Computers in Biology and Medicine 115 (2019) 103520

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: http://www.elsevier.com/locate/compbiomed

A vision-based approach for fall detection using multiple cameras and convolutional neural networks: A case study using the UP-Fall detection dataset

Ricardo Espinosa [a], Hiram Ponce [b,*], Sebastián Gutiérrez [a], Lourdes Martínez-Villaseñor [b], Jorge Brieva [b], Ernesto Moya-Albor [b]

[a] Universidad Panamericana, Facultad de Ingeniería, Josemaría Escrivá de Balaguer 101, Aguascalientes, Aguascalientes, 20290, Mexico
[b] Universidad Panamericana, Facultad de Ingeniería, Augusto Rodin 498, Ciudad de México, 03920, Mexico

ARTICLE INFO

Keywords:
Human activity recognition
Human fall detection
Machine learning
Healthcare
Computer vision

ABSTRACT

The automatic recognition of human falls is currently an important topic of research for the computer vision and artificial intelligence communities. In image analysis, it is common to use a vision-based approach for fall detection and classification systems due to the recent exponential increase in the use of cameras. Moreover, deep learning techniques have revolutionized vision-based approaches. These techniques are considered robust and reliable solutions for detection and classification problems, mostly using convolutional neural networks (CNNs). Recently, our research group released a public multimodal dataset for fall detection called the UP-Fall Detection dataset, and studies on modality approaches for fall detection and classification are required. Focusing only on a vision-based approach, in this paper, we present a fall detection system based on a 2D CNN inference method and multiple cameras. This approach analyzes images in fixed time windows and extracts features using an optical flow method that obtains information on the relative motion between two consecutive images. We tested this approach on our public dataset, and the results showed that our proposed multi-vision-based approach detects human falls and achieves an accuracy of 95.64% compared to state-of-the-art methods with a simple CNN network architecture.

1. Introduction

Human activity recognition (HAR) in the monitoring and tracking of human health is an interesting topic that has recently been growing within the research community, especially in the detection of falls among elderly people. Falls can cause injuries, bodily harm, fractures, etc. In fact, globally, falls are the second leading cause of unintentional injury and injury-related deaths among adults 65 years of age and older [1]. "Approximately 28–35% of people aged 65 and over fall each year, increasing to 32–42% for those over 70 years of age" [2]. Falls frequently cause functional dependencies in the elderly. Additionally, many fall-related deaths result from a long "laying time", which is defined as the extended period of time in which the victim remains immobile on the ground.

There are many types of falls. Oneill et al. [4] divides human falls by direction, namely, forward, backward and to the side. For instance, falls forward are the most common falls, with 38% occurring in men younger than 65 and 62% occurring in men older than 65. Similarly, falls forward occur 62% of the time in women younger than 65 and 60% of the time in women older than 65.

In 1987, the Kellogg International Working Group [3] on the prevention of falls in the elderly defined a fall as an unintentional ground impact or some lower level for reasons other than sustaining a violent blow, loss of consciousness, or the sudden onset of paralysis, as in stroke or epileptic seizure. A human fall typically starts with a short freefall period. This freefall causes the acceleration to significantly decrease below the 1G threshold. This freefall represents the period of time when the actual fall is taking place. The fall must stop, and a fall causes acceleration and a spike in the graph. The amplitude crossing an upper threshold suggests a fall [5].

It has been proven that the medical consequences of a fall are highly contingent upon the response and rescue time. In this sense, fall

* Corresponding author.
E-mail addresses: respinosa@up.edu.mx (R. Espinosa), hponce@up.edu.mx (H. Ponce), jsgutierrez@up.edu.mx (S. Gutiérrez), lmartine@up.edu.mx (L. Martínez-Villaseñor), jbrieva@up.edu.mx (J. Brieva), emoya@up.edu.mx (E. Moya-Albor).

https://doi.org/10.1016/j.compbiomed.2019.103520
Received 4 September 2019; Received in revised form 24 October 2019; Accepted 25 October 2019
Available online 30 October 2019

detection systems can improve the response time of medical professionals and decrease the medical consequences of falls.

Due to the extraordinary advances in and increased research on embedded sensor systems, mobile devices and microelectronics, Internet of Things (IoT) systems allow people to continually interact with technology. Additionally, large amounts of data about a person's daily actions are needed so that fall detection systems can allow for the rapid and appropriate assistance of elderly people.

There are many different types of fall detection systems, including sensor-based, vision-based and multimodal-based systems. For instance, sensor-based approaches make use of ambient, smart devices and wearable sensors to provide important information, such as acceleration, absence/presence of individuals, etc., while vision-based strategies use images, such as 3D reconstructions of the environment, simple 2D RGB video sequences with one or multiple cameras, or depth images acquired from 3D depth sensors, as the main input. Multimodal-based approaches collect all the information possible from cameras, microphones, wearable sensors, ambient sensors, and smart devices, among others, and they combine all this information to improve the fall detection and classification results in a practical manner.

Analytical and machine learning methods are two main approaches for detecting activities and falls [6]. Analytical methods detect falls using threshold algorithms. For example, when falling, a person hits either the ground or an obstacle. This impact results in an intense reversal of the acceleration in terms of trajectory. This change in directionality can be detected by a threshold value. With these types of methods, the most difficult task is adapting the detection to different types of falls or to different people since thresholds differ by person and/or by the type of fall [6]. To address this problem, there are other strategies, such as pattern normalization [61] and correlation-based algorithms [62], and recent investigations report the use of optimization algorithms to choose the threshold [47].

Furthermore, machine learning methods have been gaining popularity due to their flexibility to different subjects and types of falls [63]. The most well-known supervised learning techniques used for fall detection systems include multi-layer perceptron (MLP) [42], support vector machine (SVM) [39], the hidden Markov model (HMM), decision trees, random forest, k-nearest neighbors (KNN) [41], and the convolutional neural network (CNN) [40], which is a deep learning method.

Deep learning techniques are currently changing and improving the methods used to address computer vision problems. CNNs automatically learn features from training data, thus creating a feasible automatic feature extraction method for images. CNNs have been widely applied in image processing problems; for example, in Ref. [48], the authors use deep learning to detect accidents using the optical flow as the feature extraction method and then test this method using real videos. In Ref. [49], a single CNN was trained by using images to directly classify skin lesions and to detect cancer, and it achieved an area under the receiver operating characteristic curve (AUC) of 0.96%. Moreover, CNNs have been used in fall detection systems with a sensor-based approach with up to 92.3% accuracy [50] and with a wearable approach with an AUC equal to 0.75 [51].

Regarding vision-based approaches for fall recognition systems, deep learning has been successfully applied. For example, Nez-Marcos et al. [19] implemented a CNN to avoid manual feature engineering; the convolutional layers of the system extracted the most important features of the images, and a sensitivity and specificity of 94% was obtained. A CNN for a vision-based approach was also implemented in Ref. [52], in which the authors used a 3D CNN with input videos of peoples' kinematics and achieved 100% accuracy when evaluated on different datasets.

Recently, our research group released a public multimodal dataset for fall detection called the UP-Fall Detection dataset [24]. The data were gathered from different sources of information, i.e., wearable sensors, ambient sensors and cameras. Until now, we have studied this dataset using a multimodal approach [24]. However, the technical expertise and skills required for building and setting a multimodal fall detection system make it difficult to implement in the real world. Moreover, wearable and ambient sensors are conditioned by the environment, thus making portability difficult to achieve. Therefore, we are interested in creating a vision-based fall detection system using this dataset and the video recordings from multiple cameras.

Additionally, fall detection systems based on single RGB cameras are often viewpoint-dependent, according to Ref. [67]. This issue raises the need for new datasets when a camera is moved to different viewpoints and, in particular, to different heights. To address this issue, using different camera viewpoints in a dataset can help to identify whether a given method is independent of viewpoint. To this end, a fall detection system must be reliable regardless of the position of the subject while falling with respect to the camera.

Based on the above, this work presents a fall detection system based on a 2D CNN inference method and multiple cameras. As we describe later, this approach analyzes images using fixed time windows and extracts features using an optical flow method that obtains the information of the relative motion between two consecutive images from video recordings acquired from cameras with different viewpoints. We tested this approach with our public UP-Fall Detection dataset, and the results showed that our proposed multi-vision-based approach detects human falls using a simple CNN network architecture, achieving results that are comparable to those of state-of-the-art methods. In addition, the performance of the proposed approach is comparable to the performance achieved using a multimodal approach.

Even though CNN has previously been used in fall detection systems with good performance using a particular dataset, Casilari et al. [53] concluded that these systems should be trained and tested with different datasets due to the different numbers of samples, different types of falls and different time series for any type of fall. Thus, the implementation of CNN in the multicamera vision-based approach, specifically for the UP-Fall Detection dataset, might improve state-of-the-art fall detection systems.

The main contributions of this work are as follows: (i) the usage of multiple cameras with CNN for fall detection and classification, (ii) the implementation of this approach with the UP-Fall Detection database, and (iii) the comparison of the performance of this approach with those of well-known supervised learning methods. To the best of our knowledge, only one study [59] combines a CNN with a multicamera vision-based approach to recognize human falls. In contrast to our proposal, in Ref. [59], a voting strategy based on the output results from independent cameras is used; our approach uses the information from all the cameras in the same machine learning model.

The rest of the paper is organized as follows. First, we review and analyze different fall detection systems, focusing mainly on vision-based solutions. Next, we present a description of the UP-Fall Detection dataset. Then, we present the proposal in detail. We also explain the experiment as well as the results and then discuss the proposal. Finally, the conclusions are presented.

2. Fall detection systems

HAR and fall detection are difficult tasks, and there are several ways to complete these tasks due to the numerous approaches proposed in the literature. For instance, Lara et al. [8] and Noury [6] divide the HAR taxonomy into three general approaches, i.e., external, wearable and video sensing, depending on the information source. Using these approaches, there are case studies related to sensor-based [9], vision-based [10], smartphone-based [11], and multimodal-based [12] strategies to address human fall recognition, as described below.

2.1. Sensor-based fall detection systems

With the increasing accessibility of mobile sensor technology, fall detection systems have been designed for real-world purposes. Human

activity can be tracked, monitored and labeled using data from different types of sensors at various locations in the environment and in the human body. An important application of sensor-based approaches is the detection of abnormal activities from wearable sensors [9]. Abnormal activity detection methods can be applied to continuously track the movements of an individual to determine whether the person's activities are abnormal [9]. In Ref. [21], using triaxial sensors and SVM as an inference method, the authors achieved 98.33% accuracy. In Ref. [65], using acceleration and the Euler angle (yaw, pitch, and roll), the authors achieved 100% accuracy, sensitivity, and specificity. Nevertheless, some disadvantages of heterogeneous sensor networks are because human activities typically involve more than one body part. Moreover, several physiological and biomechanical studies have shown that most daily human activities are inherently multimodal [13]. Thus, different types of sensors are required for data collection.

2.2. Wearable fall detection systems

Wearable approaches are common solutions for fall detection, as they take advantage of the low cost, online tracking capability and small sizes of wearable technologies. For example, in Ref. [46], a Shimmer device was used for acquisition and transition data. The wearable device was placed on the chest and obtained 98.8% accuracy using different machine learning models. In Ref. [47], the authors used a wearable band placed on the wrist and achieved 0.95 specificity and 0.83 sensitivity using threshold-based peak recognition with SVM for classification; they optimized the threshold values for different datasets.

In wearables and smartphones, energy consumption is a difficult problem to solve since these devices need to be continuously worn to obtain tracking information from subjects. The lifetimes of wearables and smartphones are limited to the capacity of the battery, and constant recharging is necessary, which prevents the constant tracking of the patient's activities [46].

2.3. Smartphone-based fall detection systems

Currently, smartphones contain multiples integrated sensors and a large processing capacity, which has grown over the years. Smartphones can measure the movements of a controller in a nonintrusive way. Smartphone-based fall detection systems use smartphone sensors, e.g., gyroscopes, triaxial accelerometers, and altimeters, to collect data over long periods of time. Case studies using this approach can be found in Ref. [43], in which the authors use a smartphone-based triaxial accelerometer with statistical time-domain features. Then, the authors applied principal component analysis for feature selection and inferred the outputs with MLP, obtaining an accuracy of 92%. Another example is the work of Vilarinho et al. [44], which combined smartphone and smartwatch sensors and used threshold-based techniques and pattern recognition algorithms to recognize falls with an accuracy of 63% and daily activities with an accuracy of 78%.

2.4. Multimodal-based fall detection systems

Data acquisition, mainly from ambient sensors, wearables, cameras, microphones, and radio frequency identification (RFID) tags, among others, is an important task of fall detection and classification systems. Wearables are not able to distinguish a large number of fine-grained and/or complex human activities, as they have difficulty differentiating between similar activities; thus, ambient sensors are needed for context awareness. In this regard, multimodal-based approaches can combine more than one source of data to obtain information about both the environment and the user. These approaches make fall detection and classification feasible by leveraging different modes of sensing from a wide range of sources [12].

Because multimodal approaches comprise many different sources of data from subjects and environments, there are some weaknesses, as reported in Ref. [60]: (i) many information types require the application of robust feature extraction and feature selection techniques and considerations regarding machine learning approaches for different types of input data, making fall detection computationally expensive and difficult to perform, and (ii) multiple sensors with complex placement on the body (and in the environment) could lead to high costs, practical deployment difficulties, and obtrusiveness, especially for elderly people.

2.5. Vision-based fall detection systems

This work mainly focuses on vision-based methods. Traditionally, fall detection systems have been implemented by using computer vision and image processing techniques to classify activities. With recent advancements, in-depth, noninvasive imaging sensors produce high-quality images. This information is also analyzed for human tracking, monitoring and user recognition systems [14–16] and for monitoring and recognizing the daily activities of subjects in indoor environments [17].

Most of the vision-based approaches have used simple RGB cameras, web cameras, motion camera systems, or even Kinect [18]. The use of Kinect for fall detection has increased because it can obtain 3D information such as human poses or limb positions [18].

The classic vision-based fall detection and classification strategies consist of four phases [4]: (1) data acquisition from video sequences, (2) feature extraction from images, (3) feature selection and (4) learning and inference. Multiple machine learning techniques are used in the literature, such as SVM [35] or random forest [36]. Zerrouki et al. [17] proposed a fall detection system based on human silhouette variations in vision monitoring and SVM to identify postures. Then, these authors used HMM to classify the data into fall and non-fall events. Rougier et al. [7] tracked the person's silhouette along with the video sequences. Shape deformation was then quantified from these silhouettes based on shape analysis methods. Finally, falls were detected from daily activities using Gaussian mixture models (GMMs).

Vision-based systems can be divided into two categories: monocular systems and multicamera systems. Monocular-based fall detection systems depend on one viewpoint. Moving a camera to different viewpoints requires collecting new training data for that specific viewpoint and calibrating the camera sensor. However, these systems can fail because of occluding objects between the target and camera. Zhang et al. [66] proposed using multiple Kinect devices to solve that problem, as their OCCU dataset included both occluded and nonoccluded falls. Kwolek et al. [21] extracted depth maps about the environment and the person's silhouette in combination with 3-axis accelerometers and SVM as a machine learning technique. In terms of multicamera fall detection systems, Thome and Miguet [22] proposed using an HMM to distinguish falls from walking activities. The features extracted for the motion analysis were obtained from a metric image rectification in each view. Anderson et al. [23] analyzed the states of 3D objects, called the voxel of a person, obtained from two cameras. All these works construct 3D models with multiple cameras to reconstruct the environment. This task is particularly difficult because the cameras need to be calibrated to correctly compute the 3D information, which presents issues regarding the synchronization of the video sequences of each camera, making it more difficult to implement than a monocular-based approach.

Thus, from the point of view of the deployment of these systems, 2D multiple cameras are usually a good option, mainly because of their low cost and ease of implementation. It is also important to emphasize that cameras are already installed in many public places, such as airports, shops, and elderly care centers, and these cameras can also be used for fall detection systems. Thus, 2D cameras are relevant for the fall detection application domain.

Multiple studies use CNN on monocular vision-based fall detection systems with excellent results [37,50–52]. Moreover, several works use a multicamera approach with different classical machine learning models or other algorithms [54–58], and only one work uses a

multicamera approach and CNN in a fall detection system [59].

2.6. Vision-based fall detection systems using CNN

Recent works on fall recognition systems have taken advantage of the success of deep learning for recognition and classification tasks using regular images, deep images, infrared images, etc. Deep learning CNN searches for the relevant features in images, avoiding the feature engineering tasks and providing versatile automatic feature extraction, depending on the architecture of its convolutional and inference layers [25].

Some recent works on fall detection systems are reported in Ref. [70], in which the authors use rule-based filters before an input convolutional layer, combining the convolutional layer output with optical flow features to choose a good input for the inference phase of its 3D CNN architecture; these method achieved 92.67% accuracy. In Ref. [72], the authors use infrared (IR) images and a 3D CNN to find features on three color channels in real situations, taking into consideration the spatiotemporal image information; this method achieved an 85% accuracy on test video sequences.

3. Dataset Description

In this research, we use a public dataset called UP-Fall Detection [24]. This dataset was made using 17 healthy subjects without any impairments (9 males and 8 females) ranging from 18 to 24 years of age, with a mean height of 1.66 ± 0.0530 m and a mean weight of 66.8 ± 12.1182 kg; the subjects performed 11 activities and 3 trials per activity, including six simple human daily activities and five different types of human falls using a multimodal approach, with wearable sensors, ambient sensors, and vision devices. The consolidated dataset and the feature dataset are publicly available.

The activities and falls stored in this dataset are summarized in Table 1. All data were collected using the following 14 devices: five Mbientlab MetaSensor2 wearable sensors that collected the raw data from a 3-axis accelerometer, a 3-axis gyroscope, and an ambient light sensor; one electroencephalograph (EEG) NeuroSky MindWave headset was used to measure the raw brainwave signal from its unique EEG channel sensor located at the forehead; as context-aware sensors, we installed six infrared sensors as a grid 0.40 m above the floor of the room to measure the changes in interruption of the optical devices (where 0 means interruption and 1 means no interruption); and two Microsoft Life-Cam Cinema cameras, one for a lateral view and the other for a frontal view in relation to the subject, were located at 1.82 m above the floor, which is higher than the mean height of the subjects. The falls were performed from right to left according to the viewpoint of Camera 1 (lateral view). All experiments were recorded by positioning the subject at the center of the view in both cameras and at the same distance from both cameras. That is, Camera 1 and Camera 2 were 2.10 m and 1.90 m away from the center point of the mattress, respectively.

Table 1
Activities performed by subjects, adapted from Ref. [24].

Activity ID	Description	Duration (s)	Abbreviation
1	Falling forward using hands	10	FH
2	Falling forward using knees	10	FF
3	Falling backward	10	FB
4	Falling sideward	10	FS
5	Falling while attempting to sit sitting in an empty chair	10	FE
6	Walking	60	W
7	Standing	60	S
8	Sitting	60	ST
9	Picking up an object	10	P
10	Jumping	30	J
11	Laying	60	L

Additionally, all images contain at most one subject, so multiple people were not simultaneously recorded for this dataset. All these devices were located as shown in Fig. 1. For more information about the UP-Fall Detection dataset, see Ref. [24].

In this work, we use only the information from two cameras, which run at 18 fps, from the dataset, taking advantage of the multiple camera distributions.

Here, we aim to implement a fall detection and classification system using multiple cameras and to compare its performance when using only a monocular-based approach. To this end, CNN will be used as the classification model.

4. Description of the proposal

In this work, we adopted the traditional workflow for fall detection systems [8], which consists of the following steps: (i) data collection, (ii) windowing, (iii) feature extraction, and (iv) learning and inference. The workflow is shown in Fig. 2.

4.1. Data collection

One of the most challenging phases in the traditional workflow for fall detection systems and machine learning problems in general is data collection. Recently, to be correctly trained and tested, deep learning techniques require large amounts of data and other factors, such as the number of individuals, the physical characteristics of the individuals, and the number of individuals with diverse characteristics in terms of gender, age, height, weight, and health conditions [8].

As we explained in the Dataset Description section, we use the UP-Fall detection dataset. To summarize, the dataset contains information related to 17 young subjects performing 11 different activities, including 5 falls and 6 other activities. For this work, we use the information gathered from two RGB cameras positioned at different viewpoints (lateral and front views) [24].

4.2. Windowing

The windowing approaches in fall detection systems are normally used to segment the time series of performed falls. Segmentation is the process of dividing the sensor signals into smaller data segments. This process has been performed in different ways in the activity recognition field and fall detection systems. Segmentation techniques can be categorized into three main groups: activity-defined windows, event-defined windows and sliding windows [68].

We adopted a sliding window approach to capture the temporal dependency between samples. In this case, we divided all data into fixed-length time windows for each activity. Our implementation uses 1-s widows with 0.5 s of overlap due to the results reported in Ref. [24]. In that work, we experimented with 1-s, 2-s and 3-s windows with 50% overlap by applying classic machine learning methods; the best performance was achieved with the 1-s window. In this work, we employ this performance evaluation as the baseline of our multiple-camera-based proposal. To this end, we are able to analyze the fall on each window in a simple way. The result of this step was multiple 1-s window length series of images to be processed in the next step.

4.3. Feature extraction

Feature extraction is a general method in which one tries to transform the input space into a low-dimensional subspace that preserves most of the relevant information to improve data analysis [69]. In fall detection systems, there are multiple techniques used to extract relevant information depending on the data type; in vision-based approaches, the optical flow algorithm provides very rich information about the apparent movements in images, and these approaches have been used in multiple studies that combine CNN and optical flow to extract features

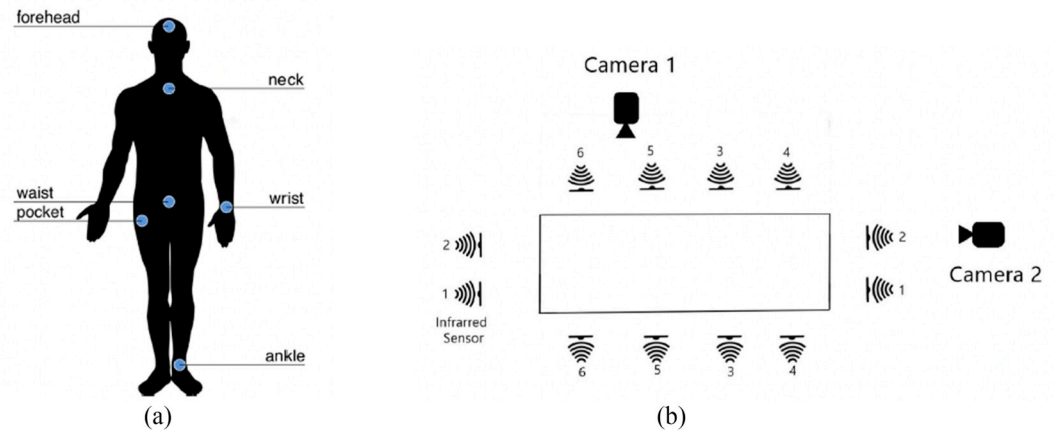

Fig. 1. Distribution of the sensors. (a) Wearable sensors and EEG helmet on the human body. (b) Layout of ambient sensors and multiple cameras. Adapted from Ref. [24].

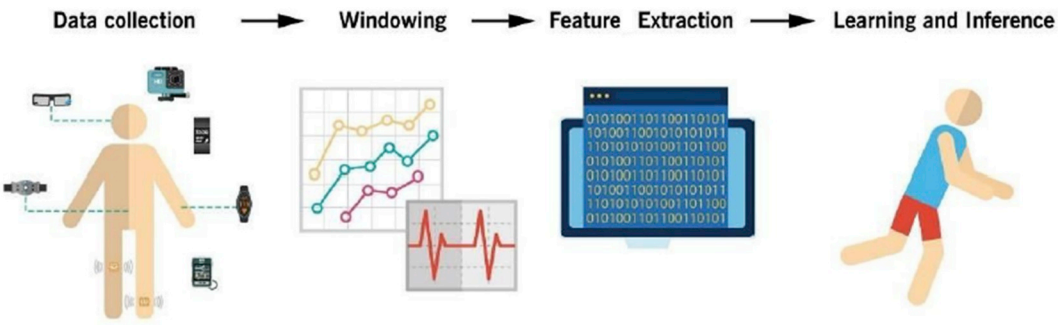

Fig. 2. Traditional workflow for fall detection systems.

from images [37,70].

For feature extraction, each window is preprocessed to obtain sufficient information for describing the activity. In this case study, the optical flow algorithm [24,26] was used for the visual features extracted from each camera. This algorithm helps us to obtain the apparent displacements between two windows and, in doing so, to distinguish movements and directions without considering the static features in the image. The obtained features are the horizontal and vertical relative movements of the pixels in the images, i.e., U and V, respectively [26]. The resultant combination, D, of these values corresponds to the magnitude of the relative movement, as shown in (1), where the resultant matrix images are the same size of the original window images, i.e., 640×480 pixels in our case study.

$$D_{i,j} = \sqrt{U_{i,j}^2 + V_{i,j}^2} \tag{1}$$

4.4. Learning and inference

In Ref. [6], there are multiple ways to achieve this phase; two of these methods are machine learning and deep learning, which have recently been used. This step searches, trains and tests the output from feature engineering to classify or predict the fall with inputs from environment sensors, wearable sensors, and, in this work, from the multicamera vision-based approach.

In deep learning, CNNs have revolutionized the way computer vision problems are addressed due to the automatic discovery of structure representations in large datasets. This method has dramatically improved the state-of-the-art methods in image processing [25].

However, finding a suitable CNN architecture is difficult [25]. To address this difficulty, the literature has reported multiple network architectures, depending on the problem to be solved. For example, in recent years, many network structures for image recognition and classification problems have been reported, such as AlexNet [27], ClarifaiNet [28], GoogLeNet [29] and VGGNet [30]. All these networks have proved to be efficient in their own problem domains, and they can also be used as pretrained models so that users can reduce the amount of time needed to retrain them for another task. However, these architectures are complex and can possibly be improved.

In this work, we design a CNN with three convolutional layers and three 2D max-pooling layers for feature extraction and three fully connected layers for fall detection. To this end, we fixed all images to 38×51 pixels; three fully connected layers were fixed and chosen because, for fully connected layers, the fixed-size constraint comes from only the fully connected layers, which exist at a deeper stage of the network [71]. The CNN receives the magnitudes D calculated from U and V, which were converted to grayscale images with 38×51 pixels, representing the optical flow features extracted. Then, these images go to the input layer, which consists of 128 convolution filters with a kernel size of 3×3. The second convolutional layer has 128 filters and the same kernel size, and the third layer has 64 convolutional filters and the same kernel. This convolutional layer architecture was selected by cross-validation and using the F1-score metric, as shown in Table 2. After each convolutional layer, 2D max-pooling layers are employed to synthetize output convolutions. Then, these results are input to three fully connected layers, i.e., 64 rectified linear units (ReLU) in the first layer, 128 ReLU in the second, 254 ReLU in the third and, finally, a 2D SoftMax layer with one output. The latter is employed to perform fall detection using fall (1) and no-fall (0) classes. Fig. 3 shows the representation of the proposed CNN.

As described above, the UP-Fall Detection dataset is integrated by information from 17 subjects performing 11 different activities/falls, with three trials for each activity. To train the CNN, we divide data

Table 2
Cross-validation for the convolutional architecture layers.

$$loss(p,t) = -(t \cdot \log p + (1-t) \cdot \log(1-p)) \quad (2)$$

CNN Architecture	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)
64 64 64	95.40	86.28	83.76	97.54	95.40
64 64 128	95.27	88.26	80.34	98.03	84.11
64 64 256	94.90	83.67	83.57	96.99	83.62
64 128 64	94.62	82.36	83.27	96.71	82.81
64 128 128	94.66	86.49	77.83	97.76	81.94
64 128 256	95.15	85.74	82.35	97.46	84.14
64 256 64	94.32	85.86	76.00	97.69	80.63
64 256 128	94.92	86.45	79.91	97.69	83.05
64 256 256	94.90	91.18	74.48	98.67	81.98
128 64 64	95.17	86.21	82.11	97.58	84.11
128 64 128	94.80	96.02	97.89	78.02	96.95
128 64 256	94.79	97.07	96.74	84.18	96.91
128 128 64	95.64	96.91	97.95	83.08	97.43
128 128 128	95.44	96.19	98.49	78.87	97.33
128 128 256	95.05	97.88	96.22	88.70	97.04
128 256 64	94.28	96.32	96.92	79.91	96.62
128 256 128	94.51	97.00	96.47	83.82	96.74
128 256 256	95.19	96.84	97.48	82.78	97.16
256 64 64	94.81	96.16	97.76	78.81	96.95
256 64 128	94.26	96.25	96.97	79.54	96.61
256 64 256	94.38	96.34	97.03	80.03	96.68
256 128 64	94.75	96.19	97.64	79.05	96.91
256 128 128	94.72	97.63	96.08	87.36	96.85
256 128 256	94.40	96.66	96.71	81.86	96.68
256 256 64	94.10	96.31	96.71	79.91	96.51
256 256 128	94.57	96.36	97.24	80.09	96.80
256 256 256	94.09	96.19	96.83	79.18	96.51

collection into trials 1 and 2 for each activity, and we use the subject as the training set (67%) and trial 3 for each activity and subject as the testing set (33%). The training dataset included 42,000 grayscale images that were 38×51 in size, and the optical flow was used for preprocessing; the testing dataset included 21,000 grayscale images with the same preprocessing flow. We trained during 50 epochs using the Adam optimizer and binary cross-entropy loss function, as defined in (2), where p is the prediction of the network, and t is the ground truth.

5. Experimentation

To analyze our proposal, the following experiments were carried out: (i) experiments to test our CNN model and to compare it with classic machine learning methods, such as SVM, random forest (RF), MPL and KNN; (ii) experiments to compare monocular with multicamera vision-based fall detection system approaches; and (iii) tests of our proposal not only for detection but also for the classification of activities and falls using the multicamera vision-based approach.

In these experiments, we used training and testing datasets with information provided from two cameras. We used the information of one camera per model and then the information from both lateral and front viewpoint cameras at the same time [8]. For windowing, 1-s windows with 0.5-s overlaps were employed. The images were treated as grayscale, and the optical flow implementation was treated as feature extraction. We resized the images to 38×51 pixels, and we performed a benchmark comparison between the classic machine learning methods (i.e., SVM, MLP, RF and KNN) and the CNN depicted in Fig. 3.

These experiments aim to explore and compare the performance of a monocular vision-based approach with multicamera vision-based fall detection systems and to create a benchmark comparison of classic machine learning methods and CNN for fall detection using the latter approach.

To evaluate the performance of our work, we use the following five metrics: accuracy, sensitivity, specificity, precision, and F1-score, as given by (3)–(7), where *TP* refers to true positives, *TN* to true negatives, *FP* to false positives, and *FN* to false negatives [32].

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$precision = \frac{TP}{TN + FP} \quad (4)$$

$$sensitivity = \frac{TP}{TP + FP} \quad (5)$$

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

$$F_1 - score = 2 \cdot \frac{precision \cdot sensitivity}{precision + sensitivity} \quad (7)$$

All experiments were conducted in Python 3.7.3 using the sklearn3 framework for classic machine learning techniques and the keras4 framework for CNN, taking advantage of its GPUs management capabilities [31].

5.1. Results and discussion

The experimental results are described in this section. Then, a discussion based on the analysis is presented.

5.1.1. Fall detection using conventional machine learning models

First, we conducted an experiment using the optical flow-based features from both cameras at the same time (Cam1 and Cam2). We trained four conventional machine learning models, namely, SVM, RF, MLP and KNN, as described above. Table 3 shows the meta-parameter settings for these models. For this experiment, we built the models using 67% for training and 33% for testing data. Table 4 summarizes the performance results using the visual features extracted in 1-s windows with 0.5 s of overlap.

From Table 4, it can be observed that the conventional machine

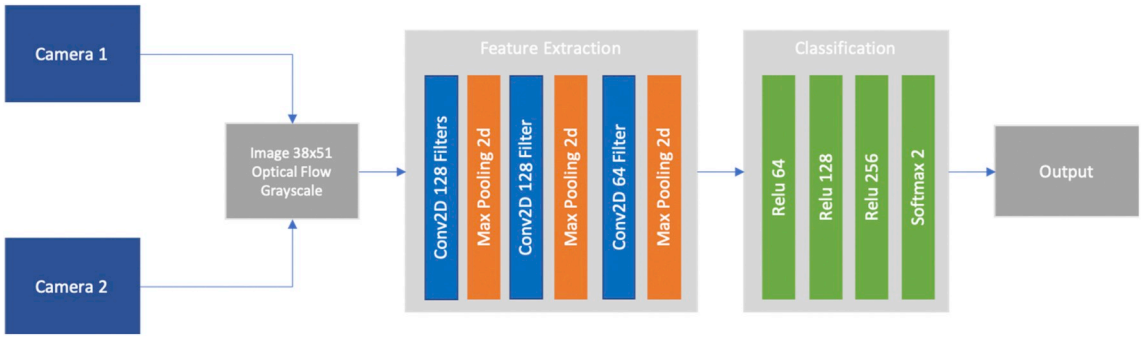

Fig. 3. Our proposal of the CNN architecture for the multicamera vision-based fall detection system.

learning models cannot predict human falls well in terms of accuracy, precision, sensitivity, specificity and F1-score. Currently, KNN seems to have the best performance based on the F1-score metric (15.27%). In terms of accuracy, SVM performs the best, with an accuracy of 32.40%. These machine learning models achieved an averaged accuracy of 29.77%. From the results, we might assume that the conventional machine learning methods using windowing and the sklearn library, as explained above, are not robust enough. To improve the performance, we implemented CNN, as described below.

5.1.2. Fall detection using CNN

In this experiment, we trained the following three CNN models: (i) a CNN model using visual features from the lateral view (Cam1), (ii) a CNN model using visual features from the front view (Cam2), and (iii) a CNN model using visual features from both cameras at the same time.

A summary of the results is shown in Table 5. As can be observed, the performances are very similar. The lateral view (Cam1) is slightly better than the frontal view, as expected [7]. However, Cam2 shows less specificity (79.67%) than Cam1 (81.58%), which could lead to misclassification. Furthermore, the combination of both views maintains the output performance of the lateral view, which is important because, if some of the cameras are occluded, it will remain feasible to detect the fall with just one camera, as supported in Ref. [7]. Furthermore, we performed an experiment using the VGG-16 CNN architecture with images from 2 cameras (frontal and lateral); the results shown in Table 5 indicate that our proposal has significantly better performance than the VGG-16 CNN architecture using UP-Fall. Thus, we conclude that our multicamera vision-based fall detection system has acceptable performance, in contrast with the conventional machine learning models and using the VGG-16 CNN architecture; additionally, our system avoids the occlusion problem as long as we do not lose sight of the subject.

We also compared our proposed method to other multiple-camera vision-based fall detection systems reported in the literature [15,35,37], considering that the latter were implemented using conventional machine learning methods.

For this comparison, we used the multicamera vision-based database, called the Multicam dataset [34]. This dataset includes 24 performances; 22 trials have at least one human fall, and the remaining two contain confounding events. Each performance was recorded from 8 different views. The same stage is used for all videos, with some

Table 3
Parameter settings used to train the classification models.

Classifier	Parameters
SVM	kernel = "radial basis function" kernel coefficient = 1 c = 1 shrinking = 1 tolerance = 0.001
RF	minimum samples split = 2 minimum samples leaf = 1 estimators = 2 bootstrap = 1
MLP	activation function = "reLU" hidden layers = 100 penalty parameter = 0.0001 batch size = min(200, num_samples) shuffle = 1 initial learning rate = 0.001 tolerance = 0.0001 exponential decay(first moment) = 0.9 exponential decay(second moment) = 0.999 regularization coefficient = 0.000000001 solver = "stochastic gradient" maximum epochs = 10
KNN	neighbors = 5 leaf size = 30 distance metric = "euclidean"

Table 4
Performance obtained by the classic ML models.

Model	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)
SVM	32.40	14.03	14.10	90.03	14.06
RF	29.30	14.45	14.30	91.26	14.37
MPL	30.08	9.05	11.03	93.65	9.94
KNN	27.30	16.32	14.35	90.96	15.27

furniture reallocation [34]. To train our proposal method, we selected two viewpoints (lateral and frontal views) from this dataset, and divide the data into training (67%) and testing (33%) sets. Table 6 summarizes the performance results in terms of sensitivity and specificity, as reported in the literature [33,35,37].

As shown in Table 6, our proposed method is competitive with state-of-the-art approaches, mainly in terms of sensitivity. In addition, our method can handle fall detection using two cameras, in contrast to the eight cameras utilized in the other approaches. Moreover, the network architecture of our proposal (Fig. 3) is very simple compared to those of other works. For example, Núñez-Marcos in Ref. [37] used a VGG-16 architecture modified to receive inputs, the authors in Ref. [33] used PCA to extract features and SVM for classification, and in Ref. [35], the authors presented a multivariate exponentially weighted moving average (MEWMA) and SVM with 2 steps for classification (see Table 6). In that sense, our system has good performance, considering that it requires much less time for training and the simplicity of its architecture.

5.1.3. Daily activities and fall classification using CNN

Finally, we conducted an experiment for daily activity and fall classification using our proposed method. In this case, each activity and type of fall recorded in the UP-Fall Detection dataset was considered, and so the CNN was converted into a multiclass classifier, as shown in Table 1.

We applied our proposed method using both cameras (Cam1 and Cam2) and the results, compared to the performance obtained in Ref. [24] using the same dataset, are depicted in Table 7. As shown, our proposed method is slightly inferior to the multimodal-based approach presented by Martínez-Villaseñor et al. [24], which is an expected result, since a multimodal approach (i.e., wearable sensors, EEG helmet and cameras) is better than a single modality approach such as ours. It is also important to note that the F1-scores of both approaches are similar, i.e., 72.94% for our proposal and 72.80% for the multimodal approach. From the results presented in Table 7, the performance obtained by our proposal can be considered competitive (e.g., similar F1-score), easier to implement (i.e., due to the number of sensors) and less obtrusive (i.e., wearable sensors) than the multimodal-based approach reported in Ref. [24].

6. Discussion

The proposed multicamera vision-based fall detection and classification system is comparable to state-of-the-art methods. The results suggest the predictive power of our proposed fall detection system (97.43% of F1-score), which outperforms conventional machine learning methods (SVM, RF, MLP and KNN) by using optical flow-based features and fewer cameras and achieves similar performance to the state-of-the-art methods reported in the literature (97.00% sensitivity and 80.00% specificity) and to a multimodal approach (72.80% F1-score).

The advantages of our proposal are as follows. Multicamera approaches offer robust solutions for fall recognition, even when occlusion occurs from one viewpoint, as long as one camera remains focused on the subject. This result can be observed in Table 5, which reports similar performance when using one camera or the other (lateral or frontal view) or both. Additionally, our proposal offers a simple CNN

Table 5
Performance of the CNN models using the lateral view, front view and both views.

Data	Method	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)
(Cam1) Lateral view	Proposed CNN	95.24	95.24	97.72	81.58	97.20
(Cam2) Frontal view	Proposed CNN	94.78	96.30	97.57	79.67	96.93
(Cam1 & Cam 2)	Proposed CNN	95.64	96.91	97.95	83.08	97.43
(Cam1 & Cam 2)	VGG-16 CNN	84.44	84.44	100	0	91.56

Table 6
Comparison between our proposal and state-of-the-art multicamera vision-based fall detection systems reported in the literature, using the Multicam dataset.

Proposal	Method	Sensitivity (%)	Specificity (%)	Cameras
Wang et al. [33]	SVM	89.20	90.30	8
Wang et al. [35]	SVM	93.70	92.00	8
Núñez et al. [37]	VGG-16 CNN	99.00	96.00	8
Ours (Combined)	CNN	97.95	83.08	2

Table 7
Comparison between our proposal and the multimodal approach reported using the UP-Fall dataset.

Data	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)
Ours	82.26	74.25	71.67	77.48	72.94
Martínez-Villaseñor et al. [24]	95.00	77.70	69.90	99.50	72.80

architecture (Fig. 3) and a low computational cost. Due to the vision-based nature of our approach, privacy must be discussed due to the nature of constant video surveillance. Our work avoids privacy concerns by analyzing only the relevant information about the fall using the optical flow information calculated from the video sequence. Therefore, the privacy of the person is not affected because the data used to recognize a fall do not contain personal information.

It is also important to consider some limitations of our proposed method. A vision-based approach always depends on the quality of the captured image, the position of the cameras, and the presence of the subject. In addition, privacy issues should be addressed before the implementation. If privacy issues are a limitation, as mentioned before, then the original images captured by the cameras should not be stored; they can be used only for extracting the optical flow features. However, pervasiveness remains an important drawback because cameras are continually acquiring videos of the subjects. In addition, the computational complexity in terms of memory and processing time should be emphasized, as this complexity hinders the scalability of a real-time fall detection system [8].

Regarding the fall and activity samples in the UP-Fall dataset, 42,958 training samples and 21,038 testing samples arranged in 1-s windows were employed in our experiments. The results were competitive with those of the state-of-the-art methods for both detection (Tables 5 and 6) and classification (Table 7) tasks. Furthermore, it is important to discuss the age of the subjects that performed the falls and activities when building the UP-Fall Detection dataset used in this work. This dataset was made using the information of 17 healthy subjects without impairments (9 males and 8 females) ranging from 18 to 24 years old. Nevertheless, in Ref. [73], it is shown that testing with a dataset built using young people does not significantly deviate from that with a dataset built using elderly people. Thus, we believe that our approach can be applied in real situations, which should be considered in future work.

In vision-based problems, the positioning of the cameras is difficult to determine in terms of the angle, height and distance between the cameras and the subject that performs the activities or falls. The UP-Fall dataset was made by recording falls and activities with fixed distances and angles and recording falls in the same direction. These aspects of the dataset could change under realistic conditions, making our approach difficult to replicate in the real world. However, in our work, we address this problem by using only the apparent movement in the image by applying the optical flow as the feature extraction method and then calculating the Euclidean distance as the apparent movement in each 1-s window. This approach helps us to precisely detect a fall, even when the subject is not positioned at the center of the images. Thus, this proposal might also be considered if the distances and angles between the cameras and the subject are different from the specifications of this dataset. As our CNN architecture receives only apparent movements as input, the angle, height and distance condition results are irrelevant to an inference of the detection only if these values change slightly. Otherwise, retraining is required. In future work, transfer learning will be applied using our proposed CNN model to analyze changes in camera positions and conditions.

The experimental results showed that our proposal is competitive with state-of-the-art multicamera vision-based approaches for detection systems and for classification (Table 7), even compared to a multimodal approach, such as that reported in Ref. [24].

7. Conclusions

In this paper, we presented a multicamera vision-based fall detection and classification system that takes advantage of CNN. In addition, we combined the CNN models with visual features extracted from sequences of images using the optical flow method. In this work, we used the UP-Fall Detection dataset as a case study. We conducted experiments to compare our proposal with conventional machine learning models, analyzed the performance of our proposal for vision-based approaches with single and multiple cameras, and extended our model for fall classification.

From the experimental results, we conclude that our proposed multicamera vision-based fall detection and classification system outperforms conventional machine learning methods, reduces computation time due to the simple CNN architecture, and is competitive with state-of-the-art and multimodal-based approaches.

Future works should implement this approach in a real-world assisted living system and analyze and propose improvements to issues related to privacy, pervasiveness, changes in environmental conditions and occlusion. In addition, we will consider testing our system in a real situation by including the transfer learning approach and different camera positions.

Declaration of competing interest

The authors have nothing to declare.

Acknowledgments

This research was funded by Universidad Panamericana through the grant “Fomento a la Investigación UP 2018”, under project code UP-CI-2018-ING-MX-04.

References

[1] Department of Health and Human Services, Fatalities and Injuries from Falls Among Older Adults - United States, 1993-2003 and 2001- 2005, November 2006, 12211224. Morbidity and Mortality Weekly Re- port.

[2] M. Schneider, Introduction to Public Health, Jones and Bartlett, Sudbury, MA, 2011.

[3] Lord, S. R., Sherrington, C., Menz, H. B., & Close, J. C. (n.d.). Strategies for Prevention. Falls in Older People,173-176. doi:10.1017/cbo9780511722233.011.

[4] T.W. Oneill, J. Varlow, A.J. Silman, J. Reeve, D.M. Reid, C. Todd, A.D. Woolf, Age and sex influences on fall characteristics, Ann. Rheum. Dis. 53 (11) (1994) 773–775, https://doi.org/10.1136/ard.53.11.773.

[5] A. Bourke, G. Lyons, A threshold-based fall-detection algorithm using a bi-axial gyroscope sensor, Med. Eng. Phys. 30 (1) (2008) 84–90, https://doi.org/10.1016/j.medengphy.2006.12.001.

[6] N. Noury, A. Fleury, P. Rumeau, A. Bourke, G.O. Laighin, V. Ri- alle, J. Lundy, Fall detection – principles and methods, in: 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2007, https://doi.org/10.1109/iembs.2007.4352627.

[7] C. Rougier, J. Meunier, A. St-Arnaud, J. Rousseau, Robust video surveillance for fall detection based on human shape deformation, IEEE Trans. Circuits Syst. Video Technol. 21 (5) (2011) 611–622, https://doi.org/10.1109/tcsvt.2011.2129370.

[8] O.D. Lara, M.A. Labrador, A survey on human activity recognition using wearable sensors, IEEE Commun. Surv. Tutor. 15 (2013), 11921209.

[9] J. Yin, Q. Yang, J. Pan, Sensor-based abnormal human- activity detection, IEEE Trans. Knowl. Data Eng. 20 (8) (2008) 1082–1090, https://doi.org/10.1109/tkde.2007.1042.

[10] X. Xu, J. Tang, X. Zhang, X. Liu, H. Zhang, Y. Qiu, Exploring techniques for vision based human activity recognition: methods, systems, and evaluation, Sensors 13 (2) (2013) 1635–1650, https://doi.org/10.3390/s130201635.

[11] T. Dungkaew, J. Suksawatchon, U. Suksawatchon, Impersonal smartphone-based activity recognition using the accelerometer sensory data, in: 2017 2nd International Conference on Information Technology (INCIT), 2017, https://doi.org/10.1109/incit.2017.8257856.

[12] P. Bharti, Complex activity recognition with multi-modal multi- positional body sensing, J. Biometrics Biostat. 08 (05) (2017), https://doi.org/10.4172/2155-6180-c1-005.

[13] G. Chetty, M. White, M. Singh, A. Mishra, Multimodal activity recognition based on automatic feature discovery, in: 2014 Inter- National Conference on Computing for Sustainable Global Development (INDIACom), 2014, https://doi.org/10.1109/indiacom.2014.6828039.

[14] P. Turaga, R. Chellappa, V.S. Subrahmanian, O. Udrea, Machine recognition of human activities: a survey, IEEE Trans. Circuits Syst. Video Technol. 18 (2008), 14731488.

[15] T.D. Raty, Survey on contemporary remote surveillance systems for public safety, IEEE Trans. Syst. Man Cybern. C Appl. Rev. 40 (2010), 493515.

[16] M. Albanese, R. Chellappa, V. Moscato, A. Picariello, V.S. Subrahmanian, P. Turaga, O. Udrea, A constrained probabilistic petri net framework for human activity detection in video, IEEE Trans. Multimed. 10 (2008), 14291443.

[17] N. Zerrouki, A. Houacine, Combined curvelets and hidden Markov models for human fall detection, Multimed. Tools Appl. 77 (5) (2017) 6405–6424, https://doi.org/10.1007/s11042-017-4549-5.

[18] E. Auvinet, F. Multon, A. Saint-Arnaud, J. Rousseau, J. Meunier, Fall detection with multiple cameras: an occlusion resistant method based on 3-D silhouette vertical distribution, IEEE Trans. Inf. Technol. Biomed. 15 (2) (2011), 290300.

[19] A. Nez-Marcos, G. Azkune, I. Arganda-Carreras, Vision-based fall detection with convolutional neural networks, Wirel. Commun. Mob. Comput. 2017 (2017), https://doi.org/10.1155/2017/9474806.

[21] B. Kwolek, M. Kepski, Human fall detection on embedded platform using depth maps and wireless accelerometer, Comput. Methods Progr. Biomed. 117 (3) (2014), 489501.

[22] N. Thome, S. Miguet, S. Ambellouis, A real-time, multi- view fall detection system: a LHMM-based approach, IEEE Trans. Circuits Syst. Video Technol. 18 (11) (2008) 1522–1532, https://doi.org/10.1109/tcsvt.2008.2005606.

[23] D. Anderson, R.H. Luke, J.M. Keller, M. Skubic, M. Rantz, M. Aud, Linguistic summarization of video for fall detection using voxel person and fuzzy logic, Comput. Vis. Image Understand. 113 (1) (2009) 80–89, https://doi.org/10.1016/j.cviu.2008.07.006.

[24] L. Martnez-Villaseor, H. Ponce, J. Brieva, E. Moya-Albor, J. Nez- Martnez, C. Peafort-Asturiano, UP-fall detection dataset: a multimodal approach, Sensors 19 (9) (2019) 1988, https://doi.org/10.3390/s19091988.

[25] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015), 436444.

[26] S.S. Beauchemin, J.L. Barron, The computation of optical flow, ACM Comput. Surv. 27 (3) (1995), 433466.

[27] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Proceedings of the 26th Annual Conference on Neural Information Processing Systems (NIPS 12), Lake Tahoe, Nev, USA, December 2012, 10971105.

[28] M.D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks, in: ECCV, 2014, 818833.

[29] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, Clin. Orthop. Relat. Res. 1 (2) (2014) 3, abs/1409.1556.

[30] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, Clin. Orthop. Relat. Res. 1 (2014) 2, abs/1409.4842.

[31] C. Francois, et al., Keras (2015). https://github.com/fchollet/keras.

[32] M. Sokolova, G. Lapalme, A systematic analysis of performance measures for classification tasks, Inf. Process. Manag. 45 (4) (2009) 427–437, https://doi.org/10.1016/j.ipm.2009.03.002.

[33] S. Wang, L. Chen, Z. Zhou, X. Sun, J. Dong, Human fall detection in surveillance video based on PCANet, Multimed. Tools Appl. 75 (19) (2015) 11603–11613, https://doi.org/10.1007/s11042-015-2698-y.

[34] Edouard Auvinet, Caroline Rougier, Jean Meunier, Alain St-Arnaud, Jacqueline Rousseau, Multiple cameras fall dataset. DIRO-Universit de Montral, Tech. Rep. 1350 (2010).

[35] I. Charfi, J. Miteran, J. Dubois, M. Atri, R. Tourki, Definition and performance evaluation of a robust SVM based fall detection solution, SITIS 12 (2012), 218224.

[36] S. Kozina, H. Gjoreski, M. Gams, M. Lutrek, Efficient Activity Recognition and Fall Detection Using Accelerometers. Communications in Computer and Information Science Evaluating AAL Systems through Competitive Benchmarking, 2013, pp. 13–23, https://doi.org/10.1007/978-3-642-41043-72.

[37] K. Wang, G. Cao, D. Meng, W. Chen, andW. Cao, Automatic fall detection of human in video using combination of features, in: Proceedings of the 2016 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2016, December 2016, 12281233. China.

[39] D. Anguita, A. Ghio, L. Oneto, X. Parra, J.L. Reyes-Ortiz, Training computationally efficient smartphone-based human activity recognition models, Lect. Notes Comput. Sci. 8131 (2013) 426433. LNCS.

[40] S. Münzner, P. Schmidt, A. Reiss, M. Hanselmann, R. Stiefelhagen, R. Dürucheb, CNN-based sensor fusion techniques for multimodal human activity recognition, in: Proceedings of the 2017 ACM International Symposium on Wearable Computers - ISWC 17, 2017, https://doi.org/10.1145/3123021.3123046.

[41] L.C. Jatoba, U. Grossmann, C. Kunze, J. Ottenbacher, W. Stork, Context-aware mobile health monitoring: evaluation of different pat- tern recognition methods for classification of physical activity, in: 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2008, 52505253.

[42] A. Jalal, S. Kamal, D. Kim, A depth video sensor- based life-logging human activity recognition system for elderly care in smart indoor environments, Sensors 14 (7) (2014) 11735–11759, https://doi.org/10.3390/s140711735.

[43] C. Torres-Huitzil, M. Nuno-Maganda, Robust smartphone- based human activity recognition using a tri-axial accelerometer, in: 2015 IEEE 6th Latin American Symposium on Circuits & Systems (LAS- CAS), 2015, https://doi.org/10.1109/lascas.2015.7250435.

[44] T. Vilarinho, B. Farshchian, D.G. Bajer, O.H. Dahl, I. Egge, S.S. Hegdal, A. Lnes, J. N. Slettevold, S.M. Weggersen, A combined smart- phone and smartwatch fall detection system, in: Proceedings of the 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Auto- Nomic and Secure Computing, Pervasive Intelligence and Computing, Liverpool, UK, October 2015, 14431448, 2628.

[46] O. Kerdjidj, N. Ramzan, K. Ghanem, A. Amira, F. Chouireb, Fall detection and human activity classification using wearable sensors and compressed sensing, J. Ambient Intell. Humanized Comput. (2019), https://doi.org/10.1007/s12652-019-01214-4.

[47] S. Khojasteh, J. Villar, C. Chira, V. Gonzlez, E.D. Cal, Improving fall detection using an on-wrist wearable accelerometer, Sensors 18 (5) (2018) 1350, https://doi.org/10.3390/s18051350.

[48] M. Bortnikov, A. Khan, A.M. Khattak, M. Ahmad, Accident recognition via 3D CNNs for automated traffic monitoring in smart cities, Adv. Intell. Syst. Comput. Adv. Comput. Vis. (2019) 256–264, https://doi.org/10.1007/978-3-030-17798-022.

[49] A. Esteva, B. Kuprel, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau, S. Thrun, Dermatologist-level classification of skin cancer with deep neural networks, Nature 542 (7639) (2017) 115–118, https://doi.org/10.1038/nature21056.

[50] A.H. Fakhrulddin, X. Fei, H. Li, Convolutional neural networks (cnn) based human fall detection on body sensor networks (bsn) sensor data, in: 2017 4th ICSAI, Nov 2017.

[51] A. Nait Aicha, G. Englebienne, K.S. van Schooten, M. Pijnappels, B. Krse, Deep learning to predict falls in older adults based on daily-life trunk accelerometry, Sensors 18 (5) (2018) 114, https://doi.org/10.3390/s18051654 (Basel, Switzerland).

[52] N. Lu, Y. Wu, L. Feng, J. Song, Deep learning for fall detection: three-dimensional CNN combined with LSTM on video kinematic data, IEEE J. Biomed. Health Inf. 23 (1) (2019) 314–323, https://doi.org/10.1109/jbhi.2018.2808281.

[53] E. Casilari, J. Santoyo-Ramn, J. Cano-Garca, Analysis of public datasets for wearable fall detection systems, Sensors 17 (7) (2017) 1513, https://doi.org/10.3390/s17071513.

[54] W. Shieh, J. Huang, Falling-incident detection and throughput enhancement in a multi-camera video- surveillance system, Med. Eng. Phys. 34 (7) (2012) 954–963, https://doi.org/10.1016/j.medengphy.2011.10.016.

[55] M.A. Mousse, C. Motamed, E.C. Ezin, Percentage of human-occupied areas for fall detection from two views, Vis. Comput. 33 (12) (2016) 1529–1540, https://doi.org/10.1007/s00371-016-1296-y.

[56] S. Zhang, Z. Li, Z. Wei, S. Wang, An automatic human fall detection approach using RGBD cameras, in: 2016 5th International Conference on Computer Science and Network Technology (ICCSNT), 2016, https://doi.org/10.1109/iccsnt.2016.8070265.

[57] M. Hekmat, Z. Mousavi, H. Aghajan, Multi-view feature fusion for activity classification, in: Proceedings of the 10th International Conference on Distributed Smart Camera - ICDSC 16, 2016, https://doi.org/10.1145/2967413.2967434.

[58] S. Su, S. Wu, S. Chen, D. Duh, S. Li, Multi-view fall detection based on spatio-temporal interest points, Multimed. Tools Appl. 75 (14) (2015) 8469–8492, https://doi.org/10.1007/s11042-015-2766-3.

[59] Y. Kong, J. Huang, S. Huang, Z. Wei, S. Wang, Learning spatiotemporal representations for human fall detection in surveillance video, J. Vis. Commun. Image Represent. 59 (2019) 215–230, https://doi.org/10.1016/j.jvcir.2019.01.024.
[60] G. Koshmak, A. Loutfi, M. Linden, Challenges and issues in multisensor fusion approach for fall detection: review paper, J. Sens. 2016 (2016) 1–12, https://doi.org/10.1155/2016/6931789.
[61] Y. Wu, Y. Su, Y. Hu, N. Yu, R. Feng, A multi-sensor fall detection system based on multivariate statistical process analysis, J. Med. Biol. Eng. 39 (3) (2019), 336351, https://doi.org/10.1007/s40846-018-0404-z.
[62] Y. Wu, Y. Su, Y. Hu, N. Yu, R. Feng, A multi-sensor fall detection system based on multivariate statistical process analysis, J. Med. Biol. Eng. 39 (3) (2019), 336351, https://doi.org/10.1007/s40846-018-0404-z.
[63] M. Mubashir, L. Shao, L. Seed, A survey on fall detection: principles and approaches, Neurocomputing 100 (2013), 144152, https://doi.org/10.1016/j.neucom.2011.09.037.
[65] A. Mao, X. Ma, Y. He, J. Luo, Highly portable, sensor- based system for human fall monitoring, Sensors 17 (9) (2017), https://doi.org/10.3390/s17092096.
[66] Z. Zhang, C. Conly, V. Athitsos, Evaluating depth-based computer vision methods for fall detection under occlusions, 196207, , 2014.
[67] Z. Zhang, C. Conly, V. Athitsos, A survey on vision-based fall detection, in: Proceedings of the 8th ACM International Conference on Pervasive Technologies Related to Assistive Environments, ACM, 2015, 2015.
[68] O. Banos, J.-M. Galvez, M. Damas, H. Pomares, I. Rojas, Window size impact in human activity recognition, Sensors 14 (4) (2014), 64746499, https://doi.org/10.3390/s140406474.
[69] S. Khalid, T. Khalil, S. Nasreen, A survey of feature selection and feature extraction techniques in machine learning, in: Proceedings of the Science and Information Conference (SAI), London, UK, August 2014, 2729.
[70] Y.-Z. Hsieh, Y.-L. Jeng, Development of home intelligent fall detection IoT system based on feedback optical flow convolutional neural network, IEEE Access 6 (2018), 60486057, https://doi.org/10.1109/ac-cess.2017.2771389.
[71] K. He, X. Zhang, S. Ren, J. Sun, Spatial pyramid pooling in deep convolutional networks for visual recognition, in: Proc. 13th Eur. Conf. Comput. Vis., 2014, p. 346361.
[72] N.V.A. Akula, A.K. Shah, R. Ghosh, A spatio-temporal deep learning approach for human action recognition in infrared videos, Opt. Photon. Inf. Proc. XII (2018), https://doi.org/10.1117/12.2502993.
[73] A. Sucerquia, J.D. Lpez, F. Vargas-Bonilla, Real- Life/Real-Time Elderly Fall Detection with a Triaxial Accelerometer, 2018, https://doi.org/10.20944/preprints201711.0087.v3.