WiFi-Assisted Human Activity Recognition

Yu Gu*, Lianghu Quan*, and Fuji Ren†
* School of Computer and Information, Hefei University of Technology, China
†Dept. of Information Science & Intelligent Systems, Univ. of Tokushima, Japan

***Abstract*—This paper investigates the indoor activity recognition issue and proposes a novel recognition framework by exploring WiFi ambient signals. The key idea is to use data mining techniques to abstract footprints of different activities on the radio signal strength (RSS) data. Our experiments show that even using a single feature and the common k-NN classifier activities such as walking, sitting and standing can be recognized with a high accuracy, i.e. 75%.**

To further improve the performance, a new feature has been abstracted to represent the fluctuation of sampled data and a novel algorithm named fusion algorithm has been specifically designed based on the classification tree. Experiments show that the proposed fusion algorithm significantly outperforms the k-NN classifier in terms of both the average recognition ratio (from 75% to 92.58%) and the computational complexity. Compared to previous solutions relying on either special hardware or the cooperation of tested subjects, the proposed recognition framework is a passive and device-free solution that could be integrated into any WLAN network with low overheads.

***Index Terms*—WiFi; ambient signals; activity recognition; fusion algorithm;**

I. Introduction

Indoor activity recognition remains a hot topic during the last decades. With the fast approach of Internet of Things (IoT) [1], it becomes an essential underlay service to provide enriched body information. Together with the location service, it potentially benefits various IoT applications such as smart home and intelligent healthcare [2].

The major challenge of activity recognition is to distinguish different activities from sensed data. Most previous research either relies on special hardware, e.g., accelerometer sensors [3] [4], vision-based approaches [5] [6], or demands for the cooperation of subjects [7] [8]. Therefore, their scalability remains questionable, especially in real-world applications.

To this end, we aim to fill in the research gap by proposing a new solution that requires neither sophisticated hardware supports nor the cooperation of tested subjects. More specifically, we explore WiFi ambient signal as the information source for the recognition in the indoor environment. In the closed indoor environment, the WLAN remains stable. Thus, if an object of different activities could block or reflect the WiFi signals in different ways and thus leaves unique RSS marks on the receiver end. If appropriate data mining techniques are applied on the data, footprints of different activities could be abstracted and thus activities could be recognized.

In detail, by exploring WiFi ambient signals in the indoor environment, we abstract distinct fingerprints of four diversified actions: namely, empty, walking, sitting, and standing. Since some activities like sitting and standing have similar patterns, it is very difficult to distinguish them if using only a single feature. Therefore, different features of the data are selected for better capturing the slight differences of activities. Experimental results indicate that the k-NN classifier has a poor performance in terms of recognition ratio, e.g.75%, even using both features. Thus we specifically design a novel fusion algorithm combining the classification tree and k-NN. Extensive experiments have been conducted and the results confirmed that the proposed algorithm significantly outperforms the k-NN in terms of recognition ratio, i.e. 92.58% on average.

Our main contributions can be summarized as follows:

1) First, we are among the first to explore WiFi ambient signals as an efficient meanings to recognize the human activities, to the best of our knowledge. Our solution demands neither special hardware supports nor the cooperation of subjects. Thus, it can be seamless integrated into existing WLAN networks with little overheads.
2) Second, we design a novel fusion algorithm, which combines k-NN algorithm with the classification tree to distinguish the human activities, has been specifically designed to improve the recognition accuracy. Experimental results confirms the superiority of the proposed algorithm over the k-NN in terms of the recognition accuracy and computational complexity.
3) Extensive real-world experiments have been conducted and important insights have been obtained. Valuable hand-on experiences are also provided to help fellow researchers to realize similar frameworks.

The rest of the paper is organized as follows. We present some related work in Section II. The system design is described in Section III.Then, we present our fusion algorithm and experimental results in Section IV. Finally, Section V concludes the paper.

II. Related Works

Activity recognition remains a hot research topic in decades, and its major objective is to explore footprints from historical information for recognizing physical activities [9] [10].

Though the information resources for the recognition could be diversified, the accelerometer sensor is the most

978-1-4799-3711-0/14/$31.00 ©2014 IEEE

Fig. 1. Experimental layout

commonly used component due to its high recognition rate. Bao and Intille developed efficient algorithms to detect human activities from accelerometer data [3]. However, their system requires objectives to wear separated sensors on different parts of the body. With the rapid development of pervasive computing, researchers realize that smartphones with built-in accelerometers are better alternatives. However, unlike sensors attached at fixed positions, smartphones could be carried along the body freely. To this end, Kan *et al.* presented a special recognition algorithm that is independent of the smartphone position [4].

Besides accelerometer sensors, vision information also captures human motions and thus can be explored for recognition [9]. Like the way human eyes detects movements, most research employs visible-light cameras and takes continuous images of motions, from which gradient-based features are extracted for activity recognition. Xia *et al.* proposed an interesting supplementation and utilised the depth information seized by the Kinect device [6] for 3D motions. A recent survey of such approaches can be found in [11].

In summary, research on activity recognition either depends on special hardwares, or requires the cooperation of tested objects. Therefore, the scalability issue becomes a major challenge for the real-world applications. To address the issue, this paper uses the WiFi ambient signals as the information source for the human activity recognition and develop a novel algorithm with high recognition ratio.

Very recently, S. Sigg *et al.* proposed a RF-based device-free recognition system that is closely related [12]. One major difference between our work and theirs is that they used a special tool GNU radio (including both software and hardware) [13] working under 900 and 82.5MHz while we used the more commonly-seen WiFi under 2.4GHz. Therefore, our work is much preferred for real-world applications. Also, we are able to use less features (one or two) to achieve similar recognition accuracy on the same activities by a specifically designed fusion algorithm.

TABLE I
SAMPLING DATE AND DETAILS

Activity	Date	# of samples	# of groups
Empty	14:00-17:00 (2013-11-7)	500	12
	19:00-21:00 (2013-11-8)	500	12
Standing	19:00-21:00 (2013-11-7)	500	12
	14:00-17:00 (2013-11-8)	500	12
Walking	14:00-17:00 (2013-11-9)	500	12
	19:00-21:00 (2013-11-10)	500	12
Sitting	19:00-21:00 (2013-11-9)	500	12
	14:00-17:00 (2013-11-10)	500	12

III. SYSTEM DESIGN

A. Experimental Settings

The experimental layout is shown in Fig.1 Our recognition system consists of one wireless AP (TP-link TL-WR845N) locating at the upper-left corner of the room under $802.11b$ (2.4GHz) and one laptop (Asus N80VM, Ubuntu 12.10) sitting on the bottom-right of the conference table. Experiments have been conducted in one conference room in our institute (room 205, $5m \times 5.5m$), where one subject has performed three activities, i.e., walking, sitting, standing. Additionally, the empty room has been considered as the baseline activity. Note that currently we have not taken the object identification issue into consideration. In other words, the activities have been performed by one selected object and the possible differences between objectives have not been considered yet. But certainly it will be our future work since it would be much more useful and meaningful to recognize different objects performing the same activities. Extensive user-related applications can be launched on the basis of this service.

To enhance the RSS alternation and improve the recognition accuracy, spatial restrictions have been employed and all actions are conducted between the AP and the laptop. To ensure a stable environment during the experiments, the door was closed to exclude the potential external interferences, e.g. passing by subjects.

B. Feature Selection

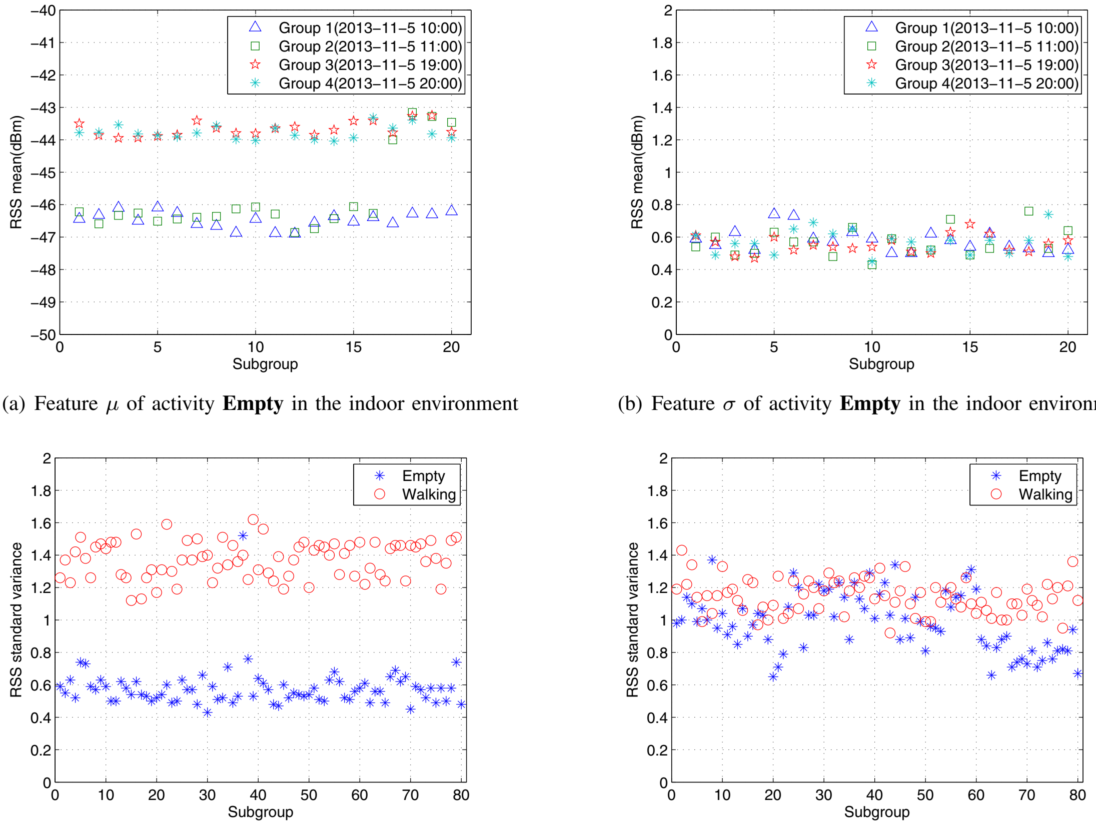
To select suitable features for the classification, we conduct a case study and collect 4 groups of data in the empty conference room over one day (2013-11-5). Each group contains 2000 RSS samples at a sampling rate of one second, corresponding to about 30 minutes. Group 2 are recorded immediately after group 1, while the time duration between group 2 and 3 is set to seven hours. To obtain explainable experimental results, we filter abnormal data, defined in Eqn.1, from the samples first.

$$RSS \ni [\mu - C, \mu + C] \tag{1}$$

where μ is the mean of one group data and C is a predefined constant value, e.g., 3 dBm.

Then, sampled RSS data are further divided into sub-groups with a size of 100, e.g. $[RSS_1, \cdots, RSS_{100}]$. The

Fig. 2. Environment and Feature Selection



(a) Feature μ of activity **Empty** in the indoor environment

(b) Feature σ of activity **Empty** in the indoor environment

(c) Feature σ of **Empty** and **Walking** in the indoor environment

(d) Feature σ of **Empty** and **Walking** in the outdoor environment

mean (μ) and standard deviation (σ) of each sub-group are calculated as the possible recognition features as follows,

$$\mu = \frac{\sum_{i=1}^{100} RSS_i}{100} \tag{2}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{100} (RSS_i - \mu)^2}{100}} \tag{3}$$

Fig.2 illustrates the experimental results. The first investigation is that for each group μ is stable over time, indicating that during the sampling period ($2000s$) the environmental alteration is limited. However, μ changes significantly between group 2 and 3, indicating the background noise could be fluctuated over a long period of time, e.g., several hours. Therefore, μ may be not the best choice. On the other hand, σ shows the potential ability of classifying different activities since it is relatively stable even in a long-term point of view, as shown in Fig.3(b). It is because σ mathematically records variations of sampled data and thus can better filter the environmental changes. In our case, though the background noise may be changing over time, leading to unstable averaged RSS measures (i.e., μ), the relative alternations of the WiFi signal caused by activities remain stable.

To further verify the investigation, experiments are conducted with the same settings. Two activities, empty and walking, are used because the environmental alteration caused by these two activities should be significantly different. One typical case is selected and illustrated in Fig.3(c). Clearly, once using feature σ it should be easier to finish the recognition due to the huge gap between them in the feature domain, and thus σ could be a qualified feature to distinguish activities that lead to significant environmental alterations. Further experimental results confirmed this investigations.

Also, to study the impact of different environments on the RSS data, experiments has been conducted to understand the differences between indoor and outdoor environments, where the laptop is placed outside the conference room (in the corridor). Feature σ is used and Fig.3(d) is a typical example for the outdoor case, where we could see that it is much harder to differentiate the two activities due to stronger interferences caused by uncontrollable factors such as passing-by objects and other WiFi APs. Therefore, for the proposed framework, the indoor environment seems to be more suitable because of its stability.

C. Recognition Process

For each activity, with the same sampling rate (once a second), we obtain 6 data groups as the training data, each consisting of 2000 samples. Abnormal samples have been removed first before the recognition process as in the last subsection. Then, we further divide each group into sub-groups with a size of 20 samples. σ of each sub-group is calculated as the recognition feature. Therefore, for the activities, we have four corresponding sets of training data, namely, $\overrightarrow{P} = \{P_1, \cdots, P_4\}$. For the observed RSS data of the unrecognized activity, we have applied the same processing procedure to get one data set, namely, U. The distances between U and $\overrightarrow{P}$ are calculated and recorded in $\overrightarrow{L} = \{L_1, \cdots, L_4\}$. We use the well-know k-nearest neighbor (k-NN) algorithm as the classifier while setting k to:

$$k = 0.8 \cdot \mu(\overrightarrow{L}) = 0.8 \cdot \frac{\sum_{i=1}^{4} L_i}{4} \tag{4}$$

For each activity, we have performed 24 experiments, and recorded the confusion matrix in Tab.II. For the activity walking, the recognition accuracy is perfect, i.e., 100%. However, other activities are much harder to differentiate: the accuracy is around 60%. For instance, sitting has a possibility of 29.17% to be misinterpreted to empty. The average accuracy for all four activities are 75%.

On one hand, activities like sitting and empty have quite similar footprints of the RSS signal and thus they are difficult to be distinguished. On the other hand, the k-NN algorithm employs one feature only, leading to a lower recognition accuracy. To deal with this issue, in the next section we will first introduce a new feature to capture the variations of the feature σ, as well as a novel fusion algorithm, combining the k-NN classifier with the classification tree, so as to improve the accuracy.

IV. Fusion Algorithm and Performance Evaluation

A. Fusion Algorithm

Since a single feature fails to meet the recognition requirement, we induce another one to enhance the performance: $\sigma^{(2)}$, which is the standard deviation of the collected σ, i.e.,

$$\mu(\sigma) = \frac{\sum_{j=1}^{N} \sigma_j}{N}$$
$$\sigma^{(2)} = \sqrt{\frac{\sum_{j=1}^{N}(\sigma_j - \mu(\sigma))^2}{N}} \tag{5}$$

where N is the number of sub-groups (25 in our case). $\sigma^{(2)}$ aims to record the variation of feature σ, so as to capture tiny differences of activities that have similar RSS footprints.

Fig.3 presents examples of both features of the training data. In general, it suggests that both features of walking

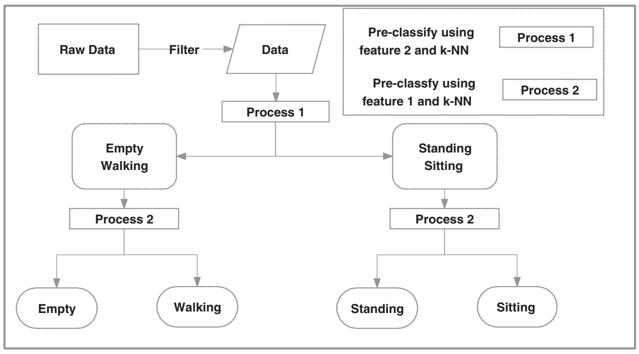


Fig. 4. The fusion algorithm

and standing are the highest while walking is slightly higher than standing on average. Therefore, it is reasonable to treat walking and standing as a cluster (C_1 = {walking, standing}) and other activities as another cluster (C_2). The fusion algorithm runs in the following steps:

S1: Apply Eqn.(3) and (5) to the sampled data and record the features.

S2: Calculate the distances of feature $\sigma^{(2)}$ between the sampled data and training data of activities, i.e. $D(\sigma^{(2)}) = D_1, D_2, D_3, D_4$; Sort $D(\sigma^{(2)})$ in the ascending order, select the first item and find out its corresponding activity. If it is activity Empty or walking, the undetermined activity belong to the cluster C_1. Then go to $S3$. Otherwise it belongs to C_2. Go to $S4$

S3: Calculate the distances of feature σ between the sampled data and training data of activities walking, standing; Then apply k-NN to pinpoint the undetermined activity.

S4: Calculate the distances of feature σ between the sampled data and training data of activities empty and sitting; Then apply k-NN to pinpoint the undetermined activity.

Fig.4 concludes the detailed procedure of the fusion algorithm.

B. Performance Evaluation

To compare with k-NN, we perform the above fusion algorithm to the same set of experimental data, and record the recognition accuracy in Tab.II. One investigation is the significant improvements of certain activities over k-NN: empty(from 79.17% to 91.67%), sitting (from 62.6% to 87.5%) and standing (from 58.33% to 91.17%).

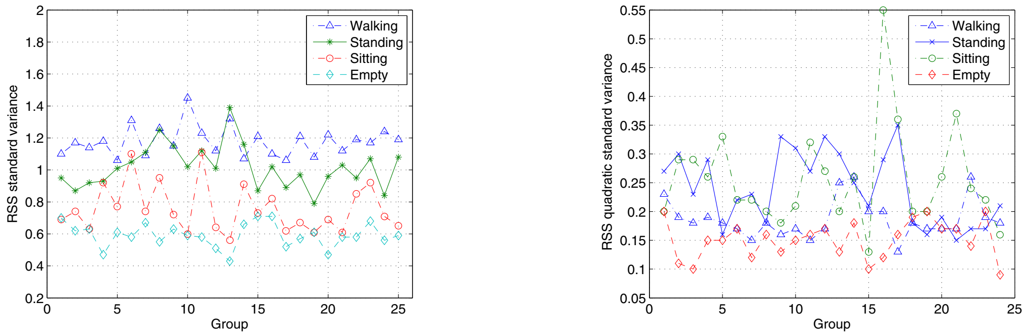
Also, we notice that the fusion algorithm can better filter interferences for one activity from similar ones by layered classifications based on two features. For instance, the activity empty can be misinterpreted into activity sitting with a possibility of 20.83% if k-NN is applied. While the fusion algorithm decreased that possibility to 8.3%.

In summary, through in-depth investigations on both features of activities, a suitable classification tree is designed

TABLE II
PERFORMANCE COMPARISON BETWEEN K-NN AND THE FUSION ALGORITHM

	Empty		Walking		Sitting		Standing	
	K-NN	Fusion	K-NN	Fusion	K-NN	Fusion	K-NN	Fusion
Empty	**79.17%**	**91.67%**	0%	0%	20.83%	8.33%	0%	0%
Walking	0%	0%	**100%**	**100%**	0%	0%	0%	0%
Sitting	29.17%	8.33%	0%	0%	**62.6%**	**87.5%**	8.33%	4.17%
Standing	0%	0%	25%	0%	16.7%	8.33%	**58.33%**	**91.17%**

Fig. 3. Feature σ and $\sigma^{(2)}$ of four activities



(a) Feature σ (b) Feature $\sigma^{(2)}$

to better capture the differences of activities, leading to better overall performance. The average recognition ratio has been improved from 75% to 92.58%.

C. Complexity Analysis

The complexity of the classifier is very crucial to the recognition process since basically it determines the overall running time. Therefore here we present the complexity analysis for the proposed algorithm. Assuming the number of training data for each activity is m and each group has been divided into n subgroups, We have the following conclusion,

Theorem 1. *The computational complexity of k-NN is* $O((4mn)^2)$*, while the computational complexity of the fusion algorithm is* $O((4m)^2 + (2mn)^2)$.

Proof. For k-NN, the number of elements needed to be sorted is $4mn$, therefore its complexity is $O((4mn)^2)$; For the fusion algorithm, we first sort $4m$ items and them sort another $2mn$ items. Therefore the overall complexity should be $O((4m)^2 + (2mn)^2)$.

To gain an intuitive impression, Fig.5 shows the complexity comparison between k-NN and the fusion algorithm, where m ranges from 1 to 100 and n ranges from 1 to 20. Although m and n are much smaller than the usual settings, the complexity differs significantly, i.e. 10×10^7 verse 2×10^7. With the increment of the problem scale, i.e. m and n, our proposed algorithm could save much time compared to the k-NN classifier.

V. CONCLUSION

This paper introduces a passive and device-free activity recognition scheme based on WiFi ambient signals. It can be easily integrated with any existing WLAN network deployed in the indoor environment with low overheads. The proposed fusion algorithm can abstract fingerprints of different activities from the RSS data and thus achieve an average recognition ratio of 92.58%. Important insights have been obtained to guide the system implementation and future research directions.

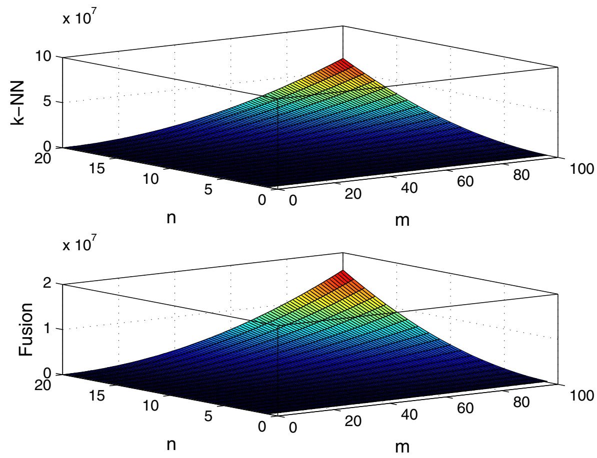


Fig. 5. Complexity comparison between k-NN and the fusion algorithm

REFERENCES

[1] J. Stankovic, “Research directions for the internet of things,” *IEEE Internet of Things Journal*, 2014.

[2] S.-C. Kim, Y.-S. Jeong, and S.-O. Park, “Rfid-based indoor location tracking to ensure the safety of the elderly in smart home environments,” *Personal and ubiquitous computing*, vol. 17, no. 8, pp. 1699–1707, 2013.

[3] L. Bao and S. S. Intille, “Activity recognition from user-annotated acceleration data,” in *Pervasive Computing*. Springer, 2004, pp. 1–17.

[4] A. M. Khan, Y.-K. Lee, S. Lee, and T.-S. Kim, “Human activity recognition via an accelerometer-enabled-smartphone using kernel discriminant analysis,” in *Proceedings of IEEE 5th International Conference on Future Information Technology (FutureTech)*, Busan, Korea, May 2010, pp. 1–6.

[5] J. Sung, C. Ponce, B. Selman, and A. Saxena, “Unstructured human activity detection from rgbd images,” in *Proceedings of IEEE International Conference on Robotics and Automation*, St. Paul, USA, May 2012, pp. 842–849.

[6] L. Xia, C.-C. Chen, and J. Aggarwal, “Human detection using depth information by kinect,” in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Portland, USA, June 2011, pp. 15–22.

[7] M. Ermes, J. Parkka, J. Mantyjarvi, and I. Korhonen, “Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 1, pp. 20–26, 2008.

[8] D. Sánchez, M. Tentori, and J. Favela, “Activity recognition for the smart hospital,” *IEEE Intelligent Systems*, vol. 23, no. 2, pp. 50–57, 2008.

[9] T. B. Moeslund and E. Granum, “A survey of computer vision-based human motion capture,” *Computer Vision and Image Understanding*, vol. 81, no. 3, pp. 231–268, 2001.

[10] P. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea, “Machine recognition of human activities: A survey,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 11, pp. 1473–1488, 2008.

[11] R. Poppe, “A survey on vision-based human action recognition,” *Image and vision computing*, vol. 28, no. 6, pp. 976–990, 2010.

[12] S. Sigg, S. Shi, F. Buesching, Y. Ji, and L. Wolf, “Leveraging rf-channel fluctuation for activity recognition,” in *Proceeding of the 11th International Conference on Advances in Mobile Computing and Multimedia (MoMM2013)*, Vienna, Austria, Dec 2013.

[13] E. Blossom, “Gnu radio: tools for exploring the radio frequency spectrum,” *Linux journal*, vol. 2004, no. 122, p. 4, 2004.