Fall Detection with Kinect in Top View: Preliminary Features Analysis and Characterization

Susanna Spinsante(✉), Manola Ricciuti, Enea Cippitelli, and Ennio Gambi

Dipartimento di Ingegneria dell'Informazione, Università Politecnica delle Marche, 60131 Ancona, Italy
s.spinsante@staff.univpm.it
http://www.tlc.dii.univpm.it

Abstract. Fall detection is a well investigated research area, for which different solutions have been designed, based on wearable or ambient sensors. Depth sensors, like Kinect, located in front view with respect to the monitored subject, are able to provide the human skeleton through the automatic identification of body joints, and are typically used for their unobtrusiveness and inherent privacy-preserving capability. This paper aims to analyze depth signals captured from a Kinect used in top view, to extract useful features for the automatic identification of falls, despite the unavailability of joints and skeleton data. This study, based on a set of signals captured over a number of test users performing different types of falls and activities, shows that the speed of falling computed over the blob identifying the person, extracted from the depth images, should be used as a feature to spot fall events in conjunction with other metrics, for a better reliability.

Keywords: Fall detection · Depth image processing · Blob Features · Speed of falling

1 Introduction and Background

In recent years, many consumer electronics devices and products, like traditional appliances, have found new and sometimes unexpected adoption as different tools, usually as sensors. In fact, in the broadest definition, a sensor is "*an electronic component, module, or subsystem whose purpose is to detect events or changes in its environment and send the information to other electronics, frequently a computer processor*" [7]. With respect to this definition, a common appliance, like a fridge, once turned into a *smart* object, may become a sensor to detect and analyse events related to Activities of Daily Living (ADLs) like cooking or having meals, and enable an unobtrusive behavioral analysis to recognize anomalous habits [4,5]. The same happened with a commercial product designed and shipped for gaming purposes by Microsoft, the Kinect, a motion

B. Guidi et al. (Eds.): GOODTECHS 2017, LNICST 233, pp. 153–162, 2018.
https://doi.org/10.1007/978-3-319-76111-4_16

sensing input device for video game consoles and Microsoft Windows PCs. Based around the concept of a webcam-style add-on peripheral, Kinect enables users to control and interact with their console or computer without the need for a game controller, just using gestures and spoken commands, i.e. through a natural user interface. Since Kinect became fully available for PC users, together with its Software Development Kit (SDK), many researchers started using it for applications not related to gaming, but pertaining to gesture, action and activity recognition and computer vision, in a broad sense, spanning from device-free interaction with other systems or devices, to gait and posture analysis, to fall detection and remote rehabilitation. Examples of the aforementioned applications may be extensively found in the literature (see [1,3,6], among others).

Among the possible fields of application of the Kinect device, this paper investigates its use in fall detection, assuming a specific constraint. In fact, differently from most of the literature available on this topic, which assumes to use the sensor in a front view configuration, we use Kinect in a top-view setting, i.e. the sensor is installed on the ceiling of a room (like the lab into which tests and experiments have been carried out). This way, the subject's skeleton, and the joints' coordinates Kinect is able to compute when used in front view, are not available for processing. In a previous paper [2], we presented an algorithm for automatic fall detection exploiting the Kinect in the same top-view configuration. That work based the fall detection capability on a number of depth image processing functions, able to recognize the blob of a person and check its geometric features against a number of anthropometric thresholds, and on comparing the relative height of the blob with respect to the floor (details are available in [2]).

In this paper, we extensively test the algorithm on different types of falls, to identify a suitable feature (generated by processing the blob data) for the purpose of automatic fall detection and classification. In fact, we focus on the speed of movement of the person's blob during the fall, and on the amount of variation of the blob's height with respect to the floor, to check if these figures may be used to discriminate between a fall or an ADL, or among different types of falls. Again, several works in the literature build upon this idea, but they exploit the subject's head joint coordinates provided by the sensor used in a front-view configuration, or the center of mass coordinates computed from the joints. We opt for a top-view setup, as it results to be far less obtrusive in a real-life scenario, and even more robust to possible occlusions due to objects, like furniture, located in the monitored environment.

The paper is organized as follows: Sect. 2 presents both materials and methods used in our study, whereas Sect. 3 describes and discusses the experiments carried out and the results obtained. Finally, Sect. 4 concludes the paper.

2 Materials and Methods

The system setup adopts a Kinect device in top view configuration, at a maximum height of 3 m from the floor, thus providing a coverage area of 8.25 m^2. To extend the monitored area, in principle the sensor could be elevated up to around 7 m (far higher than typical living environments); beyond this distance the depth data become unreliable [8]. The algorithm described in [2] works with raw depth data (given in millimeters), that are captured at a frame rate of 30 fps with a resolution of 320×240 pixels, using the Microsoft SDK v.1.5. The depth signal output by the sensor is filtered to reduce noise and to improve the blob identification. Once the person's blob is identified, the algorithm assigns it a *centroid*, the 3D coordinates of which are stored during the activity execution. The fall detection tool identifies the falls and uses different colors (yellow, red, green) to notify conditions of warning, fall and recovery, respectively. In this work, we use the 3D coordinates of the centroid and process them to obtain its speed of movement along the vertical direction (z axis).

2.1 The Fall Detection Algorithm

The fall detection algorithm, the details of which have been published in [2], relies upon some basic operations that are schematically shown in Fig. 1 and can be summarized as follows:

- preprocessing and segmentation: the incoming depth frame is pre-processed to enable the succeeding steps, and a so-called *reference frame* is generated to improve the identification of human subjects;
- distinguish object step: this component of the algorithm identifies, splits, and classifies into objects the different clusters of pixels in the reference frame;
- identification of human subject: starting from the set of separated objects, those representing a human subject are identified by checking several anthropometric relations;
- subject tracking and fall detection: the system tracks the movements of the human subjects in the depth frames, and detects if a fall occurs. The possible fusion of blobs occurring when two or more subjects get in contact is properly handled.

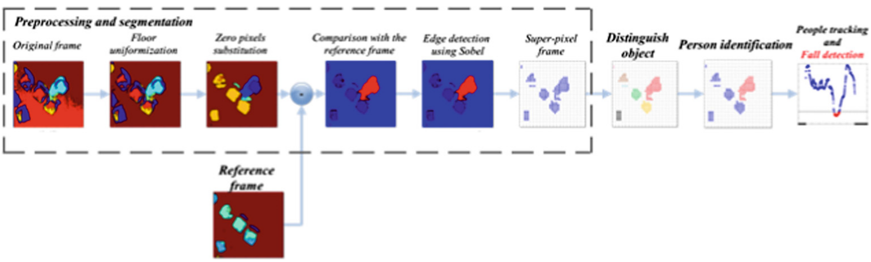

Fig. 1. Overview of the fall detection system (from [2])

The algorithm, conceived and designed with low complexity in mind, identifies the falls by comparing the depth values taken by the pixels representing the human subject's centroid to a threshold empirically set to 400 mm from the floor. When the detected height of the centroid undergoes the threshold, the algorithm triggers an alarm to notify a fall event. As shown in Fig. 2, the fall detection implemented by the algorithm depends only on the relative height of the centroid to the threshold, not on its time variation along the vertical axis. By exploiting the data provided by the algorithm (the coordinates of the centroid), and based on the knowledge of the frame rate, in this work we investigate the relation between the time variation of the centroid vertical position (i.e. its velocity along the z axis) and the type of action (fall or ADL) simulated by the subject, to possibly identify a useful feature for automatic action/fall recognition independent from the threshold value that needs to be heuristically defined.

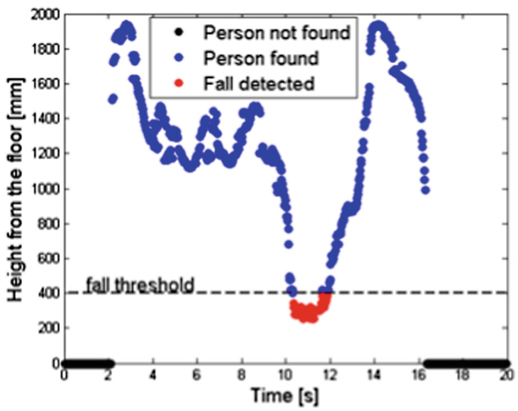

Fig. 2. Sample sequence of depth values taken by the subject's centroid, compared to the 400 mm threshold. A fall event is detected and highlighted in red (Color figure online)

2.2 Experimental Protocol

In order to evaluate the algorithm performances, several tests were carried out in a laboratory environment. Tests involved 17 healthy subjects, 3 females and 14 males, aged between 21 and 55 years. More specifically, 32 different types of falls and 8 ADLs have been simulated by each tester, for a total amount of 544 falls and 136 ADLs. The experimental protocol is reported in Table 1. The simulated falls can be divided into 8 groups: backward finishing lying, backward finishing sitting, forward finishing lying, forward finishing on knees, forward finishing on knees grabbing a chair, forward finishing on knees grabbing a sofa, left side, and right side. For each of them, 4 different situations have been considered: (i) falling from the stand position and then remaining on the ground; (ii) falling from the stand position and then recovering; (iii) falling during walking and then remaining on the ground; (iv) falling during walking and then recovering.

Regarding ADLs, the performed activities can be divided into 4 groups: pick up an object from the floor with bending, pick up an object from the floor with squatting, sit and get up from a chair, and sit and get up from a couch. For each of them, two situations have been considered: (i) from a standing position; (ii) from walking.

All testers were asked to sign an informed consensus form before starting the experiment, and let perform the tasks freely. Foam mats were used to soften the blow and protect subjects from injuries when simulating falls.

3 Results and Discussion

As a first analysis, we considered falls belonging to the classes FBESFR (backward fall, finishing sitting), FFOKFR (forward fall on the knees), FFOKSO (forward fall, on the knees, grabbing the sofa), and FFOKCH (forward fall, on the knees, grabbing the chair). Sample trajectories of the four classes in the 3-D space are shown in Figs. 3(a)–(d), for one of the subject who performed the experimental tests. As visible in the Fig. 3(a) and (b), a fall belonging to class FBESFR originates a different trajectory than a fall belonging to class FFOKFR, notably featuring an opposite direction. On the other hand, falls belonging to classes FFOKSO and FFOKCH (Fig. 3(c) and (d)) have similar patterns and differ for the value of the z coordinate in the final position (denoted by t_end).

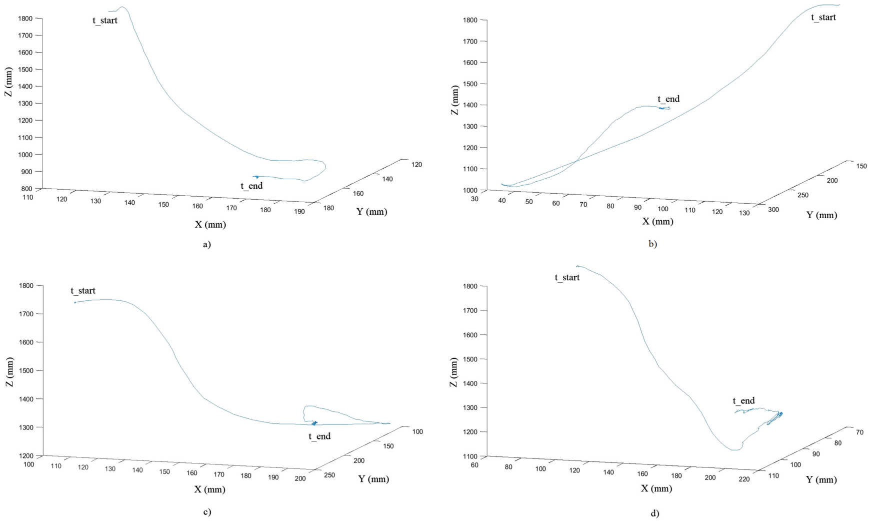

Fig. 3. Sample fall trajectories for the classes (a) FBESFR, (b) FFOKFR, (c) FFOKSO, and (d) FFOKCH for one of the subjects executing the test. Labels *t_start* and *t_end* denote the starting and ending point of the trajectory

For each type of fall within each class, the speed of fall along the z axis has been computed as the difference between the z coordinate value in two consecutive frames, captured at a rate of 30 fps, i.e. over a 33 ms time interval.

Table 1. Summary of fall detection experiments. "Subj." stands for subject.

Type	Activity name	Description
Backward fall, finishing lying	FBELFRST	Subj. is standing, falls backwards, and remains on the ground
	FBELFRSTRC	Subj. is standing, falls backwards, stays on the ground for a while and then gets up again
	FBELFRWK	Subj. walks, falls backward, and remains on the ground
	FBELFRWKRC	Subj. walks, falls backward, stays on the ground for a while and then gets up again
Backward fall, finishing sitting	FBESFRST	Subj. is standing, falls backwards, and remains on the ground
	FBESFRSTRC	Subj. is standing, falls backwards, stays on the ground for a while and then gets up again
	FBESFRWK	Subj. walks, falls backward, and remains on the ground
	FBESFRWKRC	Subj. walks, falls backward, stays on the ground for a while and then gets up again
Forward fall, finishing lying	FFELFRST	Subj. is standing, falls forwards, and remains on the ground
	FFELFRSTRC	Subj. is standing, falls forwards, stays on the ground for a while and then gets up again
	FFELFRWK	Subj. walks, falls forwards, and remains on the ground
	FFELFRWKRC	Subj. walks, falls forwards, stays on the ground for a while and then gets up again
Forward fall on the knees grabbing the chair	FFOKCHST	Subj. is standing, falls forwards, and remains on the ground, grabbing the chair
	FFOKCHSTRC	Subj. is standing, falls forwards, stays on the ground grabbing the chair for a while, and then gets up again
	FFOKCHWK	Subj. walks, falls backward, and remains on the ground, grabbing the chair
	FFOKCHWKRC	Subj. walks, falls backward, stays on the ground grabbing the chair for a while, and then gets up again
Forward fall on the knees	FFOKFRST	Subj. is standing, falls forwards, and remains on the ground
	FFOKFRSTRC	Subj. is standing, falls forwards, stays on the ground for a while and then gets up again
	FFOKFRWK	Subj. walks, falls forwards, and remains on the ground
	FFOKFRWKRC	Subj. walks, falls forwards, stays on the ground for a while and then gets up again
Forward fall on the knees grabbing the sofa	FFOKSOST	Subj. is standing, falls forwards, and remains on the ground, grabbing the sofa
	FFOKSOSTRC	Subj. is standing, falls forwards, stays on the ground grabbing the sofa for a while and then gets up again
	FFOKSOWK	Subj. walks, falls forwards, and remains on the ground, grabbing the sofa
	FFOKSOWKRC	Subj. walks, falls forwards, stays on the ground grabbing the sofa for a while and then gets up again
Left side fall	FSLEFRST	Subj. is standing, falls on the left side, and remains on the ground
	FSLEFRSTRC	Subj. is standing, falls on the left side, stays on the ground for a while and then gets up again
	FSLEFRWK	Subj. walks, falls on the left side, and remains on the ground
	FSLEFRWKRC	Subj. walks, falls on the left side, stays on the ground for a while and then gets up again.

(*continued*)

Table 1. (*continued*)

Type	Activity name	Description
Right side fall	FSRIFRST	Subj. is standing, falls on the right side, and remains on the ground
	FSRIFRSTRC	Subj. is standing, falls on the right side, stays on the ground for a while and then gets up again
	FSRIFRWK	Subj. walks, falls on the right side, and remains on the ground
	FSRIFRWKRC	Subj. walks, falls on the right side, stays on the ground for a while and then gets up again
Pick up object from floor with bending	APBEST	Subj. is standing, bends, picks up an object on the floor, and then stands up again
	APBEWK	Subj. walks, bends, picks up an object on the floor, and then stands up again
Pick up object from floor with squatting	APSQST	Subj. is standing, squats, picks up an object on the floor, and then stands up again
	APSQWK	Subj. walks, squats, picks up an object on the floor, and then stands up again
Sit and get up from the chair	ASCHST	Subj. is standing, sits on a chair, and then stands up again
	ASCHWK	Subj. walks, sits on a chair, and then stands up again
Sit and get up from the couch	ASSOST	Subj. is standing, sits on a couch, and then stands up again
	ASSOWK	Subj. walks, sits on a couch, and then stands up again

As a first possible discriminating feature, we considered the mean value of the speed of fall over all the test repetitions performed by each subject, and over all the subjects performing the same type of fall. The results obtained for the four classes considered are shown in Fig. 4. Apart from the case of FFOKCH-ST fall, that shows a quite high variability over the different 17 subjects who performed the test, with a much larger 95% confidence interval, it is not possible to say that the average speed of fall alone can be taken as a feature able to discriminate in a clear fashion the different types of falls simulated. In fact, both the mean values and the 95% confidence intervals are very similar over the four classes of falls analyzed.

We then move to analyze a different quantity, i.e. the mean difference between the ending and the starting values of the z coordinate, for each type of fall, which we call Δz. Figure 5 shows the mean value of Δz over all the test repetitions performed by each subject, and over all the subjects performing the same type of fall, again for the four classes considered before. In this case, it is quite evident that falls belonging to the class FBESFR may be grouped into a cluster (denoted as **A** in the graph) that is distinguishable from falls belonging to the classes FFOKCH, FFOKFR and FFOKSO, that are grouped into cluster **B**. Backward falls ending sitting in cluster **A** show a greater mean Δz than forward falls ending on the knees in **B**, irrespective of the different subjects' physique (the details regarding the subjects who performed the experiments are provided in Table 2). On the contrary, it is not possible to rely on the mean Δz values to discriminate among the different types of forward falls ending on the knees, that have been labeled as group **c**, **d**, and **e** in Fig. 5.

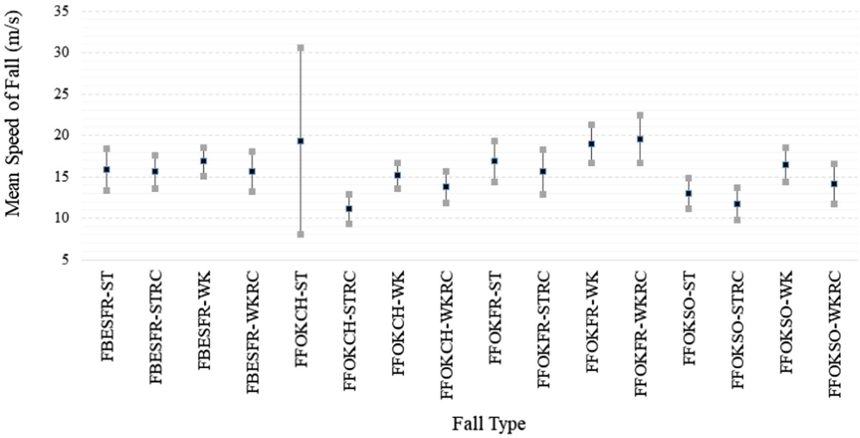

Fig. 4. Mean speed of fall along the z axis (black squares) and 95% confidence interval for the classes FBESFR, FFOKCH, FFOKFR and FFOKSO, averaged over all the subjects executing the test

Table 2. Information about the 17 voluntary users involved in the test phase.

Number	Gender	Age	Height (cm)	Weight (Kg)
1	Female	40	162	60
2	Female	29	170	74
3	Female	25	160	52
4	Male	30	176	65
5	Male	55	173	80
6	Male	21	169	58
7	Male	21	178	70
8	Male	23	175	59
9	Male	28	178	74
10	Male	28	160	76
11	Male	26	182	73
12	Male	40	187	87
13	Male	21	189	80
14	Male	22	167	64
15	Male	22	170	72
16	Male	21	188	78
17	Male	21	177	78
Mean		28	174	71
Std		9.3	9.2	9.5

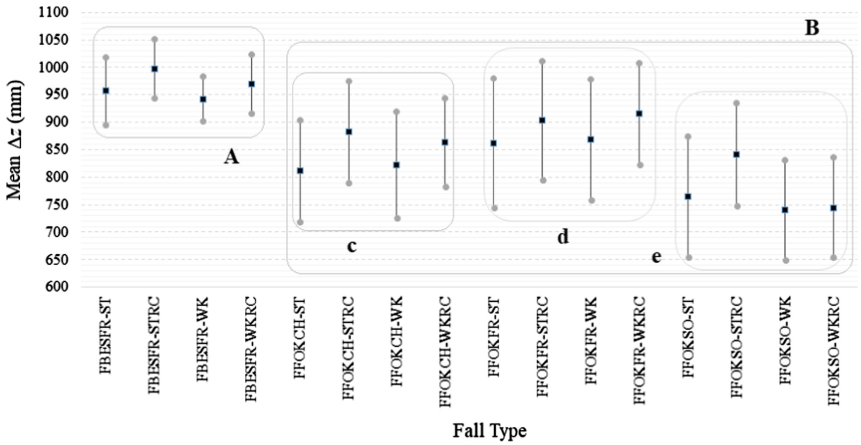

Fig. 5. Mean Δz values and 95% confidence intervals for the classes FBESFR, FFOKCH, FFOKFR and FFOKSO, averaged over all the subjects executing the test

4 Conclusion

In this paper we provided a preliminary investigation about the possibility to use the speed of fall as a feature to discriminate among different types of falls, obtained by processing the depth data provided by a Kinect device placed in top view with respect to the subject. This approach differs from those usually found in the literature, as the adoption of Kinect in top view makes not available the automatic detection of the subject's joints, which are output by the sensor when used in front view with respect to the monitored person. The speed of fall is obtained indirectly, by processing the depth frames through the identification of the person's blob in the image, and tracking of its relative distance from the sensor during the sequence of frames referred to the fall. The outcomes of this first attempt show that the average speed of fall alone is not enough to discriminate among falls, but this feature can be used in conjunction with others, like the variation of the vertical coordinate due to the fall, to improve the classification reliability. As already stated, this is just a preliminary study, but, based on the promising outcomes obtained, we aim to fully exploit the dataset collected through the experiments with 17 subjects, to improve our analysis and also include the evaluation of the activities of daily living.

Acknowledgment. This work was partly supported by Università Politecnica delle Marche under the project "PIMRA: Procedure Innovative per il Monitoraggio Remoto della Attività motoria e cardiovascolare". The authors would also like to acknowledge the contribution of the COST Action IC1303 AAPELE - Architectures, Algorithms and Platforms for Enhanced Living Environments.

References

1. Delahoz, Y.S., Labrador, M.A.: Survey on fall detection and fall prevention using wearable and external sensors. Sensors **14**(10), 19806–19842 (2014). https://doi.org/10.3390/s141019806. http://www.mdpi.com/1424-8220/14/10/19806
2. Gasparrini, S., Cippitelli, E., Spinsante, S., Gambi, E.: A depth-based fall detection system using a kinect sensor. Sensors **14**(2), 2756–2775 (2014). https://doi.org/10.3390/s140202756. http://www.mdpi.com/1424-8220/14/2/2756
3. Han, J., Shao, L., Xu, D., Shotton, J.: Enhanced computer vision with microsoft kinect sensor: a review. IEEE Trans. Cybern. **43**(5), 1318–1334 (2013). https://doi.org/10.1109/TCYB.2013.2265378
4. Spinsante, S., Cippitelli, E., De Santis, A., Gambi, E., Gasparrini, S., Montanini, L., Raffaeli, L.: Multimodal interaction in a elderly-friendly smart home: a case study. In: Agüero, R., Zinner, T., Goleva, R., Timm-Giel, A., Tran-Gia, P. (eds.) MONAMI 2014. LNICST, vol. 141, pp. 373–386. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16292-8_27
5. Spinsante, S., Gambi, E., De Santis, A., Montanini, L., Pelliccioni, G., Raffaeli, L., Rascioni, G.: Design and implementation of a smart home technological platform for the delivery of AAL services: from requirements to field experience, pp. 433–456. Healthcare Technologies. Institution of Engineering and Technology (2016). https://doi.org/10.1049/PBHE006E_ch21
6. Springer, S., Yogev Seligmann, G.: Validity of the kinect for gait assessment: a focused review. Sensors **16**(2), 1–13 (2016)
7. Wikipedia: Definition of sensor (2017). https://en.wikipedia.org/wiki/Sensor
8. Yang, L., Ren, Y., Zhang, W.: 3D depth image analysis for indoor fall detection of elderly people. Digit. Commun. Netw. **2**(1), 24–34 (2016). https://doi.org/10.1016/j.dcan.2015.12.001. http://www.sciencedirect.com/science/article/pii/S2352864815000681