Human Fall Detection Using Kinect Sensor

Michal Kepski[1] and Bogdan Kwolek[2]

[1] University of Rzeszow, 35-959 Rzeszów, Poland
mkepski@univ.rzeszow.pl
[2] Rzeszow University of Technology, 35-959 Rzeszów, Poland
bkwolek@prz.edu.pl

Abstract. Falls are major causes of mortality and morbidity in the elderly. The existing CCD-camera based solutions require time for installation, camera calibration and are not generally cheap. In this paper we show how to achieve automatic fall detection using Kinect sensor. The person is segmented on the basis of the updated depth reference images. Afterwards, the distance of the person to the ground plane is calculated. The ground plane is extracted by the RANSAC algorithm. The point cloud belonging to the floor is determined using v-disparity images and the Hough transform.

1 Introduction

Falling is an everyday possible accident that all of us are exposed to. A fall can lead to severe consequences, such as fractures, and a fallen person might need assistance at getting up again. Thus, in recent years a lot of research has been dedicated into the development of fall detection methods [16][14]. Such methods are designed to robustly detect falls and then to raise a medical alert. Medical personnel can then be dispatched to the site where the alarm was activated.

As humans become old, their bodies weaken and the risk of accidental falls raises noticeably [11]. The research results demonstrate that high percentage of injury-related hospitalizations for seniors are the results of falls [6]. Since the population of elderly is increasing dramatically in almost all countries of the world, high demand for unobtrusive and assistive technology is observed. In particular, the assistive technology can contribute toward independent living of the elderly [3][15]. However, regardless of numerous efforts undertaken to attain reliable and unobtrusive fall detection, current technology does not meet the requirements of the seniors [20]. False alarms can happen while seniors are bending over, laying down, or doing a variety of other day to day activities. In general, current technology leads to much higher rate of false alarms when compared with standard medical alert.

Applicable techniques for fall detection include a variety of methods. Most of the techniques are based on body-worn or built-in devices, which are intrusive as they require the user to wear a smart device. Such methods utilize accelerometers or both accelerometers and gyroscopes to discriminate the fall from activities of daily living (ADLs) [15]. However, very often on the basis of such sensors it is

hard to separate real falls from fall-like activities [2]. Bending over, laying down, sitting down, or even setting down a purse can all resemble a fall depending on how it was done. In consequence, these methods trigger significant number of false alarms. What's more, the detectors that are typically attached to a belt around the hip, are uncomfortable to be worn during the sleep [5]. Furthermore, their usefulness in monitoring of critical phases like getting up from the bed is relatively poor. In addition to applications presented in scientific publications, commercial fall detection systems are available as shown in a survey [16] with 7 examples of commercially available fall detection systems and over 40 patents on fall detectors.

During the recent years, a lot of research has been done on detecting falls using a wide range of sensor types [15][20], including pressure pads [19], single CCD camera [1][18], multiple cameras [4], specialized omni-directional ones [13] and stereo-pair cameras [7]. Video cameras offer several advantages over other sensors including the capability of detection of various activities. The further benefit is low intrusiveness and the possibility of remote verification of fall events. However, the currently available solutions require time for installation, camera calibration and they are not generally cheap. As a rule, CCD-camera based systems require a PC computer or a notebook for image processing. The existing video-based devices for fall detection cannot work in nightlight or low light conditions. Additionally, the lack of depth information can lead to lots of false alarms. Moreover, in most of such systems the privacy is not preserved adequately.

Recently, the Kinect sensor has been successfully used in fall detection systems [12][17][8]. It is the world's first system that at reasonable price combines an RGB camera and a depth sensor. Unlike 2D cameras, the low-cost Kinect allows tracking the body movements in 3D. Thus, if only depth images are used it can guarantee the person's privacy. The Kinect sensor is independent of external light conditions, since it is equipped with an active light source. As the Kinect uses infrared light it is able to extract depth images in a room that is dark to our eyes.

2 Motivation and Background

Depth is very useful cue to achieve reliable person detection because humans may not have consistent color and texture but must occupy an integrated region in space. However, in many home scenarios is not easy to detect a person using only depth images due to occlusions, for instance, if a person stands behind a chair being in turn in the front to the Kinect. The software called NITE, which is a binary distribution from PrimeSense offers skeleton tracking with the Kinect sensor. However, this software has been developed for human computer interaction, and not to detect the person fall. Thus, in some circumstances it has difficulties in extracting and tracking the skeleton, see Fig. 1a-b, as well as segmenting the person, see Fig. 1c, where we can see two segments belonging to the person lying on the floor.

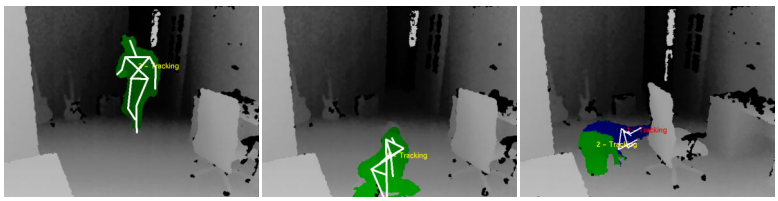

Fig. 1. NITE-based skeleton tracking during a person fall.

Because of the inconveniences mentioned above as well as the lack of the distribution of the NITE for an embedded platform we elaborated algotithm for person extraction on depth images at relatively low computational cost. In order to achieve reliable fall detection we employ both Kinect and accelerometer that complement one another [8]. We implemented the system on PandaBoard ES, which is a low-power, low-cost single-board computer development platform based on Texas Instruments OMAP4 line of processors [9]. It enables development of mobile applications. Regarding low-cost computational power of the board the person was detected on the basis of the scene reference image, which was extracted in advance. Such a fast method of person segmentation can be applied in many scenarios, for instance in fall detection systems mounted on the stairs. However, in home environments such an approach can be impractical. The main reason for this is that in case of the moved furniture, like chair or even opening the door the scene reference image contains such objects, what in turns can lead to difficulties in segmentation of the person on the basis of depth connected components.

In this work we demonstrate a method for updating the depth reference image at a low computational cost. We also demonstrate how to extract the ground plane on depth images. The ground plane is extracted automatically using the v-disparity images, Hough transform and the RANSAC algorithm.

3 Fall Detection on Embedded Platform

In this section we present the main ingredients of our embedded system for human fall detection [9]. Our fall detection system uses both data from Kinect and motion data from a wearable smart device containing accelerometer and gyroscope sensors. Data from the smart device (Sony PlayStation Move) are transmitted wirelessly via Bluetooth to the PandaBoard on which the signal processing is done, whereas Kinect is connected via USB. The device contains one tri-axial accelerometer and a tri-axial gyroscope consisting of a dual-axis gyroscope and a Z-axis gyroscope. The fall alarm is triggered by a fuzzy inference engine based on expert knowledge, which is declared explicitly by fuzzy rules and sets. As inputs the engine takes the acceleration, the angular velocity and the distance of the person's gravity center to the altitude at which the Kinect is placed. The acceleration's vector length is calculated using data provided

by the tri-axial accelerometer, whereas the angular velocity is provided by the gyroscope. The sampling rate of both sensors is equal to 60 Hz. The sensor is typically attached to trunk or lower back because such body parts represent the major component of body mass and move with most activities.

The Kinect sensor captures depth and color images simultaneously at a frame rate of about 30 fps. It consists of an infrared laser-based IR emitter, an infrared camera and a RGB camera. The IR camera and the IR projector form a stereo pair with a baseline of approximately 75 mm. Kinect depth measurement is based on structured light, making a triangulation between the dot pattern emitted and the one captured by the IR CMOS sensor. Pixels in the provided depth images indicate calibrated depth in the scene. Kinect's field of view is fifty-seven degrees horizontally and forty-three degrees vertically. The minimum range for the Kinect is about 0.6 m and the maximum range is somewhere between 4-5 m.

The depth images are acquired using OpenNI (Open Natural Interaction) library. A mean depth map is extracted in advance to delineate the foreground object at low-computational cost. It is extracted on the basis of several consecutive depth images without the subject to be monitored and then it is stored for the later use in the detection mode. In the detection mode the foreground objects are extracted through differencing the current image from such a reference depth map. Afterwards, the foreground object is determined through extracting the largest connected component in the thresholded difference map. According to the reports of the code profiler the module responsible for detection of the foreground object uses 50% of the CPU's computational power.

4 V-disparity Based Ground Plane Detection

The v-disparity images were originally proposed in [10] to achieve obstacle detection using disparity maps between two stereo images. Given a depth map extracted by the Kinect sensor, the disparity d can be calculated in the following manner:

$$d = \frac{b \cdot f}{z} \tag{1}$$

where z is the depth (in meters), b is the horizontal baseline between the cameras (in meters), f is the (common) focal length of the cameras (in pixels). For the Kinect sensor the value of b is 7.5 cm and it is the measured distance between the IR and projector lenses, whereas f is equal to 580 pixels.

Let H be a function of the disparities d such that $H(d) = I_d$. The I_d is the v-disparity image and H accumulates the pixels with the same disparity from a given line from the disparity image. Thus, in the v-disparity image each point in the line i represents the number of points with the same disparity occurring in the i-th line of the disparity image. Figure 2c illustrates the v-disparity image that corresponds to the depth image depicted on Fig. 2b. The size of the images acquired by Kinect is 640×480.

The line corresponding to the floor pixels was extracted using the Hough transform. Assuming that the Kinect is placed at height about 1 m from the

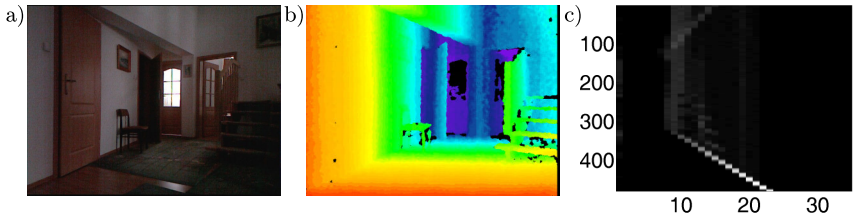

Fig. 2. V-disparity map calculated on depth images from Kinect: RGB image a), corresponding depth image b), v-disparity map c).

floor, the extracted line should begin in the disparities ranging from 21 to 25 depending on the tilt angle of the sensor. On the basis of the extracted line the pixels belonging to the floor areas were determined. Due to the measurement inaccuracies we considered also pixels in some disparity extent d_t as belonging to the ground. Assuming that d_y is a disparity in the line y, which represents the pixels belonging to the ground, we take into account the disparities from the range $d \in (d_y - d_t, d_y + d_t)$ as representing the ground. Figure 3a illustrates the point cloud corresponding images shown on Fig. 2b. On Fig. 3b we can observe the point cloud without the points belonging to the floor. Given the extracted line by the Hough transform, the points on the v-disparity image with the corresponding depth pixels were selected, and then transformed to point cloud, see Fig. 3c depicting the points cloud of the floor, which was selected using the v-disparity map.

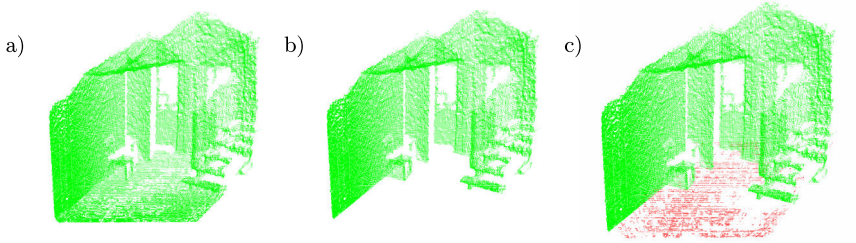

Fig. 3. Points cloud corresponding to the depth image (from Fig. 2b) a), the points cloud without the points belonging to the floor b), the points cloud from image b, with points belonging to the floor c).

After the transformation of the pixels to the 3D points cloud representing the floor, the plane described by the equation $ax+by+cx+d$ was recovered. The parameters a, b, c and d were estimated using RANSAC algorithm. The distance of the 3D centroid of the segmented person to the ground plane was determined

on the basis of the following equation:

$$D = \frac{|aX_c + bY_c + cZ_c + d|}{\sqrt{a^2 + b^2 + c^2}} \tag{2}$$

where X_c, Y_c, Z_c are coordinates of the centroid.

5 Person Segmentation

In our previous work [9], a depth reference map-based extraction of the person has been utilized. The method has been investigated mainly due to limited computational power of the PandaBoard at which the system has been implemented. The depth reference map was extracted on the basis of several consecutive depth images without the subject to be monitored and then it was stored for the later use in the person detection mode. In the detection mode the foreground objects were extracted through differencing the current image from such a reference depth map. Experimental findings demonstrated that such a technique can be applied in many scenarios, for instance in systems monitoring fall detection on stairs. However, in case of change of the scene layout, for example due to change of furniture settings some non-person objects can appear in the reference images and this in turn can lead to difficulties in segmenting the person.

In order make the system applicable in a wide range of scenarios we elaborated a fast method for updating the depth reference image. In our approach, each pixel of the depth reference image assumes the median value of several pixels from the past images. At the beginning we collect a number of images, and for each pixel we assemble a list of the pixels from the former images, which is then sorted in order the extract the median. For images of size 640×480 the computation time needed for extraction of the median is about 9 miliseconds at 2.4 GHz I7 processor running 4 threads. At the PandaBoard this operation can be completed in 0.15 sec. Given the sorted lists of pixels the depth reference image can be updated quickly by removing the oldest pixels and updating the sorted lists with the pixels from the current depth image and then extracting the median value. We found that for typical human motions, good results can be obtained using 13 depth images. For Kinect acquiring images at 25 Hz we take every tenth image.

Figure 4 illustrates some example depth reference images, which have been obtained using the discussed technique. In the image #388 we can see the opened door, which was closed to demonstrate how the algorithm updates the reference image. In frame #560 we can see that the door appears in the reference image, and then it is removed in frame #822. As we can observe, the updated reference image is free of clutter and allows us to extract the depth silhouette. In order to eliminate small objects the depth connected components were extracted. Afterwards, small artifacts were eliminated. Otherwise, the depth images can be cleaned using morphological erosion. When the person does not move the reference image is not updated. It is worth noting that the accelerometer can support the detection of periods in which the movement of the person takes

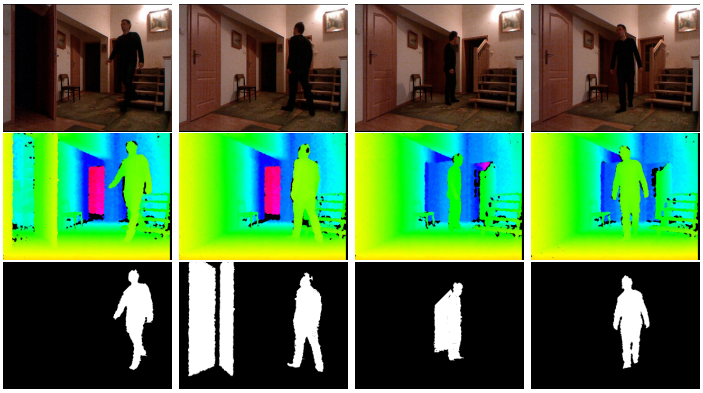

Fig. 4. Person segmentation. RGB images (upper row), the depth images (middle row) and the segmented person (bottom row).

place. As we utilize only one Kinect, an occlusion happening because of the furniture (e.g. sofa or chair) can led to difficulties in detecting the fall, when only images are employed. In such situations the decision can be made on the basis of the information provided by the accelerometer and the gyroscope.

6 Experimental Results

Four volunteers with age over 26 years attended in experiments and tests of our algorithm. A data-set of normal activities like walking, sitting down and crouching down has been composed in order to determine the threshold value, i.e. the distance of the centroid of the person to the floor below which the alarm should be triggered. Figure 5 depicts the distance D to the ground plane that has been obtained for some daily activities. As we can observe, on the basis of analysis of the motion of the centroid it is possible to discard some short-term actions, like sitting down, for which the centroid was temporally below the alarm threshold.

Intentional falls were performed in home towards a carpet with thickness of about 2 cm. Each individual performed three types of falls, namely forward, backward and lateral at least three times. Figure 6 depicts a person who has fallen and the corresponding binary map, which was obtained through differencing the current depth image from the reference depth image and then thresholding the difference image. All intentional falls were detected correctly.

The system was implemented in C/C++ and runs at 25 fps on 2.4 GHz I7 (4 cores, Hyper-Threading) notebook powered by Linux. We are planning to implement the system on the PandaBoard.

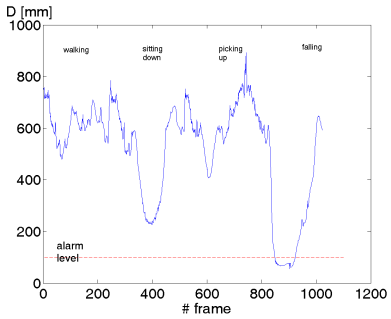

Fig. 5. Distance of the controid to the ground plane for person performing some daily activities.

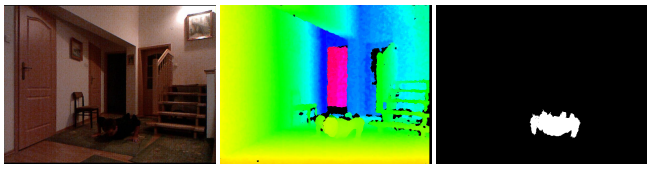

Fig. 6. Image with a fallen person (left), corresponding depth image (middle) and binary image with the extracted person (right).

7 Conclusions

In this work we demonstrated our approach to fall detection using Kinect. The detection of the fall is done on the basis segmented person in the depth images. The segmentation of the person takes place using updated depth reference images of the scene. The distance of the controid of the segmented person to the ground plane is used to trigger the fall alarm. The ground plane is extracted automatically using the v-disparity images, Hough transform and the RANSAC algorithm.

Acknowledgment

This work has been supported by the National Science Centre (NCN) within the project N N516 483240.

References

1. Anderson, D., Keller, J., Skubic, M., Chen, X., He, Z.: Recognizing falls from silhouettes. In: Annual Int. Conf. of the Engineering in Medicine and Biology Society. pp. 6388–6391 (2006)

2. Bourke, A., O'Brien, J., Lyons, G.: Evaluation of a threshold-based tri-axial accelerometer fall detection algorithm. Gait & Posture 26(2), 194–199 (2007)
3. Cook, A., Hussey, S.: Assistive Technologies Principles and Practice. Mosby, 2nd edn. (2002)
4. Cucchiara, R., Prati, A., Vezzani, R.: A multi-camera vision system for fall detection and alarm generation. Expert Systems 24(5), 334–345 (2007)
5. Degen, T., Jaeckel, H., Rufer, M., Wyss, S.: Speedy: A fall detector in a wrist watch. In: Proc. of IEEE Int. Symp. on Wearable Computers. pp. 184–187 (2003)
6. Heinrich, S., Rapp, K., Rissmann, U., Becker, C., Knig, H.H.: Cost of falls in old age: a systematic review. Osteoporosis International 21, 891–902 (2010)
7. Jansen, B., Deklerck, R.: Context aware inactivity recognition for visual fall detection. In: Proc. IEEE Pervasive Health Conference and Workshops. pp. 1–4 (2006)
8. Kepski, M., Kwolek, B., Austvoll, I.: Fuzzy inference-based reliable fall detection using kinect and accelerometer. In: The 11th Int. Conf. on Artificial Intelligence and Soft Computing. pp. 266–273. LNCS, vol. 7267, Springer-Verlag (May 2012)
9. Kepski, M., Kwolek, B.: Fall detection on embedded platform using Kinect and wireless accelerometer. In: Proc. of the 13th Int. Conf. on Computers Helping People with Special Needs - Volume Part II. pp. 407–414. LNCS, vol. 7383, Springer-Verlag, Berlin, Heidelberg (2012)
10. Labayrade, R., Aubert, D., Tarel, J.P.: Real time obstacle detection in stereovision on non flat road geometry through "v-disparity" representation. In: Intelligent Vehicle Symposium, 2002. IEEE. vol. 2, pp. 646 – 651 vol.2 (june 2002)
11. Marshall, S.W., Runyan, C.W., Yang, J., Coyne-Beasley, T., Waller, A.E., Johnson, R.M., Perkis, D.: Prevalence of selected risk and protective factors for falls in the home. American Journal of Preventive Medicine 8(1), 95–101 (2005)
12. Mastorakis, G., Makris, D.: Fall detection system using Kinect's infrared sensor. Journal of Real-Time Image Processing pp. 1–12 (2012)
13. Miaou, S.G., Sung, P.H., Huang, C.Y.: A customized human fall detection system using omni-camera images and personal information. Distributed Diagnosis and Home Healthcare pp. 39–42 (2006)
14. Mubashir, M., Shao, L., Seed, L.: A survey on fall detection: Principles and approaches. Neurocomputing 100, 144 – 152 (2013), special issue: Behaviours in video
15. Noury, N., Fleury, A., Rumeau, P., Bourke, A., Laighin, G., Rialle, V., Lundy, J.: Fall detection - principles and methods. In: Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society. pp. 1663–1666 (2007)
16. Noury, N., Rumeau, P., Bourke, A., ALaighin, G., Lundy, J.: A proposal for the classification and evaluation of fall detectors. IRBM 29(6), 340 – 349 (2008)
17. Parra-Dominguez, G., Taati, B., Mihailidis, A.: 3d human motion analysis to detect abnormal events on stairs. In: Second Int. Conf. on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT). pp. 97 –103 (Oct 2012)
18. Rougier, C., Meunier, J., St-Arnaud, A., Rousseau, J.: Monocular 3D head tracking to detect falls of elderly people. In: Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society. pp. 6384–6387 (2006)
19. Tzeng, H.W., Chen, M.Y., Chen, J.Y.: Design of fall detection system with floor pressure and infrared image. In: Int. Conf. on System Science and Engineering. pp. 131–135 (july 2010)
20. Yu, X.: Approaches and principles of fall detection for elderly and patient. In: 10th Int. Conf. on e-health Networking, Applications and Services. pp. 42–47 (2008)