Unobtrusive Fall Detection at Home using Kinect Sensor

Michal Kepski[2] and Bogdan Kwolek[1]

[1] AGH University of Science and Technology, 30 Mickiewicza Av., 30-059 Krakow, Poland
bkw@agh.edu.pl
[2] University of Rzeszow, 16c Rejtana Av., 35-959 Rzeszów, Poland
mkepski@univ.rzeszow.pl

Abstract. The existing CCD-camera based systems for fall detection require time for installation and camera calibration. They do not preserve the privacy adequately and are unable to operate in low lighting conditions. In this paper we show how to achieve automatic fall detection using only depth images. The point cloud corresponding to floor is delineated automatically using v-disparity images and Hough transform. The ground plane is extracted by the RANSAC algorithm. The detection of the person takes place on the basis of the updated on-line depth reference images. Fall detection is achieved using a classifier trained on features representing the extracted person both in depth images and in point clouds. All fall events were recognized correctly on an image set consisting of 312 images of which 110 contained the human falls. The images were acquired by two Kinect sensors placed at two different locations.

Keywords: Depth image and point cloud processing; fall detection.

1 Introduction

In almost all countries of the world the elderly population is continuously increasing. Improving the quality of life of increasingly elderly population is one of the most central challenges facing our society today. As humans become old, their bodies weaken and the risk of accidental falls raises noticeably [12]. A fall can lead to severe injuries such as broken bones, and a fallen person might need assistance at getting up again. Falls lead to losing self-confidence, a loss of independence and a higher risk of morbidity and mortality. Thus, in recent years a lot of research has been devoted to development of unobtrusive fall detection methods [15]. However, despite many efforts undertaken to achieve reliable and unobtrusive fall detection [16], the existing technology does not meet the seniors' needs [18]. The main reason is that it does not preserve the privacy and unobtrusiveness adequately. In particular, the current solutions generate too much false alarms, which in turn lead to considerable frustration of the seniors.

Most of the currently available techniques for fall detection are based on body-worn or built-in devices. They typically employ accelerometers or both accelerometers and gyroscopes [16]. However, on the basis of such sensors it is not

easy to separate real falls from fall-like activities [2]. They typically trigger significant number of false alarms. Moreover, the detectors that are typically worn on a belt around the hip, are obstructive and uncomfortable during the sleep [7]. What's more, their monitoring performance in critical phases like getting up from the bed or the chair is relatively poor.

In recent years, a lot of research has been done on detecting falls using a wide range of sensor types [16][18], including pressure pads [17], single CCD camera [1], multiple cameras [6], specialized omni-directional ones [14] and stereo-pair cameras [8]. Video cameras have several advantages over other sensors including the capability of recognition a variety of activities. Additional benefit is low intrusiveness and possibility of a remote verification of fall events. However, the solutions that are available at present require time for installation, camera calibration and in general they are not cheap. Additionally, the lack of 3D information can lead to lots of false alarms. Moreover, in vast majority of such systems the privacy is not preserved adequately.

Recently, the Kinect sensor was employed in fall detection systems [9][10][13]. It is the world's first low-cost device that combines an RGB camera and a depth sensor. Unlike 2D cameras, it allows tracking the body movements in 3D. Thus, if only depth images are used it preserves the privacy. Since it is equipped with an active light source it is independent of external light conditions. Owing to using the infrared light it is capable of extracting depth images in dark rooms.

In this work we demonstrate an approach to fall detection using only depth images. The person is detected on the basis of the depth reference image. We demonstrate a method for updating the depth reference image with a low computational cost. The ground plane is extracted automatically using the v-disparity images, Hough transform and the RANSAC algorithm. Fall detection is achieved using a classifier trained on features representing the extracted person both in depth images and in point clouds.

2 Person Detection in Depth Images

Depth is very useful cue to achieve reliable person detection because humans may not have consistent color and texture but have to occupy an integrated region in space. The depth images were acquired by the Kinect sensor using OpenNI (Open Natural Interaction) library. The sensor has 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 pattern captured by the IR CMOS sensor. The pixels in the depth images indicate calibrated depth in the scene. Kinect's angular field of view is 57° horizontally and 43° vertically. The sensor has a practical ranging limit of about 0.6-5 m. It captures depth and color images simultaneously at a frame rate of about 30 fps. The default RGB video stream has size 640×480 and 8-bit for each channel. The depth stream is 640×480 resolution and with 11-bit depth, which provides 2048 levels of sensitivity.

Due to occlusions it is not easy to detect a person using only single camera and depth images. The software called NITE from PrimeSense offers skeleton tracking on the basis of images acquired by the Kinect sensor. However, this software is targeted for supporting the human-computer interaction, and not for detecting the person fall. Thus, in many circumstances it can have difficulties in extracting and tracking the person's skeleton [10].

The person was detected on the basis of a scene reference image, which was extracted in advance and then updated on-line. In the depth reference image each pixel assumes the median value of several pixels values from the past images. In the set-up stage we collect a number of the depth images, and for each pixel we assemble a list of the pixel values from the former images, which is then sorted in order to extract the median. 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 the Kinect acquiring the images at 25 Hz we take every fifteenth image.

Figure 1 illustrates some example depth reference images, which were obtained using the discussed technique. In the image #500 we can see an office with the closed door, which was then opened to demonstrate how the algorithm updates the reference image. In frames #650 and #800 we can see that the opened door appears temporally in the binary image, and then it disappears in the frame #1000. As we can observe, the updated reference image is clutter free and allows us to extract the person's silhouette in the depth images. 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.

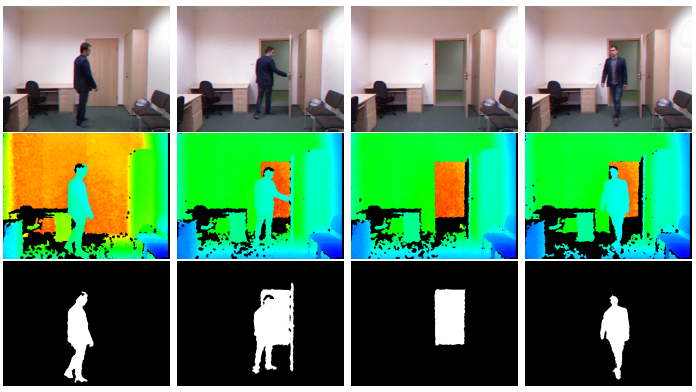

Fig. 1. Person segmentation using depth reference image. RGB images (upper row), depth (middle row) and binary images depicting the delineated person (bottom row).

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. Alternatively, the subject can be delineated using a pre-trained person detector. However, having in mind the privacy, the use of a person detector operating on depth images or point clouds leads to lower detection ratio and a higher computational cost.

3 V-disparity Based Ground Plane Extraction

In [11] a method based on v-disparity maps between two stereo images has been proposed to achieve reliable obstacle detection. Given a depth map provided by the Kinect sensor, the disparity d can be determined 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). The IR camera and the IR projector form a stereo pair with a baseline of approximately $b = 7.5$ cm, whereas the focal length 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 of 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.

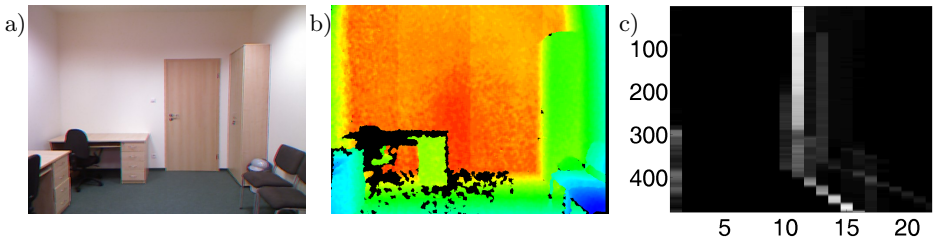

Fig. 2. V-disparity map calculated on depth images from Kinect: RGB image a), corresponding depth image b), v-disparity map c).

The line corresponding to the floor pixels in the v-disparity map was extracted using the Hough transform. Assuming that the Kinect is placed at height about 1 m from the floor, the line representing the floor should begin in the disparities ranging from 15 to 25 depending on the tilt angle of the sensor. On Fig. 3 we can see some example lines extracted on the v-disparity images, which were obtained on the basis of images acquired in typical rooms, like office, see Fig. 2c, classroom, etc.

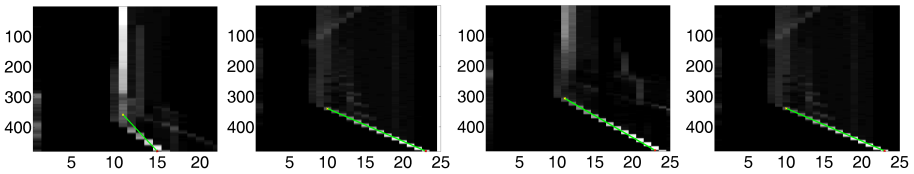

Fig. 3. Lines extracted by Hough transform on various v-disparity maps.

The line corresponding to the floor was extracted using Hough transform (HT) operating o v-disparity values and a predefined range of parameters. The accumulator was incremented by v-disparity values, see Fig. 4a. It is worth noting that ordinary HT operating on thresholded v-disparity images often gives incorrect results, see Fig. 4b where the extremum is quite close to 0 deg.

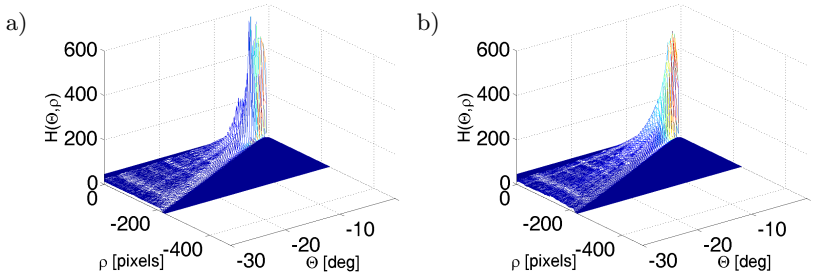

Fig. 4. Accumulator of the Hough transform: operating on v-disparity values a), thresholded v-disparity images b). The accumulator depicted on figure a) is divided by 100.

Given the extracted line in such a way, the pixels belonging to the floor areas were determined. Due to the measurement inaccuracies, pixels falling into some disparity extent d_t were also considered as belonging to the ground. Assuming that d_y is a disparity in the line y, which represents the pixels belonging to the ground plane, we take into account the disparities from the range $d \in (d_y - d_t, d_y + d_t)$ as a representation of the ground plane. Given the line extracted by the Hough transform, the points on the v-disparity image with the corresponding depth pixels were selected, and then transformed to the point cloud [10]. After the transformation of the pixels representing the floor to the 3D points cloud, the plane described by the equation $ax + by + cx + d$ was recovered. The parameters a, b, c and d were estimated using the RANSAC algorithm. The distance to the ground plane from the 3D centroid of points cloud corresponding to the segmented person 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 stand for the coordinates of the centroid.

4 Experimental Results

A data-set consisting of normal activities like walking, sitting down, crouching down and lying has been composed in order to train classifiers and to evaluate the performance of the fall detection system. Thirty five volunteers with age under 28 years attended in preparation of the data-set. The image sequences were recorded using two Kinect devices. The first Kinect was placed at a height of about one meter to the floor, whereas the second one was placed at a ceiling corner of the room. Figure 5 shows example depth images seen from two different views.

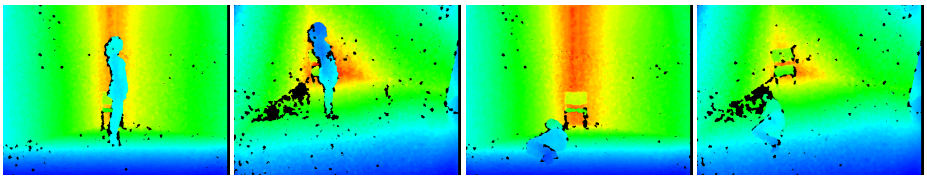

Fig. 5. Person in depth images seen from two different views.

In total 312 images representing typical human actions were selected and then utilized to extract the following features:

- h/w - a ratio of width to height of the person bounding box, calculated in the points cloud
- h/h_{max} - a ratio expressing the height of the person surrounding box in the current frame to the height of the person
- $dist$ - the distance of the person centroid to the floor, expressed in millimeters
- $max(\sigma_x, \sigma_z)$ - standard deviation from the centroid for the abscissa and the depth, respectively.

Figure 6 depicts a scatterplot matrix for the employed attributes, in which a collection of scatterplots is organized in a two-dimensional matrix simultaneously to provide correlation information among the attributes. In a single scatterplot two attributes are projected along the x-y axes of the Cartesian coordinates. As we can observe, the overlaps in the attribute space are not too significant. We considered also another attributes, for instance, a filling ratio of the rectangles making up the person bounding box. The worth of the features was evaluated on the basis of the information gain [4], which measures the dependence between the feature and the class label. In the evaluation we utilized the `InfoGainAttributeEval` procedure from the Weka [5], which is a collection of machine learning algorithms.

The classification accuracy was evaluated in 10-fold cross-validation using Weka software. The falls were classified using KStar [3], AdaBoost, SVM, multilayer perceptron (MLP), Naïve Bayes (NB) and k-NN classifiers. The KStar and

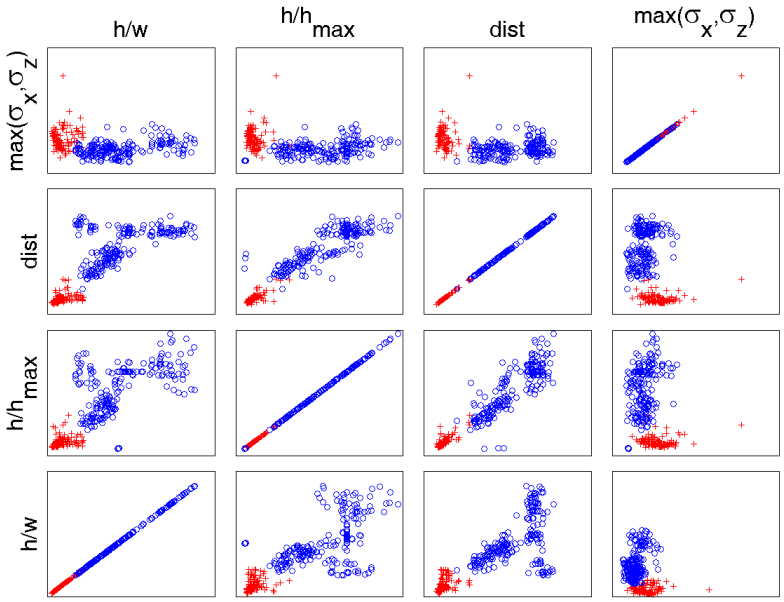

Fig. 6. Multivariate classification scatter plot.

MLP classified all falls correctly, whereas the remaining algorithms incorrectly classified 2 instances. The number of images with person fall was equal to 110.

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. The most computationally demanding operation is extraction of the depth reference image. For images of size 640×480 the computation time needed for extraction of the depth reference image is about 9 milliseconds. At the PandaBoard, which is a low-power, low-cost single-board computer development platform, this operation can be completed in 0.15 sec. We are planning to implement the whole system on the PandaBoard.

5 Conclusions

In this work we demonstrated our approach to fall detection using Kinect. The fall detection is done on the basis of the segmented person in the depth images. The segmentation of the person takes place using updated depth reference image of the scene. For person extracted in such a way the corresponding points cloud is then extracted. The ground plane is determined automatically using the v-disparity images, Hough transform and the RANSAC algorithm. The fall is detected using a classifier built on features extracted both from the depth images as well as the points cloud corresponding to the extracted person. The system achieves high detection rate. On image set consisting of 312 images of which 110 contained human falls all fall events were recognized correctly.

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. Cleary, J., Trigg, L.: An instance-based learner using an entropic distance measure. In: Int. Conf. on Machine Learning. pp. 108–114 (1995)
4. Cover, T.M., Thomas, J.A.: Elements of Information Theory. Wiley (1992)
5. Cover, T.M., Thomas, J.A.: Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, San Francisco, 2nd edn. (2005)
6. Cucchiara, R., Prati, A., Vezzani, R.: A multi-camera vision system for fall detection and alarm generation. Expert Systems 24(5), 334–345 (2007)
7. 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)
8. Jansen, B., Deklerck, R.: Context aware inactivity recognition for visual fall detection. In: Proc. IEEE Pervasive Health Conference and Workshops. pp. 1–4 (2006)
9. 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)
10. Kepski, M., Kwolek, B.: Human fall detection using Kinect sensor. In: Proc. of the 8th Int. Conf. on Computer Recognition Systems, Advances in Intelligent Systems and Computing, vol. 226, pp. 743–752. Springer (2013)
11. 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 (2002)
12. 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 J. of Preventive Medicine 8(1), 95–101 (2005)
13. Mastorakis, G., Makris, D.: Fall detection system using Kinect's infrared sensor. J. of Real-Time Image Processing pp. 1–12 (2012)
14. 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)
15. Mubashir, M., Shao, L., Seed, L.: A survey on fall detection: Principles and approaches. Neurocomputing 100, 144 – 152 (2013), special issue: Behaviours in video
16. Noury, N., Fleury, A., Rumeau, P., Bourke, A., ÓLaighin, 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)
17. 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 (2010)
18. 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)