HUMAN ACTIVITY RECOGNITION TO ASSIST ELDERLY PEOPLE

Bhargavi Boppana
bhargavi.boppana@mavs.uta.edu

Gayathri Gottipolu
gayathri.gottipolu@mavs.uta.edu

Geethika Lingamaneni
geethika.lingamaneni@mavs.uta.edu

Department of Computer Science and Engineering
The University of Texas at Arlington

ABSTRACT

As the tendency for elder people to stay alone has increased, the chances and risks of them to fall down and get injured has increased too. Though there are many wearable devices which can detect the activities performed by the adults, they do often forget to wear them. This creates the need for an alternate system. To facilitate independence of older adults in community, reduce risks and enhance quality of life at home by recognizing activities of daily living, our project is to provide a device which can be placed either to the wall or the ceiling and be able to monitor and detect the various activities that are being done by the elderly people or adults and then notifying the concerned whenever there is a fall or falling activity is detected.

KEYWORDS: *Fall-Detection, Openpose, Activity Recognition, Elderly*

1. INTRODUCTION

Falls are the subsequent driving reason for death by unplanned or accidental injuries. Fall pervasiveness increments with age comprehensively and is really viewed as a significant medical issue. They generally require quick medical help since they lead to 20–30% of mellow to extreme wounds or even demise. Falls are one of the most hazardous circumstances for older individuals at home and leading cause behind injury related hospitalization. Falls exponentially increment with age-related biological changes, which is causing a high occurrence of falls and fall related wounds in the ageing societies. Around the world, 424, 000 individuals die because of fall related wounds. Out of this number, the ones who suffer fatal falls all the more frequently are the o individuals whose age is probably more than 65 years. Individuals who lay on the floor for over 60 minutes, died in half year after the fall, despite the fact that the individual didn't endure any direct physical injury. The likelihood of a person falling increases and the chances are almost double after the person falls for the first time. The health business has a major interest for assistive innovation for older in fall detection.

A fall is characterized as an unexpected coming to ground, or some lower level not as an outcome of supporting a vicious blow, loss of cognizance, unexpected beginning of loss of motion as in stroke or an epileptic seizure". Subsequently, various kinds of falls can be recognized as falling while sleeping, for example falling of a bed, falling while sitting, for example tumbling from a seat, couch or comparable and in conclusion, tumbles from strolling or remaining, for example stumbling over the edge of a mat and other rough floors. One of the most expressive qualities of a fall is the abrupt decrease in the tallness of the head or an individual with their head near the ground plane longer for a specific time can be considered as fallen.

The fall detection and classification systems can be broadly classified into three categories. First one is wearable sensor based methods. The main disadvantage of this method is, as they have to be worn always causing discomfort and the elderly might forget to wear them. Second is the vision based systems including CCD camera, multiple cameras, specialized omnidirectional ones and stereo pair cameras. These systems offer multiple advantages over the sensor based systems. They are less intrusive and provide information regarding fall and related injury but they have their set of disadvantages including occlusion problem, privacy issues and poor lighting conditions. The Kinect camera sensor consists of RGB, depth and infrared sensor providing good information even in poor lighting conditions. The depth images preserve the privacy of elders providing just the shape information of the human silhouette. Also, Kinect camera configuration reduce the effect of occlusion to an extent. The third category is the ambient or fusion based systems employing both sensor based and vision based methods into one system, but these are a bit expensive than the other two.

This project presents a fall detection system for the elderly. The system identifies the various activities that are being done and when ever a fall is detected, an email notification or text notification is sent to the concerned people saying that some one has fell down and thus immediate action has to be taken.

2. RELATED WORK

In [1] the possibility of a novel fallen human detector framework dependent on stereo vision utilizing cutting edge 2D human individual identification, stretched out it to accomplish 3D key point data and furthermore ground plane data. We could get an understanding of fall identification utilizing sound system camera and consolidating of 2d key focuses and profundity picture information to appraise 3d present so as to distinguish the fall. In [2] they proposed a new feature representation and extraction technique utilizing a succession of depth silhouettes. Especially, we first concentrate the depth silhouettes by expelling noise and afterward separate the joints in addition to body includes as skin shading identification from joint data and multi-see body shape from profundity outlines (i.e., front and side perspectives). We consolidate the joints in addition to body shape highlights to make feature vector. These highlights have two pleasant properties incorporating invariant as for body shape or size and not sensible to small noises. Self-Organized Map (SOM) is then used to train and test the data.. This paper [3] introduces a machine learning model which helps in identifying the activities performed by human without any idea. For this reason, a LSTM - RNN was applied to three true home datasets and demonstrates the proposed approach beats existing ones as far as accuracy and precision. In this paper [4] we have presented fall detection method that combines shape features and movement. The head has been tracked using particles filter and movement was expressed by covariance of center of mass distance over time. This paper [5] showed multiple sets of features obtained from depth maps collected making use of a Kinect camera for the identification of human activities. The features obtained have details of the silhouette history which is obtained from the number of times the human silhouettes occurred in each activity and the Motion Variation data of the silhouettes calculating difference among motion in between consecutive frames. These features are used for our own annotated depth database and MSRaction3D database, respectively. In [6] a k-NN based classifier system is used to separate lying pose from other daily life activities in a ceiling mounted 3D depth camera. However the processing of depth images is initiated only after the indication for a potential fall is observed by an extra accelerometer system. Shape based features like height and width aspect ratio of a 2D bounding box are used in [7]. The system detects a fall event when the aspect ratios are less than some selected thresholding value. They used an FPGA system integrated within their system. The system in [8] contains two modules. RGB-D camera and the fall detector module based on available libraries for camera management and computer vision procedures. Firstly, the human silhouette is recognized by Kinect and the bounding rectangle is constructed around it by the algorithm. Kalman filter is used to reduce the spikes in the variation of rectangle to reduce classification errors. The features are extracted by evaluating the expansion and contraction of the width, height, depth of the 3D bounding box. The proposed system uses the shape-based features like bounding box, area, orientation, centroid and ex- tent of the silhouette for human fall classification. A binary SVM classifier performs the fall action classification from non-fall using the shape-based features. In [9], X. Yang and Y. Tian proposed a novel feature approach based on position differences of joints and eigen joints which provides human behaviour's information including static postures, motion and offset. These features deal with Naive-Bayes-nearest Neighbor classifier for multi-class action recognition followed by support vector machine (SVM) for recognition. Recent advancement in the depth sensor (i.e., Microsoft Kinect) has made it feasible to capture the depth images/videos in real-time as well as color images [10]–[12]. Moreover, depth images have appropriate resolution (i.e., 640×480) and accuracy with respect to specific subject distance. As compare to the RGB images, depth images have several advantages for the problem of activity recognition. For example, depth images show appropriate information about the shape, which can be quite different from the RGB images in a lot of problems like segmentation, identification of object and recognition of an activity. Depth images likewise give extra body shape and structure data, which have been effectively applied to recover skeleton joints from a single depth image [13]. Furthermore, depth images are insensitive to illumination changes. In this context, it seems very natural and important to use depth images/videos for computer vision problems.

3. PROPOSED METHOD

This project offers automatic monitoring and understanding of activities of patients or residents along with notification system for fall detection. The proposed method to implement this project can be explained in following words: the input to the system is a stream of video coming from a camera. Then, OpenPose [14] algorithm is adopted to detect the 18 2D human body key points in the form of human skeleton(joint positions) from each frame. The skeleton data is preprocessed and used for feature extraction and then it is fed into a classifier to recognize action at each frame. In context of real-time action recognition, identification of action is done frame by frame along the time of video and predicts a label for each video frame. And in this project, whenever a fall is detected, an email and SMS notification are sent.

3.1 Fall detection

3.1.1 Detecting human joint points from image

To detect human skeleton from the image. The main idea of OpenPose[14] is employing a Convolutional Neural Network to produce two heatmaps, one for predicting joint positions, and other for connecting the joints to human skeletons. In simple words, an image is input to the Openpose[14] and the output are the skeletons of all the humans this algorithm detects. Each skeleton has 18 joints, including head, neck, arms and legs, as shown in Figure 1. Every joint position is represented as co-ordinate values of x and y in the image, so there will be a total 36 values for each skeleton.

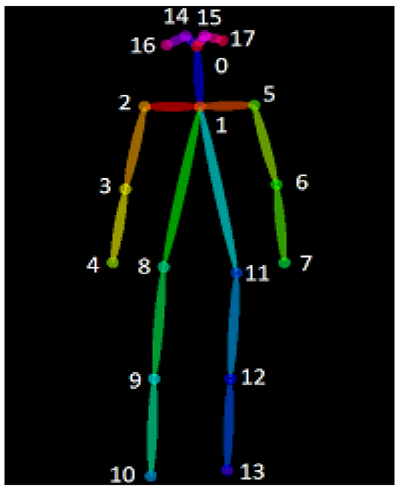


Figure 1. Openpose detected 18 human keypoints

3.1.2 Activity Recognition Classification

The total training data are split into two sets: 80% for training and 20 % for testing. Four different classifiers are experimented, including Support Vector Machine, K-Nearest Neighbor, Decision Tree and Artificial Neural Network. 5-Fold cross validation was used on all these models to ensure that every single class are present in both train and test to avoid any problem. After training our model, we then tested on live webcam to see performance of models on real time. The results are explained in Section 5.

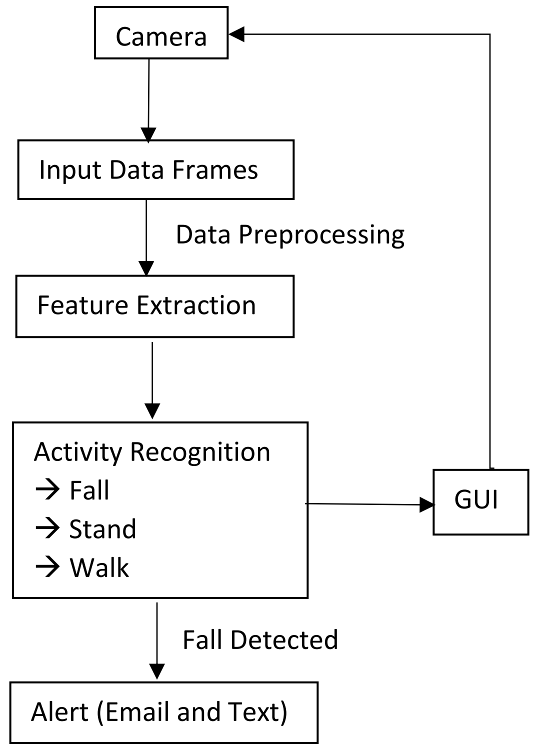
3.2 Notification

If the predicted class is fall, the system will send an alert through email as well as notification to the specified email address and mobile number. There is an alternative email option where system will send an alert to that alternate email address as well. SMTP library is used to send an alert through email and Twilio library is used to send an alert on mobile number as SMS.

3.3 GUI

A GUI is built where it provides the following options: to start the webcam, to display results of the models. It also provides an option in the toolbar to set an alternate email address and reset to default address.

3.4 System Architecture



- The person must click "Start Webcam" on GUI to open the webcam or Kinect sensor.
- The human image from the camera is preprocessed by Openpose and extracted features are fed into a classifier.
- Activities like, Fall, stand and walk are recognized by the classifier
- Once the fall is detected, an email notification and SMS are sent to the registered email address and mobile number.
- There is also an option to add alternate email in the system.

4. STUDY DESIGN AND EXPERIMENTAL SETUP

For the experimental setup, we have used Kinect camera connected to our laptop. To test the project, we have asked few of our friends to try this. Initially, participants were explained the whole process of what and how it should be done. Once they are aware of the entire process, we asked them to be in front of the experimental setup and perform the activities as instructed. We have tested this on 7 people. All the participants were asked to do the 3 activities- stand, walk and fall each 5 times. We were able to get the fall detection and notifications being sent correctly at all the times.

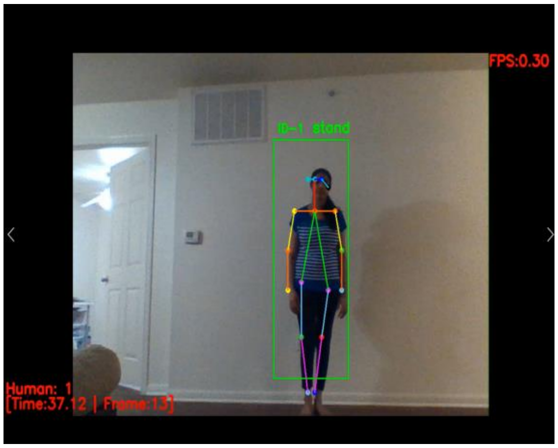


Figure 2 : Person 1 – Stand activity recognized

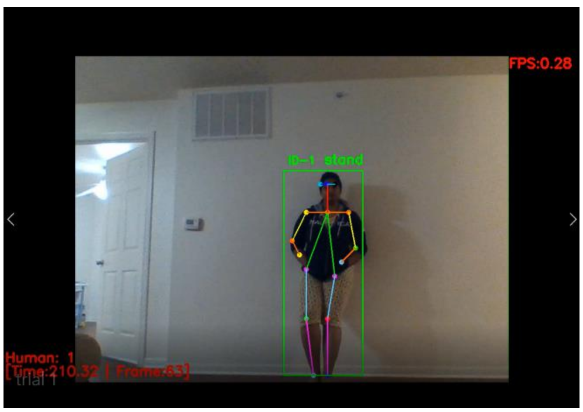


Figure 3: Person 2 - Stand activity recognized

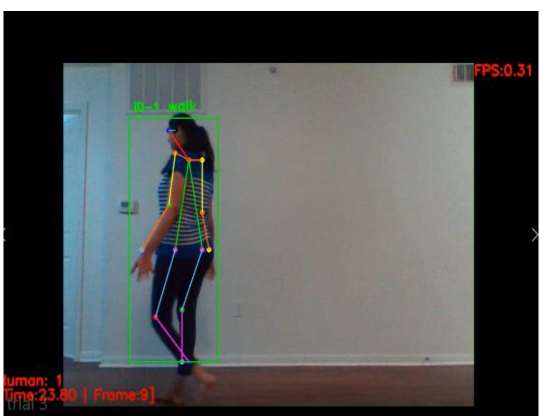


Figure 4: Person 1 – Walk activity recognized

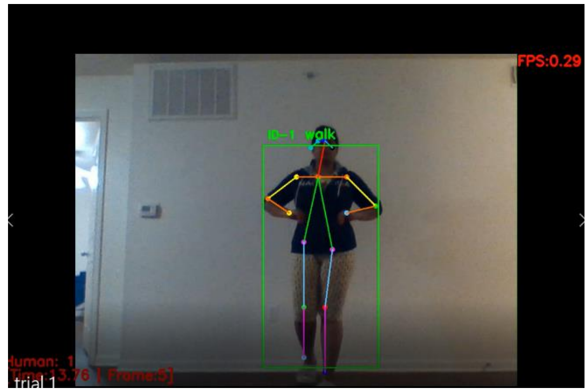


Figure 5: Person 2 - Walk activity recognized

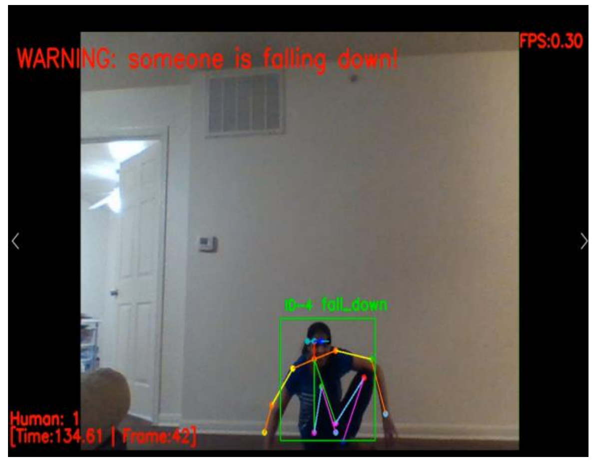


Figure 6 : Person 1 – Fall activity recognized

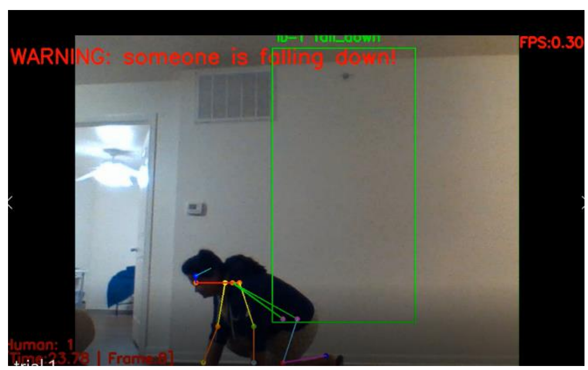


Figure 7: Person 2 - Fall activity recognized

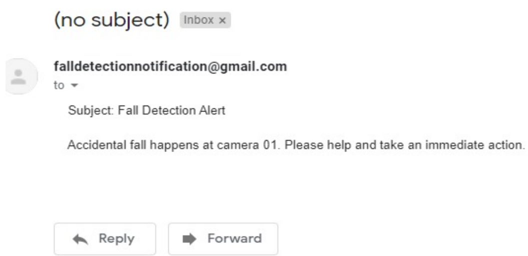


Figure 8: Alert sent through Email notification

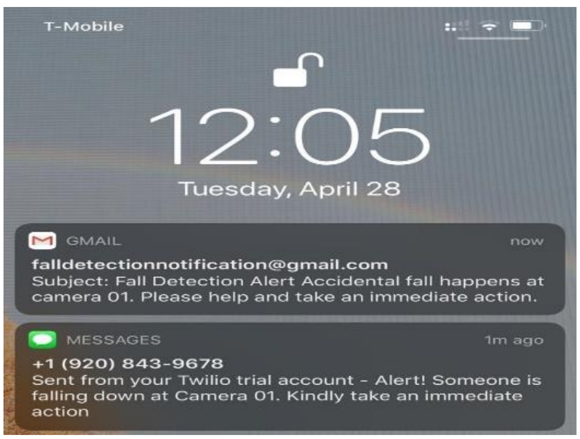


Figure 9: Alert through text message

5. EXPERIMENTAL RESULTS AND EVALUATION OF THE PROPOSED METHODS

The performance of all the models are satisfying on 5-Fold cross validation as the weighted accuracy are satisfying. We have selected K-Fold cross validation as our evaluation criteria because it best describes the performance of model. We used Stratified K-Fold to ensure that every single class are present in both train and test to avoid any problem. Below is the detailed result of each model:

5.1 Support Vector Machine (SVM) model have an accuracy of 98.9% on 80-20 train and test rule. Only 5 predictions were misclassified out of 495. On 5-Fold cross validation accuracies are 98.9%, 98.7%, 99.6%, 99.3, 100% respectively with 99.3% as weighted accuracy. Below is the Confusion Matrix of SVM.

	Stand	Walk	Fall
Stand	146	1	0
Walk	4	140	0
Fall	0	0	204

Table 1: Confusion matrix of SVM

5.2 Decision Tree (DT) model have an accuracy of 99.7% on 80-20 train and test rule. Out of 495 only 1 prediction were incorrect. On 5-Fold cross validation the accuracies are 99.8%, 99.6%, 100%, 100%, 100% respectively with weighted accuracy of 99.8%. Below is the Confusion Matrix of DT.

	Stand	Walk	Fall
Stand	146	1	0
Walk	0	144	0
Fall	0	0	204

Table 2: Confusion matrix of Decision Tree

5.3 Artificial Neural Network (ANN) model have an accuracy of 99.1% on 80-20 train and test rule. Out of 495 only 4 predictions were incorrect. On 5-Fold cross validation the accuracies are 99.8%, 99.4%, 99.1%, 99.8%, 99.5% respectively with weighted accuracy of 99.5%. Below is the Confusion Matrix of ANN.

	Stand	Walk	Fall
Stand	146	1	0
Walk	3	141	0
Fall	0	0	204

Table 3: Confusion matrix of ANN

5.4 K-Nearest Neighbor (KNN) model have an accuracy of 100% on 80-20 train and test rule. There's was not a single incorrect prediction out of 495. On 5-Fold cross validation the accuracies are 100%, 99.8%, 100%, 100%, 99.8% respectively with weighted accuracy of 99.9%. Below is the Confusion Matrix of KNN.

	Stand	Walk	Fall
Stand	147	0	0
Walk	0	144	0

Fall	0	0	204

Table 4: Confusion matrix of KNN

We have trained all the models on the same dataset and the results are satisfying and acceptable as the error rate is very low.

5.5 Real Time Results

In actual setting, we have tried to test on 7 different asked them to do each activity (fall, stand and walk) for five times each and repeated twice. We had 210 cases in total. Out of the four models we trained, Neural Network performed in predicting actions. It predicted 207 cases correctly and showed an accuracy of 98.5%. Decision Tree predicted 201 cases correctly and showed an accuracy of 95.7%. KNN predicted 199 cases correctly and showed an accuracy of 94.7%. Also, SVM showed an accuracy of 93.8% and predicted 197 cases correctly.

5.6 User Feedback Survey:

We asked our friends and roommates to take part in this experiment and give their feedback. The feedback which we got:

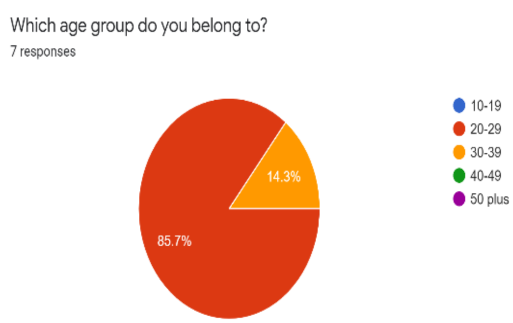


Figure 10: User feedback – age group

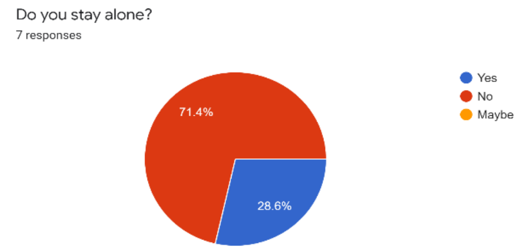


Figure 11: User feedback – Staying alone

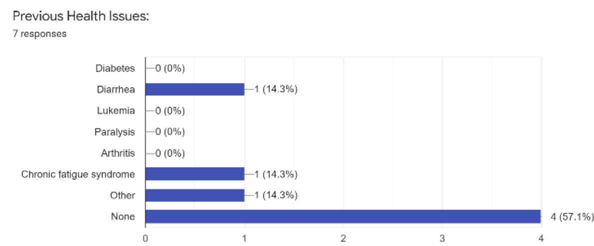


Figure 12: User feedback – Health issues

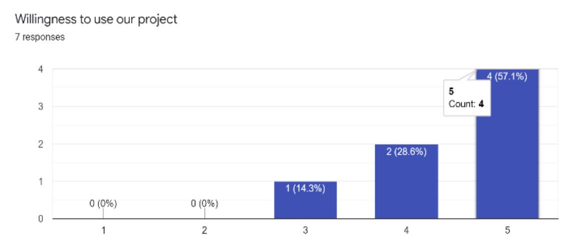


Figure 13: User feedback – Willingness

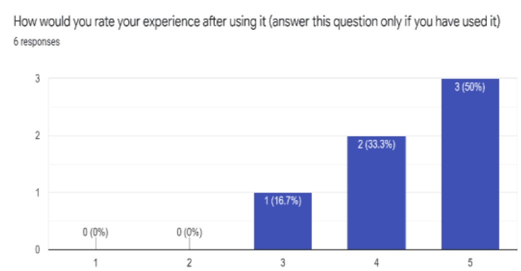


Figure 14: User feedback – User experience

6. CONCLUSION AND FUTURE WORK

In this project, Activities such as stand, walk, fall are classified in this project. These actions are shown above in Figure 2-9. And then whenever a fall is detected, a notification through email and text is sent to the specific person so that immediate help can be provided. To achieve this, we have used an open-source algorithm Openpose[14] to detect human joint points(skeleton) from each video frame and uses the skeleton as raw data for feature extraction. The classification is done by using various machine learning models like SVM, Decision Tree, KNN, ANN. To test the accuracy of this system, the results are presented using the confusion matrix and 5-fold cross validation.

From all the above models, the standout model on real time testing was Artificial Neural Network which predicted quite better than other models. As mentioned, an email and SMS notification are sent whenever there is a fall detection. Thus, we conclude saying that the action recognition can help in detect falls and report emergency to the family in time.

In the future, for a practical application, this project can be further be trained with more samples to detect a greater number of activities and a classifier which can detect action dynamically directly from the video with appropriate hardware and GPU to process them faster instead of classifying the actions framewise. Re-training of models with more samples periodically is also recommended to improve prediction accuracy from time to time.

7. REFERENCES

[1] M. D. Solbach and J. K. Tsotsos, "*Vision-Based Fallen Person Detection for the Elderly*," 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), Venice, 2017, pp. 1433-1442.

[2] A. Jalal, Y. Kim, S. Kamal, A. Farooq and D. Kim, "*Human daily activity recognition with joints plus body features representation using Kinect sensor*," 2015 International Conference on Informatics, Electronics & Vision (ICIEV), Fukuoka, 2015, pp. 1-6.

[3] Singh D. et al. (2017) "*Human Activity Recognition Using Recurrent Neural Networks*". In: Holzinger A., Kieseberg P., Tjoa A., Weippl E. (eds) Machine Learning and Knowledge Extraction. CD-MAKE 2017. Lecture Notes in Computer Science, vol 10410. Springer, Cham

[4] F. Merrouche and N. Baha, "*Depth camera based fall detection using human shape and movement*," 2016 IEEE International Conference on Signal and Image Processing (ICSIP), Beijing, 2016, pp. 586-590.

[5] A. Jalal, S. Kamal and D. Kim, "*Shape and Motion Features Approach for Activity Tracking and Recognition from Kinect Video Camera*," 2015 IEEE 29th International Conference on Advanced Information Networking and Applications Workshops, Gwangiu, 2015, pp. 445-450.

[6] Kepski, Michal and Bogdan Kwolek, *"Fall detection using ceiling-mounted 3d depth camera,"* Computer Vision Theory and Applications (VISAPP), 2014 International Conference on,Vol. 2, pp. 640- 647, 2014.

[7] Ong, Peng Shen, "*An FPGA-based hardware imple- mentation of visual based fall detection*," Region 10 Symposium, 2014 IEEE, pp. 397-402, 2014.

[8] Bevilacqua, Vitoantonio, "*Fall detection in indoor environment with Kinect sensor*," Innovations in Intelligent Systems and Applications (INISTA) Proceedings, 2014 IEEE International Symposium on, pp. 319–324, 2014.

[9] X. Yang and X. Tian, "*Eigenjoints-based action recognition using naive-bayes-nearest-neighbor*," in Proceedings of the IEEE International Conference on computer vision and pattern recognition, pp. 14–19. IEEE, 2012.

[10] A. Jalal, J. T. Kim, and T.-S. Kim, "*Human activity recognition using the labeled depth body parts information of depth silhouettes*," in Proceedings of the 6th international symposium on Sustainable Healthy Buildings, pp. 1–8, 2012.

[11] J. B.-Arie, Z. Wang, P. Pandit, and S. Rajaram, "*Human activity recognition using multidimensional indexing*," IEEE Transactions

on Pattern Analysis and Machine Intelligence, vol. 24, no. 8, pp. 1091– 1104, 2002.

[12] P. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea, "*Machine recognition of human activities: A survey*," Circuits and Systems for Video Technology, IEEE Transactions on, vol. 18, no. 11, pp. 1473– 1488, 2008.

[13] A. Jalal, N. Sarif, J. T. Kim, and T.-S. Kim, "*Human activity recognition via recognized body parts of human depth silhouettes for residents monitoring services at smart home*," Indoor and Built Environment, vol. 22, no. 1, pp. 271–279, 2013

[14] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei and Y. A. Sheikh, "*OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields*," in IEEE Transactions on Pattern Analysis and Machine Intelligence.