Securing Patient Privacy with a Sturdy Multimodal Monitoring System that Detects Abnormal Behaviour in Hospital Wards Using LWIR and mmWave

Ruhiat Sultana[1], Faizan Ahmed[2], Muhammad AasimUz Zaman[3]
[1]Department of CSE, [2,3]Student,
Lords Institute of Engineering and Technology, India[1,2,3]
[1]ruhiatsultana@lords.ac.in, [2]Mdfaizanffa@gmail.com, [3]160922733020@lords.ac.in

Abstract: In order to identify abnormal health behaviors in hospital wards, we present in this work a multimodal monitoring sensing system that combines mmWave radar and LWIR camera. Low-resolution LWIR cameras have a high degree of accuracy in posture recognition while maintaining privacy. It cannot, however, be utilized in an area that is invisible, like a ward, where curtains are closed. We want to use mmWave radar, which can sense through thin cloths, to tackle this problem. We develop and put into use a multimodal system with two sensors for environmental robustness and privacy. We suggest a multimodal black-out network dubbed COMBONet, which switches inputs based on confidence while taking conditions and background factors into account, in order to make the best use of both sensors. People who were not trained participants were used as the assessment in our implementation and experiments. It demonstrates its broad applicability in a range of settings.

Keywords: COMBONet, Convolution Neural Network, Resolution, Computer Vision.

I. INTRODUCTION

Since the advent of pandemics like as COVID-19, research on healthcare behaviour recognition—which focuses on observing and evaluating patient behaviours in hospital wards—has become increasingly prominent. Our research focuses on a system for keeping track of patients' aberrant actions, like vomiting, falling, stumbling, and limping [1]. The foundation of behaviour recognition research is a variety of sensors, which fall into two categories: wearable and non-wearable [2]. But there are drawbacks, such consumer annoyance, with wearable technology. IR and RGB vision cameras are the core components of non-wearable gadgets. However, because high-definition photographs are being continuously recorded in this instance, privacy protection is an issue. In order to address this issue, research has been done on the classification of falls and everyday activities in indoor spaces utilizing 3D Convolution Neural Network (CNN) in conjunction with low-resolution devices like Long-wavelength Infrared (LWIR) [3]. Furthermore, studies are being conducted to identify behaviour using CNN and extract point clouds utilizing mmWave with high-frequency bands [4]. Research is also being done on multimodal detection systems that use sensors like RGB, Depth, and IR and have excellent accuracy and environmental robustness [5]. In contrast to earlier research, our suggested study develops multimodal detection methods that preserve privacy by utilizing non-intrusive and non-wearable technologies.

II. PROPOSED MULTIMODAL MONITORING SYSTEM

In order to preserve privacy and environmental resilience, we suggest a multimodal monitoring system that makes use of an ISK6843 mmWave radar and a non-intrusive Lepton 3.5 LWIR camera. The proposed system's overview is displayed in Fig. 1. A deep learning server and a sensing device make up the suggested system. The 1.7-meter-tall sensor device in the ward detects a person's behaviour up to three meters in front of it. Additionally, the deep learning server suggested using the Confidence-based Multimodal Black Out Network (COMBONet) to build a system that can withstand a variety of environmental shocks.

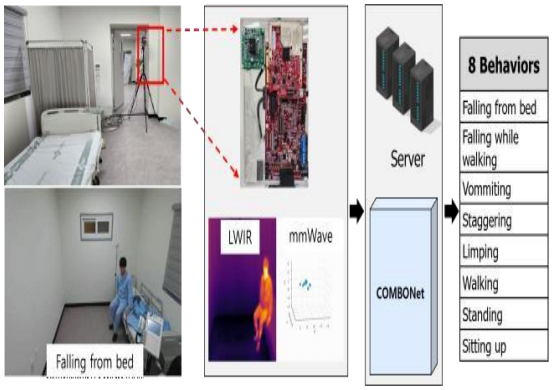


Fig.1.Theproposedmultimodalsensingsystem

A. COMBONet

As seen in Fig. 2, the suggested COMBONet is a two-stage approach. Stage 2 is the deep learning network stage for inference, and Stage 1 is the data pre-processing stage.

Using the LWIR video, we trained the YOLO-X [6] network in stage 1 to determine an individual's bounding box. The inferred confidence threshold value is utilized as a network input for deep learning using a method that resizes the sizes of different persons to generalize, provided that it is specified and greater than a given probability. In order to replicate human features and offer versatility, 3D Skeleton for mmWave is extracted by learning mmWave Point clouds over CNN based on Azure Kinect's 19 skeleton point ground truth. After that, it is

transformed into a 2D Skeleton image map so that a multimodal deep learning network can use each point's xyz value as an input.

Stage 2 involves extracting the feature vector, the converted LWIR video as a 3D CNN, and the 2D Skeleton image map as a 2D CNN. Moreover, Multi-Layer Perceptron (MLP) feature fusion is carried out using methods like Concatenation, Multiple, and Max to forward inference into the classification network.

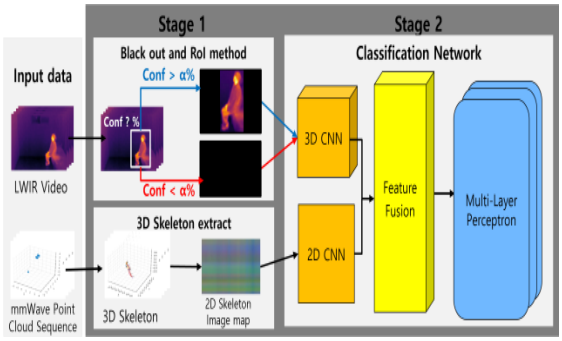


Fig.2.Theproposedtwo-stage COMBONet inference process

B. Training and Inference process

A system called COMBONet has the ability to replicate the features of a single sensing system within a multimodal sensing system. Three combination data total are inputted into learning after the raw LWIR and mmWave data are processed in ROI. As shown in Fig. 3, the two data are made with a value of zero data.

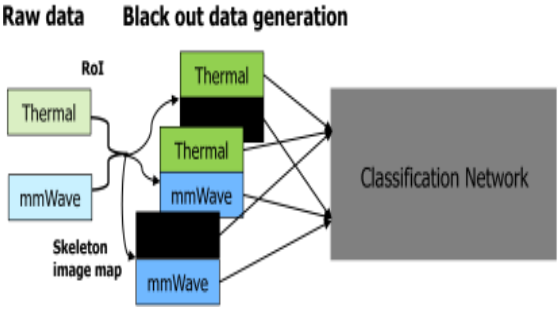
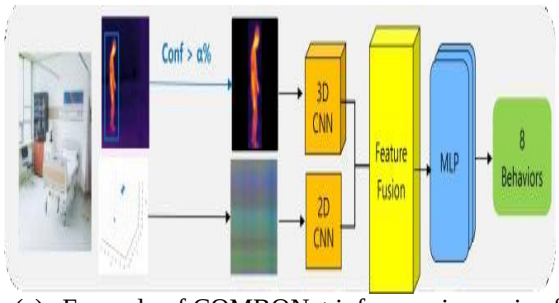
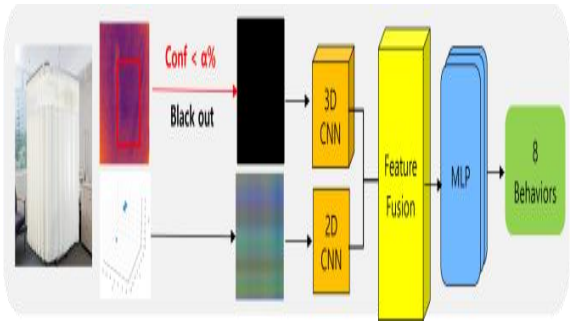


Fig.3.The proposed COMBO Net training process

Fig. 4 illustrates how the outcomes of this learning are inferred. The curtain is not drawn, as seen in Fig. 4(a), and the bounding box's dependability score for identifying the LWIR individual is high. Consequently, multimodal inference is achieved by a combo of LWIR and 2D Skeleton map, which are ROI as inference inputs. The scenario when the curtains are closed and the LWIR's bounding box is ineffective at detecting individuals is depicted in Fig. 4(b). As a result, the data is altered to a value of 0 by a predetermined threshold value, and the bounding box searching for a person has a reduced dependability rating. This allows the transformed 2D skeleton map and the value of 0 for the blackout data to be utilized together as an input for multimodal inference. Following that, the inferred result value might accurately represent the unimodal's features.



(a) Example of COMBONet inference in a noise-free environment



(b) Example of COMBONet inference in curtained non-visible environment

Fig. 4. Examples of COMBONet's inference process

III. EXPERIMENTS

A. Dataset

Six males and four women provided the dataset, which was divided into eight behaviors in total—five abnormal behavioral positions and three daily behavioural positions. Every behavioral posture was recorded for five to eight seconds at a frame rate. Three environments were divided into the data collection: one with noise, one without, and one with curtains that covered the entire space.

B. Evaluation method

Nine out of ten participants were trained to recognize people and assess the effectiveness of a robust model in an external setting. The classification accuracy of the tenth subject was assessed. Additionally, the unimodal [4] and multimodal [6] research that has already been done was compared to the baseline in order to assess the correctness of the three environment data.

C. Results

Figure 5 displays the models' accuracy in relation to the baseline system and in various situations. Unimodal, Multimodal, and the suggested system all performed similarly in the noise-free environment. The ROI approach was used to demonstrate a 24.46% performance gain over the multimodal baseline in the situation of noise. When comparing the performance of ultimately invisible to the multimodal baseline, it improved by 37.04%. This can be considered a contributing component to the black box learning method used by COMBONet.

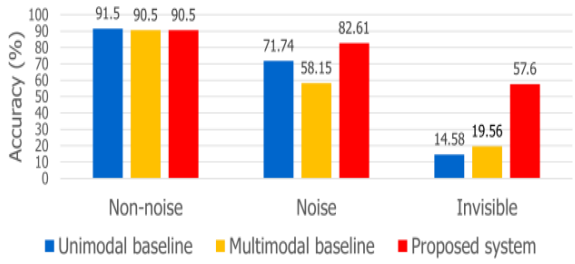


Fig.5.The accuracy comparison to baselines in 3 circumstances (Noise-free, Noise, Invisible)

IV. CONCLUSION

In order to recognize human activity, this work proposes a COMBONet resilient to human and background features employing LWIR and mmWave. In order to conduct experiments on the network resilient to background features, a new evaluation environment was built. By putting forth ROI approaches, COMBONet can achieve significant performance improvement in non-visible situations and high performance improvement over accuracy in noise environments without any RoI. This research is the first to develop a robust model in a background that can be used in a variety of situations using LWIR and mmWave as multimodal techniques.

REFERENCES

[1]. Z.Wang, V.Ramamoorthy, U.Gal, and A.Guez, "Possible Life Saver: A Review on Human Fall Detection Technology." Robotics, Vol. 9(3), pp.55, July2020.

[2]. M.H.Arshad,M.Bilal,andA.Gani,"Human Activity Recognition: Review, Taxonomy and Open Challenges." Sensors, Vol. 22(17), pp.6463, August2020.

[3]. K.Dae-Eon, J.BongKyu, and K.Dong-Soo, "3D Convolutional Neural Networks Based Fall Detection with Thermal Camera." The Journal of Korea Robotics Society13(1),45-54,February2018.

[4]. A.Sizhe, and U.Y.Ogras, "Mars: mmwave-based assistive rehabilitation system for smart healthcare." ACM Transactions on Embedded Computing Systems(TECS),vol.20(5s),pp.1-22,September2021.

[5]. A. M. De Boissiere and R. Noumeir, "Infrared and 3D Skeleton Feature Fusion for RGB-D Action Recognition," in IEEE Access, vol. 8, pp.168297-168308, September 2020.

[6]. G.Zheng, L.Songtao, W.Feng, L.Zeming, and S.Jian, "Yolox: ExceedingYoloSeriesin2021."arXivpreprintarXiv:2107.08430,August2021