A Lightweight Neural-Net with Assistive Mobile Robot for Human Fall Detection System

Wei Hong Chin*, Noel Nuo Wi Tay*†, Naoyuki Kubota*, and Chu Kiong Loo‡
* Faculty of Systems Design, Tokyo Metropolitan University, Hino, Japan
Email: {weihong, kubota}@tmu.ac.jp, taynoel@ieee.org
† Department of Artificial Intelligence, CSUN Technology Co., Ltd., Tokyo, japan
Email: zhengnuowei@dragonwake.cn
‡ Faculty of Computer Science & Information Technology, University of Malaya, Kuala Lumpur, Malaysia
Email: ckloo.um@um.edu.my

***Abstract*—Falls are a major health issue, particularly among the elderly. Increasing fall events require high service quality and dedicated medical treatment which is an economic burden. In the lack of appropriate care and support, serious injuries caused by fall will cost lives. Therefore, tracking systems with fall detection capabilities are required. Static-view sensors with machine learning techniques for human fall detection have been widely studied and achieved significant results. However, these systems unable to monitor a person if he or she is out of viewing angle which greatly impedes its performance. Mobile robots are an alternative for keeping the person in sight. However, existing mobile robots are unable to operate for a long time due to battery issues and movement constraints in complex environments. In this paper, we proposed a lightweight deep learning vision-based model for human fall detection with an assistive robot to provide assistance when a fall happens. The proposed detection system requires less computational power which can be implemented in a low-cost 2D camera and GPU board for real-time monitoring. The assistive robot equipped with various sensors that can perform SLAM, obstacle avoidance and navigation autonomously. Our proposed system integrates these two sub-systems to compensate for the weakness of each other to constitute a system that robust, adaptable, and high performance. The proposed method has been validated through a series of experiments.**

***Index Terms*—fall detection, deep learning, assistive robot**

I. Introduction

According to Noury [1], fall is described as an unintentional or sudden change of the body's position from a position upright, sitting or lying to a lower horizontal position. Falls are a dangerous situation, particularly in the elderly, which leads to further injuries, fractures, and other health problems. The World Health Organization (WHO) reports that about 28-35% of population age of 65 falls each year and the number increases as people becomes older. The detection of falls continues to be an area of active research to improve the quality of life of elderly.

In the past, there are number of significant research have been proposed using different types of sensors and methods on developing fall detection systems. According to Noury [1], Nait-charif [2] and McKenna, Mohamed et al. [3], and Khan [4], the state-of-arts can be grouped into three main groups: wearable-based technology, ambient technology and vision-based technology.

Wearable devices are increasing rapidly and rely on sensors attached to the body as accelerometers, gyroscopes, pressurized interfaces, and magnetometers [5]–[8]. These devices provide high detection rates with small and low-cost technology. However, the main drawbacks of wearable sensors is that they require active sensing by wearing the sensors in which may lead to uncomfortable for the users in long-run.

Ambient technology utilizes mounted sensors to obtain the related individual's data while they are nearby. Almost all technology uses pressure or infrared sensors that sense high pressure due to the occupant's weight. Ambient technology utilizes mounted sensors to obtain related individuals' data when they are nearby. Almost all technology uses pressure sensors to sense the occupant's location. The main disadvantage of such sensors is that users cannot distinguish if the sensing data is from the person or other objects.

Vision-based devices need no elderly assistance. During our everyday lives, today's cameras are used more and more. The location and shape of the person are analyzed in real-time using different types of algorithms combined with standard computer platforms and low-cost cameras through visualized fall detection systems. The vision-based approaches promise performance when compared to other methods because of the rapid developments in computer vision and video cameras such as affordable Microsoft Kinect as well as the advancement of artificial intelligence [9]–[11]. However, one of the drawbacks of computer vision techniques is that, they require high computational power which is not feasible for massive deployment especially for residential sectors.

Compare to static vision sensors, mobile robot is a feasible solution to keep track a person in view [12], [13]. Previous works have been proposed for tracking a person and navigating the environment [14], [15]. However, existing robot platforms are not able to operate for long hours due to battery issue.

In this paper, we proposed a human fall detection system that consists of two main sub-systems: i) a human tracking and fall detection system using a lightweight deep learning model; ii) a mobile assistive robot that communicate with the fall

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

detection system and navigate the environment autonomously. The contribution of the proposed system is that, the lightweight deep learning model can be deployed to a low-cost 2D camera and on-board GPU board for real-time human tracking and fall detection. In addition, the mobile assistive robot is used for providing assistance to a person when a fall is detected. Both sub-systems work in mutual manner in which when a person is out of view from the static camera, the assistive robot is activated to search and track for the person. On the other hand, when the person fall down is detected by the static camera, an indication is sent to the robot to look for the person and provide assistance such as call for help or activate the alarm. Thus, the proposed system overcomes each sub-systems drawback that can continually perform human tracking and fall detection.

The remainder of the paper is organized as follows. Section II describes the mathematical models of the proposed system. The experimental setup and results are showed and discussed in Section III and Section IV. Finally, conclusions and future works are highlighted in Section V.

II. Proposed System

The overview of the proposed fall detection system as shown in Figure 1. The system consists of two main sub-systems i) static camera with on-board GPU and ii) mobile assistive robot. Both systems are interacting to each other for real-time human tracking and fall detection. In this system, if the person to be monitored is in static camera view zone, the mobile assistive robot will not be activated for human tracking and monitoring. The mobile assistive robot will be triggered to track and monitor the person as soon as the person is out of the static camera view.

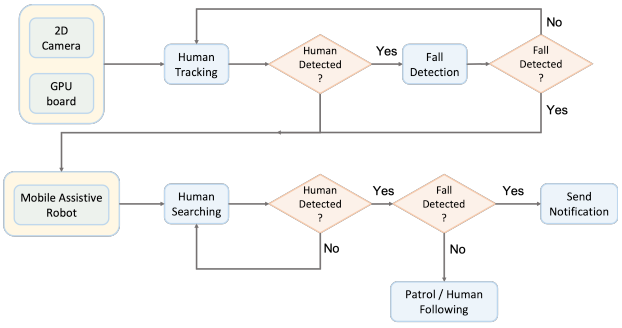

Fig. 1. Fall detection system overview

A. Fall Detection System

There are various methods to perform human fall detection, such as using wearables, vibration and sound sensors in the room, or video camera. RGB video camera is used in this work due to it 1) being non-contact 2) having abundant information on human actions 3) being able to easily perform localization 4) much cheaper compared to RGBD camera. Non-contact way of detecting fall is preferable since it is cumbersome to require everyone to have a wearable while indoors. Besides, information content from video camera is relatively larger, which also enables monitoring when fall occurs. With such information content, localization can also be performed to enable rescuers to know the exact location of the victim without the need to depend on external peripheral devices.

Methods for fall detection using video camera includes analyzing background subtracted points, optical flow and direct analysis on human actions. We choose direct human action analysis due to it being able to extend to other domains apart from fall.

Works on action recognition that utilizes deep learning can roughly be divided into two categories, which are: two stream architecture [21] and direct 3D convolution [22]. Two stream architecture network requires spatial (image itself) and temporal (dense optical flow) information, where the calculation of dense optical flow can be computationally demanding. On the other hand, original 3D convolution network for action recognition is also quite computationally intensive, but a recent work, called the SlowFast network [18], applies 3D convolution and without the need of optical flow, can achieve state-of-the-art performance with high speed. Basically, the network consists of a slow and fast stream, and that 3D convolution only started near the end of the network, making the network to be able to perform in high speed.

As this work requires computation to be done on light weight embedded systems,in this work, we further modify the network to make it more light-weight by changing the backbone to MobileNet V2 [23]. We also discard the fast stream of SlowFast Net. This is because from the original result, elimination of the fast stream resulted only in a small drop in accuracy, which is a reasonable trade-off if we require speed. All 3D convolutions are implemented using 2D convolution using group convolution approach to reduce parameter size. The network architecture is shown in Figure 2. Spatial C and temporal denotes spatial and temporal channel respectively, while s denotes stride. Input consists of a sequence of images (which traverses temporally). Convolution layer Conv0 to Conv2 are just normal 2D convolutions, which are performed separately for each image in the sequence. For Conv3, 3D convolution is done by first performing point-wise convolution on the temporal dimension, followed by spatial convolution. All this can be performed using 2D convolution by mere reshaping of the feature tensors. Finally, FC1 to FC3 are linear layers that outputs the probability of a fall.

SlowFast Net produces feature maps, where one can use them to perform various kinds of subsequent predictions. Most straight-forward is to stack it with linear classifier, or perform ROI pooling/alignment given some detections from other networks. For our network, global average pooling is performed on the temporal dimension, followed by Max pooling on the spatial dimension. The retrieved feature vector will then be passed to a fall classifier. For comparison, we also added a self-attention network to the modified SlowFast Network, where, instead of just mere max pooling, the network will choose specific regions. The additional self-attention network will take Conv3 output to produce a single channel map with the same spatial size, which acts as weights to sum over outputs of

Layers	Unless otherwise stated, the following is true: Spatial kernel =3x3 with padding=1 Temporal kernel=3 with padding=1
Conv0	Spatial C:1-16 s:2
Conv1	Spatial C:16-16 s:2
	Spatial C:16-16 s:1
	Spatial C:16-16 s:1
Conv2	Spatial C:16-16 s:2
	Spatial C:16-16 s:1
	Spatial C:16-16 s:1
	Spatial C:16-24 s:1
Conv3	Temporal C:24-24 s:1
	Spatial C:24-24 s:2
	Temporal C:24-24 s:1
	Spatial C:24-24 s:1
	Temporal C:24-24 s:1
	Spatial C:24-24 s:1
	Temporal C:24-24 s:1
	Spatial C:24-24 s:1
	Temporal C:24-24 s:1
	Spatial C:24-24 s:1
	Temporal C:24-24 s:2
	Spatial C:24-24 s:1
Conv4	Temporal C:24-24 s:1 Temporal kernel=5 with padding=0
Conv5	Spatial kernel=1x1 C:24-8
Fc1	L:392-320
FC2	L:320-100
FC3	L:320-1

Fig. 2. Network Architecture

Conv3, before feeding the result to the linear layers.

B. Assistive Robot

The assistive robot is built with an iRobot Roomba mobile base, a Hokuyo laser range finder, an Apple iPhone, and an Intel i5 processor mini PC as shown in Figure 3. Based on our previous work [16], we developed a multi-objectives motion planning that allows the robot for obstacle avoiding, SLAM, and wall-following. Thus, the robot able to patrol and localize in any indoor environment autonomously. The example of the grid map is illustrated in Figure 4.

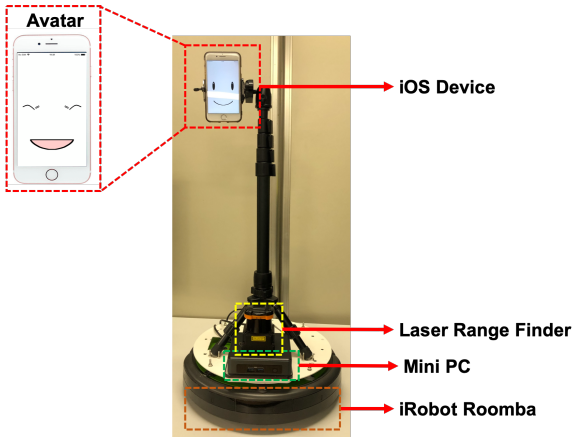

Fig. 3. Assistive Mobile Robot

For the iOS avatar chatbot application, we extended our previous work [17] by deploying the lightweight human tracking and detection model. In addition, we also added a function that will send a notification to the caregiver if fall is detected and no response from the person. As such, the assistive robot can fulfill the task in which the static camera unable to achieve. The iOS application consists of

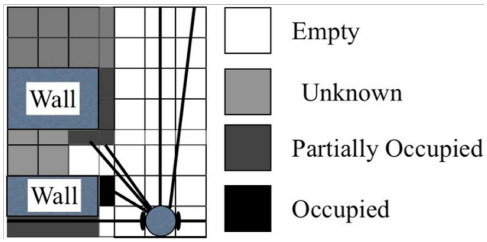

Fig. 4. Illustration of grid map building

three main components: i) Conversation module, ii) Human Tracking Module and iii) Action Module. Each of this module are interacting with each other to provide service when the robot is triggered by the static camera system. Conversation module comprises avatar animation, speech-to-text and text-to-speech function for human-robot communication. Next, human tracking module is similar to the static camera system that explained in previous section except the learned weights is converted to be compatible for iOS system by using the apple's CoreML tools. Lastly, the Action Module is to send notification to the caregiver whenever a fall is detected. In addition, the Action Module is also responsible to activate the iRobot mobile robot base to moving around whenever human tracking mode is triggered.

III. Experimental Setup & Results

The experiments were conducted in a room that located at the basement of our university. The room is to mimic a common Japanese home in which it consists of furniture, rest place and bed. The static camera was set up at the corner of the room and the assistive robot is placed in the room. The grid map of the experimental place as shown in Figure 6(a). We conduct the experiment in such environmental conditions is to validate our proposed method is reliable to work in natural environment without much setup or assumptions.

Fig. 5. Photo of the experimental place: Sensor-room

We started the experiment by commanding the robot to traverse the experimental room for mapping the environment and robot localization. After that, we perform map calibration with the static camera. The calibration is needed because when

the fall is detected by the static camera, the fall point will send it to the robot for traveling and assisting the person. Figure 7 shows the human detection system that deployed in the iPhone 6s. Figure 6(b) shows the fall point of a person that detected by the static camera system.

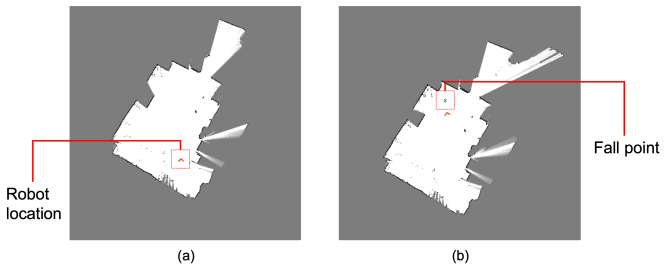

Fig. 6. (a) Example of gridmap generated by the SLAM algorithm [16], (b) Example of the fall point display on the map when a fall is detected

Fig. 7. Screenshot of iPhone of human detection and tracking

A. Fall Detection System

Experiments is setup to test the modified SlowFast neural network with its counterpart that has a self-attention network.

Camera used is supplied by CSUN Co., Ltd, which is run by RK3399 chip with 2GB RAM and Mali-T860 GPU, as shown in Figure 8. It is placed in a prototype smart home at an angle as shown in Figure 9. Though the camera is fixed during run-time, its location can be changed, as long as the angle of view doesn't differ too much. This is the constraint set to allow the classifier to be more fine-tuned to that particular viewing angle. Images are captured at 5fps, which is sufficient for fall detection. 2 seconds clip (10 images where each is resized to 112x112) is continuously fed to the neural network classifier, which has an average processing time within the range of 0.3 to 0.4 seconds.

Both neural network is trained with UR dataset [19] and Multiple Cameras Fall dataset [20], which are then fine tuned by our own collected dataset within the prototype smart home. Before that, the networks are pre-trained with UCF101 action dataset to set good initial parameters.

Fig. 8. Fall Detection Camera, courtesy of CSUN Technology Co. Ltd.

Fig. 9. Fall Detection Camera Setup

Table I shows the generalized test loss for the modified SlowFast Net and one with self-attention. From the result, the network with self-attention is selected for fall classification of the prototype smart home.

Figure 10 shows an example case. The image sequence is read from left to right, top to bottom, where only the last image is a fall. Figure 11 shows fall classification over time for the case, where one can observed a predicted fall. The classifier can also differentiate sitting down, lying and fast movements from fall.

TABLE I
TABLE TYPE STYLES

Net Type	Test Loss
Without Self-Attention	0.6281
With Self-Attention	0.4767

IV. DISCUSSION

Current fall detection system utilizes action analysis to classify a fall. It classifies based on 2 seconds clip, where if the fall occurs within this time frame, a fall is considered to occur. Given training with public database and fine-tuned with on-site videos of daily living and simulated falls, case studies demonstrates that the classifier can differentiates fast movements from falling, as well as movements they involves the person descending to the ground such as lying down or

Fig. 10. Image Sequence of a Case

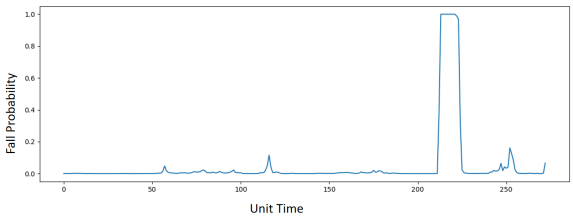

Fig. 11. Fall Classification Over Time

siting. As of now, the fall detector only responds to fall that involves action. Therefore, it will not respond to a victim who has already fallen to the ground before it is being captured.

Due to limited visual range of the robot, the fall detector camera is used by the robot as an additional eye for it to detect, find and assist fall victims. Since the detector is a camera, localization can also be achieved to enable the robot to move to the site of the incident. Such task can be easily achieved given a transformation mapping between coordinate of the fall detector camera and the robot's own map. The victim can be detected using an off-the-shelf human detector from the camera image, or one can use the salient point of self-attention network to estimate the location.

V. Conclusion

In this paper, we proposed a light-weight neural net and implemented to a low-cost 2D camera with on-board GPU for real-time human fall detection. The proposed neural net requires less computational power yet achieve a prominent performance in human fall detection. In addition, we further improved the fall detection system by integrating the assistive mobile robot to compensate the weakness of each of the subsystem.

In future work, we will improve the the neural net performance and deploy the system in more challenging environments for further validation in terms of accuracy and sensitivity aspect.

Acknowledgment

We would like to express our gratitude to CSUN Technology Co., Ltd. for their support in providing hardware as well as engineering consultation. The assistive mobile robot research was supported by the Grand Challenge Grant - HTM (Wellness): GC003A-14HTM from University of Malaya, ONRG grant (Project No: IF017-2018) from office of Naval Research Global, UK and International Collaboration Fund for project Developmental Cognitive Robot with Continual Lifelong Learning (IF0318M1006) from MESTECC, Malaysia.

References

[1] Noury N., Fleury A., Rumeau P., Bourke AK., Laighin GO., Rialle V., Lundy JE., "Fall detection–principles and methods", 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Lyon, 2007, pp. 1663-1666.

[2] McKenna, S. J. and Hammadi N.C., "Learning spatial context from tracking using penalised likelihoods", Proceedings in International Conference on Pattern Recognition, vol. 4, pp. 138–141, January 2004.

[3] Mohamed, O., Choi, H.J., Iraqi, Y., "Fall detection systems for elderly care: a survey in: New Technologies, Mobility and Security (NTMS)", 2014 6th International Conference on, IEEE. pp. 1–4.

[4] Khan, S.S. and Hoey, J. "Review of fall detection techniques: A data availability perspective", Med. Eng. Phys. 2017, 39, 12–22.

[5] Tamura, T.; Yoshimura, T.; Sekine, M.; Uchida, M.; Tanaka, O. A wearable airbag to prevent fall injuries. IEEE Trans. Inf. Technol. Biomed. 2009, 13, 910–914.

[6] Bianchi, F.; Redmond, S.J.; Narayanan, M.R.; Cerutti, S.; Lovell, N.H. Barometric pressure and triaxial accelerometry-based falls event detection. IEEE Trans. Neural Syst. Rehabil. Eng. 2010, 18, 619–627.

[7] Tmaura, T.; Zakaria, N.A.; Kuwae, Y.; Sekine, M.; Minato, K.; Yoshida, M. Quantitative analysis of the fall-risk assessment test with wearable inertia sensors. In Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Osaka, Japan, 3–7 July 2013, IEEE: Piscataway, NJ, USA, 2013; pp. 7217–7220.

[8] Rucco, R.; Sorriso, A.; Liparoti, M.; Ferraioli, G.; Sorrentino, P.; Ambrosanio, M.; Baselice, F. Type and location of wearable sensors for monitoring falls during static and dynamic tasks in healthy elderly: A review. Sensors 2018, 18, 1613

[9] Mastorakis, G.; Makris, D. Fall detection system using Kinect's infrared sensor. J. Real-Time Image Process. 2014, 9, 635–646.

[10] Stone, E.E.; Skubic, M. Fall detection in homes of older adults using the Microsoft Kinect. IEEE J. Biomed. Health Inform. 2015, 19, 290–301.

[11] Sumiya, T.; Matsubara, Y.; Nakano, M.; Sugaya, M. A mobile robot for fall detection for elderly-care. Procedia Comput. Sci. 2015, 60, 870–880.

[12] Martinelli, A.; Tomatis, N.; Siegwart, R. Simultaneous localization and odometry self calibration for mobile robot. Auton. Robot. 2007, 22, 75–85.

[13] Zhang, T.; Zhang, W.; Qi, L.; Zhang, L. Falling detection of lonely elderly people based on NAO humanoid robot. In Proceedings of the 2016 IEEE International Conference on Information and Automation (ICIA), Ningbo, China, 1–3 August 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 31–36.

[14] Mathe, K., Busoniu, L., Vision and control for UAVs: A survey of general methods and of inexpensive platforms for infrastructure inspection. Sensors 2015, 15, 14887–14916.

[15] Saturnino M.B., Cristian I.I., Pilar M.M. and Sergio L.A., "Fallen People Detection Capabilities Using Assistive Robot", Electronics, Vol. 8, No. 9, 2019.

[16] W. H. Chin, Y. Toda, N. Kubota, C. K. Loo and M. Seera, "Episodic Memory Multimodal Learning for Robot Sensorimotor Map Building and Navigation," in IEEE Transactions on Cognitive and Developmental Systems, vol. 11, no. 2, pp. 210-220, June 2019.

[17] Woo, J., and Kubota, N., A Modular Structured Architecture Using Smart Devices for Socially-Embedded Robot Partners. In M. Habib (Ed.), Handbook of Research on Advanced Mechatronic Systems and Intelligent Robotics (pp. 288-309). Hershey, PA: IGI Global, 2020.

[18] Feichtenhofer, C., Fan, H., Malik, J., and K. He, "Slowfast Networks for Video Recognition," In Proceedings of the 2019 IEEE International Conference on Computer Vision; pp. 6202-6211.
[19] Kwolek, B., and Kepski, Michal., "Human Fall Detection on Embedded Platform using Depth Maps and Wireless Accelerometer", Computer Methods and Programs in Biomedicine, vol. 117, no. 3, pp. 4890501, December 2014.
[20] Auvinet, E., Rougier, C., Meunier, J., St-Arnaud, A., and Rousseau, J., "Multiple Cameras Fall Dataset", Technical Report 1350, DIRO-Université de Montréal, Tech. Rep, July 2010.
[21] Feichtenhofer, C., Pinz, A., and Zisserman, A.,"Convolutional Two-stream Network Fusion for Video Action Recognition", In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition; pp. 1933-1941.
[22] Tran, D., Bourdev, L., Fergus, R., Torresani, L., and Paluri, M.,"Learning Spatiotemporal Features with 3D Convolutional Networks", In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition; pp. 4489-4497.
[23] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.C., "Mobilenetv2: Inverted Residuals and Linear Bottlenecks", In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition; pp. 4510-4520