2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

An Unobtrusive Fall Detection System Using Low Resolution Thermal Sensors and Convolutional Neural Networks

Ariyamehr (Mohsen) Rezaei, Michael C. Stevens, Ahmadreza Argha, Alessandro Mascheroni,

Alessandro Puiatti, and Nigel H. Lovell, *Fellow, IEEE*

Abstract— **Human activity recognition has many potential applications. In an aged care facility, it is crucial to monitor elderly patients and assist them in the case of falls or other needs. Wearable devices can be used for such a purpose. However, most of them have been proven to be obtrusive, and patients reluctate or forget to wear them. In this study, we used infrared technology to recognize certain human activities including sitting, standing, walking, laying in bed, laying down, and falling. We evaluated a system consisting of two 24×32 thermal array sensors. One infrared sensor was installed on side and another one was installed on the ceiling of an experimental room capturing the same scene. We chose side and overhead mounts to compare the performance of classifiers. We used our prototypes to collect data from healthy young volunteers while performing eight different scenarios. After that, we converted data coming from the sensors into images and applied a supervised deep learning approach. The scene was captured by a visible camera and the video from the visible camera was used as the ground truth. The deep learning network consisted of a convolutional neural network which automatically extracted features from infrared images. Overall average F1-score of all classes for the side mount was 0.9044 and for the overhead mount was 0.8893. Overall average accuracy of all classes for the side mount was 96.65% and for the overhead mount was 95.77%. Our results suggested that our infrared-based method not only could unobtrusively recognize human activities but also was reasonably accurate.**

I. INTRODUCTION

The recognition and classification of human activities have many applications such as monitoring elderly people in an aged care facility. Wearable devices can be used for such an application [1]. A survey on sensor-based activity recognition were presented in [2]. However, studies have shown that residents generally do not want or may forget to wear wearable devices [3]. Technologies like visible cameras, millimeter wave (mmWave) radar, and light detection and ranging (LiDAR) can be used for recognizing and monitoring human activities unobtrusively [4], [5]. However, these technologies are either expensive or privacy unfriendly. For instance, a visible camera is not privacy friendly and LiDAR is expensive. mmWave radar is affordable, but it generates three-dimensional point cloud data which are computationally more expensive. In this study, a low-resolution infrared sensor is used to classify human activities including sitting, standing, walking, laying in bed, laying down, and falling. Low-resolution infrared cameras are privacy friendly and unobtrusive. They produce a matrix showing a body heat signature that can be used by machine learning (ML) techniques to recognize human activities. There are two main ML approaches for human activity recognition, namely frame-based and flow-based. While in the former approach, only one frame is used [6], in the latter one, a sequence of frames is analyzed [7]. In this study, a frame-based approach was used since it is less computationally expensive. As a result, it can be executed in embedded systems. The aim of this study is to present a novel deep learning-based method for human activity recognition using low-resolution, low-cost infrared sensors, and further investigate the effect of sensor position on performance of the proposed convolutional neural network (CNN)-based model. The contributions of this study are: 1) providing two datasets by experimenting eight different scenarios on healthy young volunteers using our prototypes, 2) presenting a human activity recognition method using infrared sensors and a CNN model.

Corresponding author: Nigel H. Lovell, phone: +61-2-9385-3922; e-mail: n.lovell@unsw.edu.au.

Ariyamehr (Mohsen) Rezaei, Michael C. Stevens, Ahmadreza Argha and Nigel. H. Lovell are with the Graduate School of Biomedical Engineering, University of New South Wales, Sydney, NSW 2052, Australia (e-mail: a.rezaei@ student.unsw.edu.au; michael.stevens@unsw.edu.au; a.argha@unsw.edu.au; n.lovell@unsw.edu.au). Alessandro Mascheroni and Alessandro Puiatti are with The University of Applied Sciences and Arts of Southern Switzerland (email: alessandro.mascheroni@supsi.ch, alessandro.puiatti@supsi.ch).

II. RELATED WORK

Two low-resolution (4 by 16 pixels) MLX90621 thermal sensor were used in [8]. The collected data included 2083 and 1190 samples achieved from two and three subjects, respectively. Their system could classify four static activities namely sitting, standing, sitting on ground, and laying on ground. Using five different ML algorithms, the best achieved classification accuracy was 97.5%. A low-resolution (64 by 64 pixels) thermal sensor with ceiling position was used in [9] for recognition of 7 human actions: slow walking, fast walking, restlessness, sitting, standing, turning on a seat, and no action. The restlessness action covered all small movements around a workstation such as answering the phone, moving objects, and typing. In total,

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

the dataset consisted of 700 samples. Using a 10-fold cross validation method the F1-score achieved by their proposed model was 83%. In [10] three set of low-resolution thermal sensors (8 by 8 pixels) were used to capture human postures from x, y, and z-axes. Three thermal images were concatenated and fed to a CNN to recognize eight human postures namely stand, hand raise, akimbo, open wide arms, squat, toe touch, crawl, and lie. A total of 15063 samples were collected from four participants. Using a 10-fold cross validation method an F1-score of 99.81% was obtained. A high-resolution (320 by 240 pixels) infrared camera was used in [11]. The collected data consisted of 5278 samples which were split as 2844 for training, 1255 for validation, and 1179 for test. The classes included: walking, standing, sitting on chair, sitting on chair with desk in front, fallen on the desk in front, and fallen/lying on ground. The images were resized before entering a CCN model and the best results were achieved with an accuracy of 87.44% for size 32 by 32. Plantar inclinometer sensor was used in [12]. They detected falls of 5 subjects in forward, backward, left, and right falls and obtained the average detection rate of 85.4%.

However, no existing work attempted to classify laying in bed as a distinct action. This is of importance especially in an aged care facility, as it is required to distinguish whether the elderly person lies in bed or lies down on the floor (fall). In this study, along with other human activities, we considered a distinct category namely laying in bed, to make it possible to distinguish it from laying down on the floor.

III. EXPERIMENTAL SYSTEM

This section explains the experimental system including hardware system, position of installation, methodology, data collection, data pre-processing, and the deep learning model.

A. Hardware system

The components are EVB90640-41 board and MLX90640ESF-BAA-000-TU infrared sensor from Melexis company [13], [14]. The system can measure temperature and perform some pre-processing. It can send the measured temperatures in a csv file format to a computer. It can host infrared sensors with narrow and wide field of view, respectively. In this research, we used the wide one (110 degrees in azimuth and 75 degrees in elevation) which had accuracy of ± 1 Celsius (^{o}C). Fig. 1 shows the hardware. We set frame per second to 1, emissivity to 1, measurement pattern as chess and then logged the data.

B. Position of installation

We designed and manufactured two prototypes. They were manufactured in Polylactic Acid (PLA) material using rapid prototyping (Ultimaker Cura 3D printer). The prototypes were installed in a room of 5×5×3 m (length, depth, height) with ambient temperature from 20 to 22 ^{o}C on the side and the overhead. For the side mount, the sensor was elevated 2 m from the floor, tilted approximately 10 degrees down for better coverage and installed in the middle of the length of the room. For the overhead mount, the sensor was installed in the center of the room at 3 m elevation facing floor with no tilt. Fig. 2 shows the prototypes.

C. Methodology

The scene was simultaneously captured by the two infrared sensors and a visible camera. The video from the visible camera was used as the ground truth. The data from the sensors were sent to a computer in a csv format. In the csv file each row consisted of a timestamp and measured temperature of all 768 pixels representing a frame. Then each frame was converted into a thermal image of size 24 by 32. After performing some pre-processing, the images were split to train and validation sets and fed into a CNN model.

The experimental procedures involving human subjects described in this paper were approved by Human Research Ethics and Clinical Trials (HREC) Governance of the University of New South Wales with number HC191001.

D. Data collection

We defined eight different scenarios each between one and two minutes long and healthy young volunteers performed them. In total, 80 different experiments were performed. Table I shows the scenarios and involved activities. Ten healthy young volunteers (seven males and three females, height: 179 ± 9.4 cm, weight: 80.1 ± 17 kg, age: 26.5 ± 4.2 years) were recruited in the human trials of simulated falls and simulated activity of daily livings.

Figure 1. Hardware of EVB90640-41.

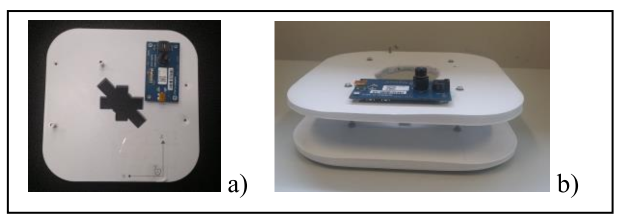

Figure 2. Prototypes. a) the sensor is installed on a platform for side mount, b) the sensor is installed on a two-story platform for overhead mount.

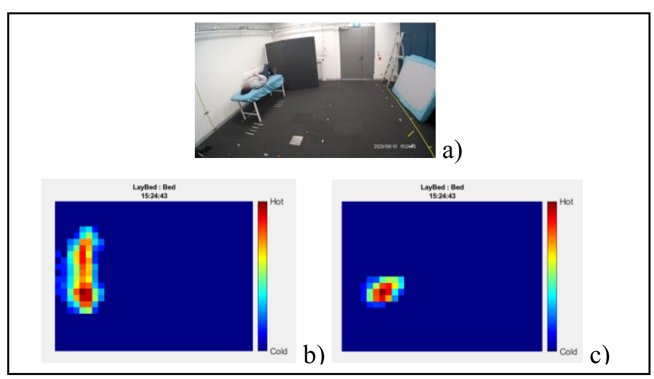

Figure 3. Visualizing a frame while a participant is laying in bed. a) shows the scene recorded by a visible camera, b) shows the thermal image from the overhead sensor, c) shows the thermal image from the side sensor.

We defined seven classes namely stand, walking, sitting, sitting in bed (SitBed), falling, laying in bed (LayBed), laying on floor (LayFloor). We then merged falling with laying on floor since both demand attention and show abnormal activities, and stand with walking, since both are similar in frame-based approach. By setting one frame per second, we collected 5117 and 5126 samples from the side and overhead sensors, respectively. The two number of samples differed because the sensors did not get samples exactly every 1 second resulting in having occasionally missed data. The obtained dataset is listed in Table II.

E. Data pre-processing

Fig. 3 visualizes a frame in which a participant is laying in bed with the corresponding thermal images from the side and overhead infrared sensors. To clean background noise such as ceiling lighting, heater, laptop, or kettle, for every participant data a time-averaged algorithm was performed on previous frames that were labelled as background and this average image was subtracted from other frames. The length of the averaging window was from the start of the previous experiment up to the current frame. The method was computationally cheap and sufficient for our purpose.

F. Deep learning model architecture

The architecture of the model consisted of an input layer, following three convolutional layers, two maxpooling layers between every two convolutional layers, one fully connected layer, and a softmax layer. The kernel size was 3×3 for all three convolutional layers. ReLu activation function was used for all the convolutional layers and fully connected layer. Stride was one by one for convolutional layers and two by two for maxpooling layers. Hyper parameters were set as optimization algorithm: adam, initial learn rate: 0.001, epoch numbers: 10, batch size: 128. A laptop with NVIDIA Quadro T2000 GPU was used for performing deep learning. The training and validation took 40 minutes over 10 epochs. Fig. 4 shows the architecture of the CNN model.

IV. RESULTS

We used MATLAB software to perform supervised deep learning on the data. To validate the results, we used leave-one-subject-out method. Tables III and IV show confusion matrices for the side and overhead mounts. We evaluated the model performance with four measures namely accuracy, recall, precision, and F1-score. Tables V and VI show the evaluation matrices. Overall average accuracy of all classes for the side mount was 96.65% and for the overhead was 95.77%. Overall average recall of all classes for the side mount was 0.9060 and for the overhead was 0.8931. Overall average precision of all classes for the side mount was 0.9029 and for the overhead was 0.8856. Overall average F1-score of all classes for the side mount was 0.9044 and for the overhead was 0.8893. For the side mount the merged falling and laying on floor, and laying in bed activities had the highest F1-score with 0.9249 and 0.9241, respectively. For the overhead mount laying in bed activity had the highest F1-score with 0.9128. For both mounts, sitting in bed activity was recognized with the lowest F1-score. For the side mount, laying in bed along with merged falling and laying down activities and for the overhead mount, laying in bed were recognized with the highest F1-score.

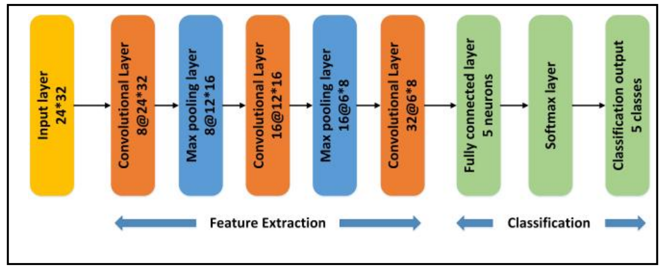

Figure 4. The architecture of the CNN model.

TABLE I. SCENARIOS AND INVOLVED ACTIVITIES.

Number	Scenario description
1	Walking and randomly stand
2	Sitting on chairs
3	Falling forward from standing position
4	Falling backward from standing position
5	Falling laterally from standing position
6	Laying in bed
7	Sitting in bed and falling forward while standing up
8	Falling from bed while laying

TABLE II. THE DATASET FOR ACTIVITY RECOGNITION.

Class	Side	Overhead
Falling+LayFloor	1701	1699
LayBed	721	726
Sit	677	679
SitBed	289	292
Stand+Walking	1729	1730

TABLE III. CONFUSION MATRIX FOR THE SIDE MOUNT.

True /Predicted Class	Falling+ LayFloor	LayBed	Sit	SitBed	Stand+ Walking
Falling+LayFloor	160	1	2	1	11
LayBed	2	67	0	2	1
Sit	2	2	62	0	3
SitBed	1	1	1	24	1
Stand+Walking	6	2	2	2	158

TABLE IV. CONFUSION MATRIX FOR THE OVERHEAD MOUNT.

True /Predicted Class	Falling+ LayFloor	LayBed	Sit	SitBed	Stand+ Walking
Falling+LayFloor	152	2	2	2	15
LayBed	1	68	1	1	2
Sit	1	2	63	1	3
SitBed	1	1	1	24	1
Stand+Walking	11	3	2	1	150

TABLE V. EVALUATION MATRIX FOR THE SIDE MOUNT.

True /Predicted Class	Accuracy	Recall	Precision	F1-score
Falling+LayFloor	0.9494	0.9143	0.9357	0.9249
LayBed	0.9786	0.9306	0.9178	0.9241
Sit	0.9767	0.8986	0.9254	0.9118
SitBed	0.9825	0.8571	0.8276	0.8421
Stand+Walking	0.9455	0.9294	0.9080	0.9186
Average value	0.9665	0.9060	0.9029	0.9044

TABLE VI. EVALUATION MATRIX FOR THE OVERHEAD MOUNT.

True /Predicted Class	Accuracy	Recall	Precision	F1-score
Falling+LayFloor	0.9315	0.8786	0.9157	0.8968
LayBed	0.9746	0.9315	0.8947	0.9128
Sit	0.9746	0.9000	0.9130	0.9065
SitBed	0.9824	0.8571	0.8276	0.8421
Stand+Walking	0.9256	0.8982	0.8772	0.8876
Average value	0.9577	0.8931	0.8856	0.8893

V. DISCUSSION

Based on the measures, the proposed CNN model slightly performed better for the side mount. However, the difference between them was quite small. Unlike [8] that considered four static activities, in this study we omitted sitting on ground activity and instead added walking, falling, laying on floor, laying in bed, and sitting in bed activities. This is because we assumed falling and laying on floor were more critical and needed more attention than sitting on ground. In [10] falling, laying down, and laying in bed activities were not investigated. For monitoring elderly people especially in an aged care facility, it is crucial to know when they are in bed and summon help in case of laying down or falling. Falling and laying on floor were investigated in [11] in which an accuracy of 87.44% was achieved by a CNN method. Using our proposed CNN model, overall average accuracy of all classes for the side was 96.65% and for the overhead was 95.77%. In addition, we calculated recall, precision, and F1-score so that the performance of the model could be seen class-wise while in [11] only accuracy was calculated. One advantage of the proposed method was it remained simple to implement and required few resources (time and space). Therefore, it could be useful for real-time application. Furthermore, we investigated the impact of the sensor position on the performance of our proposed CNN model. Moreover, the proposed system was low cost and could be installed in aged care facilities as nurse call systems.

VI. CONCLUSION

In this study, we demonstrated how low-resolution thermal sensor (24×32 thermal array) could be used to recognize human activities. We collected data from ten healthy young volunteers experimenting eight different scenarios and performing deep learning. Overall average F1-score of all classes for the side mount was 0.9044 and for the overhead was 0.8893. Overall average accuracy of all classes for the side mount was 96.65% and for the overhead was 95.77%. Our results suggested that our infrared-based method not only could unobtrusively recognize human activities but also was reasonably accurate. For future research we intend to combine the data from two sensors (stereo analysis). We also intend to compare deep learning with manually extracted features approach. Future work in this area could extend to the application of this technology to the cases where there is more than one resident. Also, higher frame rates or sensors with more resolution could be used, and rather than the devised frame-based classification method a sequence classification method could be developed. The primary limitation of our method is the need to perform offline training of the classifier. Therefore, further research could be performing online training or even using unsupervised approach. Another limitation is having falling and laying on floor classes merged which we are working to resolve.

ACKNOWLEDGMENT

We would like to thank Human Research Ethics and Clinical Trials Governance, the University of New South Wales, participants who participated in the research study as healthy young volunteer, Graduate School of Biomedical Engineering, and Neuroscience Research Australia for the experimental room, facilities and helping us to comply with the protocols of COVID-19. We are also thankful to Vitalcare Pty Ltd and the NSW Government TechVoucher program for sponsoring the project. Ariyamehr would like to thank the Australian Department of Education, Skills and Employment for the support through an Australian Government Research Training Program Scholarship.

REFERENCES

[1] C. Wang, S. J. Redmond, W. Lu, M. C. Stevens, S. R. Lord, and N. H. Lovell, "Selecting Power-Efficient Signal Features for a Low-Power Fall Detector," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 11, pp. 2729–2736, 2017, doi: 10.1109/TBME.2017.2669338.

[2] L. Chen, J. Hoey, C. D. Nugent, D. J. Cook, Z. Yu, and S. Member, "Sensor-Based Activity Recognition," vol. 42, no. 6, pp. 790–808, 2012.

[3] R. Fleming and S. Sum, "Empirical studies on the effectiveness of assistive technology in the care of people with dementia: A systematic review," *J. Assist. Technol.*, vol. 8, no. 1, pp. 14–34, 2014, doi: 10.1108/JAT-09-2012-0021.

[4] M. Moencks, V. De Silva, J. Roche, and A. Kondoz, "Adaptive Feature Processing for Robust Human Activity Recognition on a Novel Multi-Modal Dataset," 2019, [Online]. Available: http://arxiv.org/abs/1901.02858.

[5] A. D. Singh, S. S. Sandha, L. Garcia, and M. Srivastava, "Radhar: Human activity recognition from point clouds generated through a millimeter-wave radar," *Proc. Annu. Int. Conf. Mob. Comput. Networking, MOBICOM*, pp. 51–56, 2019, doi: 10.1145/3349624.3356768.

[6] L. Walmsley-Eyre and R. Cardell-Oliver, "Hierarchical Classification of Low Resolution Thermal Images for Occupancy Estimation," *Proc. - 2017 IEEE 42nd Conf. Local Comput. Networks Work. LCN Work. 2017*, pp. 9–17, 2017, doi: 10.1109/LCN.Workshops.2017.59.

[7] T. Kawashima *et al.*, "Action recognition from extremely low-resolution thermal image sequence," in *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2017*, 2017, vol. 2, doi: 10.1109/AVSS.2017.8078497.

[8] S. Singh and B. Aksanli, "Non-intrusive presence detection and position tracking for multiple people using low-resolution thermal sensors," *J. Sens. Actuator Networks*, vol. 8, no. 3, 2019, doi: 10.3390/jsan8030040.

[9] F. Polla, H. Laurent, and B. Emile, "Action recognition from low-resolution infrared sensor for indoor use: A comparative study between deep learning and classical approaches," *Proc. - IEEE Int. Conf. Mob. Data Manag.*, vol. 2019-June, no. Mdm, pp. 409–414, 2019, doi: 10.1109/MDM.2019.00-11.

[10] M. Gochoo, T. H. Tan, T. Batjargal, O. Seredin, and S. C. Huang, "Device-Free Non-Privacy Invasive Indoor Human Posture Recognition Using Low-Resolution Infrared Sensor-Based Wireless Sensor Networks and DCNN," *Proc. - 2018 IEEE Int. Conf. Syst. Man, Cybern. SMC 2018*, pp. 2311–2316, 2019, doi: 10.1109/SMC.2018.00397.

[11] A. Akula, A. K. Shah, and R. Ghosh, "Deep learning approach for human action recognition in infrared images," *Cogn. Syst. Res.*, vol. 50, pp. 146–154, 2018, doi: 10.1016/j.cogsys.2018.04.002.

[12] J. Sun, Z. Wang, L. Chen, B. Wang, C. Ji, and S. Tao, "A plantar inclinometer based approach to fall detection in open environments," *Stud. Comput. Intell.*, vol. 647, no. June 2016, pp. 1–13, 2016, doi: 10.1007/978-3-319-33353-3_1.

[13] Melexis, "MLX90640 32x24 IR array Datasheet," pp. 1–60.

[14] T. Evb and T. Evb, "1. Installation," *EVB90640-41 User Man.*, pp. 1–7.