Detecting Human Falls with 3-Axis Accelerometer and Depth Sensor

Michal Kepski[1] and Bogdan Kwolek[2]

***Abstract*— Previous work demonstrated that Kinect sensor can be very useful for fall detection. In this work we present a novel approach to fall detection that allows us to achieve reliable fall detection in larger areas through person detection and tracking in dense depth map sequences acquired by an active pan-tilt 3D camera. We demonstrate that both high sensitivity and specificity can be obtained using dense depth images acquired by a ceiling mounted Kinect and executing the proposed algorithms for lying pose detection and motion analysis. The person is extracted using depth region growing and person detection.**

I. INTRODUCTION

Falls are the considerable cause of injury death among people 65 years and over and are one of the most common cause of hospital admissions for traumatic injuries and loss of independence. More than half of all falls occur within home environment and significant percentage of them happens at nighttime [9]. An injured elderly may be lying on the ground for several hours or even days after a fall incident has occurred. Thus, various kinds of sensors were applied in fall detectors to investigate their sensing properties [3].

Very often fall detection devices use accelerometers and gyroscopes to sense sudden changes in movements. However, detectors based only on inertial sensors often fail to separate accidental falls from ordinary activities such as sitting on a chair, lying down on a bed, or any kind of sudden movements. What's more, they generate too much false alarms. This means that some activities of daily living (ADLs) are erroneously reported as falls, which in turn leads to considerable frustration of the seniors. A robust fall detector ought to classify falls as falls (sensitivity) and the non-falls as non-falls (specificity) under all real life conditions [3].

Vision systems, like other non intrusive methods, are often utilized in prototype systems for fall detection. Moreover, due to widespread availability of webcams the number of studies in vision-based systems for fall detection is still increasing. The rapid development of CCD sensors, cameras, and computer technologies make such systems feasible. However, continuous monitoring through a vision system introduces some ethical issues concerning the respect of intimacy and privacy, especially in the bedroom and the bathroom. What's more, the existing video-based devices for fall detection cannot work in nightlight or low light conditions. Additionally, the lack of 3D information can lead to a lot of false alarms.

Recently, Kinect sensor was proposed for fall detection [6], [5]. If the detection is done on depth maps only, the system preserves privacy. More recent work demonstrated that reliable fall detection can be obtained through the use of a ceiling-mounted Kinect [4]. However, the observation area in such a sensor setup is quite limited. As indicated in [4], for an overhead Kinect mounted on the height of 2.6 m from the floor the observation area is about 5.5 m^2. To extend the observation area as well as to increase the functionality of the overhead camera based fall detection, we propose to utilize a pan-tilt head for pursuing a moving subject in larger areas. To achieve such aim we propose reliable algorithms for person detection and tracking as well as a robust algorithm for fall detection under real life conditions. A dataset with both RGBD image sequences and accelerometric data for training as well as evaluating the fall detection algorithms has been developed and made publicly available [1]. The proposed algorithms achieve both very high sensitivity and specificity.

II. DATASET

Two datasets in indoor environment were recorded using both an accelerometer and Kinect sensor. The first dataset has been taken with ceiling-mounted, stationary Kinect, whereas the second one has been taken with overhead pan-tilt Kinect. They consist of sixty RGBD image sequences and the corresponding accelerometric data. The RGB and depth images were recorded by Kinect XBOX 360 with 25 frames per second, whereas the accelerometric data was recorded by x-IMU device with 256 Hz. All sequences contain falls of which half of them concerns falling from sitting on a chair. Figure 1 illustrates depth maps based person tracking in varying illumination conditions with active depth sensor.

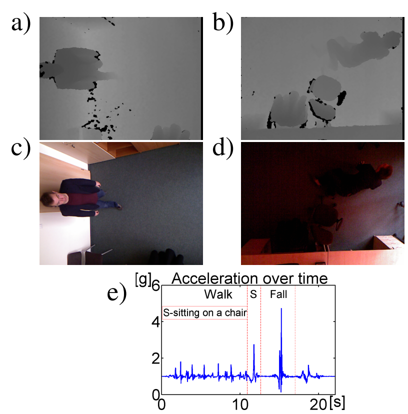

Fig. 1. Person tracking using depth maps acquired by active camera. a-b) - depth images, c-d) - corresponding RGB images, e) acceleration over time.

[1]M. Kepski is with Faculty of Mathematics and Natural Sciences, University of Rzeszow, 35-959 Rzeszów, Poland mkepski at ur.edu.pl
[2]B. Kwolek is with the Department of Computer Science, AGH University of Science and Technology, 30 Mickiewicza Av., 30-059 Krakow, Poland bkw at agh.edu.pl

[1]UR Fall Detect. Dataset http://fenix.univ.rzeszow.pl/~mkepski

978-1-4244-7929-0/14/$26.00

III. METHOD

A. Overview of the method

The algorithm begins with nearest neighbor interpolation to fill the holes in the depth map and to get the map with meaningful values for all pixels. The median filter with a 5×5 window on the depth array is executed to make the data smooth, see Fig. 2. To simplify the person extraction the algorithm extracts the floor and then removes their corresponding pixels from the depth map. Given the extracted person in the last depth frame, the region growing is executed to delineate the person in the current frame. A Support Vector Machine (SVM) based person finder is utilized then to confirm the presence of the tracked subject as well as to give his/her head location. On the basis of the person's centroid the pan-tilt head rotates the camera to keep his/her head in the central part of the depth map. Given the delineated person, a cascade classifier consisting of lying pose detector and dynamic transition detector is executed. Optionally, if an accelerometer is used, the cascade classifier is triggered only in the case of potential fall, see Fig. 2.

B. Human detection in depth sequences from active camera

The pursuing a moving person is achieved by a series of saccades of the pan-tilt head to keep the detected object in central part of the depth maps. The object position is expressed as the centroid of the delineated area. Below we detail the depth region growing that delineates the person undergoing tracking in maps acquired by an active camera. We also present an algorithm that supports locating the person's head in case of region chaining.

1) Delineation of person using region growing: The person is delineated assuming that he/she occupies an integrated region in 3D space. Owing to extracting the floor in advance, we avoid incorporating of the neighboring pixels from the floor into the person region. The developed depth region growing starts with selecting a seed point in a current frame. Assuming that there is a common depth region between regions belonging to a person in two consecutive frames, such seed region is determined using the `and` operator between the previously delineated depth region belonging to person and the current depth map. Afterwards, the algorithm repeatedly seeks all neighboring pixels of the current region. The selected pixels are sorted according to their depth similarities and are stored in a list of candidate pixels. The depth similarity is the Euclidean distance between the depth values of a pixel from such a list and its closest pixel from the current region. It is employed in order to verify if a neighboring pixel around a region pixel is allowed to be merged with the region.

2) Finding human in depth maps: Ordinary region growing algorithms suffer from the problem of region chaining (overspill), which occurs when two regions are grown into one region while they are actually separated from each other. In order to improve the delineation of the person in such situations as well as to improve the pursuing of the person by the active camera, we execute a person detector operating on window surrounding the segmented region. The detector permits also automatic initialization of person tracking. The person detection is done by a SVM for linear classification that is built on Histogram of Oriented Depths (HOD) features [7]. The HOD descriptors locally encode the orientation of depth changes and in our person finder they are calculated in sub-windows, which are scaled according to their distances to the camera. The scaling is according to the distance between the camera and the closest pixels from the sub-window. Such sub-windows of fixed size are then subdivided into cells. The descriptors are calculated for each cell and then the oriented depth gradients are collected into 1D histograms.

C. Fall indicating using body-worn accelerometer

Compared to vision-based motion analysis systems, wearable sensors offer several advantages, particularly in terms of cost, ease of use and, most importantly, portability. They are the only sensors that are used in real fall detection systems and outside laboratories. However, despite many advantages, the inertial sensors-based technology does not meet the seniors' needs, because some activities of daily living are erroneously reported as falls. The smartphones serve not only as communication and computing devices, but they also come with a rich set of embedded sensors, such as accelerometer, gyroscope and digital compass. Therefore, increasing interest on using this technology for fall detection is observed and the number of relevant papers grows considerably. Being aware of shortcomings of current solutions we believe that such technology will be significantly enhanced and in combination with small devices like smart watches it will be very usefull in fall detection. Thus, our system can optionally take data from wireless accelerometer.

A lot of different techniques for inertial sensors were proposed to achieve reliable fall detection [1]. Frequently, a single body-worn sensor (tri-axial accelerometer or gyroscope, or both embedded in an IMU) is used to indicate person fall. Tri-axial accelerometer is the most used device. The accelerometer-based algorithms raise the alarm when the signal reaches a certain threshold value. In [2] an

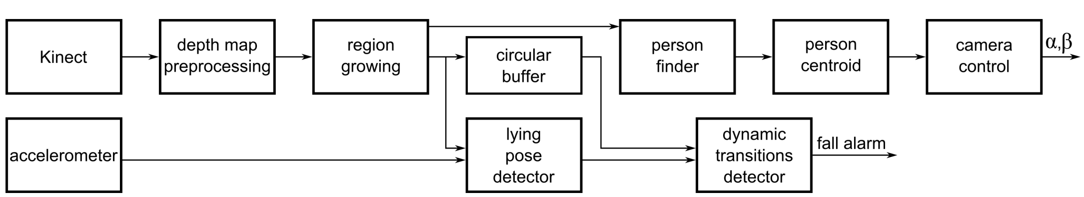

Fig. 2. Block diagram of fall detection system.

accelerometer-based algorithm, relying on change in body orientation has been proposed. It signals a potential fall if the root sum vector of the three squared accelerometer outputs exceeds an assumed threshold value.

In our algorithm a fall is indicated if the signal upper peak value (UPV) from the accelerometer is greater than 2.5 g. A survey of the relevant literature reveals that for a single inertial device the most valuable information can be obtained for devices attached near the centre of subject mass. Therefore, the accelerometer was attached near the spine on the lower back using an elastic belt around the waist.

D. Lying pose recognition

The lying pose has been distinguished from ADLs using classifiers trained on features representing the extracted person in the depth maps. We selected 214 maps from UR Fall Detection Dataset with normal activities like walking, sitting down on a chair, taking or putting an object from floor, bending right. Such representative images were then used to train a k-NN classifier and a linear SVM classifier responsible for checking whether a person is lying on the floor. Both classifiers have been trained on three features [4]: (i) H/H_{max} - a ratio of head-floor distance to the height of the person, (ii) $area$ - a ratio expressing the person's area in the image to the area at assumed distance to the camera, (iii) l/w - a ratio of major length to major width of a blob representing the person on the depth image.

E. Dynamic transitions for fall detection

Person fall entails an abrupt and significant change of head-floor distance with accompanying change from a vertical orientation to a horizontal one. The distance of the person's centroid to the floor also changes significantly and rapidly during the accidental fall period. In the images acquired from a ceiling mounted camera the area ratio also changes considerably in the case of the fall. Thus, through an analysis of the cues above mentioned we can determine whether a transition of the body is intentional or not.

The ratio $H(t)/H(t-\Delta T)$, where $H(t)$ is determined in the moment of the impact, and $H(t-\Delta T)$ is calculated ΔT before the fall, quite reliably characterizes the dynamics of the fall using a ceiling mounted Kinect. In depth images acquired by an overhead camera the peak value of $H(t)/H(t-\Delta T)$ is far below one. The ratio $H(t)/H(t-\Delta T)$ can also be determined through analysis of depth image pairs. However, the use of accelerometer as indicator of the potential fall simplifies calculation of this ratio since the time t can be determined easily and with low computational cost.

IV. Experimental results

In order to be accepted by seniors, a fall detection system should be unobtrusive and cheap, and particularly it should exhibit both high sensitivity and specificity as well as should preserve user's privacy. The proposed algorithm has been designed with regard to such factors through sedulous choice of its ingredients as well as arrangement of scenarios for training and evaluation. Below we discuss evaluation results.

A. Evaluation of the fall detector

In order to be accepted by seniors the fall detectors should be accurate and reliable. The accelerometers are the only sensors that acknowledged their effectiveness in long-time evaluations, conducted outside of the laboratory. Thus, our system can be configured to utilize both accelerometric and depth data or depth data only. In the case of using the accelerometer for signaling the potential fall, the impact is detected reliably and with low computational cost. In such a configuration the depth analysis is used to verify the fall hypothesis. Moreover, the depth map analysis is done only in the case of signaling a potential fall. The use of the accelerometer as indicator of the potential fall simplifies also the extracting of the dynamic features since the moment of the impact can be determined easily.

At the beginning of the evaluation we judged the usefulness of the accelerometer as an indicator of potential fall. The actors performed typical daily activities consisting in walking, taking or putting an object from floor, bending right or left to lift an object, sitting down on a chair, tying laces, crouching down and lying. For the carried out activities during half an hour experiment the acceleration values $2.5 - 3$g were exceeded several times. This means that within such a relatively short period of time of typical human activity, a significant number of false alarms would be generated if the fall detection was carried out only on the basis of the acceleration data. In particular, we noticed that all fall-like activities were indicated properly. In consequence, the chosen accelerometer acknowledged his usefulness as reliable indicator of the person impact.

The algorithm for lying pose recognition has been evaluated on 351 representative images from UR Fall Detection Dataset of which 61% were training examples. A linear SVM and k-NN with 5 neighbors have been trained on features discussed in Section III-D to classify falls and ADLs. As we can notice in Tab. I the lying pose detector achieves very promising results. In particular, accuracy and sensitivity are higher than 99%.

TABLE I

PERFORMANCE OF LYING POSE DETECTION [%].

	accuracy	precision	sensitivity	specificity
SVM	98.29	97.13	99.41	97.24
k-NN	99.15	98.83	99.41	98.90

In order to diminish the ratio of false alarms, and in particular, to distinguish the intentional lying on the floor from the accidental falling, we evaluated the usefulness of the dynamic feature. Figure 3 demonstrates the plots of $H(t)/H(t-\Delta T)$ for accidental fall and intentional lying on the floor. As we can observe, for ΔT equal to 600 ms the threshold that is set to 0.6 has been exceeded for the fall. Our evaluation demonstrated that a cascade classifier consisting of lying pose detector and dynamic transition detector has almost null ratio of false alarms. Very rarely the classifier can generate false alarm, mainly, due to imperfect detection of the moment of the body impact on the basis of only vision techniques. A cascade classifier extended about additional block signaling a potential fall on the basis of accelerometric data demonstrated null false alarm. In particular, all falls were detected properly on images from UR fall dataset.

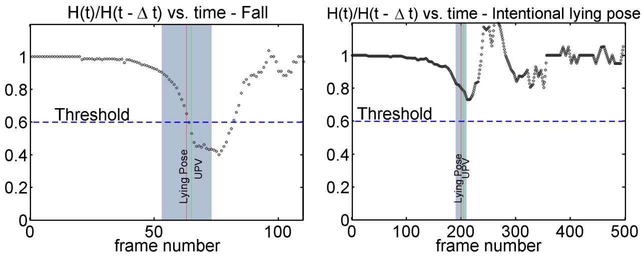

Fig. 3. $H(t)/H(t-\Delta T)$ vs. time for fall and intentional lying pose.

B. Evaluation of person detector and tracker

The performance of the fall detection depends strongly on the robustness of person detection and tracking. Therefore, a considerable attention has been paid to such issues. In particular, a hand-made two degrees-of-freedom pan-tilt head has been employed in the experiments to extend the usability of the fall detector. Thanks to pan-tilt capabilities the monitoring area of the system is extended considerably.

The region growing has been evaluated on the first dataset. The persons were extracted manually in the images starting the subsequences for the person segmentation. In all images including maps with falls all main body parts were extracted. The region growing has also been evaluated in experiments with active camera, where the aim was not only to extract in real-time the person but also to keep he/she in the central part of the depth maps. In particular, the person was extracted with sufficient precision on all images from the second dataset. Figure 4 depicts some results with delineated person, which were obtained on depth images acquired by the active camera.

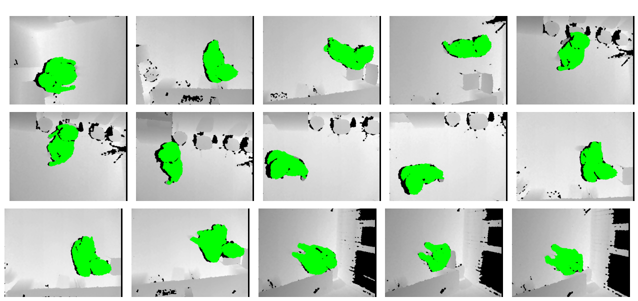

Fig. 4. Person delineation on depth maps using region growing.

The person detector has been evaluated on 254 positive samples and 638 negative samples of which 60% were used for training. The images with delineated person were scaled according to distance of his/her head to the camera. They were also rotated to a canonical pose using the axis of the person's blob. Table II illustrates results that were obtained using the detector discussed in Section III-B.2. As we can observe, the results are better if the silhouettes are rotated to the canonical pose. On the other hand, the difference is not considerable, and this means that the algorithm is quite resistant to various poses. This is because the gradients on the head in depth images seen from an overhead camera form elliptical like structures.

TABLE II

PERFORMANCE OF PERSON DETECTION [%].

	accuracy	precision	sensitivity	specificity
rotat.	99.45	98.21	100.0	99.22
no rotat.	98.91	98.18	98.18	99.22

The discussed results were obtained for HOD cell size equal to 8×8. The results are better than those presented in [8]. The system was implemented in C/C++ and runs at 30 fps on 2.4 GHz I7 (4 cores, Hyper-Threading) notebook, powered by Windows.

V. CONCLUSIONS

In this work we presented a novel approach to fall detection. Through introducing a ceiling-mounted depth sensor that is driven by a pan-tilt unit we extended the monitoring area. The proposed algorithms for real-time person detection and tracking on the basis of depth maps acquired by the active camera allow us to achieve both high sensitivity and specificity of fall detection in poor lighting conditions. Very promising results were obtained on long depth map sequences using the proposed lying pose detector together with dynamic transitions analysis. The system can optionally use an accelerometer to indicate the potential fall. In such a configuration the dynamic features can be determined more precisely and in consequence the possibility of false alarm is smaller. Thanks to the use of depth maps only, the system preserves the user privacy.

ACKNOWLEDGMENT

This work has been partially supported by the National Science Center within the project N N516 483240.

REFERENCES

[1] A. Bourke, J. O'Brien, and G. Lyons. Evaluation of a threshold-based tri-axial accelerometer fall detection algorithm. *Gait & Posture*, 26(2):194–199, 2007.

[2] J. Chen, K. Kwong, D. Chang, J. Luk, and R. Bajcsy. Wearable sensors for reliable fall detection. In *IEEE-EBMS*, pages 3551–3554, 2005.

[3] R. Igual, C. Medrano, and I. Plaza. Challenges, issues and trends in fall detection systems. *BioMedical Eng. OnLine*, 12(1):1–24, 2013.

[4] M. Kepski and B. Kwolek. Fall detection using ceiling-mounted 3d depth camera. In *Int. Conf. VISAPP*, pages II:640–647, Jan. 2014.

[5] M. Kepski, B. Kwolek, and I. Austvoll. Fuzzy inference-based reliable fall detection using Kinect and accelerometer. In *11th Int. Conf. on Art. Intell. and Soft Comp.*, LNCS, vol. 7267, pages 266–273, May 2012.

[6] G. Mastorakis and D. Makris. Fall detection system using Kinect's infrared sensor. *J. of Real-Time Image Processing*, pages 1–12, 2012.

[7] L. Spinello and K. Arras. People detection in rgb-d data. In *IEEE/RSJ Int. Conf. on IROS*, pages 3838–3843, Sept 2011.

[8] L. Xia, C.-C. Chen, and J. Aggarwal. Human detection using depth information by Kinect. In *CVPR Workshop HAU3D*, pages 15–22, 2011.

[9] Z. Zhang, U. Kapoor, M. Narayanan, N. Lovell, and S. Redmond. Design of an unobtrusive wireless sensor network for nighttime falls detection. In *IEEE-EMBS*, pages 5275–5278, Aug 2011.