Human Activities Recognition with RGB-Depth Camera using HMM

Amandine Dubois, François Charpillet

Human Activities Recognition with RGB-Depth Camera using HMM

Amandine Dubois[1,2] and François Charpillet[2,1]

Abstract— **Fall detection remains today an open issue for improving elderly people security. It is all the more pertinent today when more and more elderly people stay longer and longer at home. In this paper, we propose a method to detect fall using a system made up of RGB-Depth cameras. The major benefit of our approach is its low cost and the fact that the system is easy to distribute and install. In few words, the method is based on the detection in real time of the center of mass of any mobile object or person accurately determining its position in the 3D space and its velocity. We demonstrate in this paper that this information is adequate and robust enough for labeling the activity of a person among 8 possible situations. An evaluation has been conducted within a real smart environment with 26 subjects which were performing any of the eight activities (sitting, walking, going up, squatting, lying on a couch, falling, bending and lying down). Seven out of these eight activities were correctly detected among which falling which was detected without false positives.**

I. Introduction

One of the main problems of next years will be the ageing of the population as well as the dependence of these people which will result of it. So our aim is to allow elderly people to remain autonomous at home for the longest possible time. One of the main preoccupations concerning the security of elderly people at home is to avoid the falls. These falls are often at the origin of a reduction in mobility, of an increase of dependance because people carries out less daily life activities. We want to develop a system allowing to detect if a person fell, for thus avoiding her to remain a long time on the ground, by being able to look after her as fast as possible if necessary, of the physical wounds and also to avoid the aggravation of the psychological consequences which would result from this.

Many systems exist to detect falls. One of the categories consists in systems with sensors, that the person wears on her. These sensors are either accelerometer, gyroscopes or goniometers. These various sensors can be integrated in devices detecting the fall automatically, as shown in article of Bourke *et al.* [4] and Wu [12]. There exists also systems made up of an alarm button, in this case it's the person who must press herself on a button to alert after the fall. But these systems are restricting for the person because she must remember to wear it. Moreover in article [11], the researchers have lead a survey on the elderly people of more than 65 during 1 years. They constitute one group of "faller" (people who already fell at least once at home) and one control group. Among the nine fallers having a call system, two used it successfully after the fall, one attempted but unsuccessfully to use it and the last six didn't attempt to use it. Then an other approach to detect falls is the use of camera.

The main problem of computer vision consists firstly in extracting background to get foreground object (human). Different methods exist: running average ([7], [6], [3]), Gaussian Mixture Model [6], least median of squares [1], occupancy grid [5]. Then the second stage is to track the human in time. Several articles build a rectangular blob around the person. The blobs with smaller size are considered as artifacts and are eliminated. The blobs allow to track a person and to make the third stage of computer vision which is to recognize the behavior of the person. Some use the size of the blob to know if the person is lying, sitting or standing [7]. Or some compute the number of pixels on vertical axis for each blob [6] or for different slices of blobs [3]. For the last, they consider a person has fallen if the sum of the third slices (from of ground) exceed a threshold. Other authors considered the width to height ratio of rectangular blobs [1]. The hypothesis is that when a person is lying on the ground the ratio is much larger. Other researchers are interested in tracking only one part of the body as Rougier *et al.* [9] who track the head to detect falls.

This article is related to this last category of ambient sensor approach, we chose a RGB-Depth camera, more precisely the Kinect camera to detect falls. We made the choice to include the detection of the falls in the more general problem of activity recognition.

This paper is organized as follows. Section II is dedicated to the activity recognition method which is based on the tracking of the center of mass of the person. In Section III, we present results from an experiment undertaken on 26 subjects. Section IV and V are dedicated to discussion and conclusion.

II. Method

Our method is presented in three steps: background extraction, center of mass tracking and posture recognition.

A. Extraction of background

To extract the background we use the "running average" method [6]. This technique allows to learn the background by averaging over time the distances for each point of the depth map. Then, we identify the mobile points in the image, i.e. the point occupied by a mobile object (by a human for example). At each time we substract the background distances from the current distances to keep only mobile

[1]Université de Lorraine, LORIA, UMR 7503, Vandœuvre-lès-Nancy CEDEX, F-**54506**, France amandine.dubois at loria.fr

[2]Inria, Villers-lès-Nancy, F-**54600**, France francois.charpillet at inria.fr

points. In other terms the mobile points are the points having a different depth from the one of the background. To eliminate the noise, i.e. the points detected as mobile but not being, we use the "Erode, Dilate" fiters.

The real world coordinate system is obtained by using the Kinect factory optical parameters and we compensate the Kinect tilt angle (read from the Kinect accelerometer sensor) by a rotation on X-axis. We use OpenNI for Kinect to real world transformations.

B. Tracking a person and her center of mass

The aim of the second stage is to track the center of mass of the person. First, we gather the mobile points (2D pixels) belonging to the same object, so as to be able to distinguish several persons in the same scene, using the "Component labelling" method [10]. Then, we calculate the center of mass of the person as the average location of all the mobile points belonging to the same object.

C. Recognition of the activity of a person

In this part we present the method for recognizing the activity of a person and detecting the falls. Our method uses a Hidden Markov Model (HMM).

1) HMM models: In our model we define eight activities, postures that a person can take : walking, lying (on a bed, on a couch for example), sitting, falling, lying down, squatting, going up on an obstacle (a chair, a footboard for example) and bending. These eight postures are represented by the eight states of our HMM. The representation of this HMM is shown in Figure 1. The meaning of each state is represented by the picture. The state "walking" includes the position upright.

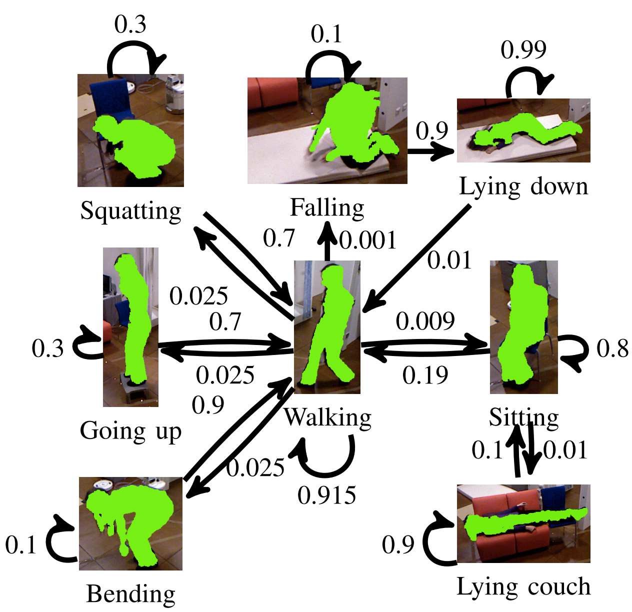


Fig. 1. HMM to eight states.

2) Observation function: In our HMM, the observations are: the vertical position of the center of mass, the vertical speed and the standard deviation of all the points belonging to the person. The observation function follows a multidimensional normal law whose parameters are calculated from the data of 16 subjects visiting the eight states of the HMM. More explanations on this experiment are given in Section III-A.

3) Inference: To calculate the probability of being in one of the eight states of the HMM, we implemented the Forward-Backward algorithm [8]. We make the hypothesis that it's possible to be in each state at the beginning of the analysis by giving the same initial probability to each state (1/8).

III. Results

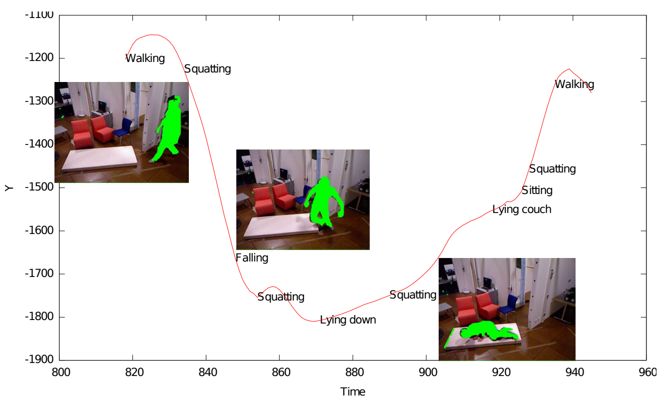
In this part we present the experimental procedure and the results.

A. Description of the experiment

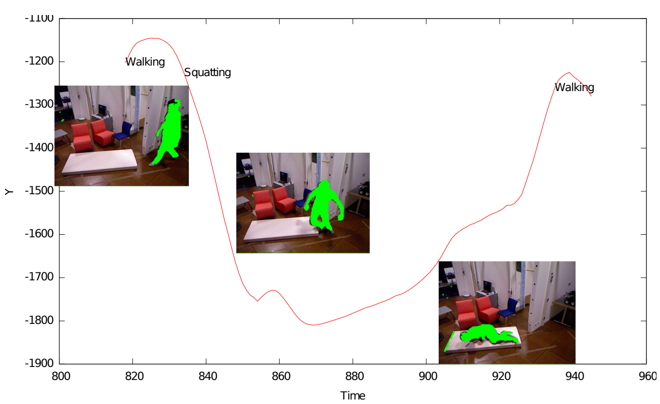
We made an experiment with 26 subjects (8 women and 18 men) of 20 at 53 years old in order to test our algorithm. Each subject realized eight situations corresponding to the eight states of the HMM. In Figure 1 we can see eight pictures corresponding to what was requested from the subjects. The state "squatting" consist in collecting a pen on the ground by putting the knee on the ground. The state "bending" consist also in collecting a pen on the ground but without putting the knee on the ground. The state "sitting" is to sit on a chair. The state "walking" is to walk across the scene. The state "falling" is the action to fall intentionally on a mattress and after when the person has fallen on the ground it's the state "lying down". The state "going up" consists in going up on a footboard. And the state "lying couch" is lain on three chairs. To learn the observation function as shown in Section II-C.2 we used the data of 16 subjects and we tested with Forward-Backward the validity of HMM model with the data of 10 other subjects.

B. Result with Forward-Backward

We have tested in which state was classified each situation of the 10 subjects not belong to the training group. The result is shown in Table I. Table I represents the number of subjects where situation is correctly classified by the state (for example 9/10 corresponds to 9 subjects out of 10 for whom the classification is good). The test has been made for Forward-Backward algorithm. We can see in Table I that there are few errors except for the state "bending" replaced for some subjects by state "sitting" and for other subjects by the state "squatting". The problem is that the observation doesn't allow to dissociate this state enough from the states "sitting" and "squatting", even visually. Another problem of classification, more important, is the error concerning the non detection of a fall. By watching the fall of the subject for whom there is the error of classification, we can notice that the subject rose immediately after the fall and that he fall with a lower speed so the fall was not realistic. We have tested with only Forward algorithm without Backward. We obtain Figure 2(a). In figure we can see the variation in the time of the center of mass on the vertical plan. We include of the images to show the behavior that the person realized in function of time. To finish in this curve we can see that the algorithm writes the state on the curve each time that this algorithm determines that there is a state transition. We can notice that the model change between "squatting" and "falling" in figure 2(a). However when we add the Backward algorithm we obtain Figure 2(b) where the model classified as "squatting" the activity. We looked at the result for other subjects with only Forward algorithm and we noticed that the model hesitates also between "squatting" and "falling" but for these other subjects the model with Backward algorithm concludes to the state "falling". The difference is that for these other subjects the following state is "lying down" for a long time and it's the reason why the model concludes "falling". Because "falling" is the only state which allows to have the state "lying down" after. The subject, where the model is false, didn't pass (or pass for less than a few seconds) by the state "lying down" and as it's impossible to be in the state "falling" without passing in the state "lying down" then the model concludes with Backward algorithm that it's not a fall but that the subject is "squatting".



(a) Analysis with Forward algorithm.



(b) Analysis with Forward-Backward algorithm.

Fig. 2. Result of the algorithm analysing a fall.

When we made the experiment a subject didn't realize the situation "lying couch" as the other subjects. He is squatted on the couch before lying. We have not put this subject in the training data but we have tested the model on this situation. The model and the other subjects in the training data pass to the state "sitting" before to be in "lying down". The result is shown in Figure 3. The algorithm has infered the right situation, the model is in the state "sitting" even if the subject didn't sit, he is squatting on the couch but this state doesn't exist in the model. This seems to indicate that our model is robust.

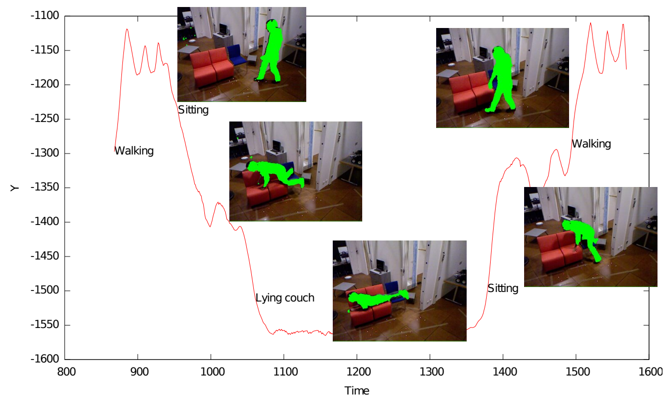


Fig. 3. The situation "lying couch" realized by a subject.

We want to know if our algorithm is really robust to the change of situations because in the experiment each person made the situation in the same place and each situation was realized one by one. So we have tested with a new situation for which we asked the subject to squat back to the camera, then to walk more further compared to the test situation and to stop at a table. Then the person must sit on the chair (placed differently compared to the previous tests) and to finish the person must lie on the couch placed perpendicularly to the Kinect (in previous test the couch was placed in front of the Kinect). The result is shown in Figure 4. We can see that the algorithm is robust because it detected all the situations even if this sequence is not in the training base.

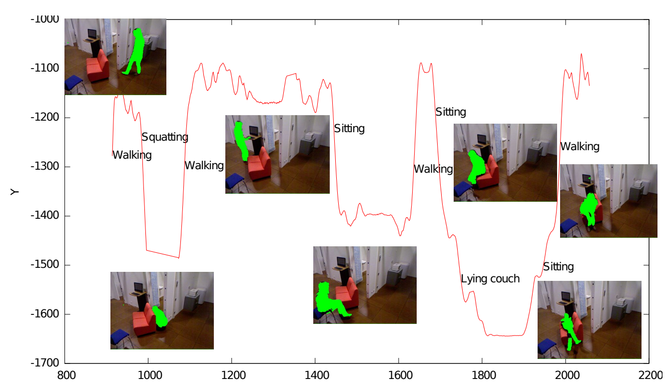


Fig. 4. Result with Forward-Backward algorithm for a situation realized in different conditions compared to training data.

	Squatting	Lying couch	Sitting	Falling	Lying down	Walking	Going up	Bending
Correct classification	10/10	10/10	10/10	9/10	9/9	10/10	10/10	0/10
Sensitivity (%)	100	100	100	90	100	100	100	0
Specificity (%)	93	100	90	100	100	100	100	100

TABLE I
SENSITIVITY AND SPECIFICITY OF THE CLASSIFICATION OF DIFFERENT ACTIVITIES USING THE FORWARD-BACKWARD ALGORITHM

IV. DISCUSSION

The results show that our algorithm allows a correct identification of the states. In Table I we calculate the sensitivity and specificity for each activity. The sensitivity is the capacity to detect a state when it is present and the specificity is the capacity of the system to detect the absence of a state when it doesn't appear. Concerning the falls the sensitivity is 90% and the specificity is 100%. Thus from the results we can conclude that our algorithm detects the falls (when it is followed by the state "lying down") and that there are no false positives i.e. other situations, as sitting, lying on a couch, could have been detected as a fall but it wasn't not the case. We can compare our algorithm to the other algorithm in the litterature using cameras. We take the review made by Auvinet *et al.* [3] of the sensitivity and specificity obtained by different authors. We have described the method of these following authors in the introduction. First the article of Rougier *et al.* [9], from a 2D camera they extract 3D information and track the head of a person. They realized 19 sequences, nine of which are different falls and 10 are normal activities as sitting down, standing up, crouching down. They obtained a sensitivity of 95,5% and a specificity of 96,4%. Anderson *et al.* [2] using several cameras and experimenting with 14 falls and 32 no falls activities. They obtained a sensitivity of 100% and a specificity of 93,75%. Auvinet *et al.* [3] deals with the problem of occlusion. From 4 cameras and the occlusions in the scene they obtain a sensitivity and specificity of 100%. They test on 22 falls and 24 other activities (crouching, sitting, lying on a sofa). We don't deal with the occlusions in our paper.

A specificity of our work compared to others in the litterature is that we have a training phase. This implies that we need more data for the experiment. Each one of the 26 subjects realized 8 situations for a total of 208 sequences. A part of our data is used to train the model (128 sequences) and the other for the validation (80 sequences).

The goal of our project is to allow elderly people to stay longer at home. One of solution is to detect the falls but also to analyse the activity of the person to detect an eventually loss of autonomy. In our work we can detect the falls but also discriminate other activities as lying on a couch or bed, sitting, walking, squatting and going up. Thus we can count the time passed sitting or lying compared to the time passed walking at home each day and detect an eventual modification of activity of the person. For example passing more time sitting compared to previous months may be a sign of modification of the behavior and maybe of loss autonomy if the person doesn't change.

V. CONCLUSION

In this paper we want to develop a system able to detect falls and the activity of elderly people using a low-cost system with one Kinect camera. Our algorithm extract the background with method "running average" to obtain only mobile pixels. After we gather these mobile pixels belonging to the person. Thus we can track this person and more precisely we track her center of mass. In this paper we want to identify which activity the person made. For distinguishing several behaviors we create a HMM with eight states corresponding to eight situations of a daily life. We realize an experiment to verify the accuracy of the model. The results show that the model provide a good classification of the situations. The algorithm detect the falls without false positives. The algorithm can also distinguish between a person sitting, walking, squatting, going up, lying on a couch and lying down. The experiment was made on healthy subjects thus our training data is built for no real fall of elderly people. In the future we plan on setting up a longitudinal experiment to obtain of real falls.

REFERENCES

[1] D. Anderson, J. M. Keller, M. Skubic, X. Chen, and Z. He. Recognizing falls from silhouettes. *IEEE EMBC*, 1:6388–91, Jan. 2006.

[2] D. Anderson, R. H. Luke, J. M. Keller, M. Skubic, M. Rantz, and M. Aud. Linguistic Summarization of Video for Fall Detection Using Voxel Person and Fuzzy Logic. *CVIU*, 113(1):80–89, 2009.

[3] E. Auvinet, F. Multon, A. Saint-Arnaud, J. Rousseau, and J. Meunier. Fall detection with multiple cameras: an occlusion-resistant method based on 3-D silhouette vertical distribution. *IEEE transactions on information technology in biomedicine : a publication of the IEEE EMBS*, 15(2):290–300, Mar. 2011.

[4] A. K. Bourke, J. V. O'Brien, and G. M. Lyons. Evaluation of a threshold-based tri-axial accelerometer fall detection algorithm. *Gait & posture*, 26(2):194–9, July 2007.

[5] A. Dubois, A. Dib, and F. Charpillet. Using HMMs for Discriminating Mobile from Static Objects in a 3D Occupancy Grid. *IEEE 23rd ICTAI*, pages 170–176, Nov. 2011.

[6] I. Gómez-Conde. Simple Human Gesture Detection and Recognition Using a Feature Vector and a Real-Time Histogram Based Algorithm. *Journal of Signal and Information Processing*, 02(04):279–286, 2011.

[7] B. Jansen, F. Temmermans, and R. Deklerck. 3D human pose recognition for home monitoring of elderly. *IEEE Engineering in Medicine and Biology Society. Conference*, 2007:4049–51, Jan. 2007.

[8] L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. In *Proceedings of the IEEE*, pages 257–286, 1989.

[9] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau. Monocular 3D head tracking to detect falls of elderly people. *IEEE EMBC*, 1:6384–7, 2006.

[10] K. Suzuki, I. Horiba, and N. Sugie. Linear-time connected-component labeling based on sequential local operations. *Computer Vision and Image Understanding*, 89(1):1–23, Jan. 2003.

[11] D. Wild, U. S. Nayak, and B. Isaacs. How dangerous are falls in old people at home? *British medical journal*, 282(6260):266–8, 1981.

[12] G. Wu. Distinguishing fall activities from normal activities by velocity characteristics. *Journal of biomechanics*, 33(11):1497–500, 2000.