Fall Detection for Elderly Using Anatomical-Plane-Based Representation

Rami Alazrai[1], Ahmad Zmily[2] and Yaser Mowafi[3]

Abstract— **Falls are a common cause of injuries and traumas for elderly and could be life threatening. Delivering a prompt medical support after a fall is essential to prevent lasting injuries. Therefore, effective fall detection could provide urgent support and dramatically reduce the risk of such mishaps. In this paper, we propose a hierarchical classification framework based on a novel anatomical-plane-based representation for elderly fall detection. The framework obtains human skeletal joints, using Microsoft Kinect sensors, and transforms them to a human representation. The representation is then utilized to classify the sensor input sequences and provide a semantic meaning of different human activities. Evaluation results of the proposed framework, using real case scenarios, demonstrate the efficacy of the framework in providing a feasible approach towards accurately detecting elderly falls.**

I. Introduction

Falls can have devastating consequences for elderly and may cause moderate to severe injuries, such as hip fractures and head traumas which could increase the risk of early death. It is estimated that one in every three adults age 65 and older falls each year[1]. According to the Centers for Disease Control and Prevention [2], emergency departments in 2011 treated 2.4 million nonfatal fall injuries among older American adults; more than 689,000 of these patients had to be hospitalized. Even if falling patients are not injured, they develop a fear of falling causing them to limit their activities, which results in reduced mobility and loss of physical fitness that in turn increase their actual risk of falling [3].

A timely response to falls is crucial and can prevent lasting injuries and save an older person's self-reliance and in some cases life. For many years, personal medical alert systems have provided help at the touch of a button. These systems help seniors when family cannot be with them in cases of emergencies such as falls. When the medical alarm is activated, the signal is transmitted to an alarm monitoring company's central station and medical personnel are dispatched to the site where the alarm was activated. Unfortunately, traditional medical alarm systems are ineffective if an elderly, after a fall, is knocked unconscious or the alarm button is out of his/her reach. In such scenarios, the person would be unable to activate the alarm to call for help.

With recent advancements in sensors, wireless networks, and smart devices, many medical alert systems are now equipped with fall detection technology built into the wearable help buttons. While this technology seems promising in cases where an individual falls and loses consciousness before pressing the help button, there are some limitations and concerns with existing products. For example, acoustic systems that use microphones to detect and measure vibrations on the floor to identify a fall might not be very accurate. Systems that rely on wearable sensors to detect and analyze movements are ineffective if the elderly during a fall is not wearing the device or the device's batteries are being recharged. Moreover, such wearable devices, if accidently dropped, can cause a false positive where the alarm is triggered but not by a fall. Visual systems that use cameras to track and learn movement patterns to detect falls suffer from invasion and privacy concerns.

In this paper, we propose a novel anatomical-plane-based representation for human-body that is utilized in a hierarchical classification framework for detecting elderly falls. We use Microsoft Kinect sensors to capture an RGB image and a depth image streams for human activity analysis to infer and track human joint positions. Based on these skeletal joint positions, we propose a view-invariant Motion-Pose Geometric Descriptor (MPGD) that consists of two profiles describing the motion and the pose of human body-parts, and capable of capturing the semantic meaning of the performed activities at each frame. The whole sequence of performed activities is then classified into falling or non-falling event. Initial evaluation of the proposed framework has been performed through extensive computer simulations on real case scenarios using Kinect sensor. The results indicate the benefits of the framework in accurately detecting elderly falls while minimizing false positive alarms.

The remainder of this paper is organized as follows: in Section II, we provide an overview of existing fall detection systems and we discuss the related research that this work is based on. Section III describes the anatomical-plane-based human representation and the hierarchical fall detection framework. Section IV presents the experimental simulation setup and results. We conclude with final comments in Section V.

II. Related Work

Existing fall detection systems are classified into wearable and non-wearable systems. Wearable systems rely on devices that utilize several kinds of sensors such as accelerometers and gyroscopes to detect falls [4], [5], [6]. Such devices depend on the elderly to wear them all the time, which might not be very convenient. In addition, those devices require periodic recharging, which make them susceptible to be forgotten to be worn. Moreover, such devices might

[1]R. Alazrai is with the Computer Engineering Department, German Jordanian University, rami.azrai@gju.edu.jo

[2]A. Zmily is with the Computer and Communication Engineering Departments, German Jordanian University, ahmad.zmily@gju.edu.jo

[3]Y. Mowafi is with the Computer Science Department, German Jordanian University, yaser.mowafi@gju.edu.jo

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

not differentiate between a fall and a regular activity like going down the stairs.

Non-wearable fall detection systems utilize environmental devices such as 2D video cameras, motion-capturing systems, and RGB-D cameras. Rougier et al. [7] proposed a method to detect falls by analyzing human shape deformation during a video sequence using Gaussian mixture models. Their approach fails to detect the fall when the person body is showing small shape deformation such as when the person is sleeping on the bed. In a different study [8], the 3D head pose was tracked using monocular 2D cameras to crate 3D trajectory of the head to distinguish falls from normal activities using 3D velocities. Auvinet et al. [9] proposed an approach to detect falls based on reconstructing the 3-D shape of an elderly person using multiple cameras. The proposed system triggers an alarm when a major part of the person's volume distribution along the vertical axis is abnormally near the floor.

In the aforementioned approaches, the input was either 2D video, or the extracted human joint positions using a motion-capturing system, or 3D-reconstruction using multiple cameras. The use of 2D videos makes the approaches sensitive to occlusions, cluttered background, shadow, variation in illumination, and view-point changes, leading to low accuracy in detecting falls. Although motion-capturing systems and multiple-cameras systems may solve the above problems, the requirement of mounting sensing devices on the people, the calibration process of the sensors, and the high cost of these equipments makes it infeasible.

Recently, Microsoft has offered Kinect sensor that combines both RGB camera and depth sensor at a reasonable low cost. Unlike 2D cameras, Kinect is capable of tracking the body-movements in 3D for up to six persons. Furthermore, using only the depth images, person's privacy can be preserved. These advantages have attracted many researchers to use Kinect sensors for human activity analysis and fall detection. Zhang et al. [10] utilized 3D depth information to construct a kinematic model for the monitored person to extract features that are fed into a hierarchy classification scheme and recognize the category of the person's activities. Huang and Pan [11] proposed a frame-by-frame fall detection system based on real-time RGB-D cameras. Despite the high accuracy rate achieve by their proposed system, the system is based on a frame-wise classification which does not take into consideration the temporal variations in the falling event. Garrido et al. [12] utilized Kinect sensors in a system that detects falls and triggers an alert. Gasparrini et al. [13] proposed an automatic, privacy-preserving, fall detection method that utilizes the Kinect depth sensor. In their suggested method, a fall is detected if the depth blob associated with a person is near the floor.

The main advantage of our proposed approach compared to the aforementioned methods is the anatomical-plane-based human representation. The proposed representation is semantically meaningful and consists of motion and pose human profiles that are constructed at each video frame to describe human activities. In addition, we developed a hierarchy classification framework that utilizes the proposed representation for analyzing the human activity in a video sequence. The framework takes into consideration the temporal variations of different activities, which reduces the false classification rate that occurs when frames are classified independently.

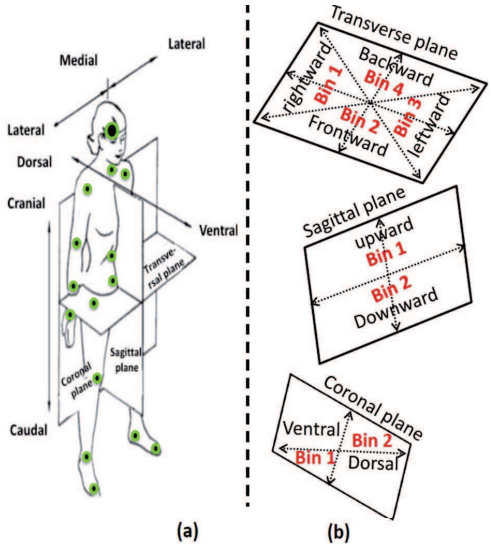

Fig. 1. Motion profile. (a) Left, a schematic diagram of the anatomical planes [14] of a human in a stand-still pose. (b) Right, quantized anatomical planes into semantically meaningful bins.

III. Methodology

A. Anatomical-Plane-Based Human Representation

Human daily life activities can be viewed as spatiotemporal movements of body-parts such as hands, arms, feet, legs, torso, and head. Thus, human activities can be defined in terms of the pose and motion profiles of the body parts, the relative temporal ordering, and the interdependency of moving body-parts. Fall detection systems heavily rely on the representation of the human activity. The more relevant spatiotemporal information that is encapsulated in the human activity representation, the higher the accuracy of the detection system. Our goal is to build a human activity representation that encapsulates both the spatiotemporal data and the associated semantic meaning in a descriptor format. The descriptor captures the motion and poses of human body-parts while preserving the temporal ordering of the moving body-parts. Furthermore, similar spatiotemporal configurations will have similar descriptors regardless of the illumination conditions or the position of the motion-capturing sensor of the person. We use Kinect sensor to capture the activity of an elderly person. We acquire 3D locations of the following twenty skeletal joints: hip center, spine, shoulder center, head, L/R hand, L/R wrist, L/R elbow, L/R shoulder, L/R hip, L/R knee, L/R ankle and L/R foot. Based on these skeletal joint positions, we build a Motion-Pose Geometric Descriptor (MPGD) that consists of two profiles describing the motion and the pose of the human body-parts. The MPGD is capable of capturing the semantic meaning of the performed activities at each frame.

1) Motion profile: The movement of human body-parts is accomplished by muscle contractions and can be viewed

relative to other body parts. For example, moving the right hand upwards can be viewed as a displacement vector between the initial and the final positions of the right hand with respect to the hip center. In anatomy science, various body-parts are described in relation to three imaginary planes (see Fig. 1): Sagittal plane (**SP**), Coronal plane (**CP**), and Transverse plane (**TP**) [15]. Based on the anatomical planes, we have developed a procedure that describes the motion profile for the limbs, torso and the head of each human as follows:

A1. We define first the hip center as the center of an attached local coordinate frame for a human in the scene. Then, the anatomical planes for a human are defined to be planes spanned by three 3D points (see Fig. 1) as follow:

$$\mathbf{SP} =< \mathbf{P}_{hc}, \mathbf{P}_{sc}, \mathbf{P}_{s} > . \quad (1)$$

$$\mathbf{CP} =< \mathbf{P}_{hc}, \mathbf{P}_{ls}, \mathbf{P}_{rs} > . \quad (2)$$

$$\mathbf{TP} =< \mathbf{P}_{hc}, \mathbf{P}_{lh}, \mathbf{P}_{rh} > . \quad (3)$$

Where $\mathbf{P}_{hc}$, $\mathbf{P}_{sc}$, $\mathbf{P}_{ls}$, $\mathbf{P}_{rs}$, $\mathbf{P}_{lh}$, $\mathbf{P}_{rh}$ and $\mathbf{P}_{s}$ represent hip-center, shoulder center, left shoulder, right shoulder, left hip, right hip, and spine points, respectively.

A2. We next construct a sliding window of size W frames over the input stream and calculate a motion profile for a human within that window. Specifically, the motion profile is constructed by computing the displacement vectors for six 3D points (head, spine, right hand, left hand, right foot and left foot) of the skeleton with respect to the hip center point in the first frame in the window. By observing that the intersection of the anatomical planes is dividing the 3D-space around a human into 8 octants (see Fig. 1), we can determine the motion direction of each of the six points by determining the octant in which the displacement vector falls. This can be determined by calculating the signed distance between each displacement vector and each of the anatomical planes to determine whether the motion is to the left, right, above, below, in front or behind a specific anatomical plane.

A3. We quantize the anatomical planes by dividing each plane into a number of bins (see Fig. 1) to determine the semantic meaning of the observed motion. Each bin corresponds to a specific motion direction such as rightward, leftward, etc. Then, we project the displacement vector onto each anatomical plane and determine which bin that contains the projected vector.

A4. We store displacement vectors, signed-distances, and bin number of the projected displacement vector for each of the six 3D points over the anatomical planes as the motion profile $\mathbf{M}_{F_t}$ for each frame in the sliding window. Then, we shift the sliding window to the right and return to step A1.

2) Pose profile: One of the challenging problems in designing a pose profile is that motion that we perceive similar is not necessary spatially similar. This results in producing a large variety of the same performed activity. Müller et al. [16] suggested a set of qualitative geometric features for efficient activity classification of motion-captured data of a single human. In this study, we utilize a subset of the qualitative geometric features to capture different fall poses.

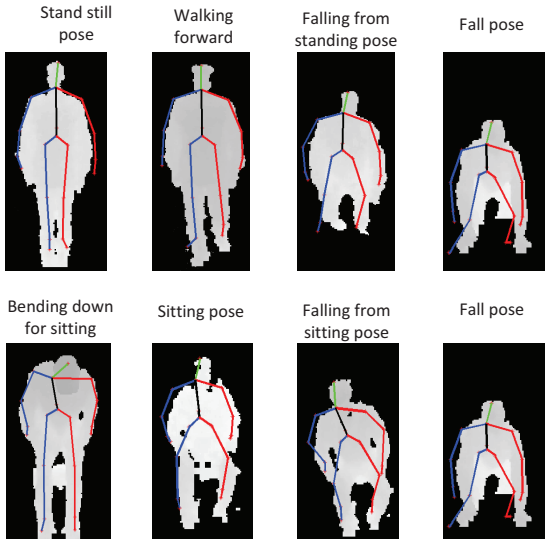

Fig. 2. Collected dataset sample images.

Our pose profile for fall detection consists of two types of relational pose features: joint-distance and angle-based features. Joint-distance pose features $\mathbf{Pose}^{jd}$ are based on calculating the Euclidean distance between the 3D joint positions for the 20 skeletal joints. Joint-distance features are calculated between a)human joints in a single pose and b)human joints at two poses separated by time. Angle-based features $\mathbf{Pose}^{Pl}$ are based on calculating the angels between a)the torso and the lower limbs and b)the line that connects the ankle with the knee, and the line that connects the knee with the hips.

After constructing the motion and pose profiles, the MPGD of a frame F_t is constructed by concatenating both motion and pose profiles into one vector as follows:

$$\mathbf{MPGD}_{F_t} = [\mathbf{M}_{F_t}, \mathbf{Pose}^{jd}_{F_t}, \mathbf{Pose}^{Pl}_{F_t}]. \quad (4)$$

B. Fall Detection

We propose a hierarchical fall detection framework that consists of a representation layer and two classification layers. At the representation layer, RGBD images are acquired from a Kinect sensor, and the 3D joint positions are estimated. Then we construct MPGD representation for each input frame. We then train a set of support-vector-machine (SVM) classifiers to classify each frame into one of different states at the first classification layer. The state of each frame describes the spatiotemporal configuration of the elderly person in that frame (e.g., the person is stand still or is stretching out his right arm). At the second classification layer, the constraint dynamic time warping (cDTW) is utilized to classify the whole sequence of states generated from the SVM classifiers into falling or non-falling activity. The use of cDTW allows processing large variation in duration of human activity video sequences efficiently.

TABLE I

THE RESULTS OF IDENTIFYING THE STATE OF FRAMES IN AN INPUT VIDEO AT THE FIRST LAYER.

State	Precision (P)	Recall (R)	F1-measure 2*P*R/(P+R)
Stand still state	97.19%	96.39%	96.79%
Walking forward	100.00%	100.00%	100.00%
Walking backward	98.99%	100.00%	99.49%
Pending down for sitting	96.85%	98.65%	97.74%
Sitting	99.45%	100.00%	99.72%
Falling down from stand still state	96.31%	96.67%	96.49%
Falling down from sitting	96.28%	96.57%	96.42%
Fall state	97.37%	97.37%	97.37%
Average	97.81%	98.21%	98.00%

IV. EXPERIMENTAL RESULTS

A Microsoft Kinect sensor was used to collect a dataset that consists of four types of human activities related to the falling event as shown in Fig. 2. The four activities include: walking, sitting on chair, falling from chair, and falling from a standing pose. Four different individuals participated in performing the activities resulting in a dataset that consists of 66 sets with different views, speeds and activity styles. Approximately, the dataset contains 14400 frames and 180 activity sequences. The length of each activity sequence is 80 frames on average. For each activity sequence, the videos were captured at a rate of 15 frames per second (fps), the RGB sequence of images at a resolution of 640×480, the depth maps at a resolution of 320×240, and the 3-dimensional coordinates of 20 person's joints at each frame.

We use 5-fold cross validation to evaluate our fall detection framework. Four folds are used for training and one for testing. The individual in testing fold does not appear in training. The window size for MPGD is set to three frames. A one-vs-one approach was used to train a set of SVM classifiers to classify each frame into one of the **8** states. For SVMs, the regularization parameter C and its radial-basis function (RBF) kernel parameter σ were selected based on the cross-validation procedure (Regularization term C=50, RBF parameter=.1).

Table I shows the precision, recall, and the F1-measure results for identifying the state of frames. Precision represents the fraction of frames classified in a specific state that are correctly classified. For example, 96.31% of the frames that were classified as "falling down from stand still" were correctly classified. Recall measures the fraction of correctly classifying frames for a specific state. For example, the system was able to classify 96.67% of the "Falling down from stand still" frames correctly. The first layer in our framework that uses MPGD was able to accurately detect the state of input video frames with an average of 97.81%, 98.21%, and 98.00% for precision, recall, and F1-measure, respectively.

The precision, recall, and F1-measure results of the second layer in classifying predicted state sequences to describe whether the performed activity represent a fall or not are 98.01%, 97.13%, and 97.57%, respectively. The results indicate the efficacy of our proposed classification framework using MPGD for fall detection.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we propose a novel elderly fall detection framework based on an anatomical-plane-based representation. The framework has been evaluated using a dataset of human activities related to fall events using Microsoft Kinect sensors. The experimental results demonstrate the accuracy in detecting elderly falls.

For future work, we plan to study the use of multiple Kinect sensors to monitor a person from different angels and different rooms. The use of more than one sensor would improve the usability and accuracy of our framework. In addition, tracking more than one person simultaneously is another future research direction.

REFERENCES

[1] A. Tromp, S. Pluijm, J. Smit, D. Deeg, L. Bouter, and P. Lips, "Fall-risk screening test: a prospective study on predictors for falls in community-dwelling elderly," *Journal of clinical epidemiology*, vol. 54, no. 8, pp. 837–844, 2001.

[2] (2014) Web based injury statistics query and reporting. Centers for Disease Control and Prevention, National Center for Injury Prevention and Control. [Online]. Available: http://www.cdc.gov/injury/wisqars/index.html

[3] B. J. Vellas, S. J. Wayne, L. J. Romero, R. N. Baumgartner, and P. J. Garry, "Fear of falling and restriction of mobility in elderly fallers," *Age and Ageing*, vol. 26, no. 3, pp. 189–193, 1997.

[4] M. Narayanan, S. Lord, M. Budge, B. Celler, and N. Lovell, "Falls management: Detection and prevention, using a waist-mounted triaxial accelerometer," in *the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug. 2007.

[5] J. Boyle and M. Karunanithi, "Simulated fall detection via accelerometers," in *International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug. 2008.

[6] C.-C. Wang, C.-Y. Chiang, P.-Y. Lin, Y.-C. Chou, I.-T. Kuo, C.-N. Huang, and C.-T. Chan, "Development of a fall detecting system for the elderly residents," in *the 2nd International Conference on Bioinformatics and Biomedical Engineering*, May 2008.

[7] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, "Robust video surveillance for fall detection based on human shape deformation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 5, pp. 611–622, May 2011.

[8] ——, "Monocular 3d head tracking to detect falls of elderly people," in *the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug. 2006.

[9] E. Auvinet, F. Multon, A. Saint-Arnaud, J. Rousseau, and J. Meunier, "Fall detection with multiple cameras: An occlusion-resistant method based on 3-d silhouette vertical distribution," *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 2, pp. 290–300, Mar. 2011.

[10] C. Zhang, Y. Tian, and E. Capezuti, "Privacy preserving automatic fall detection for elderly using RGBD cameras," in *Computers Helping People with Special Needs*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2012, vol. 7382, pp. 625–633.

[11] S.-H. Huang and Y.-C. Pan, "Learning-based human fall detection using RGB-D cameras," in *International Conference on Machine Vision Applications*, 2013.

[12] J. E. Garrido, V. M. Penichet, M. D. Lozano, and J. A. F. Valls, "Automatic detection of falls and fainting," *Journal of Universal Computer Science*, vol. 19, no. 8, pp. 1105–1122, Apr. 2013.

[13] S. Gasparrini, E. Cippitelli, S. Spinsante, and E. Gambi, "A depth-based fall detection system using a kinect sensor," *Sensors*, vol. 14, no. 2, pp. 2756–2775, 2014.

[14] (2011) The biological basis of bone & anatomical directional terms. These Bones of Mine. [Online]. Available: http://thesebonesofmine.wordpress.com/2011

[15] R. S. Snell, *Clinical Anatomy by Regions*, 9th ed. Lippincott Williams & Wilkins, Walters Kluwer, 2011.

[16] M. Müller, T. Röder, and M. Clausen, "Efficient content-based retrieval of motion capture data," *ACM Transactions on Graphics*, vol. 24, no. 3, pp. 677–685, Jul. 2005.