

Evaluation of convolutional neural networks for the classification of falls from heterogeneous thermal vision sensors

International Journal of Distributed
Sensor Networks
2020, Vol. 16(5)
© The Author(s) 2020
DOI: 10.1177/1550147720920485
journals.sagepub.com/home/dsn


Miguel Ángel López-Medina¹, Macarena Espinilla² , Chris Nugent³
and Javier Medina Quero²

Abstract

The automatic detection of falls within environments where sensors are deployed has attracted considerable research interest due to the prevalence and impact of falling people, especially the elderly. In this work, we analyze the capabilities of non-invasive thermal vision sensors to detect falls using several architectures of convolutional neural networks. First, we integrate two thermal vision sensors with different capabilities: (1) low resolution with a wide viewing angle and (2) high resolution with a central viewing angle. Second, we include fuzzy representation of thermal information. Third, we enable the generation of a large data set from a set of few images using ad hoc data augmentation, which increases the original data set size, generating new synthetic images. Fourth, we define three types of convolutional neural networks which are adapted for each thermal vision sensor in order to evaluate the impact of the architecture on fall detection performance. The results show encouraging performance in single-occupancy contexts. In multiple occupancy, the low-resolution thermal vision sensor with a wide viewing angle obtains better performance and reduction of learning time, in comparison with the high-resolution thermal vision sensors with a central viewing angle.

Keywords

Thermal vision sensors, fall detection, convolutional neural networks, fuzzy processing

Date received: 29 March 2019; accepted: 27 March 2020

Handling Editor: Joseph Rafferty

Introduction

The most recent studies of the World Bank estimated that the number of elderly people is increasing and expected to double again by 2050 worldwide. As the average age of the population continues to rise, elderly people are continuing to suffer from certain chronic diseases like dementia, hypertension, diabetes, gait issues.^{1,2}

Fall detection is a major challenge in the area of public health care, especially for the elderly, and reliable surveillance is a necessity to mitigate the effects of falls.³ In this context, an alarming 42% of people aged 70 and above are involved in falls annually, with 37.3 million

of those requiring medical attention as a result of their severity.^{1,4}

Accidental falls experienced by elderly people are a prominent cause of hospitalization, death due to the injuries sustained, and reduced independence. Several

¹Council of Health for the Andalusian Health Service, Sevilla, Spain

²Department of Computer Science, University of Jaen, Campus Las Lagunillas, Jaen, Spain

³School of Computing, Ulster University, Northern Ireland, UK

Corresponding author:

Miguel Ángel López-Medina, Council of Health for the Andalusian Health Service, Av. de la Constitución 18, 41071 Sevilla, Spain.
Email: malopezm@ephag.es



risk factors exist in relation to falls in older adults, studies mainly identifying physical frailty, poor balance, unsteady pace, poor muscle strength, and cognitive impairment. The prevalence of falls also escalates as age increases, particularly when combined with other risk factors such as chronic disease, poor sleep patterns, and diminished vision.⁵

Ambient assisted living (AAL) is becoming an important consideration to provide assistive technologies aimed at sustaining independence, well-being, and quality of life.⁶ So, there has been a growing need to promote and support “aging in place” due to demographic issues, increasing health care costs, a shortage of caregivers, and the fundamental fact that a large portion of elderly people prefer to remain independent in their own homes for as long as possible.⁷

These issues open up new research avenues for tracking activities related to elderly people’s daily routine, specifically with the aim of guaranteeing their safety. Over the last decade, interest in ubiquitous computing technologies has provided researchers with enough opportunities to design monitoring and intervention systems, which could provide continuous 24/7 real-time monitoring in environments with sensors, with the goal of improving the quality of life of elderly people.⁸

In order to evaluate the proposed methodology, a case study is presented to evaluate the methodology for fall detection using data collected by two different thermal vision sensors (TVSs) and multiple convolutional neural networks (CNNs) in two different smart labs: the smart lab of Ulster University⁹ and the smart lab of the University of Jaén.¹⁰ Moreover, the data set design for data collection includes single-occupancy as well as multi-occupancy scenes.

The article is structured as follows: in this section, we have provided a review of related research in the fields of TVSs and fall detection. The methodology for the evaluation of CNNs to classify the shapes of falls from heterogeneous TVSs is presented in “Methodology” section. The experimental setup of the case study and a discussion of the results are presented in “Experimental setup” section. Finally, in “Conclusions and ongoing works” section, conclusions and ongoing works are discussed.

Related works

The automatic detection of falls within AAL scenarios has attracted considerable research interest due to the prevalence and impact of falls in the elderly, being a crucial research area.³ Impact-related accidents in indoor environments such as falls and collisions have been identified and studied in an attempt to avoid falls or reduce aid response time.¹¹ Fall detection approaches in AAL scenarios are divided into two

categories: wearable/ambience sensors and vision sensors.^{3,12}

In approaches based on wearable/ambience sensors, sensors are attached to an inhabitant under observation—namely, wearable sensors or smart phones, or objects that make up the environment where the activity takes place—namely, dense sensing. These approaches work with time series of state changes and/or various parameter values that are usually processed through data fusion, probabilistic, or statistical analysis methods and formal knowledge technologies.¹³ The main benefit of the wearable or ambience sensor is its cost efficiency. However, two main disadvantages of this kind of sensor are intrusiveness and fixed relative relations with the object or the inhabitant that can be easily disconnected. Furthermore, installation and setup can be complex. For these reasons, this kind of device is not a very good choice for the elderly.³

Approaches based on vision sensors exploit computer vision techniques like feature extraction, structural modeling, movement segmentation, action extraction, and movement tracking to analyze visual observations for pattern recognition.¹³ In recent years, the number of approaches in this category has increased due to the fact that video cameras are commonly included in the wearable technologies or systems we use daily.³ Previously, general vision sensors entailed disadvantages concerning privacy and ethics. After the emergence of TVSs, these disadvantages can be mitigated, being an excellent alternative to find solutions for the elderly.

Exploring state-of-the-art fall detection systems, we found recent studies within vision sensor-based approaches. In the proposal presented in Bromiley et al.,¹⁴ the image stream from the thermal detector is monitored. To do so, extracted features include horizontal and vertical gradients, aspect ratio, and centroid angle to horizontal axis of the bounding box. Falls were confirmed when the angle reached a value below 45°. A fall detection system was proposed in De Miguel et al.¹⁵ based on a low-cost device comprising an embedded computer and camera, executed in a low-cost device such as Raspberry Pi, obtaining good performance values (i.e. 96% sensitivity), comparable to other systems using more expensive and more powerful hardware. An approach for unobtrusive indoor fall detection by an infrared (IR) thermal array sensor was proposed in Hayashida et al.¹⁶ The main innovation of this proposal was to perform the fall detection within the sensor node by a computationally inexpensive algorithm which notifies the server only when a fall has occurred. A method was proposed in Rougier et al.¹⁷ to detect falls by analyzing human shape deformation during a video sequence. A shape matching technique was used to track the person’s silhouette along the video sequence. The shape deformation is quantified from

these silhouettes based on shape analysis methods. In Asbjorn and Jim,¹⁸ data collected from a ceiling-mounted 80×60 thermal array were combined with an ultrasonic sensor device. This approach monitored activities, recognizing the location and posture of an individual. In Taramasco et al.,¹⁹ a non-invasive monitoring system for fall detection in older people was presented by using very low-resolution thermal sensors for classifying a fall and then alerting the care staff. Furthermore, the authors analyzed the performance of three recurrent neural networks for fall detection: long short-term memory (LSTM), gated recurrent unit, and bi-LSTM. Finally, a methodology based on CNNs to detect falls from non-invasive TVSs was presented in Medina-Quero et al.²⁰ with data augmentation techniques. The results show encouraging performance in single-occupancy contexts, with up to 92% accuracy, but a 10% reduction in accuracy in multiple-occupancy contexts.

Another work related to our proposal but without the application to fall detection was presented in Bayareh et al.,²¹ studying the diabetic foot by means of a Raspberry Pi as an embedded system and the Lepton-Flir Development Kit as an IR sensor. The IR sensor was characterized to measure the superficial temperature of the human skin radiometrically.

Most of the proposed vision-based approaches lack flexibility due to the fact that these approaches are often case-specific, depending on different scenarios and TVSs.

In this article, we present a methodology to analyze the capabilities of non-invasive TVSs²² to detect falls by means of several architectures of CNNs in different scenarios. We propose the use of the CNNs because they have provided excellent results in multiple areas such as speech recognition,²³ image classification,²⁴ or gas classification.²⁵

The learning process with CNNs requires a large amount of data.²⁶ Therefore, it is necessary to collect multiple images from different inhabitants, orientations, and cases, which takes a great effort. This process could make customization and configuration in different contexts hugely difficult. This disadvantage can be overcome by data augmentation to enlarge the number of learning cases from a limited set²⁷ and therefore reduce over-fitting.²⁸

Similar approaches have been proposed in recent works,^{29,30} where the selection of images from objects in a small number of human-annotated examples is then projected in the environmental background to provide new synthetic examples, as well as in thermal vision data sets.²⁰

In our proposal, the two studied TVSs have different capabilities. The first TVS has low resolution with a wide viewing angle and the second one has high resolution and a central viewing angle. Three types of CNN

are adapted for each TVS in order to evaluate the impact of the architecture on fall detection performance. Furthermore, a large data set is generated from a set of few images as a data source, by using ad hoc data augmentation, that is, increasing the original data set size by generating new synthetic images.

Finally, we propose to include fuzzy representation of thermal information to compute the fuzzy color of human temperature.³¹ The aim of including fuzzy processing of TVS data provides (1) a filter for irrelevant information, (2) reduction of noise from non-feasible values,³² (3) scaling and focusing the relevant data range for the CNN kernels during the learning process. The use of a fuzzy approach has been demonstrated as a successful tool to reduce uncertainty in multiple applications.^{33–36}

Methodology

In this section, we describe the methodology applied. First, in “TVSs for analyzing fall detection” section, we describe the TVSs evaluated in this work. Second, in “Fuzzy representation of thermal information” section, we define a fuzzy representation of thermal information to improve the performance of the fall detection. Third, in “Data augmentation” section, we detail an ad hoc data augmentation for fall detection in the previous learning stage. Fourth, in “Design of the CNN” section, we describe several configurations of CNNs evaluated for each TVS.

TVSs for analyzing fall detection

In this work, we have integrated two TVSs with different capabilities to evaluate their performance in analyzing fall detection:

- Low resolution with a wide viewing angle: in this case, we deployed the TVS Heimann HTPA 32×31 ,³⁷ which provides thermal vision with a 32×31 matrix, where each value defines a heat point of temperature. An effective factory calibration is integrated in the device, with no distortion by the fish-eye lens.³⁸ The data are collected from the TVS by means of a twisted Ethernet cable which is connected to the local area network. The middleware SensorCentral³⁹ integrates the TVS as a sensor source, providing the thermal sensor data within a Web Service in JSON format.
- High resolution with a central viewing angle: in this case, we deployed the Lepton LWIR module included in *FLiR Dev Kit*,⁴⁰ which provides thermal resolution with an 80×60 matrix. In addition, a Raspberry PI⁴¹ was used in order to

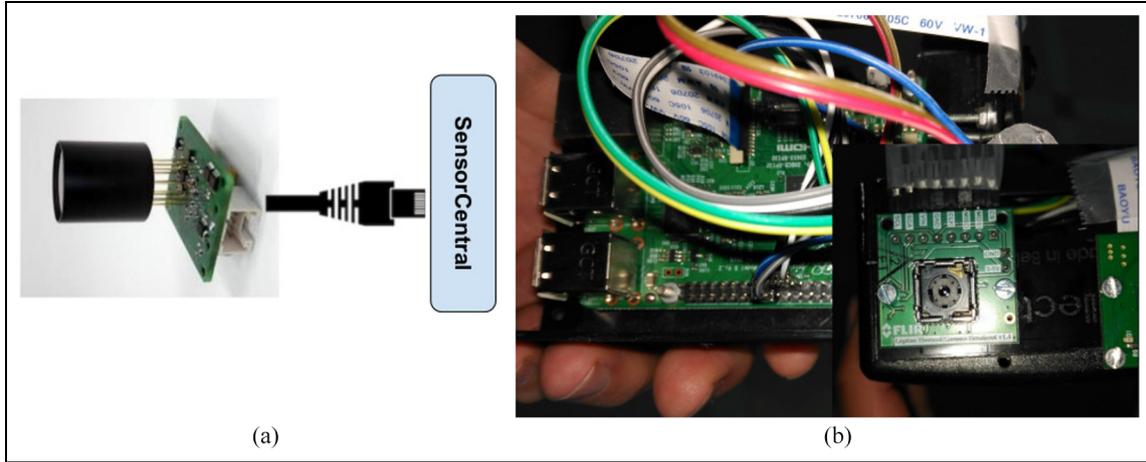


Figure 1. The thermal vision sensors for analyzing fall detection evaluated in this work: (a) Heimann HTPA and sensor central provides a thermal sensor with low resolution with wide viewing angle, (b) FLiR Dev Kit and Raspberry Pi provides a thermal sensor with high resolution and central viewing angle.

collect the information from the TVS⁴² in real time.

In a formal definition, each TVS provides a matrix $M^{w,h}$ which is formed by an array of numbers $m_{i,j}$ whose value represents a heat point of temperature. The dimensions of the matrix are defined by weight w and height h .

In Figure 1, we provide some figures on the sensors deployed and evaluated in this work.

Fuzzy representation of thermal information

The data collected by the TVS represent the heat temperature in a matrix of points. In order to provide a visual representation, a transformation function to gray scale values is required. In this work, we propose to define a fuzzy set to represent a fuzzy color⁴³ of *human temperature* by means of a membership function $\mu_M(m_{i,j})$, which relates the temperature values $m_{i,j}$ to a degree of relevance between 0 and 1

$$\mu_M(m_{i,j}) : R \rightarrow [0, 1]$$

In order to describe the fuzzy set straightforwardly, the shape of the membership function is given by a trapezoidal function which is defined as a lower limit l_1 , an upper limit l_4 , a lower support limit l_2 , and an upper support limit l_3 (see TS in the “Abbreviations” in the appendix)

$$\mu_M(m_{i,j}) = TS([l_1, l_2, l_3, l_4])$$

The aim of including fuzzy data processing from TVSs provides (1) a filter for non-relevant information, (2) the reduction of noise from non-feasible values,³²

(3) scaling and focusing the relevant data range for the CNN kernels during the learning process. In Figure 2, we show an example of the application of fuzzy representation.

Data augmentation

In this section, we propose the augmentation and enlargement of the image data from the original data set by means of image transformations. Thus, the innovation of our proposal is based on the creation of a new larger set of synthetic images to train the model. In this work, we have included the following image transformations—translation, rotation, and scale—to augment the original image data set:

- *Translation*: the original image is relocated within a maximal window size $[t_x, t_y]^+$ by using a random process, which generates a random translation transformation $[t_x, t_y]$, $t_x \in [0, t_x^+]$, $t_y \in [0, t_y^+]$.
- *Rotation scale*: the rotations are provided by two methods. First, the translated image is flipped horizontally and vertically by using a random process, which applies the transformation to a percentage of cases, defined by wH , wR respectively. Second, a rotation and scale transformation is defined by a maximal rotation angle α^+ and a scale factor s^+ , which generates a random rotation with an angle $\alpha \in [0, \alpha^+]$ and a random scale $s \in [1 - s^+, 1 + s^+]$. These transformations are then applied in the center of the image. We note that this rotation overcomes the original image size, for which reason a random scale of the image is provided.

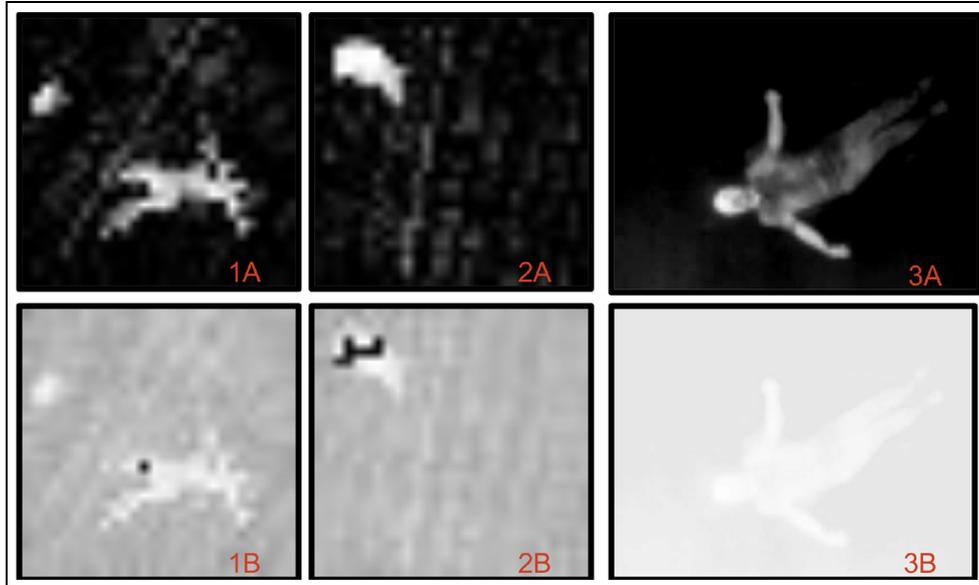


Figure 2. Images 1 and 2 show sample data from a low-resolution TVS. Image 3 shows sample data from a high-resolution TVS. Category B describes the raw TVS image data, and Category A shows the same data with fuzzy representation.

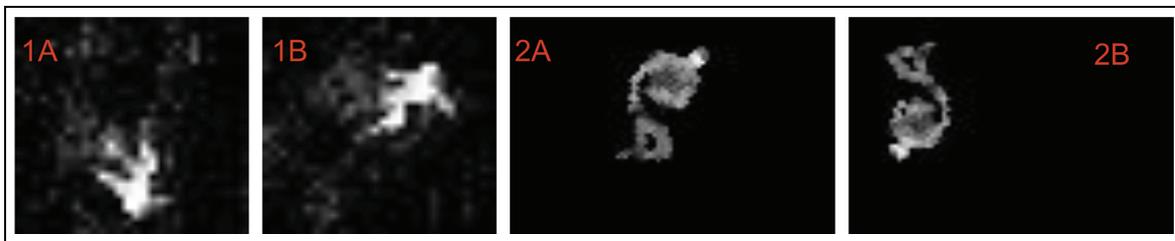


Figure 3. Images 2A and 2B show augmentation from a high-resolution TVS, and images 1A and 1B from a low-resolution TVS.

An example of new synthetic images is shown in detail in Figure 3 in order to extend the data set.

Design of the CNN

In this section, we describe several CNN architectures to classify the falls sustained by inhabitants. The two TVS devices show wide differences regarding technical characteristics and development purposes. For this reason, they are integrated within systems with different computing performance.

Regarding the low-resolution TVS, in our case a Heimann HTPA, the thermal sensor collects a smaller sized matrix of heat points which can be integrated in low-cost boards with low computing performance. For this purpose, three configurations of CNNs to classify fall detection with this kind of device are evaluated:

- CNN_2^0 : a CNN with two-kernel layers and optimized configuration for MNIST data set.⁴⁴

- CNN_2^+ : a CNN with two-kernel layers and a finer granularity configuration of kernels.
- CNN_3 : a CNN with three-kernel layers.

These three CNN configurations have been previously identified as suitable structures for fall detection,²⁰ and their details are shown in Table 1.

Regarding the high-resolution TVS, in our case, an FLiR Dev Kit and a Raspberry Pi, the matrix of heat points is wider in size, requiring deeper CNN configurations to classify fall detection. In this work, we propose three CNN configurations:

- CNN_4 : a CNN with four-kernel layers and a deeper configuration than the previous ones (see Table 2).
- $Alex_5$: based on the configuration of AlexNet, which is a five-layer CNN for large and deep CNNs with high performance in image classification.²⁸

Table 1. Configurations of convolutional neural networks for low-resolution thermal vision sensors. $[N, N] \times M$ is the convolution of window dimensions whose size is $(N \times N)$ and M is the size of filters.

CNN_2^0	CNN_2^+	CNN_3
$[5, 5] \times 16$ ReLU	$[3, 3] \times 16$ ReLU	$[3, 3] \times 16$ ReLU
$[2, 2]$, max pooling $[5, 5] \times 16$ ReLU	$[2, 2]$, max pooling $[5, 5] \times 64$ ReLU	$[2, 2]$, max pooling $[5, 5] \times 32$ ReLU
$[2, 2]$, max pooling	$[2, 2]$, max pooling	$[2, 2]$, max pooling $[7, 7] \times 16$ ReLU
Connected (1024) Dropout (0.5)	Connected (1024) Dropout (0.5)	Connected (1024) Dropout (0.5)
Connected (1024) Soft max	Connected (1024) Soft max	Connected (1024) Soft max
Cross-entropy	Cross-entropy	Cross-entropy

- *Res + Inc*: a deeper CNN with 10-kernel layers which integrates two techniques to reduce the high-dimensional hyper-parameter tuning by means of deeper architectures:

Inception, which includes multiple-sized kernels operating on the same layer.⁴⁵ In this work, we integrate convolutions by 3×3 and 1×1 .

Residual, which integrates residual blocks with the same topology ending with identity-shortcut to connect outputs from lower layers as input in upper layers.⁴⁶ The residual blocks include convolutions by 3×3 and 1×1 for a given input and output size which is defined for each layer: $\text{res_block}(\text{[in, out]})$.

The CNN architectures for the high-resolution TVS are shown in Table 2.

Experimental setup

In this section, we detail the experimental setup of the case study carried out to evaluate the fall detection methodology using data collected by two different TVSs and multiple CNNs.

The data collection design to detect falls was divided into single-occupancy and multi-occupancy. In *single-occupancy*, we included three subcategories: (1) empty room, (2) one person standing/walking, and (3) one fallen person. In *multi-occupancy*, we added two new subcategories: (4) two to three people standing/walking and (5) one fallen person with another person standing/walking. The image data from three participants were collected with the two thermal sensors. While the data were being collected, each person simulated several natural positions to simulate falls, and also took a walk around the vision area of the TVS to capture walking.

Table 2. Configurations of convolutional neural networks for high-resolution TVSs.

CNN_4	$Alex_5$	<i>Res + Inc</i>
$[5, 5] \times 64$ ReLU	$[5, 5] \times 64$ ReLU	$\text{res_block}((64, 128))$ ReLU
$[2, 2]$, max pooling $[5, 5] \times 64$ ReLU	$[2, 2]$, max pooling $[5, 5] \times 128$ ReLU	$\text{res_block}((64, 128))$ ReLU
$[2, 2]$, max pooling $[3, 3] \times 128$ ReLU	$[2, 2]$, max pooling $[3, 3] \times 128$ ReLU	$\text{res_block}((128, 256))$ ReLU
$[2, 2]$, max pooling $[3, 3] \times 128$ ReLU	$[2, 2]$, max pooling $[3, 3] \times 256$ ReLU	$\text{res_block}((128, 256))$ ReLU
$[2, 2]$, max pooling ReLU	$[3, 3] \times 256$ ReLU	$\text{res_block}((128, 256))$ ReLU
Connected (1024) Dropout (0.5)	$[2, 2]$, max pooling Connected (1024)	$\text{res_block}((128, 256))$ ReLU
Connected (1024) Soft max	Connected (1024) Dropout (0.5)	$\text{res_block}((128, 256))$ ReLU
Cross-entropy	Connected (1024) Soft max	Identity-shortcut ReLU
	Cross-entropy	Average-pool

Description of case studies

The first case study was carried out in the Smart Lab of Ulster University⁹ (<https://www.ulster.ac.uk/research/institutes/computer-science/groups/smart-environments>). The experiment was carried out in the hall of the Smart Lab. Three participants (one woman, two men) were involved in collecting data in the hall, using a TVS installed on the ceiling. The participants were 1.72, 1.68, and 1.83 m tall. The vision of the TVS in the hall was determined by a square 3.5 m bounding box (12.25 m²).

The second case study was carried out in the UJAmI smart lab of the CEATIC (Center for Advanced Studies in Information Technology and Communication) of the University of Jaen (Spain)¹⁰ (<http://ceatic.ujaen.es/ujami/>). The experiment was also developed in the hall of the Smart Lab; analogously, three participants (one woman, two men) were involved in collecting data in the hall, using a TVS installed on the ceiling. The participants were 1.88, 1.64, and 1.70 m tall. The vision of the TVS in the hall was determined by a square 2.5×2.0 m bounding box (5.0 m²).

In order to evaluate the two data sets, they were divided into 10% for testing and 90% for training by using a cross-validation (10-cross validation). Accuracy and time were collected for over 2000 learning steps for each CNN in the case of the low-resolution TVS and 200 learning steps for the high-resolution TVS. For each data set with 10 cross-validations, we computed (1) the average accuracy of the last 20 learning steps and (2) the average time wasted in all steps.

Table 3. Table summarizing the results of single- and multi-occupancy data for the low-resolution, wide viewing angle TVS.

Occupancy	CNN	Time (min)	Accuracy (%)
Single	CNN_2^0	22.48	89.3
Single	CNN_2^+	30.36	90.8
Single	CNN_3	27.86	91.9
Multi	CNN_2^0	22.53	83.9
Multi	CNN_2^+	30.36	84.0
Multi	CNN_3	28.00	85.7

CNN: convolutional neural network.

Bold value represents the highest precision values obtained.

The results presented in this work are available in following URL (<http://150.214.174.25:8052/thermal/>).

Evaluation of low-resolution TVS with wide viewing angle

In this section, we detail the results achieved with the three types of CNNs and the performance of the fuzzy representation of thermal information to detect falls from thermal vision images. From the original data set, we include the following data augmentation steps:

- *Translation*: the original images have been translated within a maximal window size, $[t_x, t_y]^+ = [3, 3]$.
- *Rotation scale*: each image is flipped horizontally and vertically by a random probability $wH = 0.5$, $wR = 0.5$ respectively, that is, horizontally in half of the cases, and vertically in the other half. Second, a rotation and scale transformation is defined by a maximal rotation angle $\alpha^+ = \pi/2$ and scale $s^+ = 0.1$. We note this configuration provides random rotation in all quadrants and angles.
- *Crop-scale*: we compute a final centered image with a window size of 28 pixels, $[s_x, s_y] = [28, 28]$, in order to fix to the bounding box of the smart lab for the case scene.

Evaluation of the best CNN configuration. In this section, we present the results from the low-resolution, wide viewing angle TVS, which was evaluated previously in

Medina-Quero et al.²⁰ to detect the best CNN configuration. In Table 3, we include the data for the single- and multi-occupancy data set. CNN_3 provides the best results in classifying fall detection with up to 91% accuracy in single-occupancy contexts and a 6% reduction in accuracy for multi-occupancy.

Evaluation of fuzzy representation of thermal information. In this section, we evaluate performance when applying fuzzy representation to the raw data of the matrix of heat points. To define the fuzzy set which represents *human temperature*, we have included the following trapezoidal membership function (TR is described in the “Abbreviations” in the appendix).

$$\mu_M(m_{i,j}) = TR([l_1, l_2])$$

where $l_1 = 219$ and $l_2 = 252$ correspond to the average temperature collected by the TVS from background and human presence, respectively. These parameters can be straightforwardly computed from a few samples in the tuning stage of the system.

In order to provide a symmetrical evaluation, both with fuzzy representation and raw data, a new augmented data set has been computed and the performance of the best configuration CNN_3 has been analyzed for both cases and the same augmented data. In Table 4, we show the results of the single- and multi-occupancy data with raw and fuzzy representation, including the evolution of accuracy while learning in Figure 4. In Figure 5, we also include a confusion matrix for the best model in single-occupancy contexts.

Evaluation of high-resolution TVS with central viewing angle

In this section, we detail the results of the three types of CNNs to detect falls from the high-resolution, central viewing angle TVS. From the original data set, we include the following data augmentation and fuzzy steps applied to previous learning data:

- *Translation*: the original image is translated within a maximal window size $[t_x, t_y]^+ = [7, 7]$.

Table 4. Table summarizing the results of single and multi-occupancy data with raw and fuzzy representation for the best configuration CNN_3 . In Figure 5, we also include a confusion matrix for best the models in single- and multi-occupancy scenarios.

Occupancy	CNN_3	F1-score	Precision	Recall	Accuracy (%)
Single	Raw	0.9	0.91	0.9	92.3
Single	Fuzzy	0.96	0.97	0.96	97.2
Multi	Raw	0.84	0.84	0.83	86.8
Multi	Fuzzy	0.93	0.93	0.93	94.3

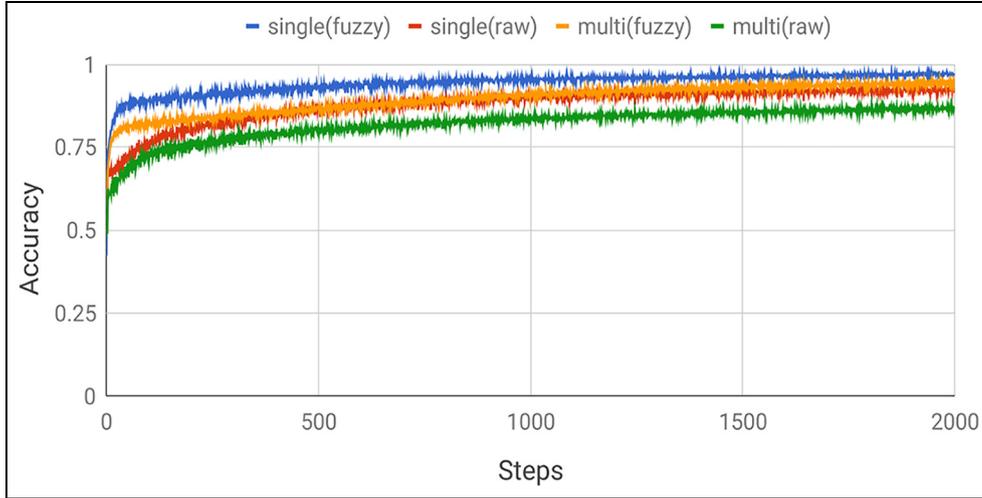


Figure 4. Evolution of accuracy of the best configuration CNN_3 for single and multi-occupancy (fuzzy vs raw representation).

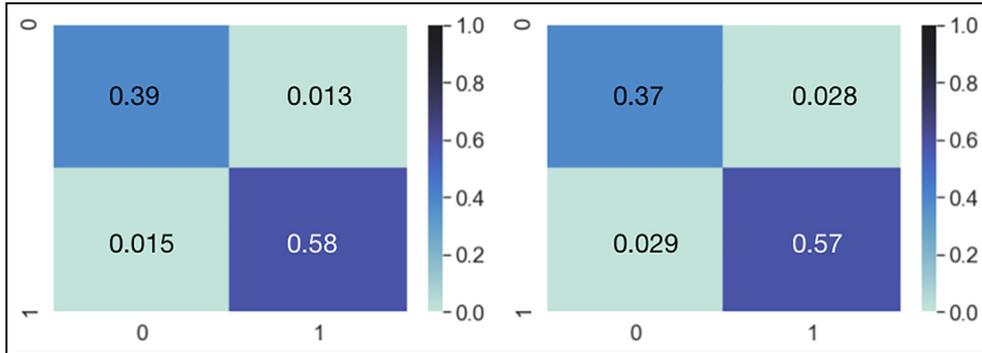


Figure 5. Confusion matrix for the best models. Fuzzy-based single- and multi-occupancy with CNN_3 .

- *Rotation scale*: each image is flipped horizontally and vertically by a random probability $wH = 0.5, wR = 0.5$ respectively, that is, horizontally in half of the cases, and vertically in the other half. Second, a rotation and scale transformation is defined by a maximal rotation angle $\alpha^+ = \pi/2$ and scale $s^+ = 0.3$.
- *Fuzzy configuration*: $\mu_M = TR([l_1, l_2])$, where $l_1 = 8150, l_2 = 8405$ which is provided as a suitable device configuration from previous works.⁴²

In Table 5, we include the data for the single- and multi-occupancy data set for each CNN configuration proposed in this work. In addition, the evolution of accuracy while learning is shown in Figure 6. In Figure 7, we also include a confusion matrix for the best model in multi-occupancy contexts.

Discussion

In this work, two TVS devices with (1) low resolution and wide viewing angle and (2) high resolution and

central viewing angle, data processing stages and different CNN architectures are proposed to classify human falls in single and multi-occupancy contexts.

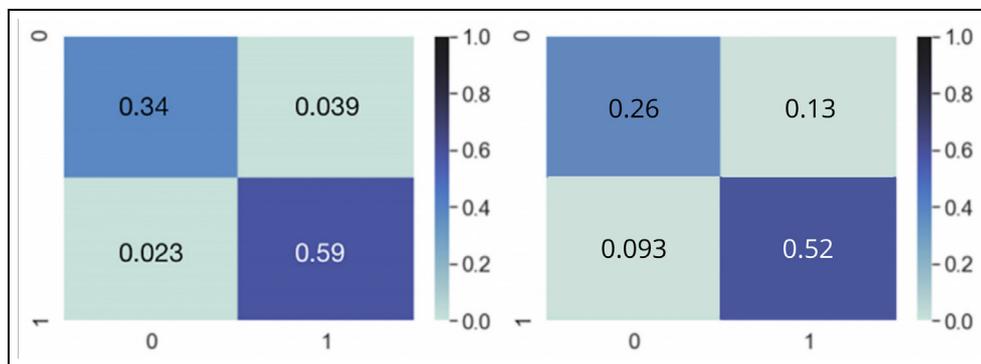
First, high performance is obtained in single-occupancy scenarios, achieving over 90% accuracy for both devices. For the low-resolution, wide viewing angle TVS, we evaluate the impact of including fuzzy representation of thermal information with previous results, which has been demonstrated to increase learning speed and accuracy notably, which with CCN_3 is increased by +5%, achieving 97.2% accuracy, and by more than +7.5%, achieving 94.3% accuracy, for single- and multi-occupancy contexts, respectively. This fact highlights the use of pre-processing the thermal data to improve both performance and learning time of CNN models.

Furthermore, despite the capabilities of CNNs to extract visual features, the initial processing of information, such as fuzzy representation, is key to obtaining encouraging results. In the case of the high-resolution TVS, different CNN architectures have been evaluated, obtaining the best performance with the configuration

Table 5. Table summarizing the results of the single and multi-occupancy data for the high-resolution TVS with a central viewing angle.

Occupancy	CNN	Time (min)	F1-score	Precision	Recall	Accuracy (%)
Single	<i>CNN</i> ₄	100	0.85	0.87	0.84	88.9
Single	<i>Alex</i> ₅	282	0.92	0.89	0.94	93.8
Single	<i>Res + Inc</i>	828	0.89	0.91	0.88	91.5
Multi	<i>CNN</i> ₄	167	0.59	0.56	0.62	69.7
Multi	<i>Alex</i> ₅	474	0.7	0.66	0.73	77.8
Multi	<i>Res + Inc</i>	738	0.61	0.59	0.62	70.6

CNN: convolutional neural networks.

**Figure 6.** Evolution of accuracy of the six types of CNN for the high-resolution TVS with a central viewing angle in single and multi-occupancy contexts.**Figure 7.** Confusion matrix for the best models. Single and multi-occupancy with *Alex*₅.

*Alex*₅ based on AlexNet²⁸ with 93.8% accuracy. We also note learning time is up to 10 times longer in *Alex*₅ than *CNN* due to the differences in the size of the matrix data (from 28×28 to 60×80).

Second, notable performance is obtained in multi-occupancy; the results show a variance of 2.3% and 7.7% of accuracy between best model and second one

in single and multi-occupancy for the high-resolution TVS, but a wide difference in performance is noted between the wide viewing angle and the central viewing angle TVS. For the low-resolution TVS with a wide viewing angle, the best performance is achieved using *CNN*₃ with an accuracy of 94.3%. For the high-resolution device, *Alex*₅ provides the best result but

with an unremarkable accuracy of 77.8%, derived from the conflicting images collected in a very limited space. Regarding the performance difference, we also note the reduction of learning steps in high-resolution approaches due to the augmentation of learning time. In this sense, a longer data set and learning time could improve these approaches, but it is outside the aim of this work, where straightforward methods for agile deployment are proposed.

It is noteworthy that one of the key reasons for this low performance derives from differences in the vision area between the two devices (12.25 and 5.0 m², respectively). The conflicting images we collected of standing and fallen people in the multi-occupancy context represent a greater visual challenge in limited spaces.

Conclusions and ongoing works

In this work, we have evaluated two TVSSs with different capabilities located in the roof of a smart environment to classify the shapes of falls. Two case studies in the Smart Lab of the University of Ulster (UK) and in the Smart Lab of the University Jaen (Spain) are examined. Several CNN configurations are evaluated for each TVSS. A low-resolution TVSS with a wide viewing angle using fuzzy representation of thermal information provides outstanding performance in single- and multi-occupancy contexts.

In future works, we will analyze the impact of temporal sequences in dynamic data sets with fall detection in natural conditions using Deep Learning approaches on temporal models, such as LSTMs.

Acknowledgements

Dr. Martin Cooney from Halmstad University for providing the thermal vision sensor (TVS) with high-resolution and central viewing angle as well as six participants from the University of Jaén who collaborated in the recording of the data set.

Author contributions

All authors contributed equally to this work.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research has received funding under the REMIND project Marie Skłodowska-Curie EU Framework for Research and Innovation Horizon 2020, under Grant Agreement No. 734355. Furthermore, this contribution has

been supported by the Program “José Castillejo” Mobility stays abroad for young doctors (CAS17/00292), the Spanish Government by RTI2018-098979-A-I00, and the University of Jaén by EI_TIC1_2019.

ORCID iD

Macarena Espinilla  <https://orcid.org/0000-0003-1118-7782>

References

1. World Health Organization. World population ageing report, http://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2017_Highlights.pdf (accessed 7 November 2017).
2. Kramarow EA, Chen LH, Hedegaard H, et al. Deaths from unintentional injury among adults aged 65 and over, United States, 2000-2013. *NCHS Data Brief* 2015; 2015: 199.
3. Mubashir M, Shao L and Seed L. A survey on fall detection: principles and approaches. *Neurocomputing* 2013; 100: 144–152.
4. Hayashida A, Moshnyaga V and Hashimoto K. New approach for indoor fall detection by infrared thermal array sensor. *Midwest Symp Circuit Syst* 2017; 196: 1410–1413.
5. Stewart Williams J, Kowal P, Hestekin H, et al. Prevalence, risk factors and disability associated with fall-related injury in older adults in low- and middle-income countries: results from the WHO Study on global AGEing and adult health (SAGE). *BMC Medicine* 2015; 13(1): 147.
6. Memon M, Wagner SR, Pedersen CF, et al. Ambient assisted living healthcare frameworks, platforms, standards, and quality attributes. *Sensors* 2014; 14(3): 4312–4341.
7. Rashidi P and Mihailidis A. A survey on ambient assisted living tools for older adults. *IEEE J Biomed Health Inform* 2013; 17(3): 579–590.
8. Rajagopalan R, Litvan I and Jung TP. Fall prediction and prevention systems: recent trends, challenges, and future research directions. *Sensors* 2017; 17(11): 2509.
9. Nugent CD, Mulvenna MD, Hong X, et al. Experiences in the development of a Smart Lab. *Int J Biomed Eng Tech* 2009; 2(4): 319–331.
10. Espinilla M, Martínez L, Medina J, et al. The experience of developing the UJAmI Smart lab. *IEEE Access* 2018; 6: 34631–34642.
11. Khojasteh SB, Villar JR, Chira C, et al. Improving fall detection using an on-wrist wearable accelerometer. *Sensors* 2018; 18(5): E1350.
12. Sathyanarayana S, Satzoda RK, Sathyanarayana S, et al. Vision-based patient monitoring: a comprehensive review of algorithms and technologies. *J Ambient Intel Humanized Comput* 2018; 9(2): 225–251.
13. Chen L, Hoey J, Nugent CD, et al. Sensor-based activity recognition. *IEEE T Syst Man Cyb Part C: Appl Rev* 2012; 42(6): 790–808.
14. Bromiley PA, Courtney P and Thacker NA. Design of a visual system for detecting natural events by the use of an independent visual estimate: a human fall detector. In:

- Christensen HI and Phillips J (eds) *Empirical Evaluation methods in Computer Vision*. Singapore: World Scientific Publishing, pp.61–87, 2001.
15. De Miguel K, Brunete A, Hernando M, et al. Home camera-based fall detection system for the elderly. *Sensors* 2017; 17(12): 2864.
 16. Hayashida A, Moshnyaga V and Hashimoto K. The use of thermal IR array sensor for indoor fall detection. In: *Proceedings of the 2017 IEEE international conference on systems, man, and cybernetics (SMC)*, Banff, AB, Canada, 5–8 October 2017, pp.594–599. New York: IEEE.
 17. Rougier C, Meunier J, St-Arnaud A, et al. Robust video surveillance for fall detection based on human shape deformation. *IEEE T Circuit Syst Video Technol* 2011; 21(5): 611–622.
 18. Asbjorn D and Jim T. Recognizing bedside events using thermal and ultrasonic readings. *Sensors* 2017; 17(6): 134.
 19. Taramasco C, Rodenas T, Martinez F, et al. A novel monitoring system for fall detection in older people. *IEEE Access* 2018; 6: 43563–43574.
 20. Medina-Quero J, Burns M, Razaq MA, et al. Detection of falls from non-invasive thermal vision sensors using convolutional neural networks. *Proceedings* 2018; 2(19): 1236.
 21. Bayareh R, Vera A, Leija L, et al. Development of a thermographic image instrument using the raspberry Pi embedded system for the study of the diabetic foot. In: *IEEE international instrumentation and measurement technology conference (I2MTC)*, Houston, TX, 14–17 May 2018, pp.1–6. New York: IEEE.
 22. Shewell C, Cleland I, Rafferty J, et al. Computer vision-based gait velocity from non-obtrusive thermal vision sensors. In: *IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, Athens, 19–23 March 2018, pp. 391–396. New York: IEEE.
 23. Abdel-Hamid O, Mohamed AR, Jiang H, et al. Convolutional neural networks for speech recognition. *IEEE T Audio Speech Lang Process* 2014; 22(10): 1533–1545.
 24. Lee SJ, Chen T, Yu L, et al. Image classification based on the boost convolutional neural network. *IEEE Access* 2018; 6: 12755–12768.
 25. Peng P, Zhao X, Pan X, et al. Gas classification using deep convolutional neural networks. *Sensors* 2018; 18(1): 157.
 26. Yamashita T, Watasue T, Yamauchi Y, et al. Improving quality of training samples through exhaustless generation and effective selection for deep convolutional neural networks. In: *International conference on computer vision theory and applications*, Berlin, 11–14 March 2015, pp.228–235. VISAPP.
 27. Ciresan DC, Meier U, Masci J, et al. High-performance neural networks for visual object classification, 2011, arXiv:1102.0183v1
 28. Krizhevsky A, Sutskever I and Hinton GE. Imagenet classification with deep convolutional neural networks. *Commun ACM* 2012; 60: 1097–1105.
 29. Dwibedi D, Misra I and Hebert M. Cut, paste and learn: Surprisingly easy synthesis for instance detection, 2017, arXiv:1708.01642v1
 30. Georgakis G, Mousavian A, Berg A, et al. Synthesizing training data for object detection in indoor scenes, 2017, arXiv:1702.07836v2
 31. Schulte S, De Witte V, Nachtegaal M, et al. Fuzzy two-step filter for impulse noise reduction from color images. *IEEE T Image Process* 2006; 15(11): 3567–3578.
 32. Schulte S, Morillas S, Gregori V, et al. A new fuzzy color correlated impulse noise reduction method. *IEEE T Image Process* 2007; 16(10): 2565–2575.
 33. Jiang Q, Jin X, Lee SJ, et al. A novel multi-focus image fusion method based on stationary wavelet transform and local features of fuzzy sets. *IEEE Access* 2017; 5: 20286–20302.
 34. Kang JK, Hong HG and Park KR. Pedestrian detection based on adaptive selection of visible light or far-infrared light camera image by fuzzy inference system and convolutional neural network-based verification. *Sensors* 2017; 17: 1598.
 35. Pelaez-Aguilera MD, Espinilla M, Fernández Olmo MR, et al. Fuzzy linguistic protoforms to summarize heart rate streams of patients with ischemic heart disease. *Complexity* 2019; 2019: 2694126.
 36. Yin S, Zhang Y and Karim S. Large scale remote sensing image segmentation based on fuzzy region competition and Gaussian mixture model. *IEEE Access* 2018; 6: 26069–26080.
 37. Heimann, March 2019, <http://www.heimannsensor.com/>
 38. Medina-Quero J, Shewell C, Cleland I, et al. E. Computer vision-based gait velocity from non-obtrusive thermal vision sensors. In: *IEEE international conference on pervasive computing and communications workshops (PerCom Workshops)*, Athens, 19–23 March 2018, pp.391–396. New York: IEEE.
 39. Rafferty J, Synnott J, Ennis A, et al. Sensorcentral: a research oriented, device agnostic, sensor data platform. In: *International conference on ubiquitous computing and ambient intelligence*, Philadelphia, PA, 7–10 November 2017, pp.97–108. Cham: Springer.
 40. Flir, March 2019, <https://www.flir.com/products/flir-one-gen-3/>
 41. Ferdoush S and Li X. Wireless sensor network system design using Raspberry Pi and Arduino for environmental monitoring applications. *Proced Comput Sci* 2014; 34: 103–110.
 42. Cooney M and Bigun J. PastVision + : thermovisual inference of recent medicine intake by detecting heated objects and cooled lips. *Front Robot AI* 2017; 4: 61.
 43. Han J and Ma KK. Fuzzy color histogram and its use in color image retrieval. *IEEE T Image Process* 2002; 11(8): 944–952.
 44. LeCun Y. The MNIST database of handwritten digits, 1998, <http://yann.lecun.com/exdb/mnist/>
 45. Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, Las Vegas, NV, 27–30 June 2016, pp.2818–2826. New York: IEEE.
 46. Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning. In: *Proceedings of the Thirty-First AAAI conference on artificial intelligence*, San Francisco, CA, 4–9 February 2017. Reston, VA: AIAA.

Appendix

Abbreviations

The following abbreviations are used in this article:

TR	$TR(x)[l_1, l_2] = \begin{cases} 1 & x \leq l_1 \\ (l_2 - x)/(l_2 - l_1) & l_1 \leq x \leq l_2 \\ 0 & l_2 \leq x \end{cases}$
TS	$TS(x)[l_1, l_2, l_3, l_4] = \begin{cases} 0 & x \leq 0 \\ (x - l_1)/(l_2 - l_1) & l_1 \leq x \leq l_2 \\ 1 & l_2 \leq x \leq l_3 \\ (l_4 - x)/(l_4 - l_3) & l_3 \leq x \leq l_4 \\ 0 & l_4 \leq x \end{cases}$
TVS	Thermal Vision Sensor
CCN	Convolutional Neural Network
