Fall Detection using Body Geometry in Video Sequences

Beddiar Djamila Romaissa[1],[2], Oussalah Mourad[2] , Nini Brahim[1] and Bounab Yazid[2]
[1]Research Laboratory on Computer Science's Complex Systems
University Laarbi Ben M'hidi, Oum El Bouaghi, Algeria
Email: ad_beddiar@esi.dz , Djamila.Beddiar@oulu.fi
[2]Center for Machine Vision and Signal Analysis,
University of Oulu, Finland

***Abstract*—According to the World Health Organization, falling of the elderly is a major health problem that causes many injuries and thousands of deaths every year. This increases pressure on health authorities to provide daily health care, reliable medical assistance, reduce fall damages and improve the elderly quality of life. For that, it is a priority to detect or predict falls accurately. In this paper, we present a fall detection approach based on human body geometry inferred from video sequence frames. We calculate the angular information between the vector formed by the head centroid of the identified facial image and the center hip of the body and the vector aligned with the horizontal axis of the center hip. Similarly, we calculate the distance between the vector formed by the head and the body center hip and the vector formed on its horizontal axis; we then construct distinctive image features. These angles and distances are then used to train a two-class SVM classifier and a Long Short-Term Memory network (LSTM) on the calculated angle sequences to classify falls and no-falls activities. We perform experiments on the Le2i fall detection dataset. The results demonstrate the effectiveness and efficiency of the developed approach.**

***Index Terms*—Fall Detection, Elderly assistance, SVM classification, Deep Learning, LSTM, Pretrained models.**

I. Introduction

Nowadays, the elderly population of over 65 years old has witnessed a steady increase, where a substantial proportion live on their own. Daily life tasks can become challenging for many of them and could be influenced by many factors, such as age-related biological changes, neuropsychological disorders, and environmental conditions. In addition to these factors, sudden loss of balance, stability, and dizziness during daily life movements are common reasons for abrupt fall that can cause damage [1]. Strictly speaking, falling is an abnormal human activity that occurs infrequently and unpredictably. It is defined by [2] as an event that results in a person coming to rest inadvertently on the ground, the floor, or any other lower level. It is acknowledged that fall is one of the major public health problems in the world that should be carefully addressed and appears to be the second leading cause of accidental or unintentional injury deaths [2]. Therefore, Fall detection and Fall prediction are recognized as important research directions in the study of falls and are among the hottest topics in healthcare policies. Indeed, the availability of efficient methods to identify and possibly predict fall occurrence can have a huge public impact since it may significantly minimize damages, enable efficient medical assistance, and provide daily health care for vulnerable population. Moreover, falling has an obvious effect on individual autonomy, independence, and life quality. [3] attests that experiencing fall may lead to Basophobia, also called fear of falling. This syndrome can cause many other disorders such as lack of mobility, loss of the ability to live independently, and social isolation. On the other hand, reducing the interval between falling and rescuing is essential to minimize falls' negative consequences. Motivated by the importance of fall detection and the observation that the vector formed by the head and the center hip of the body is aligned horizontally and in parallel to the ground during a falling posture, while it is perpendicular to the ground axis in a sitting or standing posture, as illustrated in Fig. 1 (a), Fig. 1 (b), we present a novel machine learning like approach for Fall Detection. Besides, sitting slumped to one side leads to forming an angle of around 45° or 120° between the mentioned vector and the horizontal axis, as shown in Fig. 1 (d) and Fig. 1 (c). The angle value depends on the degree of slump sitting. However, the posture is considered lying or falling when this value is close to 0° or 180°, as shown in Fig. 1 (e). Our approach relies on calculating the angle and the distance between the vector formed by the head and the body center hip and the vector formed on its horizontal axis. For each video sequence, we calculate the angle mentioned above and distance among all the frames. The computed angles and distances form the new feature set that characterizes the video sequences. Furthermore, we construct new images using these angle and distance sequences so that each video sequence is represented with one image of its corresponding angles and distances. Then, an LSTM network trained on our features and a two-class SVM is trained on these images to detect fall and no-fall activities. We use the Le2i dataset [4] to evaluate the performance of our method. The experimental results indicate that our approach is practical and achieves good accuracy in detecting falls. The rest of this paper is organized as follows. We briefly provide previous research related to vision-based Fall detection (FD) in Section II. Section III outlines our approach. Then, we describe and discuss in Section IV the

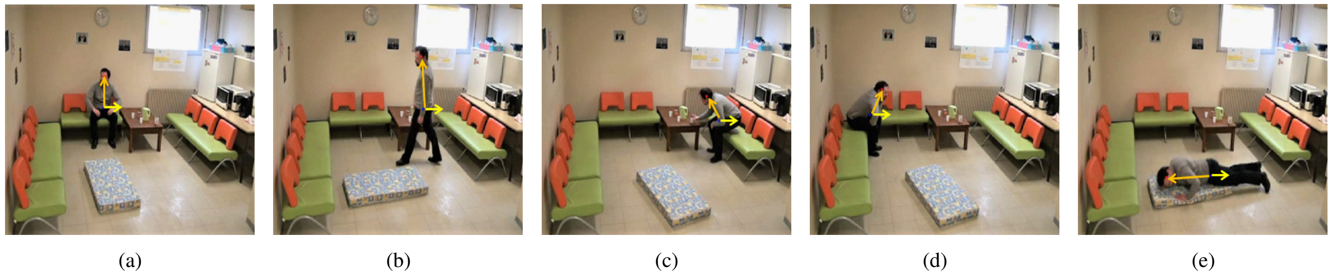
(a) (b) (c) (d) (e)

Fig. 1. Samples from the Le2i fall detection dataset representing the angle α in (a) sitting, (b) standing, (c) bending to the left, (d) bending to the right and (e) falling postures. The value of α is around 90°, 90°, 120°, 45° and 180° respectively.

experimental results of our proposal on the Le2i publicly available dataset. Finally, we conclude our paper and set future directions for fall detection in Section V.

II. Related Work

Fall detection techniques can be categorized into three major classes: ambient-based, wearable-based, and vision-based systems [1]. Ambient-based systems use light, proximity, motion, and vibration sensors to collect daily life activities data and detect falls. Wearable-based systems rely on the sensors embedded in particular devices that the subject should wear to track his/her motion [3]. Additionally, vision-based systems use RGB or depth cameras to record the subject's activities, in indoor or outdoor environments [3]. The recorded images or videos are analyzed later to detect falls. Motivated by robustness, efficiency, ease of use and installation of the last methods, the approach that we present in this paper relates to vision-based FD. Thereby, we briefly report here some of the existing vision-based FD methods. Roughly speaking, Vision-based FD approaches focus on identifying appropriate fall-related features extracted from the video frames such as silhouettes, body shape, and skeleton information. These features are then used as input to some machine learning classifiers such as SVM, KNN, Hidden Markov Models (HMM), among others, to train and later automatically detect fall and non-fall cases [5]. For instance, [6] extracts distinctive features of human silhouettes to construct new action representations. The authors model the actions using a bag-of-words and conduct the classification using an extreme learning machine (ELM). Authors in [7] suggest robust features called History Triple Features using a generalization of the Radon Transform. Furthermore, SVM based approaches have proven their efficiency for fall detection tasks in many alternative works see, for instance, [8]–[10]. In [8], five distinct features are employed (aspect ratio, change in aspect ratio, fall angle, center speed, and head speed). Authors in [9] use a normalized motion energy image to model the silhouette shape deformation features. Likewise, shape and motion features are tracked to detect falls using a single camera-based system in [11]. Authors in [12] suggest a vision-based fall detection system for elderly living alone. The system relies on the optical flow estimation to evaluate the speed of motion and to deduce the fall activity accordingly, while comparing the last positions of the target. With the advance in Deep Learning (DL) approaches, many researchers put forward DL based approaches for fall detection tasks. For instance, [13] proposes a real-time fall detection approach that allows the capture of RGB video streams, an individual's position estimation, and, thereby, fall detection likelihood, which then generates potential alert messages to caregivers with registered audio and video. In [14], the authors present a novel FD method based on Convolutional Neural Networks (CNN) using optical flow images. Moreover, transfer learning is widely used to take advantage of pre-trained models by reusing their network weights or fine-tuning the classification layers. For instance, [15] was able to detect falls using a CNN Alexnet architecture efficiently. In [16], the authors present a two-stream approach based on MobileVGG network. Similarly, the authors of [16] combine an improved lightweight VGG network and the motion characteristics of the human body. Likewise, a 3D CNN-based method combined with Long Short-Term Memory (LSTM) is also presented in [17]. The 3D CNN is used to extract motion and spatial features, while the LSTM-based spatial visual attention scheme is incorporated to locate the fall in each frame. Authors in [18] present a fall detection system based on LSTM, using location features from the group of available joints in the human body.

III. Proposed method

The starting point in our developed methodology consists in identifying relevant features that can genuinely distinguish fall from non-fall activities. In this respect, to illustrate our approach mathematically, we refer to the head centroid by the point $H(x_h,y_h)$ and to the center hip of the body by the point $B(x_b,y_b)$. It is also referred to the vector formed by H and B with $\overrightarrow{U}$, and to the vector formed between the point B and the point $C(x_c,y_c)$ with $\overrightarrow{V}$. The point C is defined such that $x_c > x_b$ and $y_c = y_b$ and the horizontal axis is defined as a straight line parallel to the *X_axis* and passing through the center hip. These notations are used along the paper.

Relying on our observation mentioned above, we calculate for each video, the angle α formed between $\overrightarrow{U}$ and $\overrightarrow{V}$ and the distance γ between the head and the center hip of the body (i.e., the magnitude of the vector $\overrightarrow{U}$) for all its frames. Each video is therefore characterized by a feature vector containing the sequence of the computed angles and distances. Fig. 2 outlines the pipeline of our proposed fall detection approach. It is summarized in four steps as follows.

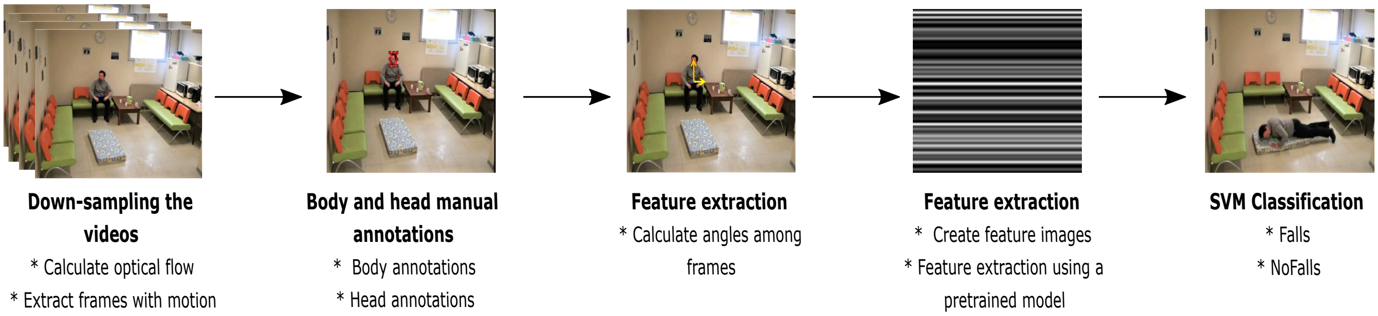

Fig. 2. The pipeline of our proposed fall detection approach

A. Down-sampling the videos

The length of the Le2i's video sequence dataset varies from 30 seconds to 4 minutes, with a frame rate of 25 frames per second. Indeed, the extraction of video frames can result in more than 1000 frames, making manual head annotations very exhausting. Therefore, instead of using all the frames, we down-sample each video to reduce the intensive computational processing. We observe that the fall, in general, occurs fast and can be characterized by a significant motion change among frames. Hence, we keep frames that represent meaningful motion change and construct re-sampled videos using these frames. We use the optical flow (OF) to estimate the motion in the video sequences using the Horn-Schunck method, which consists of resolving the constraint: $I_x.u{+}I_y.v{+}I_t = 0$. Where I_x, I_y, I_t are the Spatio-temporal image brightness derivatives, while *u* and *v* correspond to the horizontal and the vertical optical flow components respectively.

Then, we calculate the mean of both horizontal *(Vx)* and vertical *(Vy)* components of the OF, which we call *meanVx* and *meanVy* respectively. Subsequently, the mean squared normalized error performance (MSE) is computed to estimate the similarity between the horizontal\vertical components of the OF of each frame and the mean value of horizontal and vertical components of the OF respectively (*similarityVx* and *similarityVy*). Equation ((1)) demonstrates how to calculate the similarities above using the MSE. *z* refers to either *x* or *y* component. *P* corresponds to the pixels of the frame, while *i* refers to its index and *p* to a particular pixel of the frame *i*.

$$similarityVz_i = \frac{1}{P}.\sum_{p-1}^{P}(Vz_i(p) - meanVz(p)) \qquad (1)$$

Frames that have a similarity *similarityVx*$_i$ (resp. *similarityVy*$_i$) above or equal their mean similarity *meanSimVx* (resp. *meanSimVy*) are preserved while others are removed to construct the re-sampled video. Besides, the maintained frames should respect the conditions given by ((2)).

$$\begin{Bmatrix} similarityVx_i >= meanSimVx \\ similarityVy_i >= meanSimVy \end{Bmatrix} \qquad (2)$$

B. Body and head manual annotations

Once the videos are re-sampled, and since our work focuses on the features extracted from the body geometry, we manually annotate the individual's head position in the video frame and calculate its centroid. More specifically, the annotation of each frame contains the frame's index, the localization of the head presented in terms of the bounding box, and the coordinates of the head centroid. The head centroid and the center hip of the body are used later to calculate their associated distance γ and the angle α between the vector $\overrightarrow{U}$ and the vector formed by the horizontal axis corresponding to the *x* coordinate of the center hip called $\overrightarrow{V}$.

For the angle calculus, we can calculate its cosine value and deduce its corresponding value. The cosine is computed using the law of cosines (given in ((3))), and the Euclidean norm is used to calculate the magnitude of vectors. $\overrightarrow{HC}$ refers to the vector between the head centroid and the axis point *C* and $\left\|\overrightarrow{X}\right\|$ is the Euclidean norm of the vector $\overrightarrow{X}$.

$$cos(\alpha) = \frac{-\left\|\overrightarrow{HC}\right\|^2 + \left\|\overrightarrow{U}\right\|^2 + \left\|\overrightarrow{V}\right\|^2}{2.\left\|\overrightarrow{U}\right\|.\left\|\overrightarrow{V}\right\|} \qquad (3)$$

We, therefore, calculate the distance between the head and the center hip of the body among all the video frames using the Euclidean norm.

C. Feature extraction

Once these sequences of angles and distances are created for each video, we discern two scenarios. In the first one, we construct our feature vectors using angles and distances. These vectors are then fed to the classifier. In the second scenario, we use only the angles calculated above to construct the first set of images. However, we concatenate these angles and distances to construct the second set of images. Each video is characterized either by the feature vector *V* = $\{\alpha_1, \alpha_2, \alpha_3 ... \alpha_i\}$ or *V* = $\{$*[1,* α_1,γ_1*], [2,*α_2,γ_2*], [3,*α_3,γ_3*] ... [i,*α_i,γ_i*]*$\}$ where *i* is the index of the video frame. The angle sequences are used farther to construct gray-level images (1^{st} set of images) where the angle values constitute the grey level of the image. Similarly, the angle and the distance sequences are concatenated to construct RGB images (the 2^{nd} set of images) where we use the frame's index as the first channel, the angle as the second channel and the distance as the third one. Each image of both sets characterize one video. We give examples of created gray-level and RGB images of falls in Fig. 3 (a), Fig. 3 (b) and no-fall activities in Fig. 3 (c) and Fig. 3 (d).

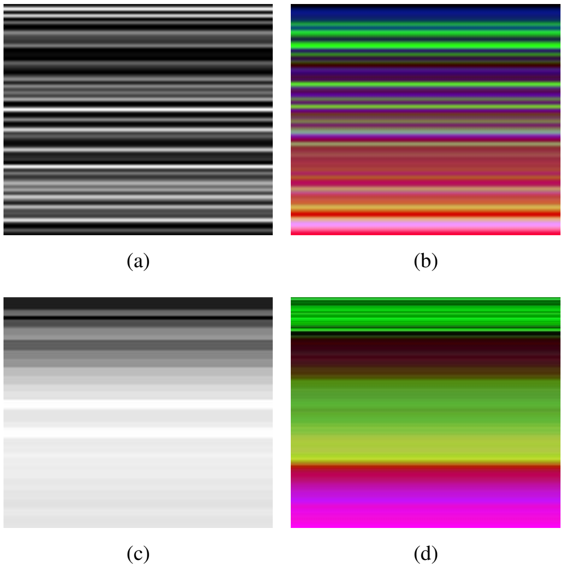
(a) (b) (c) (d)

Fig. 3. Samples of created images from only angles (a and c) and created images from angles and distances (b and d). (a) and (b) represent falls while (c) and (d) represent no-falls.

D. Classification

Feature vectors have different lengths because videos have a different number of frames. To set all these vectors to the same length, we apply a padding task at the beginning of each vector using the first frame features. A bi-LSTM short-term memory (LSTM) network is then trained using sequences of angles and distances to detect falls and no-falls in the first scenario. To detect falls in the second scenario, we extract distinctive features from our two sets of feature images using a pre-trained model. The first set consists of images constructed from angles only, while the second set includes images created using angle and distance sequences. In our approach, we use the activation of the Resnet50 network as our features. Then, we feed them to a two-class SVM classifier to distinguish between falls and daily life activities.

IV. Experimental results and discussion

A. Dataset

The Le2i fall detection dataset [4] contains 221 videos of 131 falls, and 90 daily life activities (ADL) recorded using a single fixed camera with a frame rate of 25 frames/s and a resolution of 320x240 pixels. Several actors simulate all the activities gathered at four locations: Home, Office, Coffee room, and Lecture room. The manual annotations of 191 videos were given, with extra information representing the ground-truth of the fall position and the localization of the body in the image sequence.

TABLE I
Performance results for our FD approach on the Le2i subset using a Resnet50 models for feature extraction

Features	**Accuracy**	**Precision**	**Recall**
Grey-level images + Resnet50 + SVM	0.80	0.84	0.90
RGB images + Resnet50 + SVM	0.85	**0.90**	0.90
Angle + LSTM	0.77	0.77	**1,00**
Angle + Distance + LSTM	**0.85**	0.89	0.89

B. Experiment results

We use the first protocol of evaluation P1 given in [4] to evaluate the performance of our approach. P1 consists of building the training and the test subsets from the locations ”Coffee room” and ”Home.” Hence, the subset of the Le2i dataset, defined by P1, consists of 130 videos. It contains 99 falls and 31 no-falls activities. We apply k-fold cross-validation to our LSTM and SVM models with k=10, where the Le2i subset was randomly split into k equal size subsets. At each iteration of the nine iterations, we compose the training and test sets with nine subsets and one set.

We first evaluate the results obtained from our LSTM model that is trained on two configurations of features. The first feature set is composed of angles only while the second set is composed of angles and distances. Fall detection is a binary classification problem in which the classifier should specify the existence or absence of a fall in the video. Sensitivity and specificity are most accurate to evaluate the performance of the system. We achieve a sensitivity of 100% for the first set and 89% for the second set of features. Secondly, we evaluate the results obtained from images constructed using (1) angles and (2) angles and distances, which are then trained on Resnet50 model.

Different performance metrics are, therefore calculated: Accuracy, Precision, and Recall. Table I illustrates the results obtained for both sets of images (constructed from angles only features versus angles + distances features) using the activations of the Resnet50 model as well as the results of LSTM training on the feature set. We can see from this table that the results obtained from the images constructed from the angles and the distances (RGB images) are more accurate than the results obtained from images created from angles only when using Resnet50. Similarly, the results obtained from LSTM trained on angle and distance features are more accurate than the results obtained from angle features only. Table II compares our findings to the state-of-the-art methods on the Le2i dataset. Clearly, this indicates that our results are comparable to state-of-the-art results in the field. Although, we acknowledge the lack of large scale datasets and competition in the area that would enable wide-scale state-of-art comparison of methods and foster the development of new technology for fall detection from video sequences.

V. Conclusion and Future Directions

We present in this paper an effective vision-based approach for fall detection based on angles calculation. Our approach allows us to construct gray-scale images of calculated angles

TABLE II
COMPARISON BETWEEN PERFORMANCE RESULTS OF OUR FD APPROACH WITH OTHER EXISTING APPROACHES ON THE LE2I SUBSET

Approaches	**Accuracy**	**Precision**	**Recall**	**F_score**
Gradient boosting classifier [19]	79.31%	79.41%	83.47%	0.81
GMM + PCA [20]	**86.21%**	89.13%	93.00%	0.91
OF + von Mises distribution [21]	69.23%	69.84%	**94.56%**	0.79
OF + CNN [14]	97.00%	-	99.00%	-
ours: **Angle + Distance + Resnet50 + SVM**	84.60%	90.00%	90.00%	**0.90**
ours: **Angle + Distance + LSTM**	84.60%	89.00%	89.00%	0.89

between the head, the center hip of the target subjects, and the horizontal axis passing through it. Another set of images is constructed using angles and distances between the head and the center hip of the body as well. These constructed sets of images constitute our distinctive features for the fall detection task. Next, an SVM classifier is used along with a pre-trained LSTM model to classify the created images into falls and daily life activities. We compare in this paper, the features extracted using the Resnet50 model. We use the Le2i dataset to evaluate our approach's performance in terms of accuracy, precision, and recall metrics. Experimental results show that the results of our proposed approach are comparable to the state-of-the-art fall detection methods but still need improvements to distinguish between lying and falling postures. In the future, we would like to automatically annotate the head and the body of the subjects and enhance the images we construct to get higher accuracy.

ACKNOWLEDGMENT

This work is partly supported by the Algerian Residential Training Program Abroad Outstanding National Program (PNE) that supported the first author stay at University of Oulu and European YougRes project (Ref. 823701), which are gratefully acknowledged.

REFERENCES

[1] A. Ramachandran and A. Karuppiah, "A survey on recent advances in wearable fall detection systems," *BioMed Research International*, vol. 2020, 2020.

[2] "World health organization, who global report on falls prevention in older age," Tech. Rep., 2007.

[3] S. S. Khan and J. Hoey, "Review of fall detection techniques: A data availability perspective," *Medical engineering & physics*, vol. 39, pp. 12–22, 2017.

[4] I. Charfi, J. Miteran, J. Dubois, M. Atri, and R. Tourki, "Optimized spatio-temporal descriptors for real-time fall detection: comparison of support vector machine and adaboost-based classification," *Journal of Electronic Imaging*, vol. 22, no. 4, p. 041106, 2013.

[5] D. R. Beddiar, B. Nini, M. Sabokrou, and A. Hadid, "Vision-based human activity recognition: a survey," *Multimedia Tools and Applications*, pp. 1–47, 2020.

[6] X. Ma, H. Wang, B. Xue, M. Zhou, B. Ji, and Y. Li, "Depth-based human fall detection via shape features and improved extreme learning machine," *IEEE journal of biomedical and health informatics*, vol. 18, no. 6, pp. 1915–1922, 2014.

[7] G. Goudelis, G. Tsatiris, K. Karpouzis, and S. Kollias, "Fall detection using history triple features," in *Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, 2015, pp. 1–7.

[8] G. Debard, M. Mertens, M. Deschodt, E. Vlaeyen, E. Devriendt, E. Dejaeger, K. Milisen, J. Tournoy, T. Croonenborghs, T. Goedemé *et al.*, "Camera-based fall detection using real-world versus simulated data: How far are we from the solution?" *Journal of Ambient Intelligence and Smart Environments*, vol. 8, no. 2, pp. 149–168, 2016.

[9] W. Feng, R. Liu, and M. Zhu, "Fall detection for elderly person care in a vision-based home surveillance environment using a monocular camera," *signal, image and video processing*, vol. 8, no. 6, pp. 1129–1138, 2014.

[10] A. Iscen, A. Armagan, and P. Duygulu, "What is usual in unusual videos? trajectory snippet histograms for discovering unusualness," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2014, pp. 794–799.

[11] V. A. Nguyen, T. H. Le, and T. T. Nguyen, "Single camera based fall detection using motion and human shape features," in *Proceedings of the Seventh Symposium on Information and Communication Technology*, 2016, pp. 339–344.

[12] S. Bhandari, N. Babar, P. Gupta, N. Shah, and S. Pujari, "A novel approach for fall detection in home environment," in *2017 IEEE 6th Global Conference on Consumer Electronics (GCCE)*. IEEE, 2017, pp. 1–5.

[13] L. Ciabattoni, G. Foresi, A. Monteriù, D. P. Pagnotta, and L. Tomaiuolo, "Fall detection system by using ambient intelligence and mobile robots," in *2018 Zooming Innovation in Consumer Technologies Conference (ZINC)*. IEEE, 2018, pp. 130–131.

[14] A. Nunez-Marcos, G. Azkune, and I. Arganda-Carreras, "Vision-based fall detection with convolutional neural networks," *Wireless communications and mobile computing*, vol. 2017, 2017.

[15] L. Anishchenko, "Machine learning in video surveillance for fall detection," in *2018 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT)*. IEEE, 2018, pp. 99–102.

[16] Q. Han, H. Zhao, W. Min, H. Cui, X. Zhou, K. Zuo, and R. Liu, "A two-stream approach to fall detection with mobilevgg," *IEEE Access*, vol. 8, pp. 17 556–17 566, 2020.

[17] N. Lu, Y. Wu, L. Feng, and J. Song, "Deep learning for fall detection: Three-dimensional cnn combined with lstm on video kinematic data," *IEEE journal of biomedical and health informatics*, vol. 23, no. 1, pp. 314–323, 2018.

[18] K. Adhikari, H. Bouchachia, and H. Nait-Charif, "Long short-term memory networks based fall detection using unified pose estimation," in *Twelfth International Conference on Machine Vision (ICMV 2019)*, vol. 11433. International Society for Optics and Photonics, 2020, p. 114330H.

[19] M. Chamle, K. Gunale, and K. Warhade, "Automated unusual event detection in video surveillance," in *2016 International Conference on Inventive Computation Technologies (ICICT)*, vol. 2. IEEE, 2016, pp. 1–4.

[20] A. Poonsri and W. Chiracharit, "Fall detection using gaussian mixture model and principle component analysis," in *2017 9th International Conference on Information Technology and Electrical Engineering (ICITEE)*. IEEE, 2017, pp. 1–4.

[21] A. Y. Alaoui, A. El Hassouny, R. O. H. Thami, and H. Tairi, "Video based human fall detection using von mises distribution of motion vectors," in *2017 Intelligent Systems and Computer Vision (ISCV)*. IEEE, 2017, pp. 1–5.