



Saliency based human fall detection in smart home environments using posture recognition

Health Informatics Journal

1–14

© The Author(s) 2021

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/14604582211030954

journals.sagepub.com/home/jhi**Mikael Ange Mousse** 

Institut Universitaire de Technologie, Université de Parakou, Benin

Béthel Atohoun

Ecole Supérieure de Gestion d'Informatique et des Sciences, Benin

Abstract

The implementation of people monitoring system is an evolving research theme. This paper introduces an elderly monitoring system that recognizes human posture from overlapping cameras for people fall detection in a smart home environment. In these environments, the zone of movement is limited. Our approach used this characteristic to recognize human posture fastly by proposing a region-wise modelling approach. It classifies persons pose in four groups: standing, crouching, sitting and lying on the floor. These postures are obtained by calculating an estimation of the human bounding volume. This volume is estimated by obtaining the height of the person and its surface that is in contact with the ground according to the foreground information of each camera. Using them, we distinguish each postures and differentiate lying on floor posture, which can be considered as the falling posture from other postures. The global multiview information of the scene is obtaining by using homographic projection. We test our proposed algorithm on multiple cameras based fall detection public dataset and the results prove the efficiency of our method.

Keywords

health-care, fall detection, posture recognition, medical imaging

Introduction

The importance of health-care systems is growing up. These systems are used to ensure the safety of elderly people. Indeed one of criticals problems for senior living at home is Fall. This problem causes severe injuries and some of these injuries are fatal for the people. Then automatically people falls detection is the desired part of caring for a live-alone senior. Its role is to detect quickly the

Corresponding author:

Mikael Ange Mousse, Université de Parakou, Parakou, 123, Borgou, Benin.

Email: mikael.mousse@univ-parakou.bj

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

people in a fallen position. This can reduce additional complications and allow a rapid patient management. If the fall did not result in injury, automatic detection can alert caregivers of the need for preemptive measures such as hazard elimination, physical conditioning, etc. It has been proved that the medical consequences of a fall are highly contingent upon the response and rescue time. Then, a highly-accurate auto-matic fall detection system is likely to be a significant part of the living environment for the elderly to expedite and improve the medical care provided whilst allowing people to retain autonomy for longer. A fall is characterized by a person beginning with normal behaviours followed by the person rapidly descending then laying on the ground for an extended period of time.

This paper which presents a system for fall detection consists of five sections. The related works are presented in section II. In section III, we presented the proposed algorithm for people falls detection from multiple cameras. In Section IV, we presented the experimental environment and we evaluated the performance of the system. Finally, in Section V, we ended this work with further directions.

Related works

Three categories of people fall detection systems are proposed in the state-of-art. The first category is based on wearable device, the second category used some ambience sensors and the third category is camera (vision) based system. Ren and Peng³² and Wang et al.⁷ proposed more complete reviews of fall detection system. In this paper we propose a vision based system. We prefer a computer vision system because it doesn't require that the person wears anything. Another advantage of such a system is that a camera gives more information on the motion of a person and his/her actions than an accelerometer. Thus, we can imagine a computer vision system providing information on falls, but also, checking some daily behaviours (medication intake, meal and sleep time and duration, etc.). According to this, we focus our related work on camera based method. Mubashir et al.² provide a more complete review of fall detection systems using wearable device or ambience sensor.

Concerning video based fall detection system, some of systems are based on a single camera view.^{3,6,8-10} This approach is used due to the wide availability of a single surveillance cameras and to the relative ease of implementation. Even with strong prior assumptions and no computational limitations, often it is difficult to detect efficiently falls of people from a single simple camera view. To overcome this problem, several researchers propose to use the cameras which look straight down from ceiling greatly.¹³⁻¹⁸ Some research works proposed to use multi-camera system for people fall detection. In computer vision community, the use of multi-camera takes a lot of scope. Indeed, motivations are multiple and concern various fields as monitoring and surveillance of significant protected sites, control and estimation of flows (car parks, airports, ports and motorways). Because of the fast growing of data processing, communications and instrumentation, such applications become possible. These kind of systems require more cameras to cover overall field-of-view. They reduce the effects of objects dynamic occlusion. Using several views of the same scene (multi-view) can allow to recover the information that could have been hidden in a specific view. In this category, Auvinet et al.¹⁹⁻²¹ proposed a network of multiple calibrated cameras to reconstruct the 3D shape of the person. This shape is analysed by calculate the volume distribution along the vertical axis, and an fall detection alarm is triggered when the major part of this distribution is abnormally near the floor. Anderson et al.^{22,24} used hierarchy of fuzzy logic to detect falls after the reconstruction of the 3D shape of the person. This hierarchic method consists of two levels. The first level deduces the number of states for the object at each image. The second level deals with linguistic summaries of the objects states called 'Voxel Person'. These methods provide a good

results, however, 3D reconstruction process requires more processing time. To address this drawback, some researchers have proposed methods that are not based on 3D reconstruction. Thome et al.²⁵ propose a multi-view fall detection system by which motion is modelled by a layered hidden Markov model (LHMM) with two layers for modelling motion. The first layer has two states, an upright standing pose and lying. Fall detection, in terms of sudden change, has dedicated motion features from the first layer. Three dimensional angle relationships and their image plane projections have been carefully observed. After performing an initial image metric rectification, theoretical properties are derived from binding the error angle for a standing posture during the image formation process. This simply differentiates other poses as non-standing ones. Thus falls can be accurately detected from other actions, such as walking or sitting. The proposed method uses a multi-view setting, where the low-level steps are (mainly) performed independently in each view, leading to the extraction of simple image features compatible with real-time achievement. Then, a fusion unit merges the output of each camera to provide a viewpoint-independent pose classification. Cucchiara et al.²⁶ use multiple cameras to monitor different rooms. A single room is monitored by a single camera. Multiple cameras are used to cover different rooms and the camera hand-off is treated by warping the person's appearance in the new view by means of homography. They carried out analysis of human behaviours by classifying the posture of the monitored person and consequently detecting falls. Projection histograms are calculated and compared with the stored posture maps (training). The tracking also deals with occlusions. Zweng et al.²⁷ detect falls using multiple cameras. Each of the camera inputs results in a separate fall confidence (so no external camera calibration is needed). These confidences are then combined into an overall decision. Asif et al.,¹¹ skeleton data and segmentation data of the human are extracted by a proposed human pose estimation and segmentation module with the weights pre-trained on a MS COCO Key points dataset, which is then fed into a CNN model with modality-specific layers which is trained on synthetic skeleton data and segmentation data generated in a virtual environment. Yao et al.¹² proposed a novel real-time indoor fall detection method based on computer vision by using geometric features and convolutional neural network (CNN). They applied Gaussian mixture model (GMM) to detect the human target and find out the minimum external elliptical contour. Differently from the traditional fall detection method based on geometric features, they consider the importance of the head in fall detection and propose to use two different ellipses to represent the head and the torso, respectively. Three features including the long and short axis ratio, the orientation angle and the vertical velocity are extracted from the two different ellipses in each frame, respectively, and fused into a motion feature based on time series. In addition, a shallow CNN is applied to find out the correlation between the two elliptical contour features for detecting indoor falls. Iazi et al.⁴ proposed a fall detection system for elderly people based on their postures. They suggested to recognize the postures from the human silhouette which is an advantage to preserve the privacy of the elderly. The effectiveness of their approach is demonstrated on two well-known datasets for human posture classification and three public datasets for fall detection, using a Support-Vector Machine (SVM) classifier. The experimental results show that their method can not only achieves a high fall detection rate but also a low false detection. Casilari et al.¹ proposed a convolutional deep neural network when it is applied to identify fall patterns based on the measurements collected by a transportable tri-axial accelerometer. In contrast with most works in the related literature, the evaluation is performed against a wide set of public data repositories containing the traces obtained from diverse groups of volunteers during the execution of ADLs and mimicked falls. Wang et al.⁵ proposed a visual-based fall detection approach by Dual-Channel Feature Integration. Their approach divides the fall event into two parts: falling-state and fallen-state, which describes the fall events from dynamic and static perspectives. Firstly, they used the object detection model (Yolo) and the human posture detection model (OpenPose) to obtain key points and the position information of a

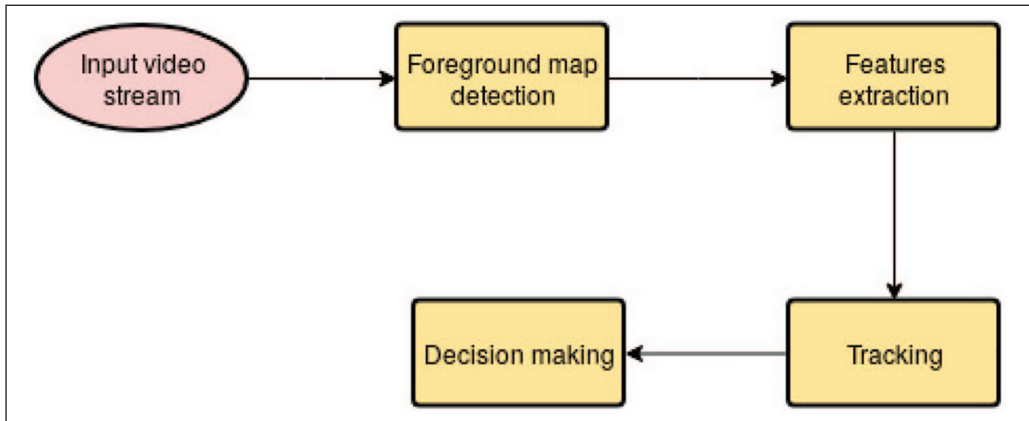


Figure 1. Architecture of the proposed system

human body. Then, a dual-channel sliding window model is designed to extract the dynamic features of the human body (centroid speed, upper limb velocity) and static features (human external ellipse). After that, they used MLP (Multilayer Perceptron) and Random Forest to classify the dynamic and static feature data separately. Finally, the classification results are combined to detect fall. Hung et al.^{28,29} propose using the measures of humans heights and occupied areas to distinguish three typical states of humans: standing, sitting and lying. Two relatively orthogonal views are utilized, in turn, simplifying the estimation of occupied areas as the product of widths of the same person, observed in two cameras. In this second subgroup of works, the fall detection decision is made by merging the decisions coming from each cameras. The final decision is strongly correlated to the decisions of each cameras and any occlusion in a view of one camera is detrimental to this final decision. To overcome this drawback, Mousse et al.²³ proposed a high-level fusion information for fall detection which is not based on 3D reconstruction.

The reduction of the computational burden of a fall detection system remains a topical challenge. This work is part of this vision. We propose an system which exploits a visual perception information to detect fall. Our system use visual information to focus the processing on the parts of the image that are near to the possible location of foreground objects. In fact, Saliency detection is a powerful tool for solving this tricky problem, due to its ability of locating the interesting content in images or videos. By ignoring the nonsalient regions of videos, the contents to be processed decrease drastically.

Proposed system for fall detection

The architecture of the proposed system is presented by Figure 1. This architecture is composed by four modules.

- **foreground map detection:** this module extracts the foreground maps of cameras;
- **features extraction:** this module proposes some metrics and rules to characterize the posture of the person.
- **tracking:** this module supports people fall detection, by keeping track of people movements and of their identities along time. The module help our system to be efficient when several persons are present on the scene
- **decision making:** this module identifies the posture of the person.

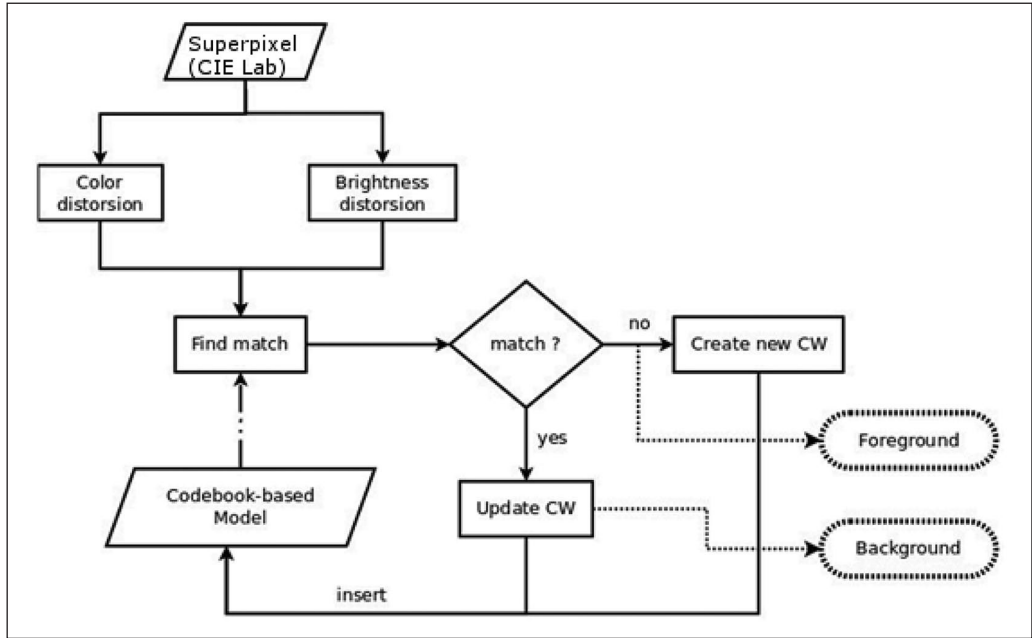


Figure 2. Flow diagram of the Codebook based algorithm (the solid lines correspond to the learning phase; dashed lines correspond to the phase of moving objects detection).

Foreground map detection

Like as other video based system, the first step of our fall detection method is the extraction of the foreground map of each camera. This extraction is done in two steps. The first step is the extraction of the foreground pixels and the second is the grouping of these pixels by using polygons. The purpose of this grouping is the reduction of information which will be processed. For the foreground pixels detection we use a background modelling based algorithm. In fact in single camera system, many algorithms about object detection exist with different purposes. These algorithms are subdivide in three categories: without background modelling, with background modelling and combined approach. Algorithms based on background modelling are recommended in case of dynamic background observed by a static camera. These algorithms are also robust in the case of illumination variation. In this work, we use the algorithm proposed by Mousse et al.³⁰ because they proved the efficiency and the efficacy of this algorithm. We chose this algorithm because it has a good performance in moving object detection. This method integrates a region based information by using superpixel segmentation algorithm into the original codebook algorithm and uses CIE $L^*a^*b^*$ colour space information. Figure 2 represents the flow diagram of the codebook based algorithm. For each frame, after the extraction of superpixels, we build a codebook background model. Let $P = \{s_1, s_2, \dots, s_k\}$ represents the K superpixels obtained after superpixels segmentation. Each superpixel $s_j, j \in \{1, 2, \dots, k\}$ is composed by m pixels. With each superpixel, we build a codebook $C = \{c_1, c_2, \dots, c_L\}$ which contains L codewords $c_i, i \in \{1, 2, \dots, L\}$. Each codewords c_i consists on an vector $v_i = (\bar{a}_i, \bar{b}_i)$ and 6-tuples $aux_i = \{\tilde{L}_i, \hat{L}_i, f_i, p_i, \lambda_i, q_i\}$ in which \tilde{L}_i, \hat{L}_i are the minimum and maximum of luminance value, f_i is the frequency at which the codeword has occurred, λ_i is the maximum negative run length defined as the longest interval during the

training period that the codeword has not recurred, p_i and q_i are the first and last access times, respectively, that the codeword has occurred. \bar{L} , \bar{a} , \bar{b} are respectively the average value of component L^* , a^* and b^* in a superpixel. The codebook model is created or updated using two criteria. The first criterion is based on colour distortion (1) whereas the second is based on brightness distortion (2).

$$\sqrt{||p_t||^2 - C_p^2} < \varepsilon_1 \quad (1)$$

$$I_{low} \leq I \leq I_{hi} \quad (2)$$

In (1), the autocorrelation value C_p^2 is given by equation (4) and $||p_t||^2$ is given by equation (3).

$$||p_t||^2 = \bar{a}^2 + \bar{b}^2 \quad (3)$$

$$C_p^2 = \frac{(\bar{a}_i \bar{a} + \bar{b}_i \bar{b})^2}{\bar{a}_i^2 + \bar{b}_i^2} \quad (4)$$

In (2), $I_{low} = \alpha \hat{L}_i$, $I_{hi} = \min\{\beta \hat{L}, \frac{\bar{L}}{\alpha}\}$ and $I = \bar{L}$.

After the extraction of foreground pixels, we group them into foreground region by finding the convex hull of all contours detected in the threshold image. The convex hull or convex envelope of a set X of points in the Euclidean plane is the convex set that contains all points of X . In order, to obtain the convex hull, we search for any point in X that enters the minimal convex hull for sure. We choose the one with the least x -coordinate (the left most one in X).

Then we project in the ground plane the polygon. This projection is performed by using planar homographic transformation.

Features extraction

We propose to extract some features to characterize the posture of the people. We use two categories of feature:

- the surfaces $\omega_i, i \in \{1, 2\}$ of the polygons obtained by projecting the polygons associated to this individual in each camera into the reference view;
- the surface σ of the polygon which is the intersection of the two projected polygons in the reference plane.

Based on these surfaces, we calculate $\varrho_1 = \frac{\sigma}{\omega_1}$ and $\varrho_2 = \frac{\sigma}{\omega_2}$. ϱ_1 and ϱ_2 respectively represent the percentage of surface in contact with the ground detected by the first camera and the second camera. These last two values allow us to assess the posture of individual. Indeed, when the individual approaches the ground (falling), the values of ϱ_1 and ϱ_2 will be greater (close to 1). Some examples of using of ϱ_1 and ϱ_2 for fall detection (based on some scenario take from the dataset which is used in this paper and described in the subsection) are showed by Figures 3 to 5.

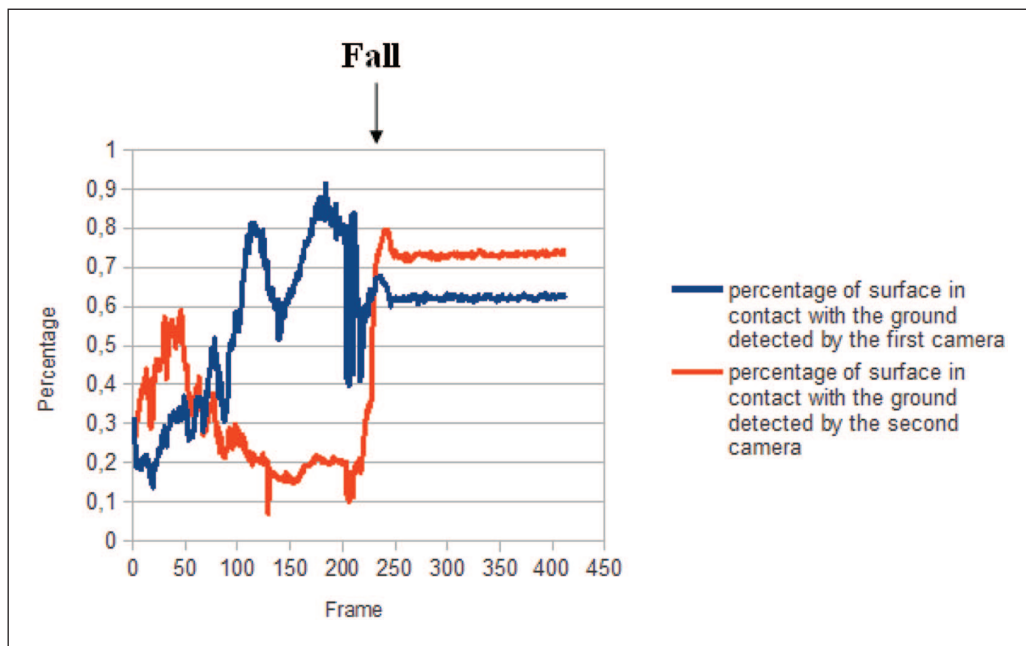


Figure 3. Example of using of the proposed features for fall detection (scenario 1).

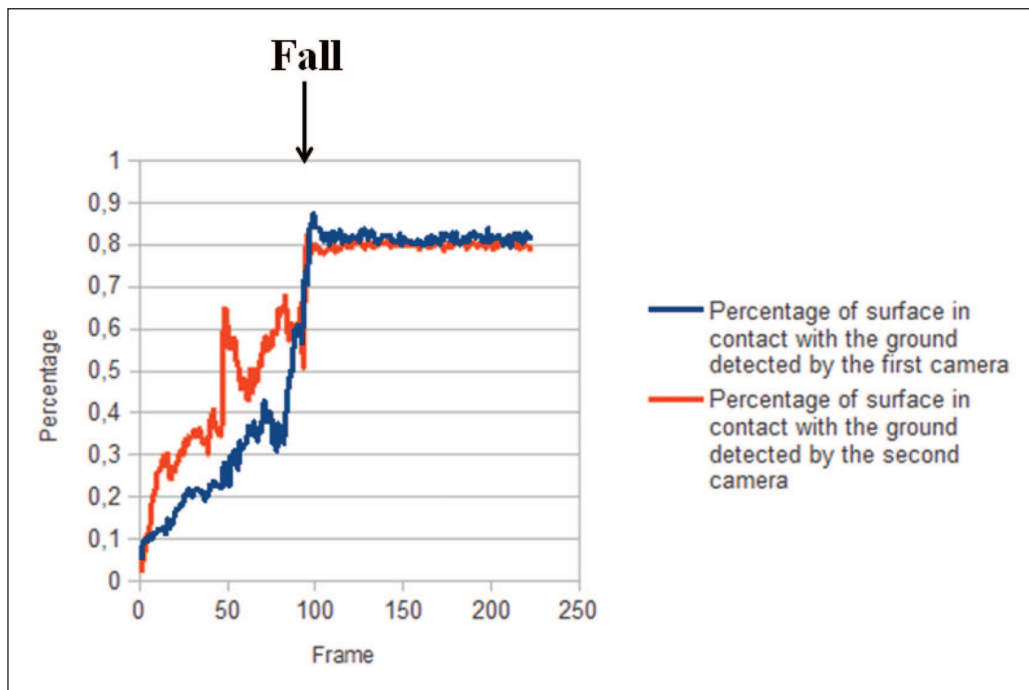


Figure 4. Example of using of the proposed features for fall detection (scenario 2).

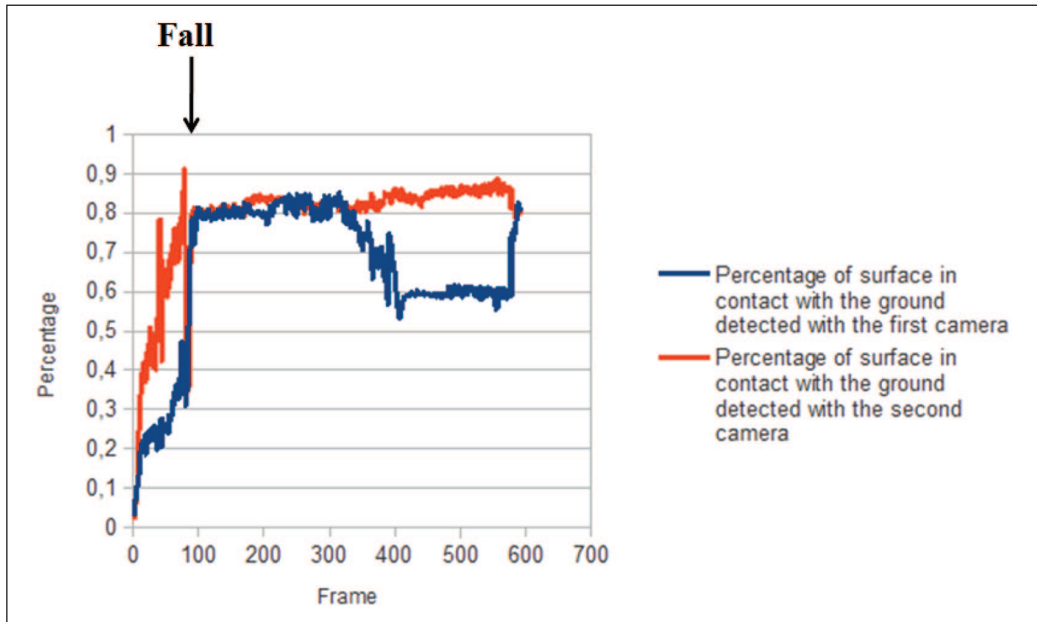


Figure 5. Example of using of the proposed features for fall detection (scenario 3).

Tracking

The object tracking is performed by the fusion centre. It is based on spatial location of objects. The adaptive strategy is introduced in this part of the system. For each foreground region, we extract some features. These features are: centre of gravity, polygon and colour histograms of the object in each view. We reduce to 32 bins per colour channel the colour histograms in order to reduce the quantity of information for each object, we define a spatial validation gate using its cinematic vitesse. The dimension of the gate is defined empirically to twice of the object bounding-box and is located around the predicted object position. This information is used to reduce region candidates. The matching is based on the spatial proximity of regions and their visual compatibilities. The algorithm evaluates explicitly the quality of each association. This information is summarized by two indicators which are Consistency and Identity indicators. These indicators are recursively updated and stored during the tracking step.

The consistency indicator of tracked objects firstly permits effective new objects to be validated after consecutive observations. Secondly, it permits to tolerate some temporary loss of the objects having a reliable track. It reacts as a robust filter at the object level. The indicator increases when no significant variation in the object features (colour histogram of associated object in each view, and bounding box size of the resulting projected polygon on the ground plane and the speed vector of the projected polygon) is perceived. Otherwise or in extreme situations when the target is lost, indicator is decreased. The dissimilarity between the object colour histograms is performed by the Bhattacharya distance. The updating process of the consistency indicator of each tracked object is controlled in terms of time delay defined by the human expert as the stability indicator. The track termination is decided for lost objects after a period of consecutive zero value consistency indicator. This delay has been fixed typically at 3 s. For each tracked object, a set of information is stored (Table 1). In Table 1 the attribute 'Object number' is a unique identifier for each object. We adopt the merging procedure (or respectively splitting

Table 1. Object information.

Object information
Object number
View 1 colour histogram (k), view 2 colour histogram (k)
Position in ground plane (k), Bounding box in ground plane (k)
Consistency indicator (k)
Speed vector (k)
Identity indicator (k)

procedure) presented in³¹ when we are in merging situation (or respectively splitting situation). The splitting situation is detected once a new object is detected close to a temporary group region whereas the merging situation is detected one object is detected close to more than one object.

Decision making

Using the feature extracted in subsection we classify the posture. This classification is done on two groups: lying down position and other positions (standing up, sitting, crouching positions). For each person, the posture is classify according the values present of ϱ_1 and ϱ_2 . Our experiments show that, if $\varrho_1 < 0.4$ or $\varrho_2 < 0.4$ then we conclude that this person is in a posture other than lying on ground position. In this case, for a more precise detection of the posture we reestimate the height h' of the detection. We compare h' to the initial height h of the person. We distinguish three 3 sub-cases:

1. if ($h' = h$ and ($\varrho_1 \leq 0.274$ or $\varrho_2 \leq 0.274$)) then the person is standing.
2. if ($h' < h$ and ($\varrho_1 \leq 0.274$ or $\varrho_2 \leq 0.274$)) then the person is crouching.
3. if ($\varrho_1 > 0.274$ or $\varrho_2 > 0.274$) then the person is sitting.

Whereas if the condition ($(\varrho_1 \geq 0.6$ and $\varrho_2 \geq 0.72)$ or $(\varrho_2 \geq 0.6$ and $\varrho_1 \geq 0.72)$) is true then the person is lying on ground plane. When a person changes rapidly from the first group of posture to lying on the ground posture, a warning state is attached to him and when he did not come quickly, an alarm is emitted to indicate the fall. In this work, we assume that the person changes rapidly posture when the delay between this change is less or equal to 1.5 s. After the warning state, we propose that the person stay down for 3 s before the alarm is triggered. This delay is important in order to manage the people who fall and get up afterwards. Also, if after the triggering of the alarm, the posture of the person is other than 'lying down', we expect 3 s before disabling the alarm. By proceeding like that, we limit the false negatives.

The system attempts to also solve the problem of occlusion by objects. Indeed, when the person is occluded by an object, the estimation of ϱ_1 and ϱ_2 will be biased.

However, it is rarely that the person is occluded by other objects and broken foreground happens in the both views of cameras.

Experimental environment and performance evaluation

Experimental environment

In this paper for our experimental results and in order to compare our performance to other methods, we use the 'Multi-view fall dataset' proposed by Auvinet et al.³³. This dataset was adopted in the

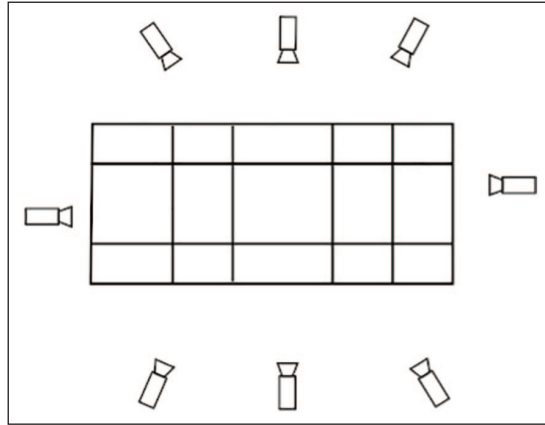


Figure 6. Experimental environment.

experiments of many research works. Then, it is possible to compare the performance between our proposed algorithm and some past research works. Auvinet et al. used eight inexpensive IP cameras with a wide angle to cover the whole room. The physical architecture of the experimental environment is presented in Figure 6. The dataset contains video of 24 scenarios showing 24 fall incidents and 24 confounding events (11 crouching, 9 sitting and 4 lying on a sofa) viewed from all the cameras and performed by one subject. These sequences include normal daily activities such as walking in different directions, housekeeping, activities with characteristics similar to falls. All types of falls are also simulated: forward falls, backward falls, falls when inappropriately sitting down, loss of balance. These falls event were performed in different directions with respect to the camera point of view. Camera settings and spatial arrangement, information of multi-camera synchronization, calibration parameters and event annotation for all scenarios are provided in their research work. The sequences contain some difficulties which can lead to segmentation errors like shadows, reflections, variable illumination, occlusions, etc. Video sequences from camera 1 and camera 3 are used in this paper since these two views are relatively orthogonal. Figure 7 presents some examples of typical simulated fall incidents and normal daily activities are shown. For single view object detection, we use parameters define in Mousse et al.³⁰ The experiment environment is Intel Core i7 CPU L 640 @ 2.13GHz × 4 processor with 4GB memory and the programming language is C++.

Performance evaluation

In this subsection, we evaluate the performance of our proposed people fall detection and compare it to other research works recent research works.^{6,21,23,28} We compute some metrics for testing the efficiency and the accuracy of our algorithm and compare it to the cited paper which are also tested on the same dataset. These metrics are sensitivity S_e and specificity S_p .

$$S_e = \frac{TP}{TP + FN} \quad (5)$$

$$S_p = \frac{TN}{TN + FP} \quad (6)$$

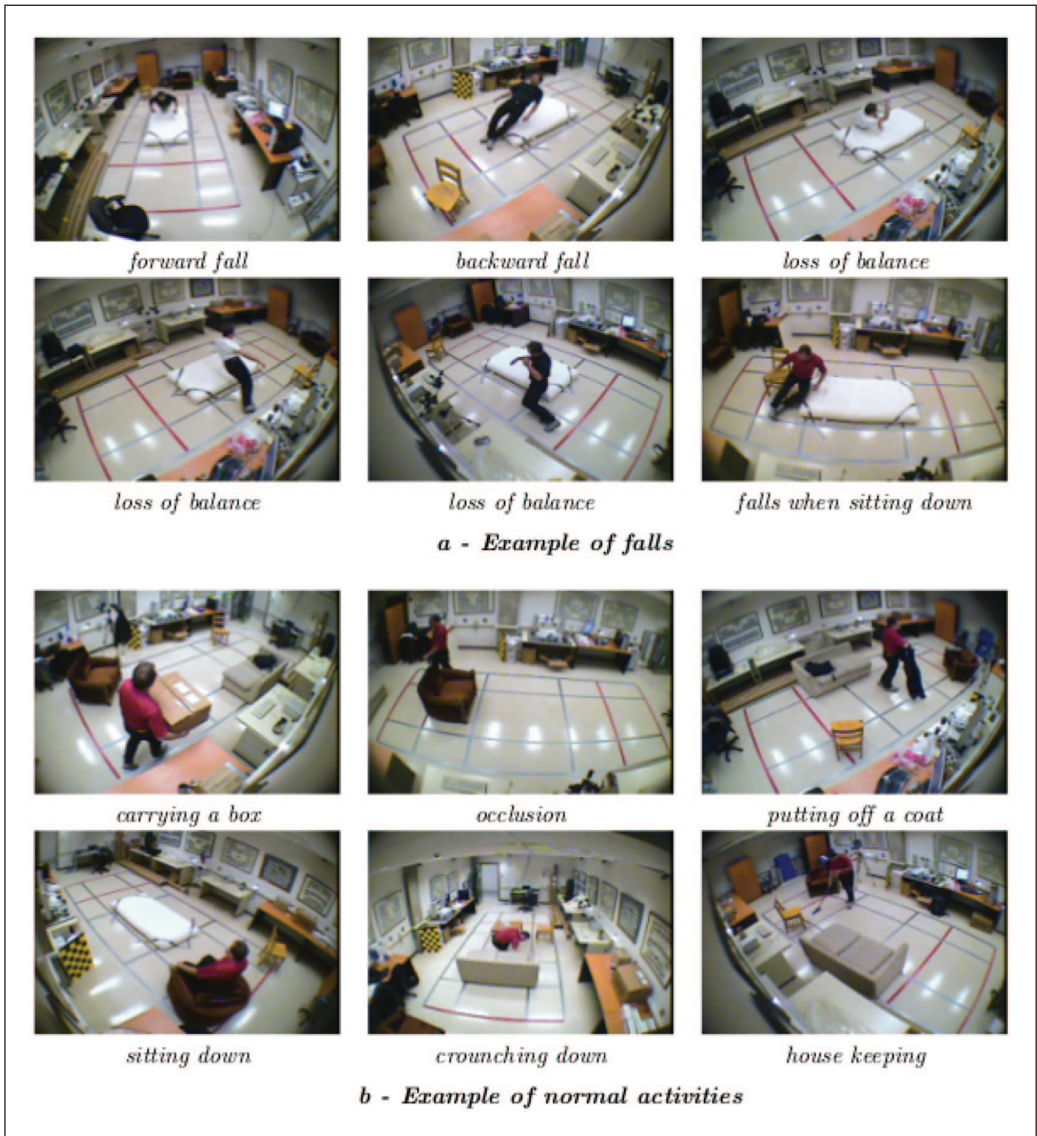


Figure 7. Examples of falls and normal daily activities.

In (5) and (6), TP is the number of falls correctly detected, FN is the number of falls not detected, TN is the number of normal activities not detected as a fall and FP is the number of normal activities detected as a fall. High sensitivity means that most fall incidents are correctly detected. Similarly, high specificity implies that most normal activities are not detected as fall events. A good fall detection method must achieve high values of sensitivity and specificity. The results are reported in Table 2. In this table, the results of the method proposed by Auvinet et al.²¹ are reported with a network of three cameras. Auvinet et al.²¹ prove that the sensitivity can be boosted to 100% if a network of more than four cameras is employed. By comparing algorithms we conclude that our algorithm has similar performance to recent algorithms. Such as Hung and Saito²⁸ and Mousse

Table 2. Performance comparison between our method and three state-of-the-art methods, tested on the same dataset.

	Sensitivity Se (%)	Specificity Sp (%)
Our method	95.8	100
Auvinet et al. ²¹	80.6	100
Rougier et al. ⁶	95.4	95.8
Hung and Saito ²⁸	95.8	100
Mousse et al. ²³	95.8	100

Table 3. People posture recognition.

	Recognition rate (%)
Standing	100
Sitting	100
Crouching	81.8

Table 4. Processing times comparison.

	Hung and Saito ²⁸	Mousse et al. ²³	Proposed method
Speed (fps)	10.95	15.25	22.25

et al.²³ algorithms, our method only fails in the 22nd scenario in which the person is sitting on a chair and suddenly slips to the floor. In addition of fall detection, our system recognizes the other postures. The algorithms proposed by Hung and Saito²⁸ and Mousse et al.²³ didn't permit this possibility. The posture recognition performance is reported in Table 3. According to Table 3, all standing and sitting poses are correctly detected. Our method fails in two cases when we try to recognize crouching poses.

Auvinet et al. and Rougier et al. methods are high computational due to the use of 3D reconstruction algorithm. Auvinet et al.²¹ presents the GPU implementation to realize their method in real time. Meanwhile, such as Hung and Saito²⁸ and Mousse et al.²³ methods, our method composing of low-cost modules is implemented in real-time in a common desktop PC and achieves very competitive performance. Then we compare our processing time to the processing times Hung and Saito²⁸ and Mousse et al.²³ This comparison is performed in Table 4 and the values are expressed in frames per second. According to this table, we conclude that our algorithm is of lower computational cost than the algorithms proposed by Hung and Saito²⁸ and Mousse et al.²³

Conclusion

This paper introduces a new video based system for fall detection. The proposed approach uses posture recognition to detect fall. For object detection, we use a codebook based method and we approximate the foreground pixel by using polygons. This approximation reduces the number of information which will be process by the fusion process. This fusion is done by using homography mapping. The fall is detected by using a set of features such as the surface of the polygon and the percentage of surface which is in contact with ground. Our experimental results using public dataset show that the proposed method can accurately detect a single falling person.

The proposed method is highly dependent on each camera foreground pixel extraction. Then the presence of false positive can influence the fall detection process. To solve this problem, we will propose in our future work, the use of other additional sensors and/or metrics. Finally errors will occur if non-human objects appear in the scene.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Mikael Ange Mousse  <https://orcid.org/0000-0002-3326-6396>

References

1. Casilari E, Lora-Rivera R and García-Lagos F. A study on the application of convolutional neural networks to fall detection evaluated with multiple public datasets. *Sensors* 2020; 20: 1466.
2. Mubashir M, Shao L and Seed L. A survey on fall detection: principles and approaches. *Neurocomput* 2012; 100: 144–152.
3. Mirmahboub B, Samavi S, Karimi N, et al. Automatic monocular system for human fall detection based on variations in silhouette area. *IEEE Trans Biomed Eng* 2013; 60: 427–436.
4. Iazzi A, Rziza M and Oulad Haj Thami R. Fall detection system-based posture-recognition for indoor environments. *J Imaging* 2021; 7: 42
5. Wang BH, Yu J, Wang K, et al. Fall detection based on dual-channel feature integration. *IEEE Access* 2020; 8: 103443103453.
6. Rougier C, Meunier J, St-Arnaud A, et al. Robust video surveillance for fall detection based on human shape deformation. *IEEE Trans Circ Syst Video Technol* 2011; 21: 611–622.
7. Wang Z, Ramamoorthy V, Gal U, et al. Possible life saver: a review on human fall detection technology. *Robotics* 2020; 9: 55.
8. Feng W, Liu R and Zhu M. Fall detection for elderly person care in a vision-based home surveillance environment using a monocular camera. *Signal Image Video P* 2014; 8(6): 1129–1138.
9. Liao YT, Huang CL and Hsu SC. Slip and fall event detection using bayesian belief network. *Pattern Recognit* 2012; 45(1): 24–32.
10. Charfi I, Miteran J, Dubois J, et al. Optimized spatio-temporal descriptors for real-time fall detection: comparison of support vector machine and adaboost-based classification. *J Electr Imaging* 2013; 22(4): 041106041106.
11. Asif U, Mashford B, Cavallar SV, et al. Privacy preserving human fall detection using video data. In: *Proceedings of the machine learning for health workshop*, 2020.
12. Yao C, Hu J, Min W, et al. Novel real-time fall detection method based on head segmentation and convolutional neural network. *J Real-Time Image Processing* 2020; 17: 19391949.
13. Rougier C, Auvinet E, Rousseau J, et al. Fall detection from depth map video sequences. *International Conference on Smart Homes and Health Telematics*, pages 121–128, 2011.
14. Kepski M and Kwolke B. Fall detection using ceiling-mounted 3d depth camera. In: *International joint conference on computer vision, imaging and computer graphics theory and applications*, Lisbon, Portugal, 2014, pp.640–647.
15. Stone E and Skubic M. Fall detection in homes of older adults using the microsoft kinect. *IEEE J Biomed Health Inform* 2014; 19: 290–301.
16. Zhang Z, Liu W, Metsis V, et al. Viewpoint-independent statistical method for fall detection. In: *IEEE Conference on Pattern Recognition*, Tsukuba Science City, Japan, 2012, pp.3626–3630.
17. Mastorakis G and Makris D. Fall detection system using Kinects infrared sensor. *J Real-Time Image Process* 2014; 9(4): 635–646.
18. Bian Z, Hou J, Chau L, et al. Fall detection based on body part tracking using a depth camera. *IEEE J Biomed Health Inform* 2015; 19(2): 430–439.

19. Auvinet E, Reveret L, St-Arnaud A, et al. Fall detection using multiple cameras. In: *2008 30th annual international conference of the ieee engineering in medicine and biology society*, Vancouver, BC, Canada, 20–25 August 2008, pp.2554–2557. IEEE.
20. Auvinet E, Multon F, St-Arnaud A, et al. Fall detection using body volume reconstruction and vertical repartition analysis. In: *International conference on Image and signal processing*, Trois-Rivières, QC, Canada, 2010, pp.376–383.
21. Auvinet E, Multon F, St-Arnaud A, et al. Fall detection with multiple cameras: an occlusion-resistant method based on 3-d silhouette vertical distribution. *IEEE Trans Inform Technol Biomed* 2011; 15: 290–300.
22. Anderson D, Luke R, Keller J, et al. Linguistic summarization of video for fall detection using voxel person and fuzzy logic. *Comput Vis Image Underst* 2009; 113(1): 80–89.
23. Mousse MA, Motamed C and Ezin EC. Percentage of human occupied areas for fall detection from two views. *Visual Comput* 2017; 33(12): 1529–1540.
24. Anderson D, Luke RH, Keller JM, et al. Modeling human activity from voxel person using fuzzy logic. *IEEE Trans Fuzzy Syst* 2009; 17(1): 39–49.
25. Thome N, Miguet S and Ambellouis S. A real-time, multiview fall detection system: a lhmm-based approach. *IEEE Trans Circuits Syst Video Technol* 2008; 18(11): 1522–1532.
26. Cucchiara R, Prati A and Vezzani RA. Multi-camera vision system for fall detection and alarm generation. *Expert Syst* 2007; 24(5): 334–345.
27. Zweng A, Zambanini S and Kampel M. Introducing a statistical behavior model into camera-based fall detection. In: *International conference on advances in visual computing*, Las Vegas, USA, 2010. pp.163–172.
28. Hung DH and Saito H. The estimation of heights and occupied areas of humans from two orthogonal views for fall detection. *IEEJ Trans Electr Inform Syst* 2013; 133(1): 117–127.
29. Hung DH, Saito H and Hsu GS. Detecting fall incidents of the elderly based on human-ground contact areas. In: *2nd IAPR Asian conference on pattern recognition*, Naha, Japan, 5–8 November 2013, pp.516–521. IEEE.
30. Mousse MA, Motamed C and Ezin EC. Fast moving object detection from overlapping cameras. In: *International conference on informatics in control, automation and robotics*, Colmar, Alsace, France, 2015, pp.296–303.
31. Motamed C. Motion detection and tracking using belief indicators for an automatic visual-surveillance system. *Image Vision Comput* 2006; 24(11): 192–201.
32. Ren L and Peng Y. Research of fall detection and fall prevention technologies: a systematic review. *IEEE Access* 2019; 7: 7770277722.
33. Auvinet E, Rougier C, Meunier J, et al. Multiple cameras fall data set Technical Report 1350. *DIRO Université de Montréal*, 2010.