Human Fall Detection by Mean Shift Combined with Depth Connected Components

Michal Kepski* and Bogdan Kwolek

Rzeszów University of Technology
Al. Powstańców Warszawy 12, 35-959 Rzeszów, Poland
bkwolek@prz.edu.pl

Abstract. Depth is very useful cue to achieve reliable fall detection since humans may not have consistent color and texture but must occupy an integrated region in space. In this work we demonstrate how to accomplish reliable fall detection using depth image sequences. The depth images are extracted by low-cost Kinect device. The person undergoing monitoring is extracted through mean-shift clustering. A depth connected component algorithm is used to delineate he/she in sequence of images. The system permits unobtrusive fall detection as well as preserves privacy of the user. The experimental results indicate high effectiveness of fall detection in indoor environments and low computational overhead of the algorithm.

1 Introduction

Falls are major causes of mortality and morbidity in the elderly. As humans become old, their bodies weaken and the risk of accidental falls increases. Many research findings show that high percentage of injury-related hospitalizations for seniors are the results of falls [8]. Thus, considerable research is devoted to the problem of fall detection, mainly due to the big demand and social values of assistive technologies [14]. Assistive technology or adaptive technology is an umbrella term that encompasses assistive, adaptive, and rehabilitative devices for people with special needs [4]. Among others, the assistive technology can contribute toward independent living of the elderly. Such an assistive device or system should detect the fall occurrence as soon as possible and to generate a warning to caregivers or an alarm to monitoring authorities. However, despite many efforts undertaken to attain reliable fall detection, the offered technology does not meet the requirements of the users [17].

A wide range of methods have been proposed for detecting a fallen person. Most of the proposed methods relies on a wearable device, which monitor the motion of an individual, recognize a fall and trigger an alarm. Such methods employ accelerometers or both accelerometers and gyroscopes to separate fall from activities of daily living (ADLs) [14]. However, on the basis of such sensors it is not easy to discriminate real falls from fall-like activities [2]. In this context, it

* M. Kepski is currently a student, doing his MSc thesis on fall detection.

is worth noting that quite a lot of ADLs like fast sitting have comparable motion patterns with real falls. In consequence, such methods can generate considerable number of false alarms. Moreover, such fall detectors, which are usually attached to a belt around the hip, are inadequate to be worn during the sleep [6]. In addition, they are not capable of monitoring in critical phases like getting up from the bed. In general, such devices are somewhat intrusive for humans since they require wearing continuously at least one smart sensor or mobile device.

During the last decade, a lot of work has been done on detecting falls using a wide range of sensor types [14][17], including pressure pads [16], single CCD camera [1][15], multiple cameras [5], specialized omni-directional ones [13] and stereo-pair cameras [9]. 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 night-light 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, in fall detection systems the time-of-flight cameras (TOF) have become more and more attractive [11]. In systems using such cameras the extrinsic calibration is restricted to the determination of the camera pose, whereas the intrinsic calibration is not required at all. TOF cameras are independent of external light conditions, since they are equipped with an active light source. It is worth noting that the TOF cameras can guarantee the person's privacy.

The existing technology permits reaching quite high performance of fall detection. However, it does not meet the requirements of the users with special needs. To make such human-assistive technology more unobtrusive and preserving privacy we developed a depth-based system for fall detection. Depth is very useful cue to achieve reliable fall detection since humans may not have consistent color and texture but must occupy an integrated region in space.

The Kinect is a revolutionary motion-sensing technology that permits sensing and tracking human position and motion. Unlike 2D cameras, Kinect allows tracking the body movements in 3D. It is the world's first system that combines an RGB camera and depth sensor. In order to achieve reliable and unobtrusive fall detection, our system employs the low-cost Kinect. The algorithm extracts the person in depth images using mean-shift algorithm [7]. Mean-shift is a general non-parametric mode finding/clustering procedure. It was extended to low-level vision problems [3], including, image segmentation and object tracking. In this work we propose to utilize the mean-shift to extract the person in depth image sequences at a low computational cost. The system can reliably distinguish the falls from activities of daily living, and thus the number of false alarms is reduced. An advantage of Kinect is that it can be put in selected places according to the user requirements. Moreover, the system operates on depth images and thus preserves privacy of people being monitored. In this context, it is worth

noting that Kinect uses infrared light and therefore it is able to extract depth images in a room that is dark to our eyes. In [10] we demonstrated an embedded system for fall detection in which a wearable motion-sensing device and Kinect complement each other. Due to limited computational power of the PandaBoard, at which the system has been implemented, a simple algorithm relying on reference map-based extraction of the person has been employed. However, after changing the setting of furniture the performance of the person detection drops. In such conditions, the wearable device contributes more to the decision of a fuzzy inference engine. The contribution of this work is an algorithm for person extraction, which delineates of individual even in case of changing the setting of furniture. It has low computational demands and runs in real-time.

2 The System for Fall Detection

At the beginning of this section we discuss usefulness of the Kinect for the detection of human fall. Afterwards, the extraction of the object of interest in depth images is presented. In the last part a proposed low-cost algorithm for person delineation is discussed.

2.1 Depth Images

The hardware employed in this work is Kinect, a motion sensing device developed by PrimeSense company. It is composed of an RGB camera, infrared laser-based IR emitter, an infrared camera, a multi-array microphone and a motorized tilt. The IR camera and the IR projector comprise a stereo pair with a baseline of approximately 75 mm. In this work, only the RGB and depth sensors are utilized to provide the input data. The device simultaneously captures depth and color images at a frame rate of about 30 fps. The resolution of color images is 640×480 pixels with 8 bits for every color channel. The Kinect projects the structured light code with an infrared laser onto the scene. Such a pattern is then read by an infrared camera and the 3D information is reconstructed from the distortions of the pattern. The disparity measurements are supplied in VGA resolution (640×480 pixels) as 11-bit integers, where 1 bit is reserved to mark the pixels for which no disparity is available. Since depth is inversely proportional to disparity, the depth resolution is also inversely related to the disparity. Thus, the depth resolution is not fixed and drops off with the distance growing to the sensor. For example, the depth resolution is about 1 cm at 2 m distance, whereas at 5 m distance one disparity level corresponds to about 7 cm depth resolution. 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. Thanks to the Kinect's capability to extract the depth images in unlit or dark rooms, the fall detection can be performed in the late evening or even in the night.

The depth images can be acquired using OpenNI (Open Natural Interaction) library [1]. The OpenNI framework supplies an application programming interface

[1] Available at: `http://www.openni.org/`

(API) as well as it provides the interface for physical devices and for middleware components. The NITE middleware, which is supplied by PrimeSense is a perception component enabling extraction of a person and tracking his/her skeleton. The discussed engine that enables natural human-computer interaction is protected by copyright and does not run on embedded platforms like PandaBoard.

2.2 Extraction of the Object of Interest

In our previous work [10] we presented an embedded system for person fall detection, in which a wearable motion-sensing device and Kinect complemented each other. Due to limited computational power of the PandaBoard, at which the system has been implemented a reference map-based extraction of the person has been employed. 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.

The procedure responsible for extraction of the object of interest in the depth reference images has low computational cost. However, the disadvantage of such an approach is that in the case of modification of the furniture settings, the furniture will appear in the depth difference images as foreground objects. Therefore, to cope with such an undesirable effect the depth difference images were segmented using mean-shift algorithm. Mean-shift is a nonparametric estimator of probability density [3]. The mean-shift vector always points toward the direction of the maximum increase in the density. The main idea behind mean-shift is to treat the points in the d-dimensional feature space as an empirical probability density function, where dense regions (or clusters) in the feature space correspond to the modes (or local maxima) of the underlying distribution. In contrast to the classic k-means clustering approach, it does not require prior knowledge on the shape of the distribution nor the number of modes or clusters.

Given n data points $\mathbf{x}_k, k = 1, \ldots, n$ on a d-dimensional space R^d the mean-shift is calculated on the basis of the following equation:

$$m(\mathbf{x}) = \frac{\sum_{k=1}^{n} \mathbf{x}_k g(||\frac{\mathbf{x} - \mathbf{x}_k}{h}||^2)}{\sum_{k=1}^{n} g(||\frac{\mathbf{x} - \mathbf{x}_k}{h}||^2)} - \mathbf{x} \tag{1}$$

where g is a kernel function and h denotes bandwidth. The mean-shift algorithm clusters a d-dimensional data set by associating each point with a peak of the data set's probability density. Each pixel is assigned a feature point $\mathbf{x}_k$. To account for the pixel position and depth, 3-d vectors were used. For each image pixel the mean-shift computes its associated peak by first defining a spherical window at the data point and computing the mean of the pixel values that

lie within the window. The algorithm then shifts the window to the mean and repeats until convergence, i.e., until the shift is less than a threshold.

In the assumed model the depth of the point k in the object space depends from the observed disparity as follows:

$$Z_k = \frac{Z_o}{1 + \frac{Z_o}{fb} d} \tag{2}$$

where b is the base length, f denotes the focal length of the IR camera, d is the observed disparity in the image space, whereas Z_o is the distance of the reference pattern. The parameters b, f, Z_o can be obtained by calibration of the camera.

Figure 1 illustrates the main steps of the extraction of the object of interest using mean-shift. In the upper row are shown the disparity images, in the middle row are depicted the segmented images, whereas on the bottom row are shown images with the extracted object of the interest. The mean-shift based

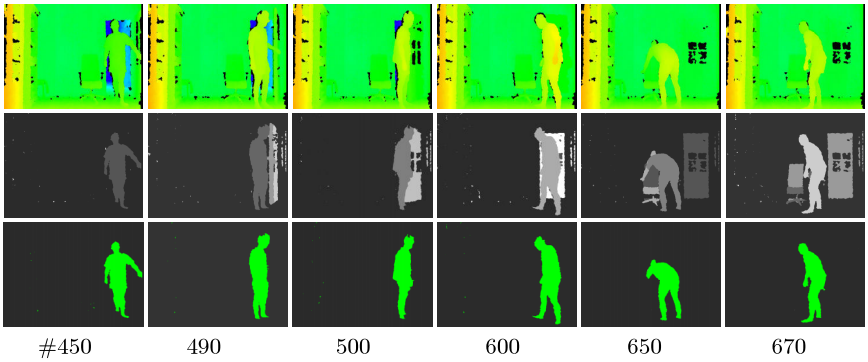

Fig. 1. Person detection. Disparity images (upper row), the segmented images (middle row) and the extracted person (bottom row).

segmentation was done on the difference image, which was obtained through differencing the current disparity image with the reference image. The images were then median filtered to remove small artefacts. Afterwards, the foreground object was extracted through determining the largest connected component in the segmented map. A correspondence between the segments was taken into account at this stage. It is also assumed that the depth values on the surface of a human object are continuous and vary only within a specific range. Finally, the center of gravity of the object of interest was calculated. As we can observe in frames #490 - 600 the person opens the door, which then appears as a separate segment in the image. Afterwards, in frames #650 - 670 the person moves the chair, which also appears in the segmented image. As we can see on images depicted on the bottom row, the person was extracted correctly in the considered image sequence despite change of the furniture settings.

As one can notice in the above images, the modifications of the furniture settings contribute towards clutter in the segmented image, which in turn can lead to difficulties in extraction of the object of interest, see also images in middle row on Fig. 1. Thus, in our approach we update the depth reference image on the basis of median filtered collection of images. Figure 2 illustrates the depth reference image (left), image on which a chair is moved (middle), and the updated reference image (right). As we can observe, the chair is at new location and no person is present on the extracted depth reference image. The updated depth reference image has been obtained on the basis of 30 depth images. Every 20th depth image was included in the collection of the images.

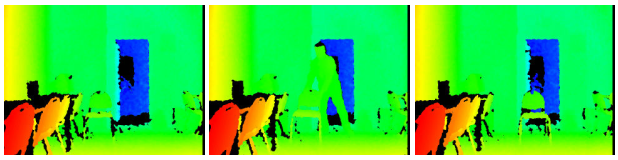

Fig. 2. Update of the depth reference image.

Using the discussed algorithm, the fall alarm was triggered on the basis of the distance of the person's gravity center to the altitude at which the Kinect was placed. In a more sophisticated approach a height/width ratio of the person's bounding box in k-nearest neighbor classifier can be used to detect the fall [12].

2.3 Low-cost Algorithm for Person Segmentation in Depth Images

The mean-shift is too slow to be used for image segmentation, where each pixel is a data point. In particular, the computational overhead is too large to make possible real-time image segmentation on mobile/embedded platforms. On mobile platforms the computational cost should be as low as possible to reduce the energy demand. Thus, in order to decrease the computational cost the mean-shift based person extraction was done in images with reduced spatial resolution by factor 5, i.e. on depth images of size 128×96. On Fig. 3 b-c) are shown the images that were segmented using the mean-shift. On the discussed images the ground plane is not shown since it was removed using a method similar to [18]. As one can observe the mean-shift is able to delineate the whole person. However, in some images the person is represented by several components, see Fig. 3c. Thus, in the next stage the segmented images are further refined using a connected components algorithm. Its aim is to connect at a low computational cost the neighboring depth segments possessing similar depth, see Fig. 3d. By starting from a seed connected component, which is located near the person's gravity center in the previous frame and simultaneously has the most similar depth to the depth at the gravity center, and then linking the neighboring components with similar depth the algorithm is able to extract the person in long image sequences, say several hundred of images, see also Fig. 3e.

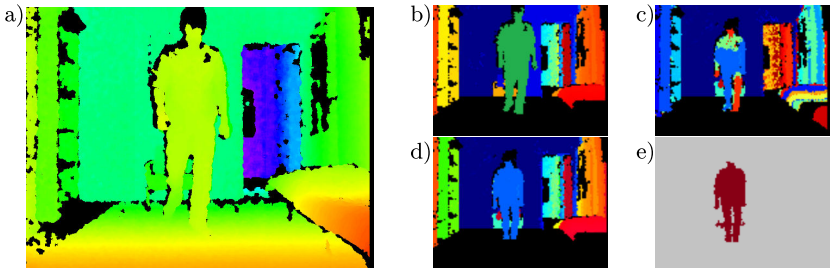

Fig. 3. Person detection. Depth image a), segmented images using mean-shift b), c), refined segmentation using connected components d), extracted person e). For visualization purposes the images b-e were resized to the half of the size of image a.

3 Experimental Results

Five volunteers with age over 26 years attended in evaluation of our developed algorithm and the system. Intentional falls were performed in home by four persons 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 4 depicts a person who has fallen and some daily activities that can be distinguished from the fall. All intentional falls were detected correctly.

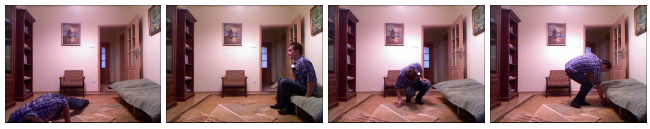

Fig. 4. A fallen person and some activities of daily living: sitting down, squatting/picking up objects from the floor, bending down, which were shot by Kinect.

The system has been implemented in C/C++ and runs at 25 fps on 2 Duo T8100 (2.1 GHz) notebook powered by Linux. The execution time needed for mean-shift segmentation of depth images of size 128×96 is about 40 ms. We are planning to implement the system on the PandaBoard.

4 Conclusions

In this work we demonstrated how to achieve reliable fall detection using Kinect. A mean-shift based algorithm with low computational burden was proposed to extract a person in depth image sequences. A depth connected component algorithm is used to extract the person in sequence of images. A fall alarm is generated on the basis of the gravity center of a connected component representing the extracted person. The system permits unobtrusive fall detection and preserves privacy of the user.

Acknowledgments. 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. Comaniciu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis. IEEE Trans. Pattern Anal. Mach. Intell. 24(5), 603–619 (2002)
4. Cook, A., Hussey, S.: Assistive Technologies Principles and Practice. Mosby, 2nd edn. (2002)
5. Cucchiara, R., Prati, A., Vezzani, R.: A multi-camera vision system for fall detection and alarm generation. Expert Systems 24(5), 334–345 (2007)
6. 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)
7. Fukunaga, K., Hostetler, L.D.: The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Tr. Inf. Theory 21(1), 32–40 (1975)
8. 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)
9. Jansen, B., Deklerck, R.: Context aware inactivity recognition for visual fall detection. In: Proc. IEEE Pervasive Health Conference and Workshops. pp. 1–4 (2006)
10. Kepski, M., Kwolek, B.: Fall detection on embedded platform using kinect and wireless accelerometer. In: 13th Int. Conf. on Computers Helping People with Special Needs. pp. II:407–414. LNCS, vol. 7383, Springer-Verlag (2012)
11. Leone, A., Diraco, G., Siciliano, P.: Detecting falls with 3d range camera in ambient assisted living applications: A preliminary study. Medical Engineering & Physics 33(6), 770–781 (2011)
12. Liu, C.L., Lee, C.H., Lin, P.M.: A fall detection system using k-nearest neighbor classifier. Expert Syst. Appl. 37(10), 7174–7181 (2010)
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. 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)
15. 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)
16. 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)
17. 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)
18. Zhao, J., Katupitiya, J., Ward, J.: Global correlation based ground plane estimation using v-disparity image. In: IEEE Int. Conf. on Robotics and Automation. pp. 529–534 (2007)