Real-time Application for Monitoring Human Daily Activity and Risk Situations in Robot-Assisted Living

Mário Vieira, Diego R. Faria and Urbano Nunes*

Institute of Systems and Robotics, Department of Electrical and Computer Engineering, University of Coimbra, Portugal

Abstract. In this work, we present a real-time application in the scope of human daily activity recognition for robot-assisted living as an extension of our previous work [1]. We implemented our approach using Robot Operating System (ROS) environment, combining different modules to enable a robot to perceive the environment using different sensor modalities. Thus, the robot can move around, detect, track and follow a person to monitor daily activities wherever the person is. We focus our attention mainly on the robotic application by integrating several ROS modules for navigation, activity recognition and decision making. Reported results show that our framework accurately recognizes human activities in a real time application, triggering proper robot (re)actions, including spoken feedback for warnings and/or appropriate robot navigation tasks. Results evidence the potential of our approach for robot-assisted living applications.

1 Introduction

Mobile robots endowed with cognitive skills are able to help and support humans in an indoor environment, providing increased availability, awareness and access, as compared to static systems. Thus, a robot can act not only as assistant in the context of robot-assisted living, but also offer social and entertaining interaction experiences between humans and robots. For that, the robot needs to be able to understand human behaviours, distinguishing human daily routine from potential risk situations in order to react in accordance. In this work, we focus our attention on the domain of human-centered robot application, more precisely, for monitoring tasks, where a robot can recognize daily activities and unusual behaviours to react according to the situation. In this context, a robot that can recognize human activities will be useful for assisted care, such as human-robot or child-robot interaction and also monitoring elderly and disabled people regarding strange or unusual behaviours. We use a robot with an RGB-D sensor (Microsoft Kinect) on-board to detect and track the human skeleton in order to extract

* This work was supported by the Portuguese Foundation for Science and Technology (FCT) under the Grant AMS-HMI12: RECI/EEI-AUT/0181/2012. The authors are with Institute of Systems and Robotics, Department of Electrical and Computer Engineering, University of Coimbra, Polo II, 3030-290 Coimbra, Portugal (emails: mvieira, diego, urbano@isr.uc.pt).

motion patterns for activity recognition. We present an application that combines different modules, allowing the robot localization and navigation in an indoor environment, and also to detect obstacles and human skeleton for motion tracking. In addition, we use modules for voice synthesizer and recognition, that will be triggered by our activity recognition module. The activity recognition module uses a Dynamic Bayesian Mixture Model (DBMM) [2] [1] for inference, in order to classify each activity, enabling the mobile robot to make a decision to react accordingly. The main contributions of this work are:

- Combining different ROS modules (navigation, classification and reaction module), towards a real time robot-assisted living application.
- Extending the use of DBMM to real-time applications using proposed discriminative 3D skeleton-based features, which can successfully characterize different daily activities.
- Assessment and validation: (i) leave-one-out cross validation of the activity recognition using our training dataset; (ii) comparison of different classification models using our proposed features; (iii) online validation of the integrated artificial cognitive system.

The remainder of this paper is organized as follows. Section 2 covers selected related work. Section 3 introduces our approach, detailing the proposed 3D skeleton-based features as well as the classification method. Section 4 describes how the approach is implemented in ROS. In section 5, the performance of the proposed application is presented. Finally, Section 6 brings the conclusion of this research pointing future directions.

2 Related Work

In order to have a fully operational robot-assisted living application, it is essential that the robot can recognize daily activities in real scenarios, in real-time. In spite of some proposed works that use inertial sensors for human activity recognition [3] [4], the most common approaches use vision-based depth sensors, even more nowadays, with low cost vision sensors (e.g. RGB-D sensors [5] [6]) that can track the entire human body accurately. In [7], a Microsoft Kinect sensor is used to track the skeleton and posteriorly extract the features. The action recognition is done using first order Hidden Markov Models (HMMs) and for every hidden state, the observations were modelled as a mixture of Gaussians. The work presented in [8] uses depth motion maps as features for activity recognition. Other works on the recognition of human activities focus their research on how to extract the right features in order to obtain better classification performance [9] [10] [11]. In the context of robot assisted living, [12] describes a behaviour-based navigation system in assisted living environments, using the mobile robot ARTOS. In [13] a PR2 robot is used to assist a person. The robot detects the activity being performed as well as the object affordances, enabling the robot to figure out how to interact with objects and plan actions. In [14], a mobile robot is used in a home environment to recognize activities in real-time by continuously tracking the pose and motion of the user and combining them with structural knowledge like the current

room or objects in proximity. In our work, we use a Nomad Scout with a laser Hokuyo to assist the localization and navigation module, and an RGB-D sensor on-board to detect and track a person. It is a small mobile robot that monitors a person in an indoor environment, recognizing daily and risky activities and reacts with defined actions, assisting the person if needed. Our activity recognition module is based on the framework proposed in [1], where the features are also skeleton-based, however, herein we model different skeleton-based features, and in addition, we use a new collected dataset.

3 Activity Recognition Framework

3.1 Extraction of 3D Skeleton-based Features

We have used a Microsoft Kinect sensor and the OpenNi's tracker package for ROS to detect and track the human skeleton. This package allows the skeleton tracking at 30 frames per second, providing the three-dimensional Euclidean coordinates of fifteen joints of the human body with respect to the sensor. Using this information, we compute a set of features as follows:

- Euclidean distances among the joints, all relative to the torso centroid, obtaining a 15×15 symmetric matrix with a null diagonal. Let *(x,y,z)* be the 3D coordinates of two body joints b_j with $j = 1,2,...,15$ and b_i with $i = 1,2,...,15$, then $\forall\ \{b_i, b_j\}$, the distances were computed as follows:

$$\delta(b_j, b_i) = \sqrt{(b_j^x - b_i^x)^2 + (b_j^y - b_i^y)^2 + (b_j^z - b_i^z)^2} \tag{1}$$

  Subsequently, we removed the null diagonal, obtaining a 14×15 matrix $\mathbf{M}$ to compute its *log-covariance* as follows:

$$\mathbf{M}_{lc} = \mathbf{U}(\log(\mathrm{cov}(\mathbf{M}))), \tag{2}$$

  where $\mathrm{cov}(\mathbf{M}_{i,j}) = (M_i - \mu_i)(M_j - \mu_j)$; $\log(\cdot)$ is the matrix logarithm function (logm) and $\mathrm{U}(\cdot)$ returns the upper triangle matrix composed by 120 feature elements. The rational behind of log-covariance is the mapping of the convex cone of a covariance matrix to the vector space by using the matrix logarithm as proposed in [15]. A covariance matrix form a convex cone, so that it does not lie in Euclidean space, e.g., the covariance matrix space is not closed under multiplication with negative scalers. The idea of log-covariance is based on [16], where examples of manifold Riemannian metrics and log-covariance applied in 2D image features for activity recognition were used.
- The global skeleton velocities, assuming the 3D coordinates of 14 joints in the case of having the torso centroid as origin; and 15 joints in the case of having the sensor frame as origin were computed as follows:

$$v_j = \frac{\sqrt{(b_{j_x}^t - b_{j_x}^{t-t_w})^2 + (b_{j_y}^t - b_{j_y}^{t-t_w})^2 + (b_{j_z}^t - b_{j_z}^{t-t_w})^2}}{f_{rate} \times t_w}, \tag{3}$$

  where v_j is the velocity of a specific skeleton joint j; b_{j_d} represents the position $d = \{x,y,z\}$ of a skeleton body joint j in the current time t, and $t - t_w$ represents some preceding frames, herein $t_w = 10$; the frame rate is set to $f_{rate} = 1/30$.

– Differently of the aforementioned velocities in the torso frame of reference, herein, relative to the sensor frame, for all joints, for each dimension individually, we computed the difference $\delta(b^t_{j_d}, b^{t-t_w}_{j_d})$ between the position at a given frame and the preceding 10^{th} frame. Using these values, we computed the velocities of the same joints for each dimension individually, $v_j = \frac{b^t_{j_d} - b^{t-t_w}_{j_d}}{f_{rate} \times t_w}$, obtaining additional 45 features.
– The angles variation of certain joints play a crucial role in carrying out many activities. We are interested in knowing whether a person is sitting or standing, so we compute the angles of both right and left elbows in the triangle formed by the hands, elbows and shoulders. We also compute the angles of the hip joints in the triangle formed by the shoulders, hips and knees and the angles of the knees in the triangles formed by the feet, knees and hips. The angle θ_i is given by:

$$\theta_i = \arccos\left(\frac{(\delta_{j_{12}})^2 + (\delta_{j_{23}})^2 - (\delta_{j_{13}})^2}{2 \times \delta_{j_{12}} \times \delta_{j_{23}}}\right), \tag{4}$$

where $\delta_{j_{12}}$ is the distance between two joints, e.g. j_1 and j_2, that are forming a triangle in the skeleton. We have 2+2+2=6 features for angles, since we are considering the left and right side for the body joints. In addition, we compute the difference between these angles at a current frame and the preceding 10^{th} frame, $\theta_{v_i} = \theta^t_i - \theta^{t-10}_i$, obtaining additional 2+2+2=6 features.

Thus, in total, we attained a set with 206 spatio-temporal skeleton-based features, useful to discriminate different classes of activities.

Features pre-processing: Before using the features set in the classification module, we perform a pre-processing step. Normalization, standardization or filtering may be a requirement for many machine learning estimators, as they can behave badly if no pre-processing is applied to the features set. So, in the dataset case, we apply a moving average filter with 5 neighbours data points to filter the noise, smoothing the data. Subsequently, a normalization step is applied in such a way that the values of minimum and maximum obtained during the training stage were applied on the testing set as follows:

$$\mathbf{F}_{tr_i} = \frac{\mathbf{F}_{tr_i} - \min(\mathbf{F}_{tr})}{\max(\mathbf{F}_{tr}) - \min(\mathbf{F}_{tr})}, \text{ and } \mathbf{F}_{te_i} = \frac{\mathbf{F}_{te_i} - \min(\mathbf{F}_{tr})}{\max(\mathbf{F}_{tr}) - \min(\mathbf{F}_{tr})}, \tag{5}$$

where $\mathbf{F}_{tr}$ is the set of features for training and $\mathbf{F}_{te}$ is the set of features for test; i is an index to describe a set of features in a specific frame; $\max(\cdot)$ and $\min(\cdot)$ are functions to get the global maximum and minimum value of a feature set. In the real-time case, we did not apply the moving average filter because it returns worse results. The normalization step is done in the same way as in the offline tests because we keep the maximum and minimum values of the training set.

3.2 Probabilistic Classification Model

In this work, we adopt an ensemble of classifiers called Dynamic Bayesian Mixture Model (DBMM) proposed in [2] [1]. DBMM uses the concept of mixture models in a

dynamic form in order to combine conditional probability outputs (likelihoods) from different single classifiers, either generative or discriminative models. A weight is assigned to each classifier, according to previous knowledge (learning process), using an uncertainty measure as a confidence level, and can be updated locally during the online classification. The local weight update assigns priority to the base classifier with more confidence along the temporal classification, since they can vary along the different frames. The key motivation of using a fusion model is because we are taking into consideration that an ensemble of classifiers is designed to obtain better performance than any of their individual classifiers, once there is diversity of the single components. Beyond of employing this classification model in an on-the-fly robot-assisted living application, we also compare the activity classification results with different well-known state-of-the-art classification models, such as Naive Bayes Classifier (NBC), Support Vector Machines (SVM) and k-Nearest Neighbours (k-NN). The DBMM general model for each class C is given by:

$$P(C|A) \;=\; \beta \;\times\; \underbrace{P(C^t|C^{t-1})}_{\text{dynamic transitions}} \;\times\; \underbrace{\sum_{i=1}^{n} w_i^t \times P_i(A|C^t),}_{\text{mixture model with dynamic w}} \qquad (6)$$

$$\text{with} \begin{cases} P(C^t|C^{t-1}) \equiv \frac{1}{C} \;\; (\text{uniform}), \, t = 1 \\ P(C^t|C^{t-1}) = P(C^{t-1}|A), \quad t > 1 \end{cases},$$

where $P(C^t \mid C^{t-1})$ is the transition probability distribution among class variables over time, which a class C^t is conditioned to C^{t-1}. This means a non-stationary behavior applied recursively, then reinforcing the classification at time t; $P_i(A|C^t)$ is the posterior result of each i^{th} base classifier at time t, becoming the likelihood in the DBMM model. The weight w_i^t in the model for each base classifier is initially estimated using an entropy-based confidence on the training set (offline), and afterwards ($t > 5$) it is updated as explained in our previous work [1]; $\beta = \frac{1}{\sum_j \left(P(C_j^t|C_j^{t-1}) \times \sum_{i=1}^{n} w_i \times P_i(A|C_j^t) \right)}$ is a normalization factor, ensuring numerical stability once continuous update of belief is done.

Base Classifiers for DBMM In this work, we have used the NBC, SVM and k-NN as base classifiers for the DBMM fusion. The NBC assumes the features are independent from each other given a class, $P(C_i|A) = \alpha P(C_i) \prod_{j=1}^{m} P(A_j|C_i)$. For the linear-kernel multiclass SVM implementation, we adopted the LibSVM package [17], trained according to the 'one-against-one' strategy, and classification outputs were given in terms of probability estimates. A k-NN was also combined into the DBMM fusion. An object is classified by a majority vote of its neighbours, with the object being assigned to the class most common among its k nearest neighbours. The classification outputs of the adopted k-NN were given in terms of probability estimates as well.

4 Robot-Assisted Living Architecture in ROS

The proposed artificial cognitive system was implemented in ROS and comprises three main modules, as shown in Figure 1: classification, navigation and reaction modules.

In order to properly test the system in real scenarios, a mobile robot is used. Therefore, a personal robot endowed with cognitive skills, capable of monitoring the behaviours of a person should be able to autonomously navigate in an indoor environment. The navigation module uses odometry and laser scans from the robot to map the environment and self-localization, randomly navigating, avoiding obstacles. We use the navigation stack available in ROS distributions, more specifically, the *move_base* package to generate an appropriate collision free trajectory. For simultaneous localization and mapping (SLAM) the *hector_slam* package is used. While the robot is navigating, the MS-Kinect sensor is sending RGB-D data to the classification module. Once a skeleton is detected, the robot stops and the feature extraction process starts. Then, classification is done using the DBMM and an activity is recognized. Once the system knows the human activity being performed, the reaction module is in charge to select what the robot should do next. For each human activity, a predefined reaction in a lookup-table was associated, including warnings, questions or changes in navigation (Figure 2). In the event of a person telling the robot to follow him/her, a safe distance of 2.5 meters is maintained. A Kalman filter is used to estimate the trajectory of the person one second ahead in order to avoid collision between the robot and the human. If a collision trajectory is estimated, the robot will step away, in order to the person walk through safely. For the prediction of the human motion, a position model was adopted, where the state includes position $(x(k); y(k))$ of the human target:

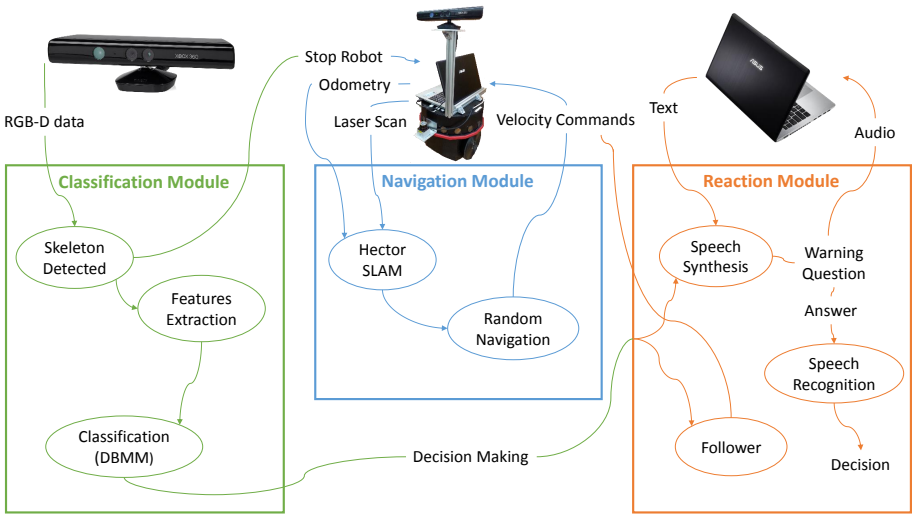

Fig. 1: System overview.

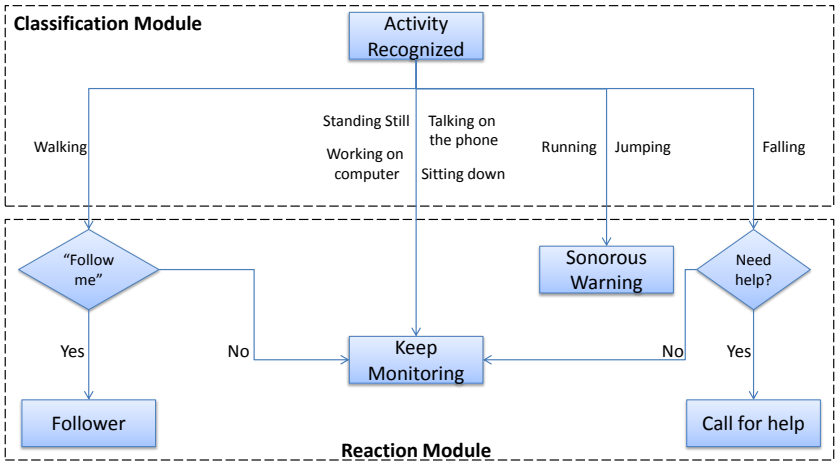

Fig. 2: Decision tree in reaction module.

$$\begin{cases} x(k) = x(k-1) + v_x(k-1) \times \Delta t \\ y(k) = y(k-1) + v_y(k-1) \times \Delta t \end{cases} \tag{7}$$

with $\Delta t = t(k) - t(k-1)$. Using the torso coordinates as measures, it is possible to compute the x velocity v_x and y velocity v_y. For speech synthesis, we use the *sound_play* package that given a text input, it will be synthesized into sound output. For speech recognition, we use the *pocketsphinx* package. This package recognizes a single word or a stream of words from a vocabulary file previously created. In our work, the vocabulary comprises the following words: "no", "yes", "please", "help", "follow", "me". The package can recognize combinations of these words, such as "please help me".

5 Experimental Results

5.1 Performance on collected dataset

A new dataset of daily activities and risk situations, more complete and challenging than the one used in our previous work [1], was collected to train the activity recognition module. This dataset (Figure 3) comprises video sequences of two male subjects and two female subjects performing eight different activities in a living room. The daily activities are: *1-walking, 2-standing still, 3-working on computer, 4-talking on the phone, 5 sitting down*; and the unusual or risk situations are: *6-jumping, 7-falling down, 8- running*. This dataset is a challenging one, once there is significant intra-class variation among different realizations of the same activity. For example, the phone is held with the left or right hand. Another challenging feature is that the activity sequences are registered from different views, i.e., from the front, back, left side, and so on. The classification results are presented in a confusion matrix and with the measures of Accuracy, Precision, Recall of the four tests. The idea is to verify the capacity of generalization of

Fig. 3: Few examples of the dataset (RGB with skeleton joints and depth images) which was created to learn some daily and risk situations.

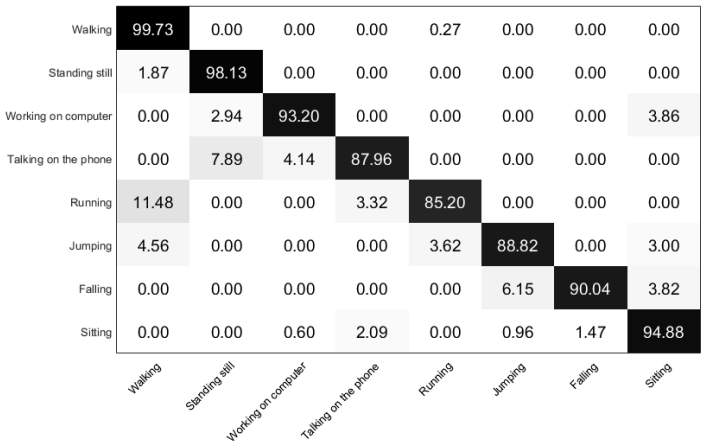

Fig. 4: Confusion matrix obtained from the DBMM classification applied on the dataset

the classifier by using the strategy of ”new person”, i.e., learning from different persons and testing with an unseen person. Figure 4 shows the results in a single confusion matrix. Table 1 shows the performance in terms of Precision (Prec) and Recall (Rec) of this approach for each activity. The results show that using DBMM, improvements in the classification were obtained in comparison with using the base classifiers alone. The overall results attained were: accuracy 93.41%, precision 93.61% and recall 92.25%.

Table 1: Performance on the dataset ("new person"). Results are reported in terms of Precision (Prec) and Recall (Rec).

Activity	DBMM Prec	DBMM Rec
walking	89.63%	99.73%
standing still	94.86%	98.13%
working on computer	95.93%	93.20%
talking on the phone	93.64%	87.96%
running	92.81%	85.20%
jumping	92.52%	88.83%
falling down	97.24%	90.04%
sitting down	92.27%	94.88%
Average	**93.61%**	**92.25%**

Table 2: Global results using single classifiers, a simple average ensemble (AV) and the DBMM.

Method	Acc.	Prec.	Rec.
NBC	82.90%	85.79%	82.67%
SVM	88.47%	89.02%	87.62%
k-NN	87.98%	90.09%	87.06%
AV	85.29%	87.74%	84.68%
DBMM	**93.41%**	**93.61%**	**92.25%**

For comparison purposes, Table 2 summarizes the results from single classifiers and an average ensemble compared with DBMM, showing the improvement achieved using the described skeleton-based features. The SVM was trained with *soft margin* (or Cost) parameter set to 1.0, and the *k*-NN was trained using 20 neighbours.

5.2 Performance on-the-fly using a mobile robot

The experimental tests using the proposed approach for a real time application is a little bit different than the experimental tests on the dataset. In this case, the robot will acquire 5 seconds of RGB-D sensor data for features extraction and classification. Only the NBC and SVM were used as base classifiers for the DBMM fusion, because they are enough for obtaining good results, thus, avoiding spending more processing time using other base classifiers. After 5 seconds of frames classification, a final decision is made for activity recognition to trigger a proper robot reaction. The proposed framework is capable of recognizing different activities transitions that happens sequentially in case of a person transit from one activity to another one, e.g., a person that is standing and sequentially pass to a sitting down position and consequentially working on the computer. Figure 5 shows some examples of tests of daily activities and unusual or risk situations that the mobile robot correctly recognized. Three tests were carried out for each activity with three different subjects. One of the subjects was already "seen" in the training, while the rest are "unseen" subjects.

Fig. 5: Shots of tests of activity recognition (‘unseen” person) using a mobile robot.

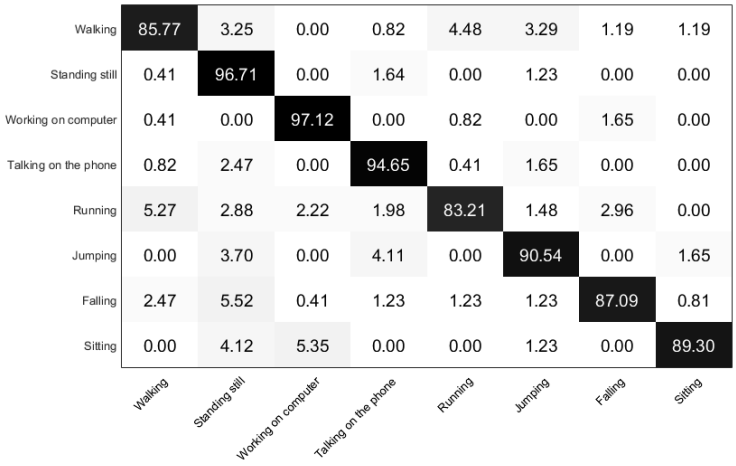

Fig. 6: DBMM on-the-fly classification confidence (average) presented in a confusion matrix

All activities were correctly classified, so that the overall performance of classification is shown in Figure 6. The overall (average) results attained in real-time experiments were: accuracy 90.55%, precision 90.84% and recall 90.55%. Table 3 shows the results in terms of recall of each test for each subject. Looking at the results attained, it is possible to conclude, as expected, that the best performance is achieved for the ”seen” person (subject 1). However, the difference of results between subjects is not very significant, which indicates that the fact of being or not a ”seen” person is not a key factor for the performance of the classification. The most important factor in a real-time application is that in the end, the activity being performed is correctly recognized. Since the robot correctly classified the activity performed, it also successfully reacted accordingly to the situation. Figure 7 shows a sequence of events from an activity that is being recognized (in this case falling) to react according to this activity. First, the skeleton of a person is detected and tracked, initiating the monitoring stage. Then, the person falls on the floor and the robot correctly recognizes the risk situation ”falling”. Detecting such a behaviour, the robot asks if the person needs help. The robot receives an affirmative answer from the person, recognizes the command and immediately calls for help.

Table 3: On-the-fly results in terms of recall for 3 different subjects. One subject seen and two unseen.

		Activity								Overall
	Test	walking	standing still	working on computer	talking on the phone	running	jumping	falling down	sitting down	
Subject 1 (seen)	1	96.30	100	100	59.26	85.19	85.19	85.19	96.30	88.43
	2	96.30	100	100	100	85.19	88.89	95.45	96.30	95.27
	3	92.59	100	92.59	100	85.19	88.89	92.86	96.30	93.55
	Average	95.06	100	97.53	86.42	85.19	87.65	91.17	96.30	**92.42**
Subject 2 (unseen)	1	66.67	100	96.30	100	96.30	81.48	74.07	70.37	85.65
	2	81.48	85.19	96.30	92.59	85.19	92.59	74.07	92.59	87.50
	3	81.48	100	88.89	100	85.19	92.59	95.45	92.59	92-02
	Average	76.54	95.06	93.83	97.53	88.89	88.89	81.20	85.18	**88.39**
Subject 3 (unseen)	1	82.14	96.30	100	100	73.33	96.30	85.19	88.89	90.27
	2	92.86	96.30	100	100	80.00	92.60	100	85.19	93.37
	3	82.14	92.59	100	100	73.33	96.30	81.48	85.19	88.88
	Average	85.71	95.06	100	100	75.55	95.07	88.89	86.42	**90.84**
Overall Average		**85.77**	**96.71**	**97.12**	**94.65**	**83.21**	**90.54**	**87.09**	**89.30**	**90.55**

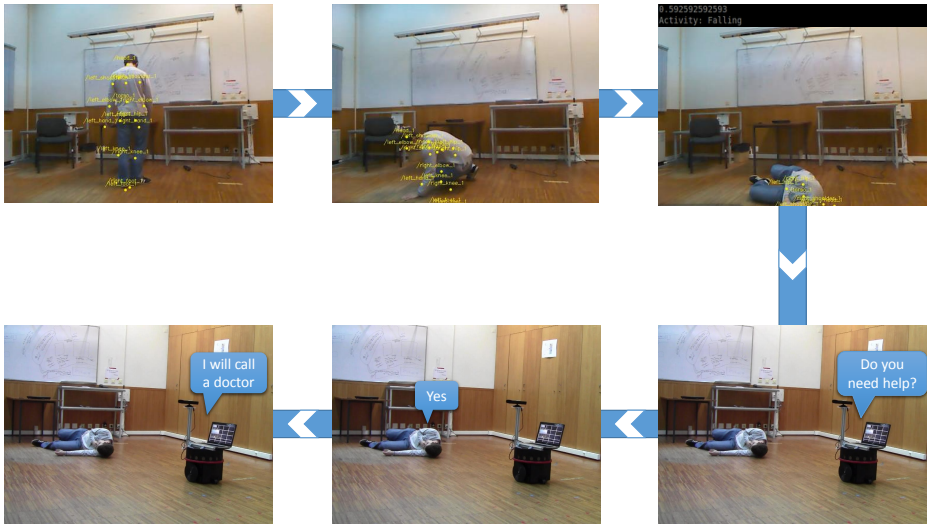

Fig. 7: Sequence of events on detecting a person falling and reacting

6 Conclusions and Future Work

The main contribution of this work is a robotic application for real-time monitoring of daily activities and risk situations in indoor environments. A dynamic probabilistic ensemble of classifiers (DBMM) was used for daily activity recognition using a proposed spatio-temporal 3D skeleton-based features. We collected a dataset to endow a robot to recognize daily activities, and we used this dataset to compare our approach with other

state-of-the-art classifiers. Using our proposed skeleton-based features, we attained relevant results using the DBMM classification, outperforming other single classifiers in terms of overall accuracy, precision and recall measures. More importantly, the experimental tests using a mobile robot presented good performance on the activity classification, allowing the robot to take appropriate actions to assist the human in case of risk situations, showing our framework has good potential for robot-assisted living. Future work will address addition of contextual information, such as "who", "where", "when" in order to fully understand human behaviours, as well as exploitation of our approach with more daily activities, risk situations and robot reactions.

References

1. D. R. Faria, M. Vieira, C. Premebida, and U. Nunes, "Probabilistic human daily activity recognition towards robot-assisted living," in *IEEE RO-MAN'15*, 2015.
2. D. R. Faria, C. Premebida, and U. Nunes, "A probalistic approach for human everyday activities recognition using body motion from RGB-D images," in *IEEE RO-MAN'14, * Kazuo Tanie Award Finalist*, 2014.
3. C. Zhu and W. Sheng, "Realtime human daily activity recognition through fusion of motion and location data," in *IEEE International Conference on Information and Automation*, 2010.
4. ——, "Human daily activity recognition in robot-assisted living using multi-sensor fusion," in *IEEE ICRA'09*, 2009.
5. "Microsoft kinect, Website: https://www.microsoft.com/en-us/kinectforwindows/, accessed on June/2015."
6. "Asus xtion, Website: http://www.asus.com/multimedia/xtion_pro_live/, accessed on June/2015."
7. G. T. Papadopoulos, A. Axenopoulos, and P. Daras, *Real-time Skeleton-tracking-based Human Action Recognition Using Kinect Data*. Springer International Publishing, 2014, ch. 3D and Augmented Reality, pp. 473–483.
8. C. Chen, K. Liu, and N. Kehtarnavaz, "Real-time human action recognition based on depth motion maps," *Journal of Real-Time Image Processing*, 2013.
9. J. Sung, C. Ponce, B. Selman, and A. Saxena, "Unstructured human activity detection from RGBD images," in *ICRA'12*, 2012.
10. L. Xia and J. Aggarwal, "Spatio-temporal depth cuboid similarity feature for activity recognition using depth camera," in *CVPR*, 2013.
11. Y. Zhu, W. Chen, and G. Guo, "Evaluating spatiotemporal interest point features for depth-based action recognition," *Image and Vision Computing*, 2014.
12. S. A. Mehdi, C. Armbrust, J. Koch, and K. Berns, "Methodology for robot mapping and navigation in assisted living environments," in *2nd International Conference on Pervasive Technologies Related to Assistive Environments*, 2009.
13. H. S. Koppula, R. Gupta, and A. Saxena, "Learning human activities and object affordances from RGB-D videos," in *IJRR journal*, 2012.
14. M. Volkhardt, S. Mller, C. Schrter, and H.-M. Gross, "Real-time activity recognition on a mobile companion robot," in *55th Int. Scientific Colloquium*, 2010.
15. V. Arsigny, P. Fillard, X. Pennec, and N. A. 5, "Log-euclidean metrics for fast and simple calculus on diffusion tensors," *Magnetic Resonance in Medicine, 56(2):411–421.*, 2006.
16. K. Guo, "Action recognition using log-covariance matrices of silhouette and optical-flow features," Ph.D. dissertation, Boston University, College of Engineering, 2012.
17. C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM TIST*, 2011, http://www.csie.ntu.edu.tw/ cjlin/libsvm.