Multisensor Low-Cost System for Real Time Human Detection and Remote Respiration Monitoring

Arturs Ivanovs* †, Agris Nikitenko*, Mario Di Castro† ‡, Toms Torims* †, Alessandro Masi† and Manuel Ferre‡
*Riga Technical University, Kalku iela 1, LV-1658, Riga, Latvia
Email: agris.nikitenko@rtu.lv, toms.torims@rtu.lv
†European Organization for Nuclear Research, Route de Meyrin, 1211 Geneva, Switzerland
Email: arturs.ivanovs@cern.ch, mario.di.castro@cern.ch, alessandro.masi@cern.ch
‡Centro de Automatica y Robotica-CAR UPM-CSIC, Universidad Politecnica de Madrid, Madrid 28006, Spain
Email: m.ferre@upm.es

***Abstract*—The most important features of autonomous Search And Rescue robots are abilities to autonomously detect victims and assess their basic vital parameters, such as respiration and heartbeat status, by using their on-board sensors to classify survivors according to their need of medical care. This paper presents a novel sensor composition for autonomous victim detection and non-contact respiration monitoring with SAR robots having limited on-board computational power, using a combination of commercial low-cost components - a visual sensor and Ultra-Wide Band radar. In the proposed method, a pre-trained neural network (*MobileNet*) is used to process camera frames and detect human presence in real-time. Once the victim is localized, the radar is used to perform respiration monitoring. The proposed method is evaluated by building a prototype and performing measurements on volunteers in different positions, clothing and amount of subjects in frame.**

***Index Terms*—UWB, human detection, vital sign, respiration detection, search and rescue**

I. Introduction

Rapid algorithm and sensor development for robotics applications has attracted attention to develop autonomous Search And Rescue (SAR) robots for various tasks such as mapping, object detection and medical assessment. Use of autonomous reconnaissance robots for victim localization could be a useful extension for SAR teams. After the position of victim has been estimated, robot could activate its on-board sensors for victim status monitoring and automatically send data (e.g. respiration rate, temperature distribution) to the command center for evaluation.

Despite the attempts, autonomous human-seeking SAR robots are still in prototyping phase. Most pressing shortcomings according to explored literature are following:

1) limited on-board computational power limits use of computationally expensive algorithms, furthermore, upgrades to higher resolution equipment will exponentially increase requirement for computational power (i.e. computational complexity of Fast Fourier Transform for radar data processing is $O = n\log(n)$);
2) disaster environments are simulated for testing purposes and often lack the real life complexity;
3) absence of sizable, publicly accessible training data taken in real disaster environments prevents researchers to deliver more robust solutions.

In this study we address the aforementioned problems by presenting a work-flow for SAR robots to perform computationally efficient autonomous human detection and remote vital parameter acquisition using a combination of commercially available visual sensor and radar. A prototype of proposed system is built to demonstrate its ability to work on mobile platforms with computational, size and energy restrictions by combining the best features from visual sensors and radars. The contributions of this work include work-flow for autonomous human detection and remote vital sign acquisition using a combination of visual sensor and Ultra-Wide Band (UWB) radar and system performance evaluation using a demonstration prototype for different human postures, clothing and distances.

II. State of the Art

The most used sensors for object detection tasks are visual sensors, RGB-D sensors, infrared cameras and radars.

A. Visual and Infrared Based Detection

Meisel [1] proposed a real-time detector based on Haar features for face detection in RGB images. The approach provided considerable results in face detection, however amount of fake detections was dependent on background texture. Setjo et al. [2] used the same approach for infrared images. Their results were highly dependent on distance (intensity of infrared emitted drops across distance), face orientation and reflectivity of background (can create fake detections).

Andriluka et al. [3] used a combination of two Histogram of Oriented Gradient (HOG) detectors to perform autonomous exploration of disaster sites using RGB images from Unmanned Aerial Vehicles (UAV's). It was concluded, that a combination of given detectors performed better than each of them separately, while providing real-time performance with limited computational resources and low amount of necessary training data. Qi et al. [4] used HOG detector for pedestrian detection in infrared images and achieved 65 % log-average miss rate. However, HOG feature utilization has limited use in SAR tasks due to high variety of camera orientations and postures the victim may have.

Saputra et al. [5] proposed ROI selection and victim body detection based on template matching using RGB-D sensor, however an important drawback this approach is the need for an extensive library with different body configurations for a reliable detection.

Oliveira et al. [6] used Convolutional Neural Networks (CNN) for victim body part detection in RGB images, whereas John et al. [7] applied the same technique for pedestrian detection in infrared images. The computational complexity of CNN is much higher than for any traditional detectors, however it can be justified by the high performance of detectors (in [6] the worst detection rate is for the legs with 64.91 %, whereas in [7] log-average miss rate is 34 %). Rapid computational power cost decrease and technology miniaturization makes neural network based algorithms the best choice for human detection in SAR applications.

B. Radar Based Detection

Kocur et al. [8] demonstrated the advantage of using a mobile robot equipped with UWB radar, that navigated towards a lying victim and detected respiration signs inside an obstacle area filled with smoke. Nezirovic et al. [9] used UWB radar to perform person presence detection at 1.5m based on respiration activity. Respiratory data were compared for different postures, where the best signal was acquired when human was with chest towards radar. Both papers confirm feasibility of human detection using UWB radar at a several meter distance. Chen et al. [10] reported detection of heartbeat and lung activity at a distance more than 30m across free space and also behind a cinder block wall using a 10GHz CW radar. A major drawback of CW architecture, however, is inability to detect distance to the object, also simultaneous multiple target acquisition requires complex signal processing. A important feature for radars is micro-Doppler signal processing for target classification with machine learning algorithms. However, a drawback is the necessity for data accumulation (up to couple of seconds) before the time-dependent Doppler signal analysis can be performed.

Reviewed literature identifies the problems associated with one type sensor usage, therefore encourages sensor and data fusion as one of solutions to overcome limitations set for victim detection in complex environments.

C. Remote Vital Sign Acquisition Techniques

The most used remote vital sign acquisition techniques are done via thermography, video- and radar-based motion analysis. The first method acquires data about blood and respiration flow by observing frequency of temperature oscillation around regions of interest, i.e. neck and nose area. However, the infrared camera needs to have high accuracy and resolution to record small periodic temperature changes, which increases the costs of using this method. Video-based motion analysis exploits the same phenomena as radars - body vibration due to respiration or heartbeat, however the quantification of vibration is done with a visual sensor, which makes the measurement accuracy highly dependent on surrounding environment conditions (i.e. lighting conditions, particle presence in the air). Radar-based lung and heart activity monitoring methods do not possess aforementioned drawbacks and due to rapid radar advancement, there are simple and low cost respiration sensors are already available in market, which makes radars the most suitable technology for vital sign acquisition in SAR scenarios. However, radar-based respiration sensors are not yet medically approved, which leads to conclusion, that their performance is still limited.

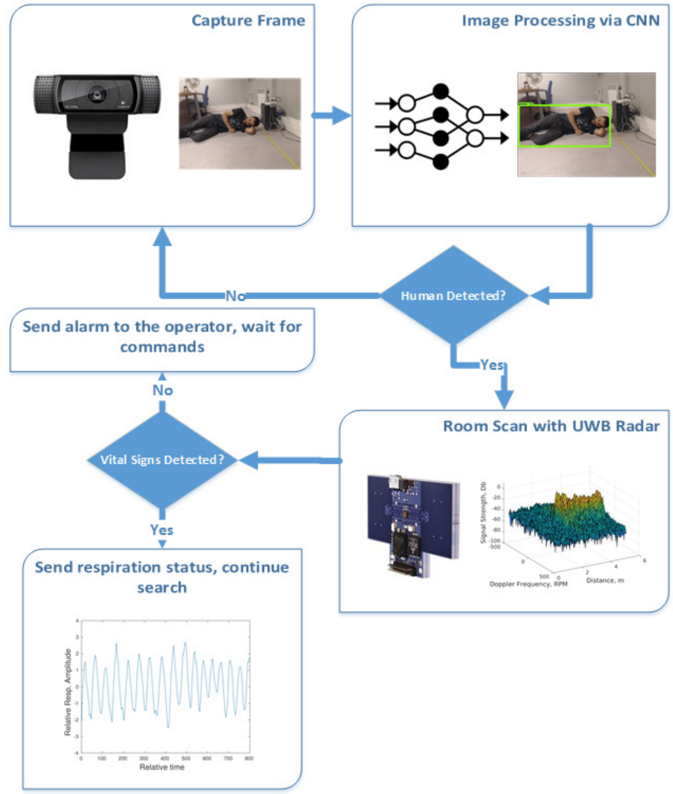

Fig. 1: Work-flow of the proposed method.

III. Proposed Approach and Experiment Setup

In this paper, we suggest the work-flow (figure 1), consisting of two main tasks: human detection and respiration monitoring. We propose to use *MobileNet*, which is a *light-weight* version of CNN [11], to achieve the best human detection performance in varying positions with low computational complexity. Tensorflow repository [12] offers a variety of pre-trained *MobileNet* models on ILSVRC-2012-CLS dataset. In our experiment, the *MobileNet_v1_1.0_224* was used. Visual frame acquisition was done with a *Logitech C920* web-cam.

When human presence in the frame is detected, the radar performs a room scan. The most used radar architectures for respiration monitoring are CW and UWB. We propose to use the latter, due to its ability to deliver not only complex signal phase movements, but also the distance to the oscillating object. For practical evaluation we used a commercially available low-cost System on Chip (SoC) respiration sensor *X4M200* with existing algorithm for human respiration detection. The sensor is very recommended for getting started with radar data processing to explore the possibilities and shortcomings of technology. It also offers easy integration into already existing systems. Detailed specifications of *X4M200* respiration sensor can be found in [13].

A set of experiments was conducted in office environment with a person laying on floor to simulate an unconscious victim in SAR scenario indoors. The volunteer was instructed to breath normally and lay in different positions, which can be classified into five large groups - on spine, on side towards the camera, on side away from it, with head towards camera and then with legs towards at distances from 1 to 5 meters with 1 meter iteration. Additionally, performance was tested with multiple victims in the field of view and also impact of clothing was explored for respiration detection. Sensor system was installed on a tripod at 60 cm height to simulate the view from a ground robot. To achieve the best performance from respiration sensor, before experiments started, respiration sensor was left inside the empty room for 120 seconds to generate a noise map and sensitivity of sensor was put to maximum.

IV. Results

Ratio of positive detection amount over total frame amount per posture and distance is used to quantify the performance of *MobileNet*. Figure 2 presents the overall results, where max contribution of each posture at every distance is 20 % to illustrate overall performance across distance. Only the detections with more than 50% confidence threshold were counted as positive. Overall *MobileNet* performs fairly well (detection rate up to 90 %), considering that model was pre-trained from a general purpose object detection database, whereas experiment was made with humans lying on floor in different positions. The effective recognition distance (detection rate higher than 50 %) is 4m with some positions less favourable than others. This problem could be mitigated by creating a specific dataset of people on the floor. *MobileNet* also performed well when camera captured only a part of body (due to close distance) and in presence of irregularities such as shadow on the leg domain and chair blocking the view on lower part of body. In case of two persons inside the frame, detector had some trouble with adjusting bounding boxes properly, however the number of detected persons was almost always correct.

In parallel to visual frame acquisition, data from the respiration sensor were also recorded. The small lung movements were detected by *X4M200* in all cases. However, table I presents, that reconstruction of respiration profile is highly dependent on human posture. The best results where achieved, when lung movement was directed perpendicularly towards the radar as in figure 3A. By comparison of the Pulse-Doppler maps (fig. 3 C and F) it is possible to confirm, that signal

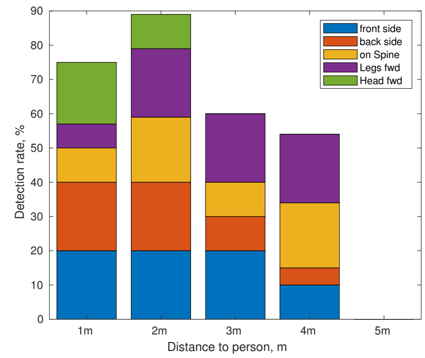

Fig. 2: *MobileNet* detection performance

	1m	2m	3m	4m	5m
back side	24.1 %	23.6 %	56.8 %	2.1 %	8.7 %
front side	72.4 %	85.9 %	50.7 %	75.8 %	0 %
on spine	41.3 %	40.8 %	6.1 %	0 %	0.8 %
legs fwd	8.3 %	0 %	0 %	0 %	0 %
head fwd	15.7 %	9.8 %	10.0 %	0 %	0 %

TABLE I: Respiration sensor detection performance.

strength in the first case is higher than surrounding noise, and has a distinct peak at 2m distance (which is distance to subject's chest), whereas in the latter case respiration signal is almost in the same level as noise. According to manufacturer [13], the maximal respiration sensor detection distance is 5m, which fits well with our experiment results (effective respiration detection is 4m). To evaluate the effect of clothing on the respiration detection, same experiments were conducted again, but this time with the same person wearing two layers of clothing (shirt and a jacket). However, the acquired results with thick clothing mirrored the data with a T-shirt. Respiration sensor was also tested in case of two persons at different distances in front of device. According to manufacturer, the built-in respiration detection algorithm should be able to identify multiple target presence and monitor breathing of only the closest subject, however in our experiments device could not achieve reliable respiration monitoring. On the other hand, visual identification of two peaks was possible in the Pulse-Doppler map with both persons about 1 meter away from each other.

V. Conclusion and Future Work

In this paper we introduced a low-cost system for autonomous human detection and remote respiration activity monitoring based on visual and UWB sensors, that could be used in SAR scenarios. We developed a prototype, based on pre-trained CNN to perform victim detection with a visual sensor and monitor respiration with a UWB radar in real-time. Experiments were carried out to validate the proposed method.

Firstly, we concluded, that *MobileNet* delivers acceptable initial results, considering that model was pre-trained from a general purpose object detection database, whereas experiment was made with humans lying on floor in different positions. The effective recognition distance (detection rate higher than

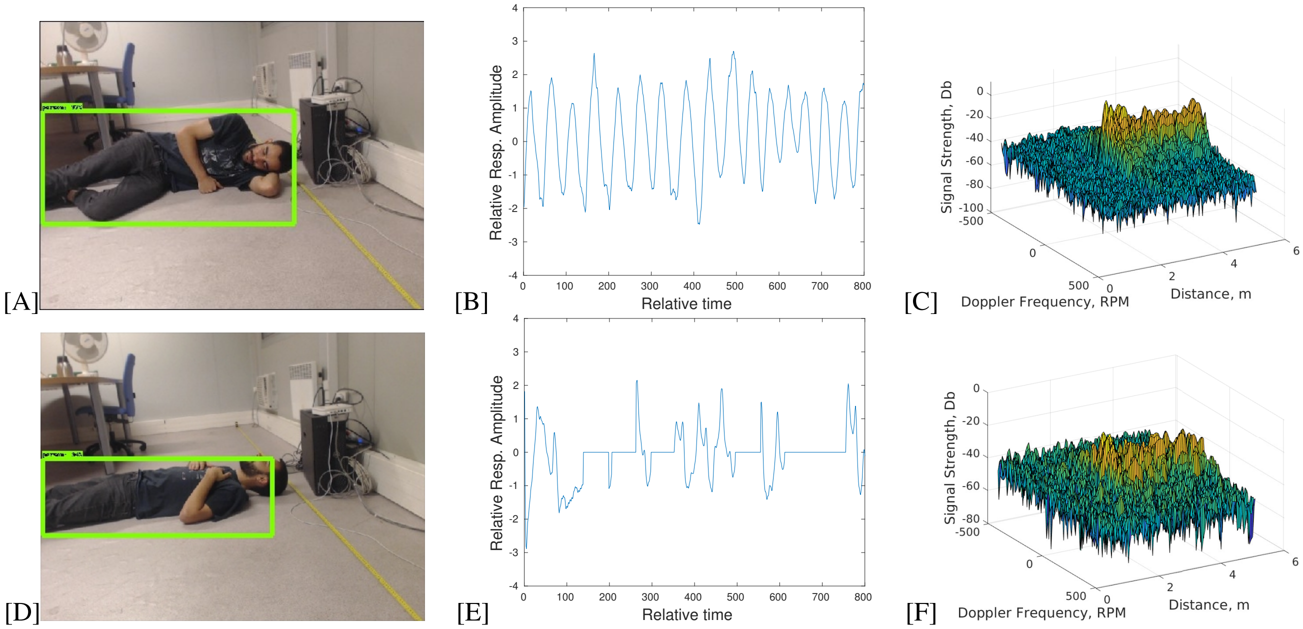

Fig. 3: A, D - different victim postures 2m away from sensor, bounding boxes generated by *MobileNet*. B, E - recorded respiration in time. C, F - Pulse-Doppler map snapshot.

50 %) is 4m with some positions less favourable than others, which could be mitigated by creating a specific dataset of people on the floor and re-training the CNN. Secondly, the victim posture plays a major role not only in visual human detection, but also in respiration monitoring. UWB radar performance is the most reliable in case when person's chest is turned towards radar. A solution for this problem could be using visual sensor for human posture recognition and automatically relocating the radar to a more favourable position for reliable monitoring. Another important shortcoming of the *X4M200* is, that, before being able to detect respiration activity, the device requires about 120 seconds to initialize a noise-map (remove static surrounding objects and filter out movement in the field of view). This problem could be at least partly mitigated by fusing visual sensor information together with radar. Finally, our experiments lead us to a conclusion, that both sensors have their own advantages as well as shortcomings, and, as already stated by multiple researchers, combinations of microwave technologies with other sensors, such as optical or infrared, can provide robust capabilities by performing sensor fusion. No single detection method or system can perform perfectly in all conditions, and for SAR scenarios, where false detections are very costly, the combination of sensors holds a significant potential.

In the future we are planning to focus on visual, infrared and radar data fusion to achieve more robust and computationally efficient real-time human detection and medical field diagnostics in harsh environments for robotics applications.

References

[1] A. Meisel, "Feature based Sensor Fusion for Victim Detection in the Rescue Robotics Domain," no. May, 2014.

[2] C. H. Setjo, B. Achmad, and Faridah, "Thermal image human detection using haar-cascade classifier," in *2017 7th International Annual Engineering Seminar (InAES)*, Aug 2017, pp. 1–6.

[3] M. Andriluka, P. Schnitzspan, J. Meyer, S. Kohlbrecher, K. Petersen, O. Von Stryk, S. Roth, and B. Schiele, "Vision based victim detection from unmanned aerial vehicles," *IEEE/RSJ 2010 International Conference on Intelligent Robots and Systems, IROS 2010 - Conference Proceedings*, pp. 1740–1747, 2010.

[4] B. Qi, V. John, Z. Liu, and S. Mita, "Use of sparse representation for pedestrian detection in thermal images," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 274–280, 2014.

[5] R. P. Saputra and P. Kormushev, "Casualty detection from 3d point cloud data for autonomous ground mobile rescue robots," in *2018 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, Aug 2018, pp. 1–7.

[6] G. L. Oliveira, A. Valada, C. Bollen, W. Burgard, and T. Brox, "Deep learning for human part discovery in images," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, May 2016, pp. 1634–1641.

[7] V. John, S. Mita, Z. Liu, and B. Qi, "Pedestrian detection in thermal images using adaptive fuzzy C-means clustering and convolutional neural networks," *2015 14th IAPR International Conference on Machine Vision Applications (MVA)*, pp. 246–249.

[8] D. Kocur, P. Kažimír, J. Fortes, D. Novák, M. Drutarovský, P. Galajda, and R. Zetik, "Short-range UWB radar: Surveillance robot equipment of the future," *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, no. Jan., pp. 3767–3772, 2014.

[9] A. Nezirovic, A. G. Yarovoy, and L. P. Ligthart, "Experimental study on human being detection using UWB radar," *2006 International Radar Symposium*, 2006.

[10] K. M. Chen, D. Misra, H. Wang, H. R. Chuang, and E. Postow, "An X-Band Microwave Life-Detection System," *IEEE Transactions on Biomedical Engineering*, vol. BME-33, no. 7, pp. 697–701, 1986.

[11] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications."

[12] Mobilenet github repository, accessed 10/10/2018. [Online]. Available: www.github.com/tensorflow/models/blob/master/research/slim/nets/

[13] X4m200 respiration sensor description, accessed 10/10/2018. [Online]. Available: https://www.xethru.com/x4m200-respiration-sensor.html