Activity Recognition for Indoor Fall Detection in 360-degree videos using Deep Learning techniques

Dhiraj[1], Raunak Manekar[1], Sumeet Saurav[1], Somsukla Maiti[1], Sanjay Singh[1], Santanu Chaudhury[1]

Neeraj[2], Ravi Kumar[2], and Kamal Chaudhary[2]
[1] CSIR-CEERI, Pilani, INDIA
[2] Samsung Research India, New Delhi, INDIA
dhiraj@ceeri.res.in

Abstract. Human activity recognition(HAR) targets the methodologies to recognize the different actions from a sequence of observations. Vision-based activity recognition is among the most popular unobtrusive technique for activity recognition. Caring for elderly, living alone from remotely is one of the biggest challenges of modern human society and is an area of active research. The usage of smart homes with increasing number of cameras in our daily environment provides the platform to use that technology for activity recognition also. The usage of omnidirectional cameras for fall detection activity minimizes the requirement of multiple cameras for fall detection in the living scenario. Consequently, we propose a vision-based solution using convolutional neural networks and long short-term memory networks using 360-degree videos for human fall detection. We have constructed an omnidirectional video dataset by recording a set of activities performed by different people as no such 360-degree video dataset is available for human activity recognition in public domain. Our results prove the suitability of 3DCNN and LSTM techniques for fall detection activity for even omnidirectional videos.

Keywords: Activity Recognition, Omnidirectional Video, 3D Convolutional networks, Long short-term memory networks. Fall Activity, Daily Activity.

1 Introduction

Human Fall is one of the major health risks in modern life style particularly for old people who are living alone, which may cause death in some situations if proper medication is not followed after the event. Along with this, it may result in post-fall syndrome such as permanent immobilization, depression, etc. which further restrict the movement. As reported, approximately 40 % of deaths due to injury are primarily caused due to fall only. So, early detection of fall is an important step so as to timely support the elderly by warning or informing their family members. The fall accidents cannot be completely prevented but fall detection system can save lives if it can identify a fall event and an alert has been generated instantaneously.

In this paper, we focus on vision-based approaches for fall detection. The cameras have now become omnipresent as they provide very rich information about persons and their environment. The complete view of living room is provided using single omnidirectional camera installed in room ceiling with downward facing arrangement. So, vision-based fall detection systems prove to be reliable and play a major role in future health care and assistance systems. Due to rapid rise in neural network architectures, deep learning has improved the results obtained in many techniques such as image classification, object detection, segmentation, and image captioning. In this paper, we present an approach using 3DCNN and LSTM techniques for fall detection. More specifically, we propose architectures which detect fall in 360 degree videos which take advantage of the capacity of CNN's to be sequentially trained on the generated 360-degree dataset. First of all, we have generated an Omnidirectional dataset in lab environment. It is then subsequently used to acquire the relevant features for activity recognition. As an outcome of this research carried out, this paper presents the following main contributions:

1. To the best of our knowledge, this is the first time that fall activity has been targeted on 360 videos using deep learning approach. In that respect, it is crucial to address this problem as no such video database exists in public domain which boost up the activity recognition.
2. The techniques such as Frame differencing(FD), Dense optical flow(OF) and normal RGB has been tested and compared for their suitability for fall detection in 360 degree videos.
3. As fall detection is a sequential activity, so techniques which can model the temporal relation between frames such as 3DCNN and LSTM have been used and hyper-tuned for their best utilization in this activity.

2 Related Work

The fall detection activity has been targeted by three main approaches, namely ambient device based, wearable device based and video based. The ambient based approaches majorly use pressure sensors due to its cost-effective and less intrusive nature, but it has its inherent disadvantages also in terms of false alarms and volume quantity usage for better signal acquisition [1]. The wearable sensor devices rely on sensors such as accelerometers [2] or posture sensors [3] to detect the motion and location of person but are very inconvenient for elderly fall detection due to its intrinsic intrusive nature and burden of time for sensor placement. In case of fall activity, measures such as vertical acceleration are strictly distinct from other normal daily activities or confounding events such as bending over or squatting which allows us to separate fall from other similar events. Video-based methods are nowadays most popular detection strategies due to their unobtrusive nature, e.g., easy availability, and wide environment suitability. In most of the video-based techniques, video capture, its analysis and inference display or communication are majorly involved. In those techniques only frames obtained from video provide the relevant information about activity performed. From them, the useful features in the form of bounding boxes or silhouette are extracted to facilitate the activity detection. In some techniques such as Gaussian Mixture Model(GMM), Support Vector Machine(SVM) or Multilayer Perceptron(MLP) those features are used as input to the classifier to trigger the automatic detection of fall activity and can be further extended to tracking also as proposed by Lee [4]. The silhouette information can also be used to detect positions and in combination with a matching system, the deformation in the shape of the human body can be utilized [5]. The person's head can also be tracked to improve the results along with Gaussian classifier [6]. Vallejo et al. [7] and A Sengto et al. [8] proposes to fed the data of 3-axis accelerometer to a Multilayer Perceptron (MLP) for activity recognition. Kepski [9] used the depth maps derived from Kinect in conjunction with the inertial measurement unit and classifier in the form of a SVM. F. Harrou[10] used vision-based approach along with sensor techniques only to ascertain the results. So, out of three feasible approaches video-based activity recognition is the most widely used by several researchers. For example, human shape variation is used to detect fall by Foroughi[11]. Features extracted in form of the best fit ellipse around the human body, histograms of silhouette and temporal changes in head pose, are fed to a traditional classifier such as SVM [12] or an MLP ANN [13] for classification of motion and detection of fall action. People have targeted the fall detection using combinations of time motion images and also using Eigen space methods [14]. The bounding boxes of the objects have also been computed to find out if they contain a human and the fall event has been detected by means of features extracted from it [15,16]. Wang [17] formed a PCAnet to generate features from RGB images which are then used the input to a SVM classifier to detect falls. A similar approach has been targeted by vishwakarma[18] in which bounding boxes are used to compute the aspect ratio followed by horizontal and vertical gradients of a human along with the fall angle and are fed into a GMM for final event detection.

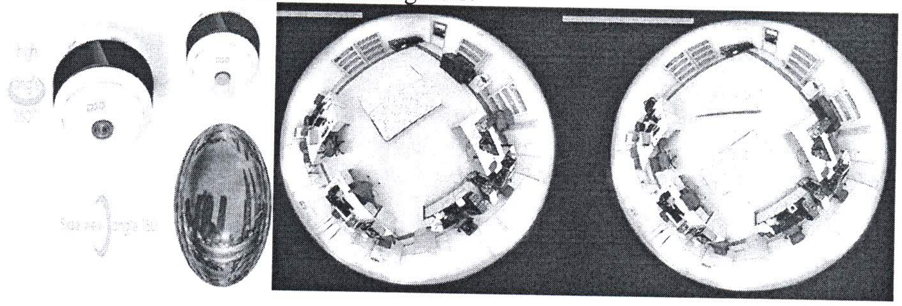
Fall activity has also been targeted by using supervised learning methods in which features are extracted from raw data and using them as input to a classifier to learn class decision for example charfi[19] derived 14 features from raw data and applied some transformations to them and used SVM as a classifier. The occupancy areas around body's gravity center followed by angle extraction and subsequently to SVM is proposed by zerrouki [20]. Curvelet coefficients are also used in the form of features and HMM is used to model different body poses as proposed by zerrouki[20]. Vision-based fall detection techniques also include using 3D information about the environment in which fall event takes place by using depth cameras such as Microsoft Kinect or time-of-flight cameras. 3D silhouette based analysis of events have been proposed by Auvinet et al [21] for subsequent analysis in the form of volume distribution along an axis. Extraction of 3D features using Kinect and then using it in a tracking system to detect the falls is also proposed by Gasparrini et al [22] for detecting the orientation of major axis and depth maps to compute 3D features by Diraco et al [23]. People such as mastorakis [24] have adapted the 3D bounding box strategy to harness the 3D information. The major drawback of 3D based approaches is in terms of the system deployment due to their narrow field of view and limited depth and requirement of multiple synchronized cameras focused on the same area.

The present research focus is mainly directed along usage of deep learning based methods for vision-based inputs in several applications such as feature combination in form of Histogram of Oriented Gradient(HoG), Local Binary Patterns(LBP) and features generated from caffe[25] neural networks are used to recognize a silhouette and then used as SVM classifier. The motion in the video was also targeted by using optical flow algorithm and to avoid the influence of static image features [26]. However, instead of doing feature engineering ourselves the deep learning based approaches provides end to end learning in terms of feature extraction and classification to be handled on their own.

3 Database and Techniques

3.1 Dataset

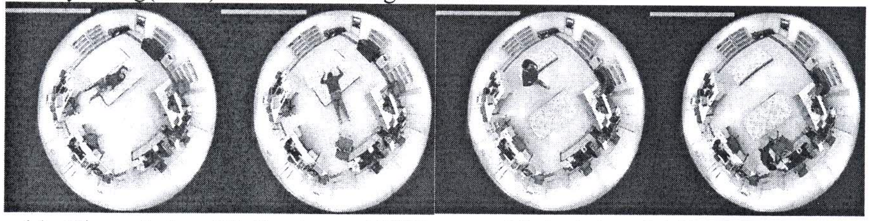
To facilitate the learning and testing of any model and classifier we need a large amount of data. As fall event is multi-frame based activity so a set of frames are required for learning by a model. Based on current literature, it was found that out of the vision based databases which are publically available none of those belong to 360-degree category. 360-degree fisheye cameras are a popular topic in the security industry today and offer a 360-degree view from just single lens, and give users the ability to digitally pan, tilt and zoom in the live and stored video along with the widest field-of-view available without blind spots making them ideal for wide areas. Since 360-degree cameras have no moving parts, they avoid the lag (latency) associated with PTZ devices. In addition, fisheye cameras tend to have a longer life, since all of their parts are stationary and do not wear out as quickly. The specifications of camera used in database preparation are: Camera Model: D3D 360 Degree, Height: 960, Width: 1280, Frame Rate: 20 fps.
However, we prefer 360 videos based fall activity analysis due to its inherent advantages in terms of very wide angle view thus providing complete living space view possible to be framed using the only single camera. The database was recorded in a controlled environment as shown in Figure 1.



(a) Camera installation (b) Periphery data collection (c) Centralized data collection

Fig. 1. Database preparation environment

The total number of video sequences thus generated is 2718 with 1327 sequences of AFC and 1391 of ADL. The recording of Activity of Fall category(AFC) and Activity of daily living(ADL) are shown in Figure 2.



(a) The activity of Fall category(AFC) (b) Activity of daily living(ADL)

Fig. 2. Human Actions representing Activity of Fall Category and Activity of Daily Living

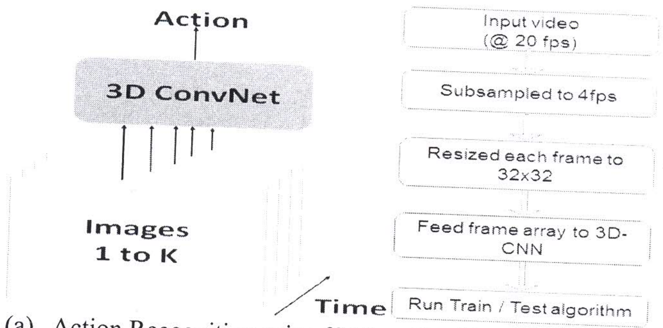
From the videos captured during database preparation, it has been found that on average a person takes around 3-4 seconds during the fall event. So it has been decided video clips should be of 5-second duration for uniformity across all instances. All composite videos of each subject have been segmented in frames and then as per the ADL and AFC instances regrouped in 5 sec durations.

3.2 Techniques Adopted

The design of our fall detection architecture was driven by objectives such as to minimize the hand-engineered image processing steps and to make the system generic and deployable on a low computational power machine. The first requirement was met by developing a system which works on human motion without correlating the image appearance such as fall action can be represented by a stack of frames in which each frame represents a pose or an instance in fall action. Another requirement of minimizing the hand-engineered image processing steps was met by using models which proves to be versatile in automatic feature extraction such as CNN [27] and LSTM [28].

3D Convolutional Neural Networks (3DCNN)

The convolution based neural networks have proved to be convenient tools to achieve generic features. The convolutional operation can be armed with temporal information such as 3D convolutional networks [29,30]. Since fall detection is a time based event so time management is crucial and a way to cope with time and motion has to be added to CNN's. To incorporate the motion information in video analysis, 3D convolution is performed in the convolutional layers of CNN so that discriminative features along both the spatial and temporal dimensions are captured as shown in Figure 3, where (a) shows the architectural details and (b) shows the process flow details.



(a) Action Recognition using 3DConvNet (b) Process Flow

Fig. 3. 3D-ConvNet

It can be strengthened by the fact that by applying multiple distinct convolutional operations at the same location on the input, multiple types of features can be extracted.The training methodology has to be adopted by keep in mind about the generality of the learned features for different fall scenarios. The details about the architecture of the proposed 3D CNN algorithm are mentioned in Figure 4. It consists of 4 convolution layers with characteristics as kernel size = (3,3,3), padding= same, activation = relu/softmax. The max pooling layer is having a pool size of (3,3,3) and dropout =0.25. The output of the fourth convolution layer is further flattened to 1D array. The final dense layer produces two class outputs: Fall or Not Fall.

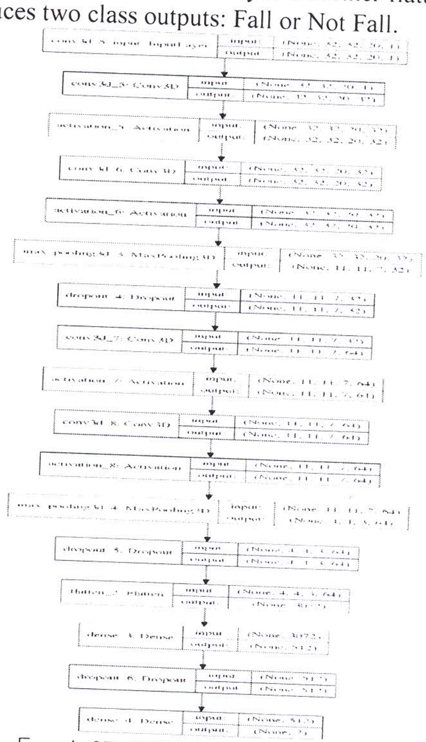
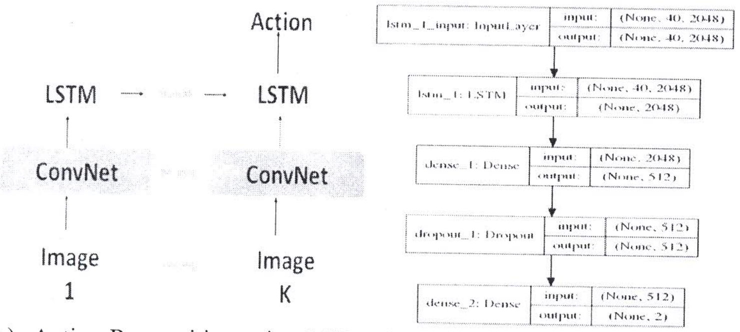


Fig. 4. 3D CNN details: Layer wise

The hyper parameters used in training are Learning rate: 10e-4, loss function = categorical cross entropy and train-test split with 60-20-20 proportion for training, validation and testing.

Long Short-Term Memory Networks (LSTM)

CNN has demonstrated a great ability to produce a rich representation of input and achieving promising results on object detection. Recurrent neural networks(RNNs) are a class of artificial neural networks, where the recurrent architecture allows the networks to demonstrate the dynamic temporal behavior. So, many researchers [31,32] have combined RNN with CNN for action recognition with promising results. In particular, LSTM [33] has been proved by Donahue et al [34], Shi et al. [35] and Ng et al. [36] to be useful to activity recognition specifically from videos. The proposed model consists of two parts convolution based (Inception V3) feature extraction and long short term memory(LSTM) based sequence modeling. CNN provides a rich representation of the input image by embedding it into a fixed feature vector. Since the video is a sequence of frames, a model with RNN is established and after extracting the features in videos, all features X_t are fed as input in LSTMs. Feature lengths of 40 and 20 sequences are experimented but superior results are obtained with sequences length of 20. The block level composition of LSTM is shown in Figure 5, where (a) shows the architectural details and (b) shows layer-wise composition and input output dimensions of the proposed LSTM model.



(b) Action Recognition using LSTM (b) Process Flow

Fig. 5. LSTM Model for Action Recognition

4 Results and Discussion

The deep learning based models which can process the frame sequence based video events such as 3DCNN and LSTM model has been proposed for fall activity recognition in this paper. In 3DCNN, the model constructs feature from both spatial and temporal dimensions by performing the 3D convolutions. The developed deep architecture constructs multiple channels of information from neighboring frames and performs convolution and subsampling separately in each channel. The final feature representation is obtained by combining information from all channels.

4.1 3DCNN performance on 360^0 videos.

Input parameter: Input video Size: 640 x 480, FPS: 30
X shape 1804 x 32x32x5x1
Y shape 1804 x 2
Random state = 46,44
Training parameter: Learning rate: 0.0001, Batch size: 128
The mean Accuracy obtained with RGB videos is 95.3 %.
The model accuracy and loss are shown in Figure 6.

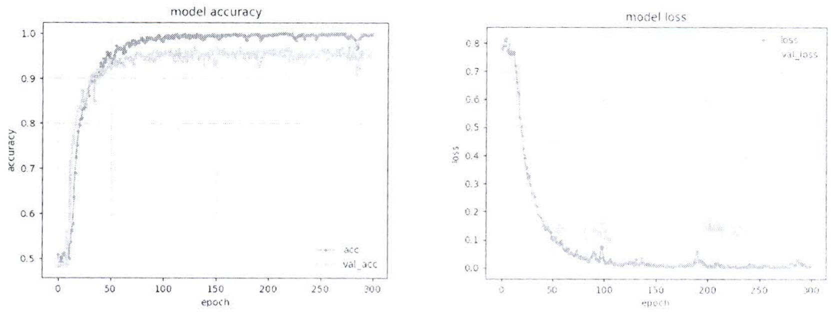


Fig. 6. 3DCNN model performance

The classification results of 3DCNN are shown in Figure 7 for ADL and AFC category.

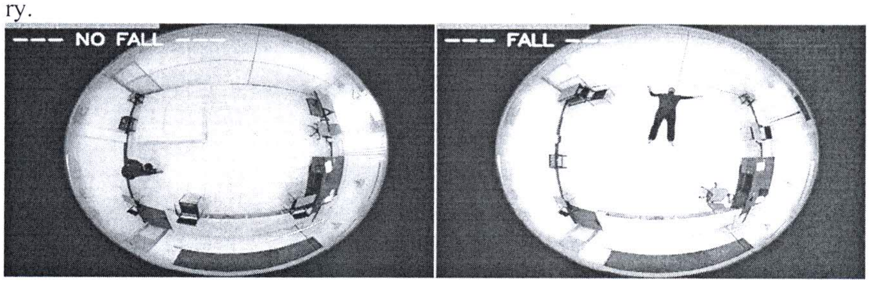


Fig. 7. Activity classification performance of 3DCNN

4.2 LSTM performance on 360^0 videos

Input parameter: Input video size: 640 x 480, FPS:30
Feature extraction of size 20 by using InsceptionV3
Input Shape 756x20x2048 for LSTM
Early stopper: Patience = 5
Training parameter: Learning rate = 1e-5, epoch = 60, Random state = 841.
The mean Accuracy obtained with RGB videos is 91.62 %.
The training model accuracy and loss are shown in Figure 8.

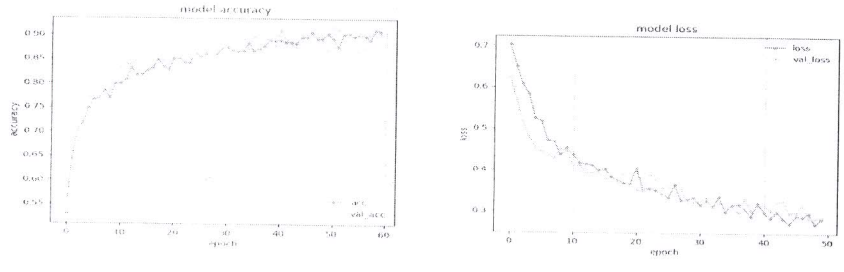


Fig. 8. LSTM model performance

The classification results of 3DCNN are shown in Figure 9 for ADL and AFC category.

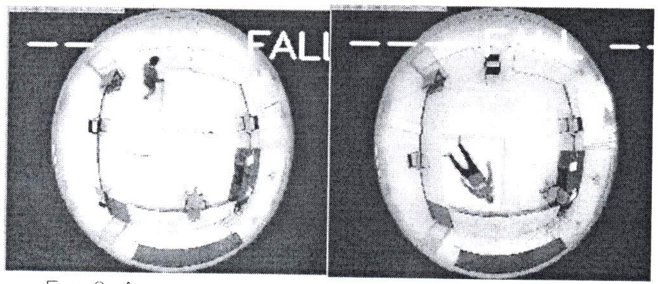


Fig. 9. Activity classification performance of LSTM

The comparative analysis between the two approaches for activity classification is presented in Table 1.

Table 1. Performance comparison between 3DCNN and LSTM.

Model	Depth	Learning Rate	Number of Epochs	Validation Accuracy
3DCNN	20	1e-4	300	95.3 %
LSTM	40,20	1e-5	60	91.62 %

5 Conclusions and Discussions

In this paper, we presented a successful application of convolutional networks and Long short-term memory networks to 360-degree videos for activity detection to create vision based fall detector system which obtained state-of-art results on benchmark database created in-house with 22 participants. To the best of our knowledge, this is the first effort to generate such omnidirectional video database of fall activity. We have used indoor scenario illustrating activities in real life of elderly people. The CNN are applicable to wide range of tasks and because of their relatively uniform architecture. ConvNets are ideal models for hardware implementation and embedded applications. The ability for real-time processing can be acceptable. The proposed

3DCNN and LSTM models have been tested with different scenarios of fall event such as front fall, backward fall, side fall, fall during standing up or sitting on chair, imbalance resulting in fall and daily activities events such as lying down, sitting on chair, picking up object and squatting to evaluate its performance. The experiment results showed that the 3DCNN model trained and validated on both ADL and AFC activity videos provide 95.3 % accuracy on test set. The LSTM model also suits well to fall event detection with 91.62 % accuracy for RGB omnidirectional video samples. Both models are not affected by the position of the person and can face any side, angle and can also walk around as opposed to the normal RGB camera where better accuracies are mainly claimed in situations when camera faces the person from front. In this paper, we have considered 3DCNN and RNN for activity recognition. Apart from this, there are other deep learning architectures, such as deep belief networks [37,38] with promising results claimed in object recognition domain. The proposed models were trained using supervised learning technique which requires a large number of labeled samples. Previous references in the literature [39] suggests that number of labeled samples can be significantly reduced if the model is pretrained using unsupervised algorithms. Future research will be targeted to explore the utilization of unsupervised training of 3DCNN and LSTM models.

Acknowledgments

The authors gratefully acknowledge the support of Director, CSIR-CEERI, Pilani for his guidance and making the resources available which make this research possible.

References

1. 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", IEEE International Conference on Information & Communication Technologies(ICITA), pp. 1003-1007, 2006.
2. M.A. Estudillo-Valderrama, L.M. Roa, J. Reina-Tosina, and D. Naranjo-Hernandez, "Design and implementation of a distributed fall detection system - personal server," IEEE Trans. Inf. Technol. Biomed., vol. 13, pp. 874–881, 2009.
3. J.M. Kang, T. Yoo, and H.C. Kim, "A wrist-worn integrated health monitoring instrument with a tele-reporting device for telemedicine and telecare," IEEE Trans. Instrum. Meas., vol. 55, pp. 1655–1661, 2006.
4. T. Lee and A. Mihailidis, "An intelligent emergency response system: Preliminary development and testing of automated fall detection," Journal of Telemedicine and Telecare, vol. 11, no. 4, pp. 194–198, 2005.
5. 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, 2011.
6. M. Mubashir, L. Shao, and L. Seed, "A survey on fall detection: principles and approaches," Neurocomputing, vol. 100, pp. 144–152, 2013.
7. M. Vallejo, C. V. Isaza, and J. D. Lopez, "Artificial Neural Networks as an alternative to traditional fall detection methods," in Proceedings of the 2013 35th Annual International

Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2013, pp. 1648–1651, Japan, July 2013.

8. A. Sengto and T. Leauhatong, "Human falling detection algorithm using back propagation neural network," in Proceedings of the 5th 2012 Biomedical Engineering International Conference, BMEiCON 2012, Thailand, December 2012.
9. B. Kwolek and M. Kepski, "Human fall detection on embedded platform using depth maps and wireless accelerometer," Computer Methods and Programs in Biomedicine, vol. 117, no. 3, pp. 489–501, 2014.
10. F. Harrou, N. Zerrouki, Y. Sun, and A. Houacine, "Statistical control chart and neural network classification for improving human fall detection," in Proceedings of the 8th International Conference on Modelling, Identification and Control, ICMIC 2016, pp. 1060–1064, Algeria, November 2016.
11. H. Foroughi, A. Rezvanian, and A. Paziraee, "Robust Fall Detection using Human Shape an Multi-Class Support Vector machine", IEEE 6th Indian conference on Computer Vision, Graphics & Image Processsing (ICVGIP), pp. 413-420, 2008.
12. H. Foroughi, H.S. Yazdi, H. Pourreza, and M. Javidi, "An Eigenspace-based Approach for Human Fall Detection Using Integrated Time Motion Image and Multi-class Support Vector Machine," IEEE 4th International Conference on Intelligent Computer Communication and Processing (ICCP), pp. 83-90, 2008
13. H. Foroughi, B.S. Aski, and H. Pourreza, "Intelligent Video Surveillance for Monitoring Fall Detection of Elderly in Home Environments," IEEE 11th International Conference on Computer and Information Technology (ICCIT), pp. 24-27, 2008
14. H. Foroughi, A. Naseri, A. Saberi, and H.S. Yazdi, "An Eigenspace-based Approach for Human Fall Detection Using Integrated Time Motion Image and Neural Network," IEEE 9th International Conference on Signal Processing (ICSP), pp. 1499-1503, 2008.
15. S.-G. Miaou, P.-H. Sung, and C.-Y. Huang, "A customized human fall detection system using omni-camera images and personal information," in Proceedings of the 1st Transdisciplinary Conference on Distributed Diagnosis and Home Healthcare, D2H2 2006, pp. 39–42, USA, April 2006.
16. C.-L. Liu, C.-H. Lee, and P.-M. Lin, "A fall detection system using k-nearest neighbor classifier," Expert Systems with Applications, vol. 37, no. 10, pp. 7174–7181, 2010.
17. S. Wang, L. Chen, Z. Zhou, X. Sun, and J. Dong, "Human fall detection in surveillance video based on PCANet," Multimedia Tools and Applications, vol. 75, no. 19, pp. 11603–11613, 2015.
18. V.Vishwakarma, C.Mandal, and S. Sural, in Pattern Recognition and Machine Intelligence, Automatic detection of human fall in video, Ed., pp. 616–623, 2007.
19. I. Charfi, J. Miteran, J. Dubois, M. Atri, and R. Tourki, "Definition and performance evaluation of a robust SVM based fall detection solution," in Proceedings of the 8th International Conference on Signal Image Technology and Internet Based Systems, SITIS 2012, pp. 218–224, Italy, November 2012.
20. N. Zerrouki, F.Harrou, A.Houacine, and Y. Sun, "Fall detection using supervised machine learning algorithms: A comparative study," in Proceedings of the 8th International Conference on Modelling, Identification and Control, ICMIC 2016, pp. 665–670, Algeria, November 2016.
21. 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, 2011.

22. 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.
23. G. Diraco, A. Leone, and P. Siciliano, "An active vision system for fall detection and posture recognition in elderly healthcare," in Proceedings of the Design, Automation and Test in Europe Conference and Exhibition, DATE 2010, pp. 1536–1541, deu, March 2010.
24. G. Mastorakis and D. Makris, "Fall detection system using Kinect's infrared sensor," Journal of Real-Time Image Processing, vol. 9, no. 4, pp. 635–646, 2012.
25. Y. Jia, E. Shelhamer, J. Donahue et al., "Caffe: convolutional architecture for fast feature embedding," in Proceedings of the ACM International Conference on Multimedia, pp. 675–678, ACM, Orlando, Fla, USA, November 2014.
26. S. S. Beauchemin and J. L. Barron, "The Computation of Optical Flow," ACM Computing Surveys, vol. 27, no. 3, pp.433–466, 1995.
27. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.
28. Zhenyang Li, Kirill Gavrilyuk, Efstratios Gavves, Mihir Jain, Cees G.M. Snoek,Video LSTM convolves, attends and flows for action recognition, Computer Vision and Image Understanding,Volume 166, pp. 41-50, 2018.
29. Ji, S.; Xu, W.; Yang, M.; Yu, K. 3D convolutional neural networks for human action recognition. IEEE Trans. Pattern Anal. Mach. Intell. 2013, 35, 221–231.
30. Karpathy, A.; Toderici, G.; Shetty, S.; Leung, T.; Sukthankar, R.; Fei, L.F. Large scale video classification with convolutional neural networks. In Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014; pp. 1725–1732
31. Srivastava, N.; Mansimov, E.; Salakhutdinov, R. Unsupervised learning of video representations using LSTM. arXiv 2015, arXiv:1502.04681.
32. Sharma, S.; Kiros, R.; Salakhutdinov, R. Action recognition using visual attention. arXiv 2015, arXiv:1511.04119.
33. Hochreiter, S.; Schmid Huber, J. Long short-term memory. Neural Comput. 1997, 9, 1735–1780.
34. Donahue, J.; Hendricks, L.A.; Guadarrama, S.; Rohrbach, M.; Venugopalan, S.; Saenko, K.; Dar-rell, T.Long-term recurrent convolutional networks for visual recognition and description. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 8–10 June 2015; pp. 2625–2634.
35. Shi, Y.; Tian, Y.; Wang, Y.; Huang, T. Sequential deep trajectory descriptor for action recognition with three-stream cnn. IEEE Trans. Multimed. 2017, 19, 1510–1520.
36. Ng, J.Y.; Hausknecht, M.; Vijayanarasimhan, S.; Vinyals, O.; Monga, R.; Toderici, G. Beyond short snippets:Deep networks for video classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 8–10 June 2015; pp. 4694–4702.
37. V. Jain, J.F. Murray, F. Roth, S. Turaga, V. Zhigulin, K.L. Briggman, M.N. Helmstaedter, W. Denk, H.S. Seung, "Supervised Learning of Image Restoration with Convolutional Networks", Proc. 11th IEEE Int'l Conf. Computer Vision, 2007.
38. V. Jain, S. Seung, "Natural Image Denoising with Convolutional Networks", Proc. Advances in Neural Information Processing Systems 21, pp. 769-776, 2009.
39. M. Ranzato, F. Huang, Y. Boureau, Y. LeCun, "Unsupervised Learning of Invariant Feature Hierarchies with Applications to Object Recognition", Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2007.