspond to the estimated positions. The rectangle is the search window

IV. MULTIMODAL: LASER RANGE SCANNER- GAIT CYCLE DETECTION

A. Gait Analysis

Our next objective is the analysis and classification of walking patterns. Normal gait has periodic transitions of each foot from one position of support to the next. The gait cycle describes the period of time during which one leg leaves the ground for the first time to perform a forward motion till when the same leg contacts the ground again, [15]. Each gait cycle has two phases: stance and swing. In stance the foot is in contact with the ground; 60% of the gait cycle. In swing the foot is in the air performing a ballistic motion; 40% of the gait cycle. The gait cycle is divided into eight events:

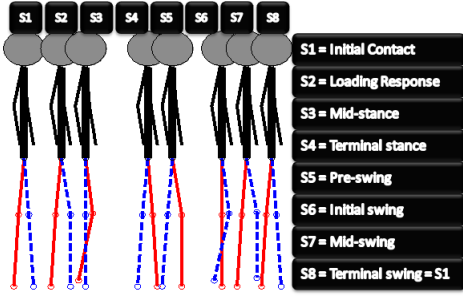


Fig. 4. Internal states of normal gait cycle (Left Leg: blue dashed line, Right Leg: red solid line).

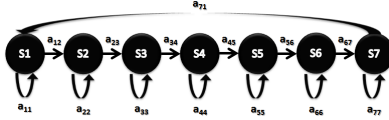


Fig. 5. Left-to-right HMM used for gait analysis and recognition.

- 1) *IC* - Initial Contact: 0% (Heel Strike)
- 2) *LR* - Loading Response: 0-10% (Foot Flat)
- 3) *MS* - Midstance: 10-30% (Opposite Toe Off)
- 4) *TS* - Terminal Stance: 30-50% (Heel rise)
- 5) *PW* - Preswing: 50-60% (Opposite Initial Contact)
- 6) *IW* - Initial Swing: 60-70% (Toe Off)
- 7) *MW* - Midswing: 70-85% (Feet Adjacent)
- 8) *TW* - Terminal Swing: 85-100% (Tibia Vertical)

B. Detection and Tracking through laser data

The recognised actions initialize gait processing. The detection and tracking of user's limbs using the range data precedes the feature extraction. Towards this end we apply the following process, i.e. a combination of K-means clustering along with Kalman Filtering (KF):

- 1) An initial orthogonal search window in the x-y plane is set, as the most probable area of detecting the user's limbs.
- 2) A background extraction based on thresholding excludes the outliers and the K-means algorithm assigns two clusters to the corresponding legs. A least squares nonlinear circle fitting process is used to approach the legs' centers. Those centers are the observation vector that enters a constant acceleration KF.
- 3) The KF tracks the central positions of the limbs by stochastically estimating their position according to a motion and an observation model.
- 4) The predicted positions are seed to the K-means algorithm for the next laser frame, while their variances are used to adjust the initial search window.
- 5) The process is repeated until no more laser measurements enter the system.

In Fig. 3 we show an example of a laser frame with the detected legs' centers and their estimated positions. From the results of the detection and tracking process the 9×1 feature vector per laser frame is used for gait cycle classification. The feature vector consists of the left and right leg positions and velocities in the x,y coordinates along with the interleg distance. In Fig. 7

we show a few strides of normal gait with respect to the laser frame along with a plot of the distance between the legs.

C. Detection of Gait Cycle

The dynamic properties of walking led to the usage of Hidden Markov Models. These modeling generative sequences, where an observable variable results from a combination of hidden states, are stochastic processes, commonly employed in speech processing, as well as human activity analysis, [16], due to their ability to model temporal transitions.

We employ a HMM for gait recognition with seven states corresponding to the seven events of normal walking; the TW phase is excluded as equivalent to the IC that triggers a new gait cycle – see Fig. 4. Our observations are based on the feature vector presented in Sec. IV-B. A left-to-right HMM is defined (see Fig. 5) and the Baum-Welch algorithm is used for the estimation of the model parameters, [14]. Initial results obtained by applying this model in normal human gait data are promising, [13].

V. EXPERIMENTAL RESULTS

Dataset: For the experiments in this preliminary work we employ data from subjects who take part in the following scenario: the subject (patient) sits in a chair, stands up, performs straight walking and then sits back. For the case of visual modality processing and action recognition we employ data from six patients performing the above three actions (SU, WALK, SD), while performing several gestures. For the laser data we employ data from 10 subjects (non-patients, i.e. subjects who performed normal walking) who took part in the same scenario. The multimodal data set is acquired by two sensors: the upper Kinect (KFW) and the rear laser scanner facing the user's legs. The KFW captures RGB and depth data on a 30Hz sampling frequency. The laser scanner has an angular resolution of 0.36° , scan time 28msec and accuracy 10mm for distances till 1m.

The evaluation of the proposed approach for action recognition is applied on the above real experimental data. Results on the detection and classification of actions on a continuous video via the probabilistic outputs of SVMs is shown in Fig. 6; these include three actions and a background model. Then, by

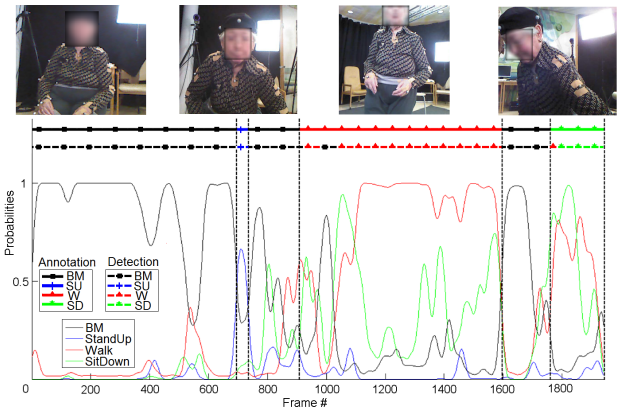


Fig. 6. Probabilistic outputs of 3 actions + BM, the annotated and recognized labels for our action recognition framework.

taking into account the recognized actions, the gait analysis can be initialized at the detection of ‘StandUp’, translated as the patient’s intention to start walking.

For the evaluation of HMM-based gait detection, we employ experimental data from ten subjects with normal gait. HMM training incorporates only a part of subjects data, excluding the recorded data of one subject for evaluation. In the evaluation phase we estimate the sequence of state transitions that occur in the unseen test data.

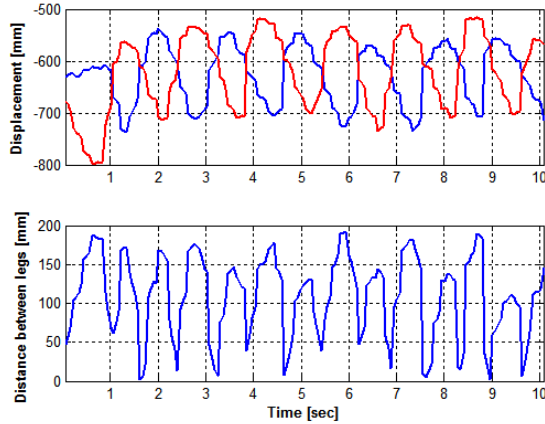


Fig. 7. Real experimental normal gait data for testing and evaluation of HMM. Top: Left (blue data) and Right (red data) legs displacement. Bottom: legs distance in the sagittal plane.

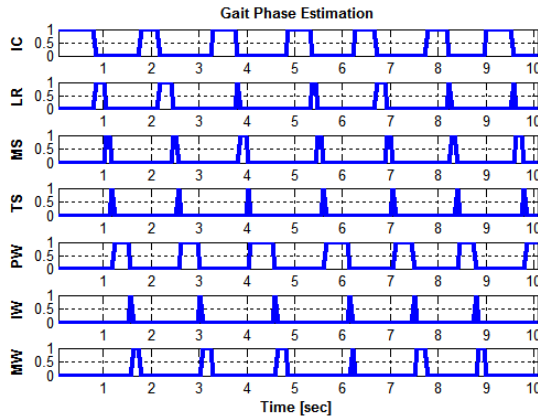


Fig. 8. Estimated sequence of gait phases based on the constructed model (HMM), w.r.t. time.

Preliminary results are presented in Fig. 8. These show that the evolution of the gait states after testing the trained model on the unseen experimental data of Fig. 7, match the general evolution of the human gait model that is represented by the HMMs; i.e. the gait phases appear sequentially with the correct order, and the time frame of each state is within the general bounds as discussed in Sec. IV-A.

VI. CONCLUSIONS AND FUTURE WORK

We presented our objectives on developing an intelligent robotic walker that provides optimal support to the users in indoor environments by integrating sensorial data from visual and laser range sensors. The first prototype consists of a

passive rollator, used for data acquisition with elderly people. Initial results have been presented regarding recognition of human actions. The action recognition subsystem employing visual data can trigger the process for detection of user’s walking patterns based on the laser range data.

Our ongoing research is to develop an extended action recognition system that will provide the ability of user-robot communication based on audio-gestural recognition. An advanced framework for the recognition of gait patterns and their classification into normal or pathological ones is also currently being developed. Finally, we aim at integrating all modalities into a general human-state recognition system that aims at designing a human-adaptive mobility aid with cognitive support according to the patient’s needs in the context of assisted living.

REFERENCES

- [1] P. D. Foundation, “Statistics for parkinson’s disease,” 2010. [Online]. Available: <http://www.pdf.org>
- [2] S. Center, “Stroke statistics,” 2010. [Online]. Available: <http://www.strokecenter.org>
- [3] USCensus, “The elderly population,” 2010. [Online]. Available: <http://www.census.gov>
- [4] M. Montemerlo, J. Pineau, N. Roy, S. Thrun, and V. Verma, “Experiences with a mobile robotic guide for the elderly,” in *Proc. of AAAI National Conf. on Artificial Intelligence*, Canada, 2002.
- [5] X. Papageorgiou, C. Tzafestas, P. Maragos, G. Pavlakos, G. Chalvatzaki, G. Moustiris, I. Kokkinos, A. Peer, B. Stanczyk, E.-S. Fotinea, and E. Efthimiou, “Advances in intelligent mobility assistance robot integrating multimodal sensory processing,” in *Universal Access in Human-Computer Interaction. Aging and Assistive Environments*, ser. Lecture Notes in Computer Science, C. Stephanidis and M. Antona, Eds. Springer Int’l Publ., 2014, vol. 8515, pp. 692–703.
- [6] H. Van der Loos and D. Reinkensmeyer, “Rehabilitation and health care robotics,” in *Springer Handbook of Robotics*, B. Siciliano and O. Khatib, Eds. Springer Berlin Heidelberg, 2008, pp. 1223–1251.
- [7] M. M. Martins, C. P. Santos, A. Frizera-Neto, and R. Ceres, “Assistive mobility devices focusing on smart walkers: Classification and review,” *Robotics and Autonomous Systems*, vol. 60, no. 4, pp. 548 – 562, 2012.
- [8] C.-H. Ko and S. Agrawal, “Walk-assist robot: A novel approach to gain selection of a braking controller using differential flatness,” in *American Control Conf.*, 2010, pp. 2799–2804.
- [9] N. Belloto and H. Hu, “Multisensor-based human detection and tracking for mobile service robots,” *IEEE Trans. on systems, Man and cybernetics, Part B: Cybernetics*, vol. 39, no. 1, February 2009.
- [10] C. Cifuentes, C. Rodriguez, A. Frizera-Neto, T. Bastos-Filho, and R. Carelli, “Multimodal human-robot interaction for walker-assisted gait,” pp. 1–11, 2014.
- [11] G. Pavlakos, S. Theodorakis, V. Pitsikalis, S. Katsamanis, and P. Maragos, “Kinect-based multimodal gesture recognition using a two-pass fusion scheme,” in *Proc. Int’l Conf. on Image Processing*, 2014, (to appear).
- [12] K. Maninis, P. Koutras, and P. Maragos, “Advances on action recognition in videos using an interest point detector based on multiband spatio-temporal energies,” in *Proc. Int’l Conf. on Image Processing*, 2014, (to appear).
- [13] X. Papageorgiou, G. Chalvatzaki, C. Tzafestas, and P. Maragos, “Hidden markov modeling of human normal gait using laser range finder for a mobility assistance robot,” in *Proc. IEEE ICRA*, 2014.
- [14] L. R. Rabiner and B. Juang, “An introduction to hidden markov models,” *ASSP Magazine, IEEE*, vol. 3, pp. 4–16, jan 1986.
- [15] M. Jacquelin Perry, *Gait Analysis: Normal and Pathological Function*, 1st ed. SLACK Incorporated, 1992.
- [16] P. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea, “Machine recognition of human activities: A survey,” *IEEE Trans. Circuits and Systems for Video Technology*, vol. 18, no. 11, pp. 1473–1488, 2008.