AUTOMATIC FALL DETECTION FOR ELDERLY BY USING FEATURES EXTRACTED FROM SKELETAL DATA

Amir Davari[1], Tarkan Aydin[2], Tanju Erdem[1]

[1] Department of Electrical Engineering, Ozyegin University, Turkey
{amir.davari, tanju.erdem}@ozyegin.edu.tr
[2] Department of Computer Engineering, Bahcesehir University, Turkey
tarkan.aydin@bahcesehir.edu.tr

ABSTRACT

Automatic detection of unusual events such as falls is very important especially for elderly people living alone. Real-time detection of these events can reduce the health risks associated with a fall. In this paper, we propose a novel method for automatic detection of fall event by using depth cameras. Depth images generated by these cameras are used in computing the skeletal data of a person. Our contribution is to use features extracted from the skeletal data to form a strong set of features which can help us achieve an increased precision at low redundancy. Our findings indicate that our features, which are derived from skeletal data, are moderately powerful for detecting unusual events such as fall.

Index Terms— fall detection, event detection

1. INTRODUCTION

Recently, there has been an increase in the number of computer vision and machine learning techniques proposed to detect and recognize human activities. One of the potential applications of these techniques is in the area of automatic detection of unusual events for elderly living alone. One of the most critical events that need to be detected for elderly is fall because it could imply severe medical complications depending on the time to rescue. According to recent research, almost half of the elderly people who fall at home and remain on the ground for more than an hour without any treatment usually die within six months [1]. Automatic detection methods of such events can significantly reduce the health risks for elderly.

An extensive research has been studied for detecting and identifying unusual events for elderly care using motion, audio or video data collected from the subject and/or the surrounding environment. Most of the proposed techniques have focused on fall detection. Among them, sensor based approaches usually employ wearable motion sensors consisting of accelerometers, gyroscopes and magnetic sensors which are attached to various body parts of the user, such as the foot, wrist, or waist [2][3]. The data generated by these sensors are analyzed with several machine learning and pattern recognition techniques to identify and detect fall events. The main problem with such an approach is that the subject has to wear the sensors, which can be easily forgotten or ignored. Although use of non-wearable sensors based solutions such as floor vibration based fall detectors [4] are also proposed, these approaches tend to produce false alarms frequently.

The limitations associated with the sensor based fall detection approaches are addressed by the video based fall detection systems. Initial systems analyzed the features of a foreground object that represent a moving person in RGB images. Most commonly used features are the velocity of the bounding box [5], aspect ratio of the person [6], and the trajectory of a fitted ellipse [7]. Rougier et.al. proposed to extract features from Motion History Image (MHI) which accumulates motion of a foreground object within finite time duration [8]. However, The MHI suffers from the problem of generating many false positives. As an alternative, Integrated Spatiotemporal Energy (ISTE) map was proposed in [9], which fits a blob to the moving region in an RGB frame and computes the temporal variance of its orientation. A high temporal variance of the orientation of the blob is interpreted as a slip and a low value of the length of the major axis in proportion to that of the minor axis is interpreted as fall. Performance of these methods depends on the extracted 2D silhouette of the subject. Because the same 2D silhouette may be generated with different 3D poses, the features computed from 2D silhouette are not reliable for differentiating certain events.

There has been an increased use of depth data in computer vision problems with the recent availability of inexpensive depth cameras, and especially after the release of Microsoft Kinect Depth camera [10]. Recent research has shown that the availability of depth data results in considerable performance improvement for the solution of several vision problems such as human pose estimation [11], and action recognition [12]. One advantage of depth cameras in comparison to RGB cameras is that depth cameras preserve privacy. [13] employed a time-of-flight depth camera for fall detection. They calculated the distance of the centroid of the moving region to the ground plane and the orientation of the person's spine. Falls are identified by thresholding on this distance. Another work using centroid and velocity of the moving objects is studied by [14] using Kinect. [15] computed 3-D bounding box of the moving regions to detect falls. Their features are compared with a temporally varying threshold found by a search technique. Their work produced acceptable results but searching can be computationally tedious and no basis for the frame duration required for searching was given.

Skeletal data can be readily obtained from depth cameras using third party libraries. [16] took advantage of the skeletal data and fit a line to the locations of head, neck torso joints and mean position of the knees. The line fitted to these points forms a major axis. Length and orientation of this major axis form their features. The problem with their method is that the orientation of fall in the whole data set is static and perpendicular to Kinect camera view, which can prevent the classifier from generalizing to falls in other orientations. [17] introduced a structural cost function where the norm of the 3-D inter-joint angles was considered. The mean and variance of the value of this cost function along with maximum and minimum values of the height of the skeleton formed a 4-D kinematic feature. They also used features based on the histogram of width to height ratio of a blob representing the moving person. Additionally, they employed RGB data when depth was not available. The main drawback of this method is the high dimensionality of the proposed features and the high redundancy among them.

Although several features were used to detect falls, none of them were designed based on the major reason of fall. A recent medical observation study [18] shows that the majority of falls in elderly people are due to the problems they have in weight shifting. This study proposes an automatic real-time fall detection algorithm for elderly care by considering weight shifting problems. The method utilizes only the extracted skeleton data of the subject. The features are optimized to detect falls related to weight shifting problem. However other kind of falls like collapse and slip are perfectly detected.

2. METHODOLOGY

2.1. Data preprocessing

Skeletal data is extracted via OpenNI framework [19] which represents the skeletal with 3D coordinates of 24 joints. Skeletal data is nothing but a fitted virtual skeleton pattern on the body depth map. The position of joints is defined in camera coordinate system. Therefore, it is a function of camera viewing angle. In order to eliminate this dependency, skeleton should be translated from camera coordinates to world coordinates by applying a rigid transformation. This transformation also provide the parameters of Ground plane G in world coordinates, which is simply $y = 0$ plane. Figure 1 shows sample depth map of a person, skeleton and its joints fitted on the depth map and ground plane.

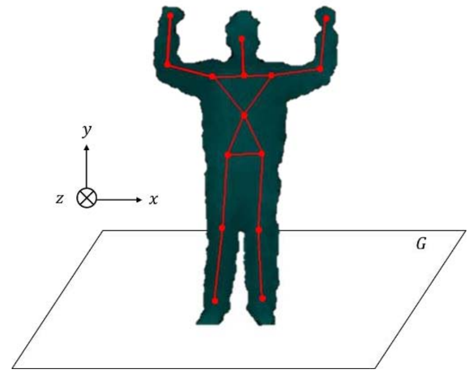

Figure 1 Skeleton joints fitted on depth data and ground plane G. All joints are given in 3D world coordinates; therefore, y component of any coordinates represents distance to the ground plane.

2.2. Features

Let the body be represented by 24 skeletal joints positions $\{\boldsymbol{s}_i^n(x, y, z) \in R^3\}$ where i is the skeletal joint index and n denotes the time index i.e. the frame number. Five features extracted from joint positions to identify fall event.

1) Height: One of the indicators of body pose is the body height. Falling follows by a considerable change in body height and can be a suitable feature for fall detection. Skeleton data provided by OpenNI is noisy in the end points of the body i.e. head, hands and legs whereas central joints like shoulders, neck, torso, hips and knees are more stable. Therefore, average of the shoulders heights were used for body height approximation;

$$f_1(n) = (s_6^n(y) + s_{12}^n(y))/2,$$

where "6" and "12" are indexes of shoulders. Since joints are transformed into world coordinates, y component of a joint position were used as height of the joint.

2) Height Vertical Temporal Gradient: Vertical speed of the upper body is also a major indicator in fall definition;

$$f_2(n) = f_1(n) - f_1(n-1)$$

3) Body Orientation: The main orientation of the body can be estimated by a line fitted on it. In this study, shoulders, torso and hips 3D positions were selected to find the body orientation. A line l was fitted through these 3D points and the angle θ between the line and ground plain was calculated as the third feature;

$$f_3(n) = \theta(n)$$

4) Body Orientation Temporal Gradient:

Fall is following by a big change is body orientation. Therefore the temporal gradient of θ was chosen as the fourth feature;

$$f_4(n) = \theta(n) - \theta(n-1)$$

5) Distance Between Center of Mass (COM) and Center of Body Support (COS): The last feature is based on the main contribution of this research study, i.e. weight shifting. Body balance is kept by feet. In a well-balanced

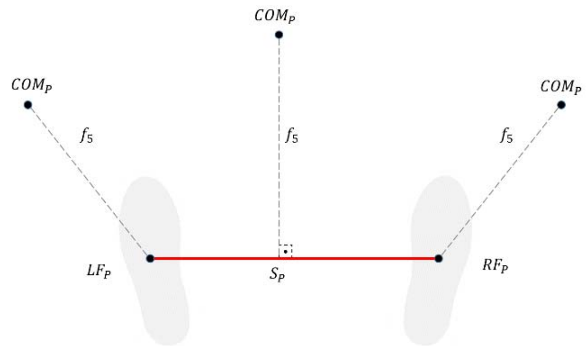

Figure 2 Top view of skeleton and three cases of COM projection. Computation of the feature f_5 depends on the position of the COM_P, which is simply expressed as the distance between COM_P to line segment S_p.

situation, COM projection on the ground (COM_P) is between the feet and has the least possible distance to them.

Now consider the following to possible fall instances;

1) Front/Backward fall in which the distance from (COM_P) to the line passing the projections of feet joints (LF_P and RF_P) will increase.

2) Side fall in which this distance will remains almost unchanged while the distance between projection of center of mass and projections of feet joints will increase.

By considering the cases above, last feature were defined as the distance between the body center of mass (COM) and body supports i.e. line segment connecting the feet projections. In this study, the joint representing torso was chosen as an approximation for COM. To write the mathematical expression of this feature, we need to introduce the following variables:

COM_P : Projection of COM on the ground plain
RF_P : Projection of right foot on the ground plain
LF_P : Projection of left foot on the ground plain
S_P : Line segment $\overline{RF_P LF_P}$ between two endpoints RF_P and LF_P

$$f_5(n) = dist(COM_P, S_P)$$

Figure 2. shows the cases for different fall scenarios and the computation of the feature f_5.

Sample feature-time graphs of each features computed for 160 frames long video can be seen in Fig.4. Note that there is a fall event between 80th and 100th frames

The extracted features are applied a classification to identify fall events from non-fall events. In this study, we use Linear Support Vector Machine (SVM) as the classifier. We employed a Matlab interface for SVM named LIBSVM [20] to classify our data set.

3. EXPERIMENTAL RESULTS AND DISCUSSION

We employed OpenNI [19] framework to extract depth and body joints 3D positions (Skeleton data). Data is recorded in 30 fps. Our data set consists of two main activities; fall and Normal (not fall). Normal events include walking, standing, sitting and sleeping. Falling was recorded in different orientations to make sure about generalization. Our training data set consists of 47310 frames, 42 fall events. 5124 frames (11%) were labeled as fall and the rest 89% was labeled as Normal. In the data set, the frame number of each observation is known. As we know the frame rate of the recorded skeleton data, i.e. 30 fps, we can draw a one to one correspondence between the skeletal data and class labels corresponding to events.

In order to eliminate the effect of noise added during data recording feature vectors were applied a post processing step. Windows with one second length were pushed on the feature vectors in the steps of 0.5 second. In each window firstly the feature values were sorted. The first and last 10% were removed. The average of the remaining values in the window was selected to as features.

To evaluate performance of the proposed system, we use two famous criteria that are widely used in fall detection systems. "Detection" is the percentage of falls which were truly detected as fall and "False alarm" is the percentage of non-fall windows which were wrongly detected as fall.

As the fourth and fifth features are novel features with respect to the state of art, we did the experiments with and without using them in the feature vector for classification.

In the first experiment we used just the first three features. By feeding this 3-dimensional feature vector, we achieved 80.7% detection rate and 6.8% false alarm rate.

As the second experiment we used the first four features to see the effect of adding using the fourth feature. The results were 84.7% detection rate and 6.5% false alarm rate.

By feeding the first three and the fifth feature to the classifier, we obtained 83.8% detection rate and 6.3% false alarm rate.

In the last experiment we fed all five features to the classifier which gave us 89.1% detection rate and 4.5% false alarm rate.

Figure 3 shows error rates of our results. As it is shown in the figure, by adding the fourth and fifth feature individually, around 4% improvement in detection rate was achieved. Using both these two features along with the first there ones, not only made 8.4% improvement in detection rate, but also improved the false alarm rate by 2.8%. These improvements both in fall detection rate and false alarm rate shows this fact that the proposed two novel features in this work have powerful information for fall detection and have the potential to make the fall detection systems more accurate and acceptable.

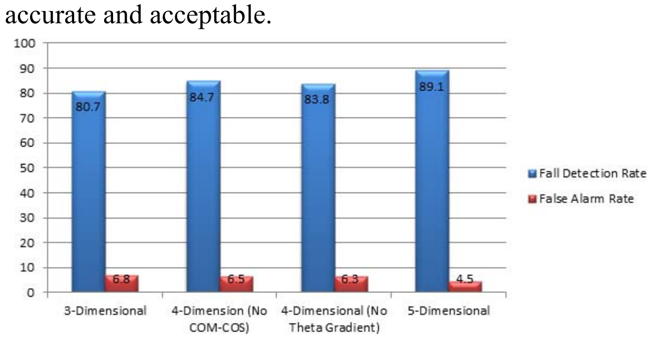

Figure 3. Detection rate and false alarm rate in the four experiments

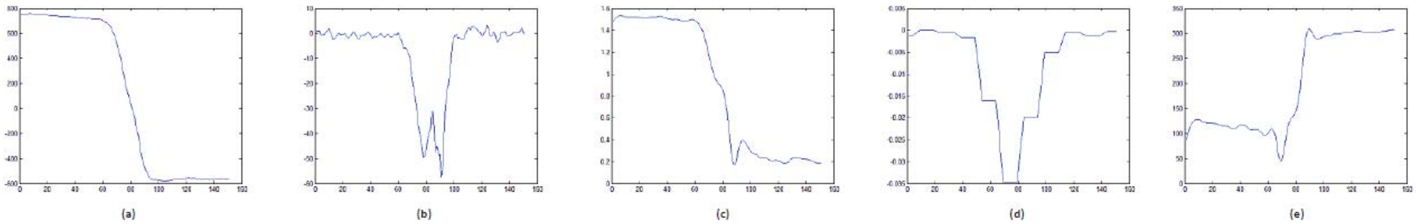

Figure 4. Features plotted vs. time for a sample fall event, (a) Body Height: During Fall, body height is expected to be decreased (b) Body Height Vertical Temporal Derivative: Fall follows by a sudden decrease in body height (c) Body Orientation: In a Fall, angle between body and ground is expected to go from 90 to 0 (d) Body Orientation Temporal Derivative: A sudden change in body orientation is expected in Fall (e) Distance between COM and body supports i.e. feet

4. CONCLUSION

Automatic fall detection is very important problem especially for elderly people living alone. We introduce a novel method for this problem. Contrary to features in the existing approaches, our features are designed by considering reasons of falls in elderly. Experimental results admit the strength of our features and processing algorithm for human fall detection. In comparison to the state of the art, false alarm rate decreased while detection rate remained comparable.

There are limitations in using skeleton data. It is not always reliable. Noise and limited range is its major limitations. Hence we must look for alternative features which are more reliable and available for detecting various events.

5. ACKNOWLEDGEMENT

This work was supported by Turk Telekom Argela under Grant Number 11580-03

6. REFERENCES

[1] S.R. Lord, C. Sherrington, H.B. Menz, and J.C.T. Close. Falls in Older People: Risk Factors and Strategies for Prevention. Cambridge University Press, 2007.

[2] J. Chen, Karric Kwong, D. Chang, J. Luk, and R. Bajcsy. Wearable sensors for reliable fall detection. In Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the, pages 3551 –3554, jan. 2005.

[3] A K Bourke and G M Lyons. A threshold-based fall-detection algorithm using a bi-axial gyroscope sensor. Med Eng Phys, 30(1):84–90, 2008.

[4] M. Alwan, P.J. Rajendran, S. Kell, D. Mack, S. Dalal, M. Wolfe, and R. Felder. A smart and passive floor-vibration based fall detector for elderly. In ICTTA ’06. 2nd, volume 1, pages 1003 –1007, 0-0 2006.

[5] G. Wu. Distinguishing fall activities from normal activities by velocity characteristics. J Biomech, 33(11):1497–500, 2000.

[6] Ji Tao, Mukherjee Turjo, Mun-Fei Wong, Mengdi Wang, and Yap-Peng Tan. Fall incidents detection for intelligent video surveillance. In Information, Communications and Signal Processing, 2005 Fifth International Conference on, pages 1590 –1594, 0-0 2005.

[7] H. Nait-Charif and S.J. McKenna. Activity summarisation and fall detection in a supportive home environment. In Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on, volume 4, pages 323 – 326 Vol.4, aug. 2004.

[8] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau. Fall detection from human shape and motion history using video surveillance. In Advanced Information Networking and Applications Workshops, 2007, AINAW ’07. 21st International Conference on, volume 2, pages 875–880, May 2007.

[9] Tim Liao and Chung-Lin Huang. Slip and Fall Events Detection by Analyzing the Integrated Spatiotemporal Energy Map. 2010 20th International Conference on Pattern Recognition, August 2010.

[10] MS Kinect. http://www.microsoft.com/en-us/Kinectforwindows. Last visited on 4 Feb 2013.

[11] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from single depth images. In CVPR 2011, pages 1297–1304.

[12] Jaeyong Sung, Colin Ponce, Bart Selman, and Ashutosh Saxena. Unstructured Human Activity Detection from RGBD Images. In ICRA, 2011.

[13] G. Diraco, A. Leone, and P. Siciliano. An active vision system for fall detection and posture recognition in elderly healthcare. In Design, Automation Test in Europe Conference Exhibition (DATE), 2010, pages 1536 –1541, march 2010.

[14] Caroline Rougier, Edouard Auvinet, Jacqueline Rousseau, Max Mignotte, and Jean Meunier. Fall Detection from Depth Map Video Sequences. 9th International Conference on Smart Homes and Health Telematics (ICOST), pages 121–128, 2011.

[15] Georgios Mastorakis and Dimitrios Makris. Fall detection system using Kinects infrared sensor. Journal of Real-Time Image Processing, March 2012.

[16] Rainer Planinc and Martin Kampel. Introducing the use of depth data for fall detection. Personal and Ubiquitous Computing, May 2012.

[17] Zhong Zhang, Weihua Liu, Vangelis Metsis, and Vassilis Athitsos. A Viewpoint-Independent Statistical Method for Fall Detection. In ICPR, 2012.

[18] Robinovitch S. Video capture of the circumstances of falls in elderly people residing in long-term care: an observational study. Lancet, 2012.

[19] OpenNI. http://www.openni.org. Last visited on 4 Feb 2013.

[20] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1–27:27, 2011.