2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)

Activity Recognition from Sensor Fusion on Fireman's Helmet

Sean Hackett, Yang Cai and Mel Siegel

Carnegie Mellon University
4720 Forbes Ave. Pittsburgh, PA 15213, USA
Contact email: ycai@cmu.edu

Abstract - **Recognizing human activities in emergency situations is critical for first responders to ensure their safety and well-being. In many cases, the thick smoke in a burning building impairs computer vision algorithms for activity recognition. Here we present a helmet-based sensor fusion method with IMU and time-of-fly laser distance sensor. We use a Decision Tree to as a classifier and select the most significant features. Our test shows that the method can recognize over seven activities: walking, running, crawling, duck walking, standing, walking upstairs and downstairs, with an accuracy between 81.7%. and 93.6%. With limited training data and a lightweight requirement for implementation on the fireman's helmet the Decision Tree provided an accurate and reliable result. The use of the 1-D Lidar, which is not feasible in typical activity recognition application but essential for the helmet, combined with the 10-DOF IMU sensors improved the robustness of the classifier. We found this sensor fusion approach needs much less training data, compared to methods such as Deep Learning. Once implemented on the helmet the activity recognition is executed in real-time at sampling rate at 50 Hz within a 2-second window.**

Keywords: activity recognition; gesture recognition; wearable sensors; sensor fusion; augmented reality; helmet; IMU; pitch; time-of-fly laser; distance measurement; first response; fire-fighting; decision tree

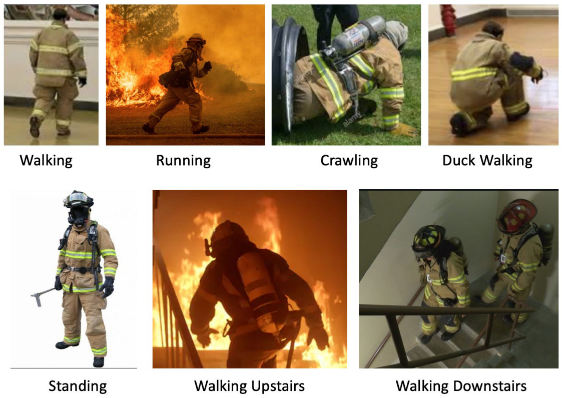
Index Terms: human-computer interaction; sensor; decision-making; classification; pattern recognition

1 Introduction

Firefighting takes place in an extremely hazardous environment where visibility and communication are often very poor. Recognizing firemen's activities is critical to ensure their personal safety and mission status. It also helps to assess firemen's health condition and location. In rescue missions, firemen often move forward in a low profile, such as crawling and duck walking, in order to see and reach victims. Combining normal activities such as walking, running, standing, downstairs and upstairs, we have at least seven activities in this study. Figure 1 illustrates the seven basic activities of a fireman.

In many cases, the thick smoke in a burning building impairs computer vision algorithms for activity recognition. Wearable inertial motion unit (IMU) sensors are a reasonable alternative. There are many locations where we can embed motion sensors, for example, on knee caps, gloves, shoes, or belt. In this study, we embed the sensors in the most common utility of firemen: helmet.

We have developed a prototype smart helmet, called "hyper-reality helmet." The term "hyper" means it provides more real-time information than ordinary systems. The helmet contains multiple sensors such as an IMU sensors, thermal camera, color camera, Lidar, an onboard processor, and a heads-up display (HUD) that provides real-time on-demand information to the fireman and data communication to other first responders. In contrast to prevailing augmented reality technologies, which overlays virtual reality objects on top of the real world scene, this hyper-reality helmet fuses multiple real-time sensory data and projects the data on the HUD. Utilizing these sensors on the hyper-reality helmet enables mission critical information from the firefighter to be shared with other first responders monitoring the scene so they can make more informed decisions during missions and improve the safety of the firefighters and victims. This could help reduce the 72 firefighters deaths and 64,929 injuries that occur every year worldwide [1].



Walking Running Crawling Duck Walking

Standing Walking Upstairs Walking Downstairs

Figure 1. Activities of emergency responders: 1) walking, 2) running, 3) crawling, 4) duck walking, 5) standing, 6) walking upstairs, and 7) walking downstairs

Real-time activity recognition with multiple sensors is a challenging task because of variation of poses, body shapes, heights, and calibration methods. In contrast to prevailing activity recognition algorithms such as machine learning, we use sensory fusion approach in this study. Our empirical results show that the proposed method can

978-1-7281-4852-6/19/$31.00 ©2019 IEEE

recognize the seven firemen's activities with a decent accuracy. It needs less training data than deep learning algorithms and it is relatively robust in terms of diversity of body types, genders, and spatial and temporal dynamics.

2 Related Work

Activity Recognition has been applied in healthcare for monitoring the elderly and in the fitness sector with peoples rising desire to track their workouts and calorie expenditure [2]. These devices are typically worn on the arm or wrist to increase sensitivity and to minimize intrusion to the wearer [3].

Typical methods for activity recognition involves the use of 3 axis accelerometers for data collection with some work also using gyroscopes or complete IMU's. Work in [4] uses a belt mounted 9-DOF IMU to classify pedestrian activity using a simple linear classifier achieving an accuracy of 91%. There are many classification methods for activity recognition. Among the most popular classification methods are Support Vector Machines (SVMs) and they are often used as a baseline for other methods [5]. Using the multiple sensors available in IMU's and applying sensor fusion can improve the accuracy of activity classification [6].

Dynamic Time Warping (DTW) is a method which can be integrated into other classification methods to improve accuracy, as used in [7]. It can be used to detect similarities in sequences occurring at different speeds, e.g two people walking at different paces. However it does have a high computational cost. DTW has been frequently used for comparing gestures in terms of spatial and temporal data sequences in 2D videos or 3D kinetic scan data [14]. The 3D data representations often use a skeleton to describe the dynamics of joints and connections [15], which provides gesture or activity features that contain much less data than the raw data such as pixels or voxels. Relatively speaking, *gesture recognition is an essential part of activity recognition*. A gesture is like a phrase; an activity is like a sentence, which contains more information and takes longer time to recognize. Gesture or activity recognition in a normal environments such as the home and office have been fairly successful, especially with MS XBox Kinect ™ and its variations such as Intel's RealSense ™. However, they are hardly survivable in low visibility environments.

Recently, there has been an increasing popularity in Deep Learning as a classification tool in the area of activity recognition [8]. Activity features are extracted automatically from the data instead of being manually calculated e.g. statistical measures such as *mean*, *standard deviation*, *peak-to-peak*, *root mean squared*, etc. This makes it a more a more general purpose classification tool. However classification with Deep Learning requires large data sets which is problematic for a developing project, and the reason for classification may not be clear or explainable to humans.

Sensor fusion has been a classic topic in system engineering, such as robotics and remote sensing [10-11]. Typical algorithms include Voting Logic Fusion, Bayesian Inference, Dampster-Shafer Evidence Theory, and Artificial Neural Network. In order to fuse the imperfect sensory data, fuzzy logic, fuzzy neural networks, and coherent processing techniques are developed [10]. In multi-modal sensor fusion applications, Mutual Information method has been wided adopted, especially in multi-modal medical image registration, recognition, and tracking, e.g. fusing CT data with MRI data [12-13]. Mutual Information method is based on Information Theory that seeks reduction of entropy of the multiple data sources. The methods discussed above normally need significant large size of data to build the sensory fusion models.

In this study, we have very limited size of data about the seven activities but many sensors available on the hyper-reality helmet. Our goal is to fuse multi-modal sensors on the helmet for activity recognition in real-time.

3 Hyper-Reality Helmet and Multi-Modal Sensors

The hyper-reality helmet sensory system contains a 1-D Lidar for distance measurement up to 40 m, a 10 DOF IMU sensor, including 3 axis gyro, 3 axis accelerometer, 3 axis compass, and altimeter, shown in Fig. 2. The hyper-reality helmet system also includes a quad core GPU processor and a projection heads-up display. The sensors are placed on the front brim of the fireman's helmet. The Lidar and relative axis of the accelerometer and gyroscopes are shown in Fig. 3. Fig. 4 is the helmet prototype.

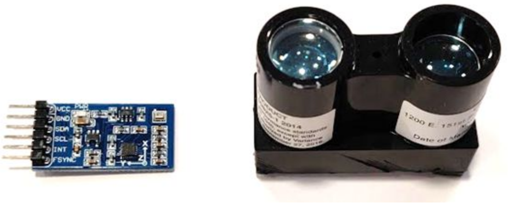


Figure 2. WaveShare 10-DOF IMU sensor (left) and time-of-flight 1D Lidar distance sensor (right)

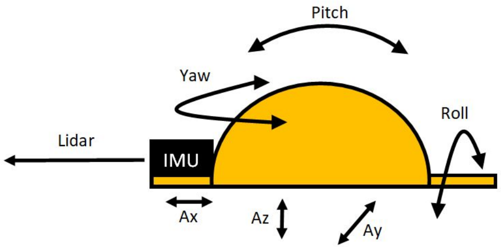


Figure 3. Position of 3-axis Accelerometer, 3 axis Gyroscope and Lidar in relation to the helmet

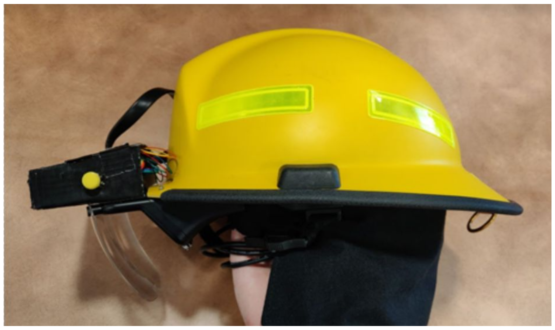


Figure 4. Hyper-Reality Helmet prototype with embedded multiple sensors, heads-up display and onboard processor

4 Data Collection Procedure

The data collection took place in an indoor environment at the Collaborative Innovation Centre at Carnegie Mellon University. A total of 6 participants completed the following activities; *walk*, *crawl*, *upstairs*, *downstairs*, *run*, *stand*, *duck walk* for an approximate duration of 30 – 45 secs for each activity. A participant collecting data while crawling is shown in Fig. 5 The participants completed the activities at their own pace. A push button on the helmet was pressed at the beginning and end of each activity to act as a marker in the data file for future labelling. The sensor data was collected at 50 Hz on the helmet. These were then transferred to the computer for data analysis. After the analysis, we obtain the recognition model for real-time activity recognition on the helmet.

Figure 5. Participant performing crawling activity for data collection

5 Activity Recognition Methods

The raw sensor data used included the *Pitch*, *Yaw* and *Roll* from the Gyroscope, *Ax*, *Ay*, *Az* from the Accelerometer, and the Lidar Distance. The sensor data was split into 2 second windows of 100 samples with a 50% overlap. The 2 second window was chosen as a good compromise between accurately detecting the activity and a fast update rate. The mean and standard deviation were then calculated from the sliding window and used as predictors for the activity recognition. This resulted in 14 predictors that would be used for the activity recognition. The chosen activity recognition method would need to be implemented on the helmets low power embedded system, so an easily implemented, lightweight method was preferred. ưu tiên These included Linear and Quadratic Discriminants and Decision Trees.

5.1 Quadratic Discriminant

Initial activity recognition focused on categorising walking and crawling as these were the 2 most common activities for a firefighter. It was found that the Pitch and Lidar distance where the 2 most distinguishing features and could be used to accurately separate the classes as shown in Fig. 6.

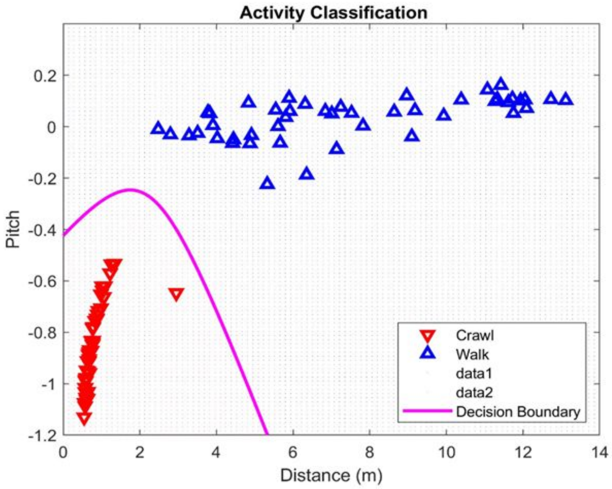


Figure 6. Quadratic Discriminant separating walking and crawling classes

5.2 Decision Tree

With the addition of the remaining 5 of the 7 activities the use of linear and quadratic discriminants became problematic as more sensor data needed to be utilized than pitch and Lidar distance alone. A Decision Tree was chosen as a simple, lightweight, and more scalable solution and that has been successfully used in many activity recognition applications [5]. The Decision Tree features 2 branches per node.

Multiple Decision Trees were created with increasing numbers of branches. *K*-fold cross validation was used for training and testing of the data, with the cross validation error calculated at each stage. From this Decision Tree with the minimum error and lowest number of branches was chosen and a confusion plot generated.

LIDAR: máy quét phát ra các tia hồng ngoại để xác định vị trí của mục tiêu = cách đo lường time mà các tia sáng đó phản xạ lại bằng cảm biến.
Nếu time phản xạ lâu => mục tiêu ở xa & ngược lại.
Tập hợp kết quả các tia sáng và cảm biến trong máy quét sẽ nhận diện được độ nông sâu của môi trường xung quanh nó.

The pseudo code of the process is as following:

```
Import raw Helmet Sensor Data
For person i = 1 to n
    Label the data with activity
    Create sliding window for each activity, 2 second, 50% overlap
    Calculate mean, standard deviation of window
Combine statistics for n users
Plot statistics for each activity
Create Activity tree for i = 1 to 50 branches
    Perform Cross Validation
    Calculate error for each tree
Plot Cross Validation Error and highlight minimum
Plot Confusion Matrix
Plot Predictor Importance
```

6 Results and Analysis

The predictors of mean Lidar distance, pitch, and standard deviation of *Ay*, *Az* were selected as having the most important effect on the activity recognition shown in Fig. 7-10. Many of the distinguishing features between the different activities can be easily visually identified.

6.1 Predictors

The mean Lidar distance in Fig. 7 tends to be low for crawling and duck walking due to been lower and typically facing the ground. The Lidar distance is lower when walking up/down stairs compared to walking as stair wells are more confined than the office environment were walking took place. The distance is higher when going downstairs as you will be facing the bottom of the stairs further ahead rather than the steps directly in front of you when going upstairs.

hướng xuống đất)

hạn chế

quay mặt về phía dưới cầu thang phía trc

các bậc ngay trc mặt

The *pitch* in Fig. 8 is especially useful for identifying crawling as the head is facing toward the floor, compared to upright for all other activities. The pitch is also higher when walking upstairs due to looking up compared to downstairs.

The *y* direction of the accelerometer in Fig. 9 detects side to side motion of the user. This is useful for identifying duck walking due to the large weight transfer to each foot in the squatting position. This is similar but not as extreme in running. There is little to no Ay acceleration detected when standing.

ngồi xổm

The *z* direction of the accelerometer in Fig. 10 detects up and down motion of the user. This is most useful for identifying running. There is little to no *Az* acceleration detected when standing. A higher *Az* for going downstairs vs upstairs and walking due to increased speed that can be achieved.

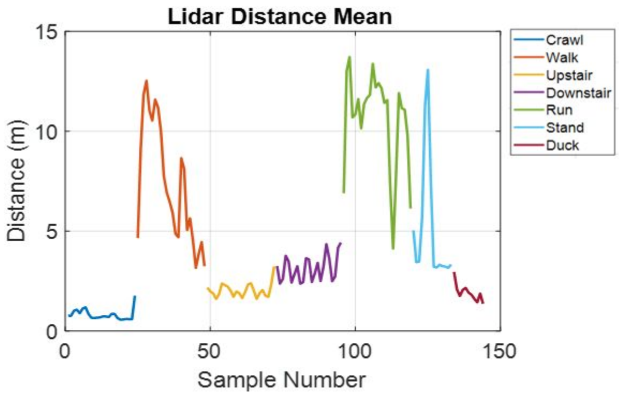


Figure 7. Lidar Distance sliding window mean for activities

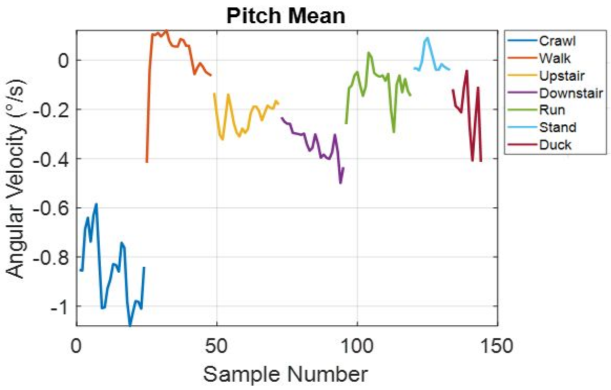


Figure 8. Pitch sliding window mean for activities

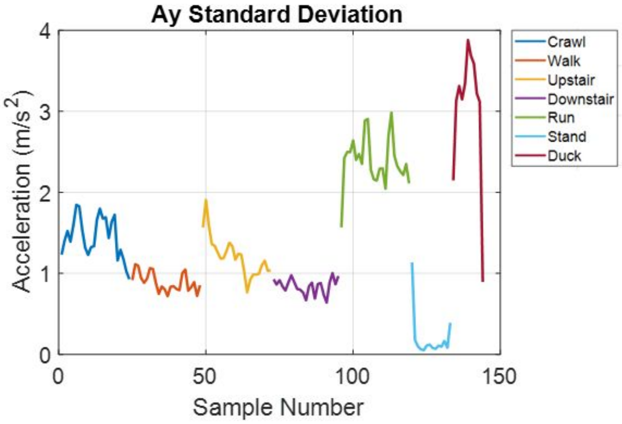


Figure 9. *Ay* sliding window standard deviation for activities

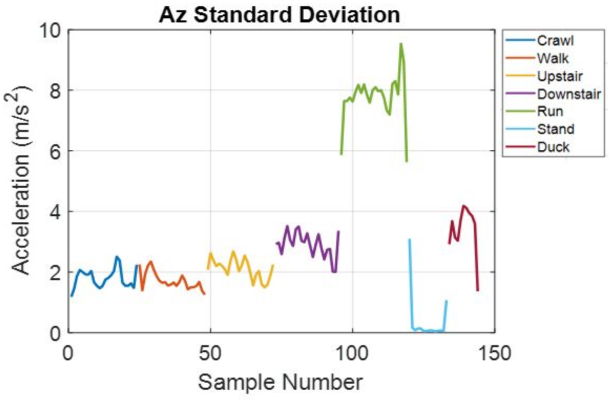


Figure 10. *Az* sliding window standard deviation for activities

The corresponding importance of the predictors is shown in Fig. 11 with Mean Lidar Distance, Pitch, and Standard Deviation of *Ay, Az* were selected as having the most important effect on the activity recognition. Mean yaw, Standard Deviation of yaw, Lidar distance are not included as they had little to no predictor importance. A version of the 8 branch Decision Tree is shown in Fig.12 but a multitude of different combinations of branches could produce similar results.

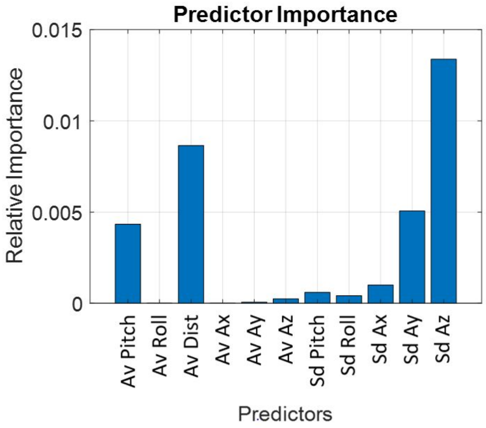


Figure 11. Predictor Importance with Mean Pitch and Distance, Standard Deviation of *Ay* and *Az* the most important

6.2 Accuracy and Error

The 10-fold cross validation was used to train and test the Decision Tree due to small recorded data set. The cross validation error shown in Fig. 13 is used to find the minimum error vs the branch cost. After 8 branches there is no significant decrease in the error which reaches a minimum error of approx 16% or accuracy of 84%.

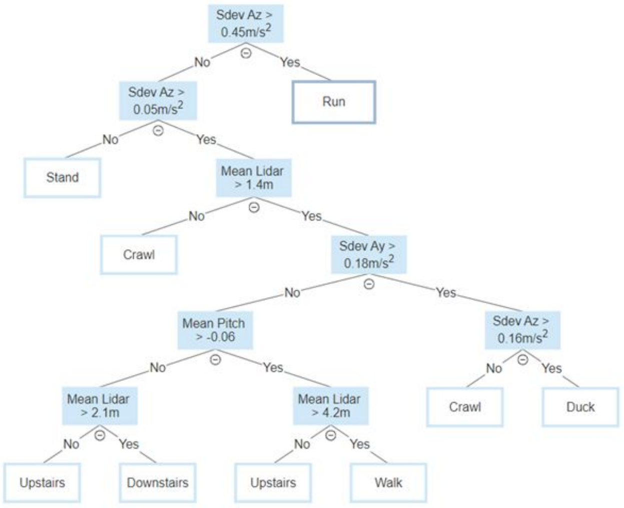


Figure 12. Final 8 branch Decision Tree for the activity recognition

The *confusion matrix* for the data set is shown in Fig 14. The accuracy here of 86% is slightly higher than from the 10-fold cross validation of approx. 84% due to overfitting. Standing, Running, and Crawling are easiest to identify with the highest accuracy. Walking upstairs is most difficult to Classify (81%) as it can be confused with walking and downstairs. The addition of the barometric pressure sensor would be a useful addition due to the pressure decrease as altitude increases.

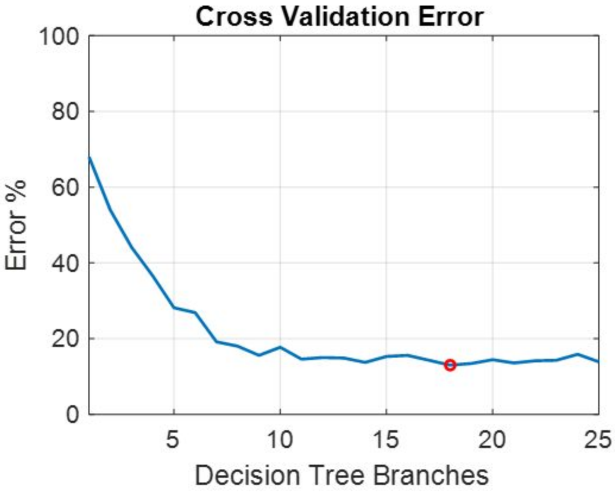
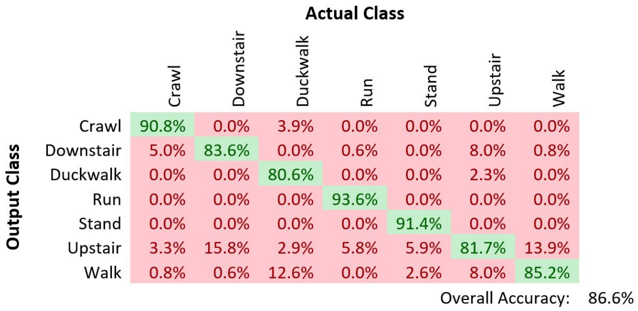


Figure 13 10-fold Cross validation error with a minimum at 18 branches, 8 branches chosen for simplicity as error is sufficiently close to minimum



Actual Class

Output Class	Crawl	Downstair	Duckwalk	Run	Stand	Upstair	Walk
Crawl	90.8%	0.0%	3.9%	0.0%	0.0%	0.0%	0.0%
Downstair	5.0%	83.6%	0.0%	0.6%	0.0%	8.0%	0.8%
Duckwalk	0.0%	0.0%	80.6%	0.0%	0.0%	2.3%	0.0%
Run	0.0%	0.0%	0.0%	93.6%	0.0%	0.0%	0.0%
Stand	0.0%	0.0%	0.0%	0.0%	91.4%	0.0%	0.0%
Upstair	3.3%	15.8%	2.9%	5.8%	5.9%	81.7%	13.9%
Walk	0.8%	0.6%	12.6%	0.0%	2.6%	8.0%	85.2%

Overall Accuracy: 86.6%

Figure 14. Confusion Matrix showing the accuracy of class prediction for the 7 activities

7 Conclusions

In this paper, we presented a Decision Tree classifier utilising sensor fusion to predict 7 different activities with an accuracy between 81.7% and 93.6%. With limited training data and a lightweight requirement for implementation on the fireman's helmet the Decision Tree provided an accurate and reliable result which would not have been possible with a deep learning approach. The use of the 1-D Lidar, which is not feasible in typical activity recognition application but essential for the helmet, combined with the 10-DOF IMU sensors improved the robustness of the classifier.

We found this sensor fusion approach needed much less training data, compared to Deep Learning method. It could be simply implemented on the helmet and executed in real-time at a sampling rate of 50 Hz within a 2-second window.

8 Discussion & Future Work

One of the limitations of the system was the difficulty in detecting the difference between moving upstairs and downstairs. The use of an altimeter would make this task more reliable. The 10 DOF-IMU sensor does feature an altimeter but due to time constraints this could not be implemented for this version of the activity classification. The altimeter will be utilised in the future but will face challenges such as the calibration due to temperature changes affecting the barometric pressure. This is especially important in a fire fighting scenario.

The data collected in the office environment by participants in regular clothes does not replicate a firefighter in full gear in a live scenario. New data will need to be collected for the Decision Tree classifier to be effective here. In addition to the 7 activities in this paper we will also need to add other firefighter activities such as sitting, lying, rolling, carrying a victim, dragging a victim and so on. If these are not possible to identify they should be represented as an additional class “Others”. The addition of these activities to the Decision Tree should be straightforward as it is scalable.

Acknowledgement

This work was performed under the financial assistance award 70NANB17H173 from U.S. Department of Commerce, National Institute of Standards and Technology, PSCR Division and PSIA Program. This project is also funded in part by CMU’s Mobility21 National University Transportation Center, sponsored by the US Department of Transportation. The authors are grateful to the NIST PSCR Program Manager Jeb Benson for his comments and suggestions about the technical development of the hyper-reality helmet system. We also thank Benjamin Graham for his work on the helmet and Weizhe Guo’s assistance with the data collection.

References

[1] N. N. Brushlinsky, M. Ahrens, S. V. Sokolov, P. Wagner, “World Fire Statistics. Technical report 24” Center of Fire Statistics Int. Assoc. of Fire and Rescue Services, 2019

[2] A. Pentland, "Healthwear: medical technology becomes wearable", *Computer*, vol. 37, pp. 42-49, May 2004.

[3] L. Atallah, B. Lo, R. King and G.-Z. Yang, "Sensor Positioning for Activity Recognition Using Wearable Accelerometers," *IEEE Transactions on Biomedical Circuits and Systems,* vol. 5, no. 4, pp. 320 - 329, 2011.

[4] M. Bocksch, J. Seitz, J. Jahn, "Pedestrian Activity Classification to Improve Human Tracking and Localization", Fourth International Conference on Indoor Positioning and Indoor Navigation (IPIN2013), pp. 510-513, Nov. 19, 2013.

[5] T.-K. Woodstock, R. J. Radke and A. C. Sanderson, "Sensor fusion for occupancy detection and activity recognition using time-of-flight sensors," 2016 19th International Conference on Information Fusion (FUSION).

[6] Z.-Y. He and L.-W. Jin, "Activity recognition from acceleration data using AR model representation and SVM," International Conference on Machine Learning and Cybernetics, 2008.

[7] S. Seto, W. Zhang and Y. Zhou, "Multivariate Time Series Classification Using Dynamic Time Warping Template Selection for Human Activity Recognition," 2015 IEEE Symposium Series on Computational Intelligence.

[8] D. Ravi, C. Wong, B. Lo and G.-Z. Yang, "Deep learning for human activity recognition: A resource efficient implementation on low-power devices," 2016 IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN).

[9] O. Lara, M. Labrado, "A survey on human activity recognition using wearable sensors", *IEEE Commun. Surveys Tutorials*, vol. 15, no. 3, pp. 1192-1209, 2013.

[10] L.A. Klein. Sensor and Data Fusion.SPIE, 2004

[11] Y. Cai. Ambient Diagnostics. CRC Press and Taylor and Francis Publisher, 2014

[12] T.M. Cover; J.A.Thomas. Elements of Information Theory (Wiley ed.). ISBN 978-0-471-24195-9

[13] M. Hultter. Distribution of Mutual Information. Advances in Neural Information Processing Systems, 2001.

[14] P. Doliotis, et al. Comparing gesture recognition accuracy using color and depth information. In Proceedings of PETRA 2011, Crete, Greece, 2011

[15] S. Celebi, et al. Gesture recognition using skeleton data with weighted dynamic time warping. In Proceedings of VISAPP 2013, 2013