International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 06 | June - 2021 ISSN: 2582-3930

Fall Detection based on Machine Learning using PoseNet model

K Saurabha Bhat[1], Samreen Bai[2], Sarah Jacob[3], Shubhangi Hallale[4], Dr. Shashikumar D.R[5]

[1,2,3,4] *UG Student, Department of Computer Science and Engineering, Cambridge Institute of Technology, Bengaluru, Karnataka, India*

[5] Head of Department, *Department of Computer Science and Engineering, Cambridge Institute of Technology, Bengaluru, Karnataka, India*

---***---

Abstract- Research done by WHO estimates fall to be the second leading cause for fatal injuries worldwide and it's estimated to be 80 %. South East Asia account for 60 % of fall related injuries. Among elderly, around 3.7 crore falls need medical attention and around 6.5 lakh individuals die due to fall [1,2]. In this paper, we propose a method to detect a fall. We extract the spatial locations of key body joints (key points) of human skeleton using PoseNet and determine the fall using several parameters. We also estimate if a person can stand back up after a fall. An alarm is activated when a fall is detected. The accuracy obtained using this model is 86%.

Key Words: Fall Detection, PoseNet, pose estimation

1. INTRODUCTION

Falls is the second leading cause of unintended injuries worldwide. Children who are in their growing stages, workers in factories, and the elderly are 3 main categories prone to fall injuries [1]. Research has shown that sooner the people are assisted, the risk of death is reduced by 80% after a fall. Also, the requirement of long – term treatment is reduced by 26% [2]. Thus, fall detection systems are important. There are mainly three types of fall detection system: Wearable Sensor based, Environment / Ambient based, and Computer Vision based systems.

In Wearable Sensor based systems, the bearer wears the sensor - based device on wrists like a watch, under or top of clothing. These systems usually have accelerometers, gyroscopes to obtain patient's position information, heart rate sensors, to detect the fall [4]. There are certain shortcomings to wearable devices; Some people forget to wear them; some are injured due to the device upon fall, false alarms are high [5].

Environmental / Ambient based systems have microphones and infrared sensors in the environment where fall have to be detected. The sensor data is sent to the personal computer where analysis of fall is done. Floor vibrations, infrared images and sound along with machine learning techniques is used to determine a fall.

This method is said to be less accurate than wearable device as the sensing range is more and is prone to environmental interference. These systems are expensive as more equipment has to be installed in sensing environment [6].

In Computer Vision based systems, information from raw images is used for predicting a fall. These raw images can be obtained by video sequences and perspective view from multiple cameras. Different machine learning techniques is used with probabilistic models to predict the fall behaviour [7]. These systems are low cost and high accuracy.

This paper proposes a method where every frame of the video is extracted for determining a fall. We use PoseNet algorithm to determine key body joints (key points) of a human skeleton for every person present in the frame. In the plane coordinate system, the x - and y- coordinates are used to represent each key point. We then apply 3 critical conditions (Velocity of descent, angle between the ground and human body centreline, Height to Width ratio of human) to determine the fall. We also determine if the person can stand back up after the fall.

The rest of the paper is divided as follows: Section II reviews the current systems present for Fall Detection; Section III explains in detail the methodology used in this paper; Section IV presents the results obtained and Section V concludes the discussion with possible Future Works.

2. LITERATURE REVIEW

The human skeleton is defined as torso, head and limb [8]. The positions of the joints in a human skeleton determines the posture of a body. Hui Tang et al. has proposed a method for 3D Human Pose Estimation. RGBD camera is used to obtain both colour images and depth images. First, to obtain joint points in colour images a convolutional neural network is used for 2D human pose estimation. Second, the returned result is then mapped onto corresponding depth image which gives the 3D joint point information. In the mapping process, SURF algorithm based sparse feature point matching method is used which determines the calibration parameters of both colour images and depth images [9].

Weiming Chen et al. reorganised accidental falls based on symmetry principle. First, OpenPose algorithm to extract skeleton information of the human body. Next, three parameters (Velocity, angle and bounding box ratio) were used to determine the fall. They also checked if the person could stand back up after the fall [10].

Raul Igual et al. has explored the challenges, issues and trends among the developed fall detection systems. First, an introduction to the three types of fall detection systems is shown. It is noticed that the studies regarding ambience / environmental based systems are increasing. A new trend of integrating fall detection systems into smartphones and using machine learning techniques to detect falls is seen. The challenges discovered in the survey was: The performance of the systems in real life conditions is decreasing; Usability of the system were less; Acceptability of this technology was very little. These systems have some common issues such as: Limitations in smartphones; there are privacy concerns with respect to the data collected [4].

3. METHODOLOGY

The workflow of our proposed method comprises of using the Posenet model to get the skeleton information of the human body in the form of key- points. On constant analysis of these key-points three Decision Parameters have been formulated to detect and classify the action as fall or not fall. The following are the Decision conditions: 1) The Velocity of descent of the person (2) The angle between the head points of the first frame and the 3rd consecutive frame (3) The width to height ratio of the human body's bounding box. Lastly, the person's posture is checked that is, if he stands up after falling or not. An alarm is raised based on these derived results. The flowchart of implementation of our proposed approach is as shown in Figure 1.

3.1 PoseNet

Key points are spatial locations in the image that determine what is interesting or stands out in that image. The joints present in the human skeletal body are considered as key-points which help in determining the posture formed or movements made by the human body. The skeletal structure comprises of the following key-points: knees, ankles, elbows, shoulder, wrists and the orientation of limb. Machine Learning technique uses the joints of the human body as key points to predict human posture and action.

To obtain these key points, joints extraction of a human body is done using deep learning algorithms. When video records are used as a data source, key points are detected from a sequence of frames, not just a single picture. It allows us to achieve more accuracy as the system analyses an actual movement of a person, not a steady position. There are several pretrained models available to extract these key points and estimate the pose through the skeleton figure.

OpenPose and PoseNet are two such models. OpenPose is the first real-time and multi-person system to detect human body, facial, hand and foot key points (i.e., in total 135 key points) on single image. PoseNet is a TensorFlow model which allows you to estimate and track human by detecting body parts such as, hips, wrists, elbows, knees, and ankles. It is a Convolutional Neural Network (CNN) architecture that focuses on key point mapping. In our project we used PoseNet to obtain the key points.

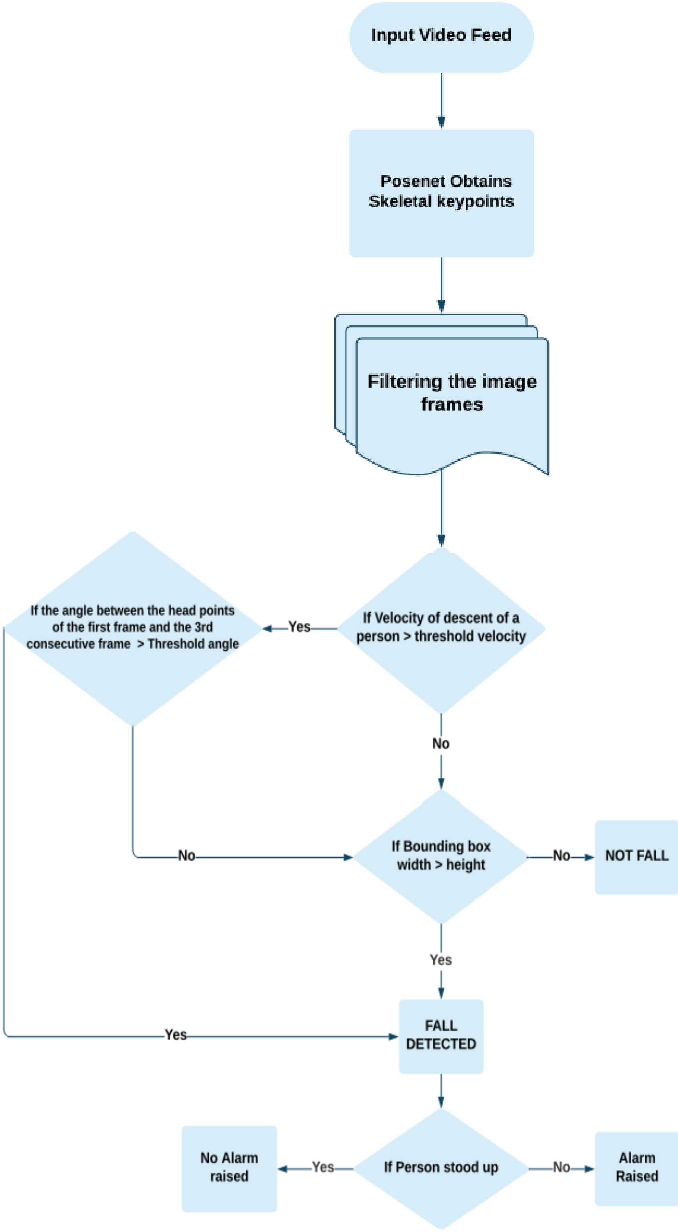

Figure 3.1: Workflow of Project.

Twelve key points are taken in total. They represent the left and right elbows, shoulders, hands, hips, knees and foot. These key points are used by the parameters, that will be mentioned later on in the paper, to determine the fall.

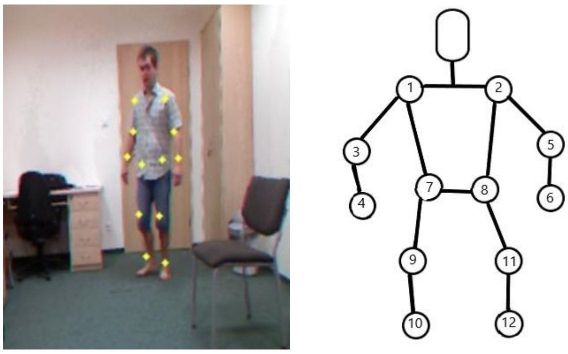

Figure 3.2: Key points obtained from PoseNet

3.2 Filtering

Models do not always give accurate key points and hence a deformed skeleton figure might be formed. This is due to various limitations such as: different camera angle, other objects in the frame that cause as distraction etc. For some images, the accuracy of the coordinate values of the key points is not very ideal to analyse for fall. This problem may be due to the defects of PoseNet algorithm itself, but the deviation of these key points has some effect on the recognition of the whole fall action.

As shown in the Figure 3.3, various objects in the room have confused the model and hence returned wrong key point values. Therefore, the skeleton figure is randomly drawn across the entire image. Just like the image show in Figure 3.3, almost 90% of the images that had wrong skeleton figures, their edges were longer than half the image itself. Using this as a parameter, we filtered out images that had skeleton edges greater than half the image.

Figure 3.3: Image with incorrect skeleton figure

Hence, it is important to eliminate such images because they disturb the calculations applied to detect the fall.

3.3 First Decision Parameter

Velocity is the first parameter considered for fall detection in rate of change of its position with respect to a frame of reference. When a person falls, the velocity of descent is much higher compared to when doing he's doing any day-to-day activities/ non-fall activities.

As show in Figure 3.4, the no of frames it takes for a person to fall is 3 frames and in most of the videos in the dataset it ranges from 1 to maximum of 5 frames. In Figure 3.5, the target takes more than 25 frames to lie down. This difference in frames shows the speed difference at which both the actions were done. And hence this parameter was used to calculate the fall.

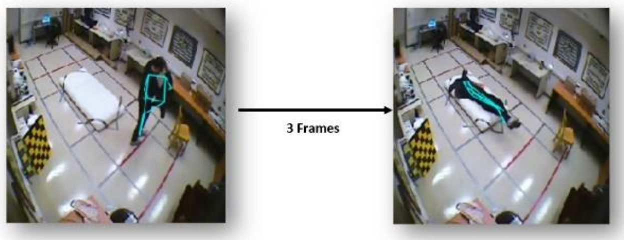

Figure 3.4: Number of frames taken for a person to fall

Figure 3.5: Number of frames taken for a person to lie down

To calculate the velocity, we have taken the hip point (mean of points 7 and 8 in Figure 3.2) as the centre of gravity that changes vertically as a person falls. The hip point of the person before and after he falls, that is at time t_1 and t_2, are taken (A difference of three frame is considered in our model).

Taking the coordinates of the hip at frame 't_1' and 't_2' as (x_1, y_1) and (x_2, y_2), we calc the velocity using the formula (1)–

$$v = \frac{|y_2 - y_1|}{t_2 - t_1} \qquad (1)$$

We take into consideration only the y coordinate since we want to know the displacement horizontally. If the calculated velocity 'v' is greater than a set threshold value 'v_0' then we conclude the person has fallen. If not, the person has not fallen. (The threshold value taken in our model is 25).

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 06 | June - 2021 ISSN: 2582-3930

3.4 Second Decision Parameter

Angle is the second decision parameter. In the fall process, the most obvious and observable feature is the change in the angle of the body. In order to reflect this characteristic of the body's continuous change in inclination while falling, an angle is taken into consideration by using three points. The 3 points are as follows:

1) The head point of the first frame(A)

2) The midpoint of the 2-foot points (B) and

3) The head point of the 3rd consecutive frame(C).

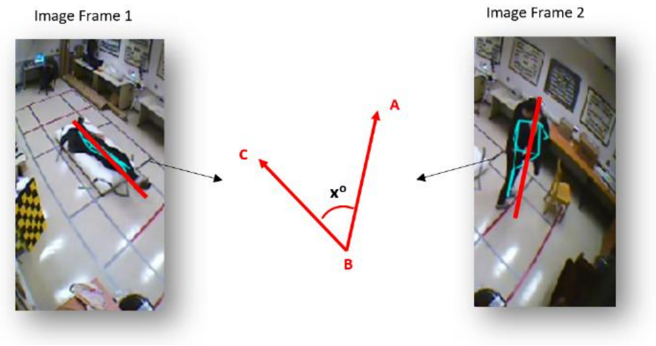

Figure 3.6: Angle of a human before and after a fall

The above Figure 3.6 shows the two frames - before and after the fall, that are taken to calculate the angle of the fall. The head point in image frame 2 i.e., the frame before the fall is taken as point A. The mean of the foot points in both frame 1 and 2 is taken as point B. And lastly the head point in frame 1 – frame after fall, is taken as point C. We used the formulae mentioned below to calculate the angle formed by these three points –

$$|\overrightarrow{BA}| = A - B \quad (2)$$

$$|\overrightarrow{BC}| = C - B \quad (3)$$

Dot Product of BA and BC:

$$\cos\theta = \frac{|\overrightarrow{BA}|.|\overrightarrow{BC}|}{|\overrightarrow{BA}| * |\overrightarrow{BC}|} \quad (4)$$

$$\theta = \cos^{-1}\frac{|\overrightarrow{BA}|.|\overrightarrow{BC}|}{|\overrightarrow{BA}| * |\overrightarrow{BC}|} \quad (5)$$

If the angle θ is greater than a set threshold angle θ_0 then it is categorised as fall else not fall. (The threshold value taken in our model is 50).

3.5 Third Decision Parameter

The final decision parameter considered for the prediction of the extreme condition fall is the bounding box. The bounding box is an imaginary box which encloses the object of interest (person) in it. This box gives a rectangular outline of the person in the frame using the person's height and width. When one falls, the change in these contours is maximum and visibly notable. Hence, it is an important parameter which can be used along with the previous two parameters.

Figure 3.7: Bounding box enclosing target

In actions where a person is standing or walking as shown in the above figure, the height of the bounding box is always greater than the width. This ratio, B=width/height is always less than one where the bounding box height is greater than width. In extreme conditions like fall, the outer rectangle of the person will drastically change where the width is greater than the height and value of the ratio B will be greater than one. Hence, in our model if B is greater than a threshold value, it is concluded as a fall.

The width and height of the target is calculated by finding the minimum and maximum values of each of the x and y coordinates from all the key points, in the skeleton figure. With these calculated values we determine the width and height using the formula:

$$w = max\_x - min\_x \quad (6)$$

$$h = max\_y - min\_y \quad (7)$$

3.6 Fall Detection

Hence to finally detect the fall, we clubbed the first two parameters, velocity and angle, such that if both the parameters return as fall, then we conclude the person has fallen. If the fall was not detected by these parameters, the third parameter i.e., the bounding box is checked.

Now, it is important to check whether the person actually requires help because there are scenarios where the person might be able to recover and get up by themselves within a given period of time. Hence not all falls need to activate an alarm for help. To detect this, our model checks whether the target has managed to get up within a certain period of time The below image shows a series of frames of the person standing up after a fall.

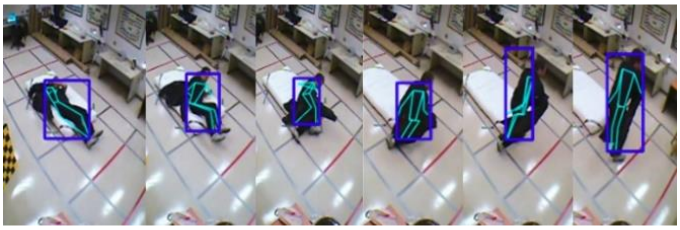

Figure 3.8: Sequence of frames of target standing up

The model checks two factors to see if the target has stood up – the angle and the bounding box. The same method used to detect the fall can be applied to detect whether the target is standing up. With respect to the bounding box, if the height is greater than the width after a certain number of frames, we conclude that the person has stood up. Hence if the model detects the person standing after the fall, the alarm is not activated. But if the person is not able to stand an alarm is activated for immediate assistance.

4. RESULTS

To determine the fall event which is the extreme condition in our paper, we propose 3 decision parameters: The velocity of descent, angle of inclination and the bounding box. In order to verify the effectiveness of this proposed method, the fall event is tested with a set of videos. The test videos consist of actions performed by a single person in different environments and different positions of the camera. The actions collected for testing are divided into 3 categories, namely fall actions (fall, stand up after the fall), actions similar to falling (squat, stoop, crouch), actions involved in daily activities (walk, sitting down, lying down). A total of 50 actions are collected, including 25 falling actions and 25 non-falling actions, each lasting about 4–11 secs.

Actions	**No of videos taken**	**Videos correctly classified**
Fall	20	18
Stand after falling	5	3
Walking	10	10
Other Actions	15	12

Table 4.1: Test Results

The accuracy according to the final result is 86%. From the results above, it is observed that the accuracy is not perfect, the reason for this wrong discrimination of the following: (a) The lack of accuracy in key points estimation by the PoseNet model affected the recognition of the correct posture which finally gave false results. (b) The two thresholds selected for velocity and angle in the experiment are not necessarily optimal for all instances. (c) The placement of the camera to analyse the video is also a major factor considered as the results are not favourable in extreme positions for an instance, over the head of a person. (d) Since the videos are primarily made for dataset purposes, they may be different from real falls.

5. CONCLUSIONS

As the elderly population will increase in the nearing future, it is necessary to determine characteristics of fall to build an accurate Fall Detection System. Droghini et al. predicted a fall by capturing the sound waves transmitted on the fall. But this method is highly vulnerable to interference from external noise and the environment can be limited [11]. Shahzad et al. uses sensors in smartphones to predict the fall. This system is noticed to cause a lot of false positives and it requires the person to wear the phone at all times [12]. Quadros et al. used machine learning and a threshold method together to fuse multiple signals that predicts a fall. But then again, the person needs to wear the device for a very long time [13]. Hence, we used Vision Based method.

PoseNet algorithm is used to get the key body joints of human skeleton. The key points of the human skeleton are used to recognise falling motion based on three critical parameters; 1. The angle between human body and ground, 2. Width to height ratio of human body and 3. Speed of descent at the hip joint. We also check if the person can stand back up after the fall which is the opposite process of falling. Our method is simple and low cost.

REFERENCES

[1] Fall Statistics: https://www.who.int/news-room/fact-heets/detail/falls

[2] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6625258/

[3] Romeo, L.; Marani, R.; Petitti, A.; Milella, A.; D'Orazio, T.; Cicirelli, G. Image-Based Mobility Assessment in Elderly People from Low-Cost Systems of Cameras: A Skeletal Dataset for Experimental Evaluations. In Proceedings of the 2020 International Conference on Ad-Hoc Networks and Wireless, ADHOC-NOW 2020, Ad-Hoc, Mobile, and Wireless Networks, Bari, Italy, 19–20 October 2020; Lecture Notes in Computer Science. Volume 12338, pp. 125–130.

[4] Igual, R., Medrano, C. & Plaza, I. Challenges, issues and trends in fall detection systems. BioMed Eng OnLine 12,66(2013). https://doi.org/10.1186/1475-925X-12-66.

[5] Noury N, Fleury A, Rumeau P, Bourke AK, Laighin GO, Rialle V, Lundy JE: Fall detection - principles and methods.In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Lyon: Institute of Electrical and Electronics Engineers;2007:1663–1666. doi:10.1109/IEMBS.2007.4352627

[6] Lin, C.-B.; Dong, Z.; Kuan, W.-K.; Huang, Y.-F. A Framework for Fall Detection Based on OpenPose Skeleton and LSTM/GRU Models. *Appl. Sci.* **2021**, *11*, 329.

[7] http://njppp.com/fulltext/281433399870.pdf

[8] Johansson, G. Visual perception of biological motion and a model for its analysis. Perception & Psychophysics 14, 201–211 (1973). https://doi.org/10.3758/BF03212378

[9] H. Tang, Q. Wang and H. Chen, "Research on 3D Human Pose Estimation Using RGBD Camera," 2019 IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC), 2019, pp. 538-541, doi: 10.1109/ICEIEC.2019.8784591.

[10] Chen, W.; Jiang, Z.; Guo, H.; Ni, X. Fall Detection Based on Key Points of Human-Skeleton Using OpenPose. Symmetry 2020, 12, 744. https://doi.org/10.3390/sym12050744

[11] Droghini, D.; Principi, E.; Squartini, S.; Olivetti, P.; Piazza, F. Human fall detection by using an innovative floor acoustic sensor. In Multidisciplinary Approaches to Neural Computing; Springer: Berlin/Heidelberg, Germany, 2018; pp. 97–107.

[12] Shahzad, A.; Kim, K. FallDroid: An automated smart-phone-based fall detection system using multiple kernel learning. IEEE Trans. Ind. Inform. 2018, 15, 35–44

[13] De Quadros, T.; Lazzaretti, A.E.; Schneider, F.K. A movement decomposition and machine learning-based fall detection system using wrist wearable device. IEEE Sens. J. 2018, 18, 5082–5089

[14] https://manivannan-ai.medium.com/find-the-angle-between-three-points-from-2d-using-python-348c513e2cd