Activity recognition from Radio Frequency data: Multi-stage recognition and features

Shuyu Shi
National Institute of Informatics
Email: shi-sy@nii.ac.jp

Stephan Sigg
National Institute of Informatics
Email: sigg@nii.ac.jp

Yusheng Ji
National Institute of Informatics
Email: kei@nii.ac.jp

***Abstract*—We introduce a novel activity recognition method based on the RF-signal originated from ambient FM radio source. For the purpose of classifying activities, we utilise a two stage approach which can initially distinguish between coarse-grained activities, then make further fine-grained recognition. Additionally, a study on features is conducted to investigate the most suitable combination to achieve the highest accuracy on the detection of activities. By comparing to a one stage classification process, the experimental results demonstrate the advantage of our designed approach.**

I. Introduction

In context and activity awareness, a challenge is to assign suitable labels for activities by extracting features out of data streams from sensors monitoring the entities performing these activities [5]. The attributes and properties of sensors naturally impact the precision of detecting various activities. Prominent examples are sensors for motion detection such as acceleration-type sensors [2], [6], [15] as well as video or RFID-based installations [8], [16]. Alternatively, to avoid the burden of carrying a sensor device, Patel et al. [9] introduced an infrastructure-mediated sensing approach that mitigated the necessity to equip users and also the frequent recalibration of the system. Fluctuations induced by human actions to electromagnetic systems naturally installed in an environment are utilised as a source to establish environmental awareness. In this case, however, the area covered for activity detection is restricted to indoor settings.

Given these restrictions of contemporary sensing technology, an ubiquitously available sensor class from which we can detect activities of individuals without equipping these entities might be the radio transceiver of electronic equipment. We propose to use the RF-channel as a sensing source to detect activities of individuals by monitoring peculiarities of RF-based features induced by these very activities. It is well established that the wireless channel can be utilised for motion detection [7], [21] and localisation [12], [20]. Some preliminary studies have been conducted to classify static environmental alterations such as opened or closed doors, presence of an individual, location and count of persons [11], [10], [13]. In these implementations, a transmitter is deployed to constantly generate the signal received by other nodes. The novelty of our activity classification method stems from relaxation of the requirement for an explicit transmitting node. In this work, we will demonstrate the feasibility of activity recognition by utilising RF amplitude-based features of signals derived from an ambient FM radio source.

Besides the nature of sensing devices, various algorithms, utilised to extract features from the raw data, impact the efficiency of activity recognition as well. Figo et al. [4] have presented a survey for extracting activity information from raw accelerometer data. Similar to his findings for signals from an accelerometer, the amplitude value of RF signals gathered by USRP SDR devices[1] can not be used directly for activity recognition either. In order to support researchers in their study of suitable features for RF-based recognition, we will introduce common features based on the RF signal's amplitude in time and frequency domain. To evaluate which features can characterise the physical activities best, a variety of feature combinations are exploited. All features are chosen with the ultimate aim to optimise the accuracy of activity recognition utilising ubiquitously available FM radio signals.

II. Related Work

We will give a brief overview on the state of the art for the detection of situations and activities based on RF signals. A discussion on the mathematical metrics of processing sensor data is also conducted.

A. RF-based context awareness

In Pervasive Computing, recently several groups focus on detecting situations and activities utilising RF signals following the seminal work of Woyach et al. [19], where the authors detailed the difference of RSSI fluctuations of signals when an object either moved between a transmitter and a receiver, shadowing the signal paths, or moved nearby without crossing the line-of-sight between the transmitter and receiver. Also, it was observed that the trajectory of an object, network topology and geometry of the environment impacted the RSSI fluctuations considered. Finally, after reverting any kind of temporary change in the environment, the authors observed that the variance of the RSSI fluctuation returned to the same measurement values seen prior to the change. Building on these results, several authors [20], [21], [17] have investigated the possibility to detect the location of passive objects by RF-transceivers. Among them, Youssef et al. detected the presence and location of individuals by exploiting 802.11b nodes deployed in the

[1]http://www.ettus.com

978-1-4673-1881-5/12/$31.00 ©2012 IEEE

corners of a indoor environment, continuously transmitting packets [20]. Zhang et al. investigated the accuracy for locating moving objects by RSSI estimates [21]. In that study, they succeeded in locating a moving entity with an accuracy of about 1m and an update frequency of 3.8Hz. Wilson and Patwari utilized radio tomographic imaging on the two-way RSSI variance [17] or RSSI mean fluctuations [18] between nodes arranged in a rectangle surrounding the monitored area for robust localisation of up to two individuals simultaneously.

Rather than location, as acquired in the work above, RF-based detection of activities is conducted in [1], [14]. We can differentiate this work into studies in which the monitored entities are equipped with sensing devices and other studies in which monitored entities are passive. In the former domain, for instance, Anderson et al. [1] implemented activity recognition systems utilizing the fluctuation in GSM signal strength. Alternatively, the first study to report activity detection from passive entities was presented in 2011 by Scholz et al. [11]. The authors describe a system to detect walking, talking on a mobile phone and the state of the door in a typical office room with two software defined radio nodes (transmitter and receiver at 900MHz) placed on both sides of the door of a typical office room. In this paper, we realise activity recognition without an active transmitting component. The novelty that separates our work from previous mentioned studies is that no transmitter needs to be installed in an environment to generate the signal picked up at a receiver. Furthermore, we do not equip the monitored entities with any active transmit or receive components.

B. FM radio source

In our implementation and case studies we decided to use FM radio signals since we believe that they are most widely available both indoors and outdoors. FM broadcasting has a long history of more than 90 years and its coverage spreads almost every country in the world [2]. Also, due to its integration with other wireless technologies on a single chipset and due to the price advantage, the FM radio is becoming increasingly available on mobile devices such as smartphones, tablets and media players. For instance, the Windows Phone 7 platform embraces the FM radio as well as many Android-OS-based mobile phones or the Apple iPhone and iPod Touch devices. Therefore, due to the wide deployment of FM-capable devices and FM broadcasting stations, FM radio signals can be seen as an ubiquitously available signal source so that activity recognition based on FM-Signal sensing is a promising approach for widespread, practical applications.

III. A STUDY ON RF-BASED FEATURES

Utilising the USRP SDR device, we can aggregate raw data containing various characteristics induced by specific activities. However, the raw data can not be directly classified to distinguish various activities. Extracting and interpreting suitable features is essential to the recognition accuracy. We have attempted to extract a wide spread of different, reasonable and important features. Given these features, we explored their impact, either isolated or combined on the precision of recognition. All features can mainly be categorised into time domain and frequency domain.

A. Time domain features

The metrics in time domain are often used to extract signal information from raw data in many practical activity recognition systems. As for our system, these feature metrics include Mean(μ), Median(Med), Variance(Var), Standard Deviation(Std), Root Mean Square(RMS), Central Moments(CM_i) and the count of values that deviate from the Mean by more than twice the RMS in a given time window (Θ). The i^{th} Central Moment can be calculated as

$$CM_i = \mathrm{E}[\zeta_{\text{rec}}(t) - \mathrm{E}(\zeta_{\text{rec}}(t))]^i. \tag{1}$$

Θ can be computed by the following formula

$$\begin{aligned} \theta(t) &= \begin{cases} 0 & ,\mu - 2 \cdot \text{RMS} < |\zeta_{\text{rec}}(t)| < \mu + 2 \cdot \text{RMS} \\ 1 & ,else. \end{cases} \\ \Theta &= \sum_{t=1}^{n} \theta(t). \end{aligned} \tag{2}$$

In both, equation (1) and equation (2), the $\zeta_{\text{rec}}(t)$ stands for the amplitude of superposed signals which the transceiver obtained at the time t. In equation (1), the window size used to calculate the mathematical expectation is set as n as well. In this work, we only utilised the 2^{nd} and 3^{rd} Central Moment. The Mean and Median can both smooth the overall dataset and describe a kind of average value of the raw data. Variance, Standard Deviation and RMS can represent the variability of the dataset and imply the stability of the signal. Θ reflects the frequency at which peak values occur in a certain time interval.

B. Frequency domain features

Frequency domain techniques have been widely used to capture the periodic nature of activities, such as walking, running and cycling. For example, if the time domain pattern of a signal is repetitively centralised around at a fixed interval T_c, the Fourier coefficient will approximately concentrate on $\frac{1}{T_c}$Hz in the frequency domain. In our system, we try to calculate three features in the frequency domain, namely,Direct Current component(DC component), Spectral Energy and Information Entropy. To compute them, the first step is to transform the amplitude from the time domain into the spectral values in frequency domain using, for instance, the Fast Fourier Transform (FFT). The FFT can be represented as

$$\mathrm{X}(i,n) = \sum_{m=1}^{M} x(m,n) e^{-j\frac{2\pi}{N}im}. \tag{3}$$

In equation (3) the value M is the quantity of the samples in the FFT and n represents the n^{th} frame. The value $\mathrm{X}(i,n)$ denotes the i^{th} frequency component in the n^{th} frame.

The DC component is the first coefficient in the spectral representation of a signal and can be calculated as the mean

[2]http://en.wikipedia.org/wiki/FM_broadcasting

value of the signal over the frame. Based on the results of the FFT transformation, the DC component of the n^{th} frame can be calculated as

$$\mathrm{DC}(n) = \frac{\sum_{i=1}^{M} \mathrm{X}(i,n)}{M}. \tag{4}$$

The normalized energy of a signal can be computed as the squared sum of its probability density of spectrum in each frame. The probability of each spectral $\mathrm{X}(i,n)$ band can be computed as

$$\mathrm{P}(i,n) = \frac{\mathrm{X}(i,n)^2}{\sum_{m=1}^{M/2} \mathrm{X}(m,n)^2}. \tag{5}$$

Hence, the formula to calculate the normalized energy can be defined as

$$\mathrm{E}(n) = \sum_{m=1}^{M/2} \mathrm{P}(i,n)^2. \tag{6}$$

The metric of the information entropy in the frequency domain can be used to discriminate activities which have similar energy values. The following equation defines the normalized spectral entropy

$$\mathrm{T}(n) = -\sum_{m=1}^{M/2} \mathrm{P}(i,n) log_2 \mathrm{P}(i,n). \tag{7}$$

IV. Experimental setup

The activities our system recognises can be classified into two coarse-grained categories: dynamic and stationary. Activities that include major movement of the entity, such as walking, crawling or running are dynamic and more static activities such as standing, lying and an empty room are stationary activities. The environmental setting in which we gather the raw data is a seminar room (approximately 3m×5m) with typical furniture, such as table, chairs and a whiteboard. As depicted in figure 1, an USRP device treated as receiver is deployed next to the door and tuned to an appropriate spectral band while all of the six activities were performed in the marked area (1m×2m).

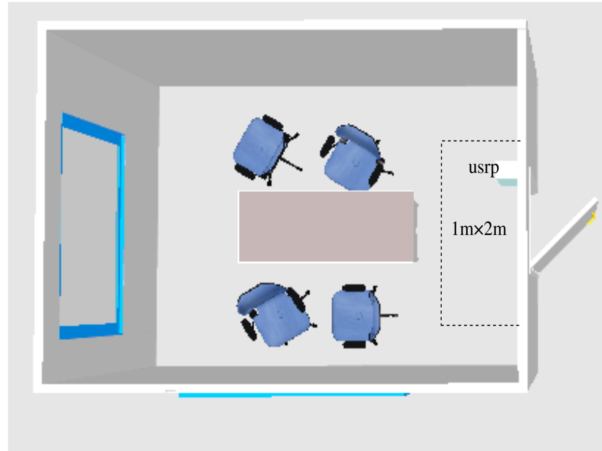

Fig. 1: Schematic illustration of the seminar room used in our case study

The system utilises the WBX[3] daughter board mounted onto a USRP N210[4] SDR device. The antenna utilised is the VERT900[5] model with 3dBi antenna gain. The USRP SDR employs a general purpose software configurable FPGA-based digitizer equipped with a 12-bit dual-channel Analog-Digital Converter (ADC). Every activity was continuously conducted for 3 minutes while the USRP SDR device, configured to monitor the FM radio channel at 82.5MHz was sampling the signal amplitude at a rate of 256kHz. After aggregating raw data derived from the ambient FM radio station, we preprocess it by calculating the average amplitude with a window size of 2^{12}, to reduce the noise of the signal. After the preprocessing, the system extracts from instances of average amplitude features at 64 samples per second.

In contrast to many other activity recognition approaches, which classify all the activities using a single set of classifiers, we leveraged a two stage recognition approach which has been argued by Berchtold et al. [3] to be superior to the one stage approaches. In the first stage, by extracting suitable features, the system is designed to distinguish between the two coarse-grained activities, namely, dynamic and stationary. The window size for extracting features from the preprocessed dataset is 32. Based on the first stage recognition results, the second stage recognition is in charge of classifying these activities which are within the same coarse-grained activity only.

To be specific, during training we first need to identify suitable features to distinguish between the coarse grained activities. In a second step, the feature extraction is repeated and possibly new features are identified for the fine-grained activity recognition.

After the two independent classifications in these two steps, the results are merged for recognition accuracy statistics. For a set of k activities $\mathcal{A} = \{a_1, \ldots, a_k\}$ let $\mathcal{I}(a_i)$ be the total number of feature samples for activity a_i and $\mathcal{I}_{\text{cor}}(a_i)$ the number of correctly classified instances for this activity in which the classification matches the ground truth. We then define the accuracy by which an activity a_i could be detected as

$$\mathcal{ACC}(a_i) = \frac{\mathcal{I}_{\text{cor}}(a_i)}{\mathcal{I}(a_i)}. \tag{8}$$

As described above, three activity recognition processes are realised. For all the processes, we utilise a 5-fold cross validation approach to generate two data-subsets. Only one is labelled and used to train classifiers while the other is used for classification. Support Vector Machine(SVM), Decision Tree (DT) and k-Nearest-Neighbor (k-NN) with its default configuration from the Orange data mining toolkit[6] are used as

[3]https://www.ettus.com/product/details/WBX

[4]https://www.ettus.com/content/files/2987_Ettus_N200-210_DS_FINAL_1.27.12.pdf

[5]https://www.ettus.com/product/details/VERT900

[6]http://orange.biolab.si/

	features	Mean Accuracy
coarse-grained	Var, Std, RMS	99.4%
dynamic	Mean, CM_2, Energy	84%
static	Mean, Std, RMS	98.7%

TABLE I: The highest recognition rate achieved by utilising most suitable feature combinations for coarse-grained, dynamic and static activities respectively

classifiers. The features which are utilised to obtain the highest accuracy in every classification are detailed in the following section.

V. Evaluation

In this section, we discuss the classification accuracy achieved by combining some features for coarse-grained activities, dynamic activities and stationary activities respectively. It demonstrates that various features contain information on different aspects of activities. Also, for verifying the performance of the two stage classification algorithm, we compare the classification results for one stage and two stage approaches respectively.

A. Accuracy for various activity recognitions

As depicted in section III, we explored as many as 13 mathematical approaches to explore the efficiency of RF-based activity recognition. For each classification, 3 features are extracted to calculate the recognition accuracy. Therefore, the diversity is $\binom{13}{3}$. We experimentally tested all possible combinations of features for coarse-grained, dynamic and stationary activities and the results are available online [7]. However, for space restrictions, we can only present the highest mean accuracy of each of the classifications utilising the SVM, k-NN and DT algorithms in table I.

For distinguishing coarse-grained activities between movement and stationary, the recognition precision can be achieved over 99% by the most suitable features combination of Variance, Standard Deviation and Root Mean Square. To classify crawling, walking and running which are within the dynamic category, 84% of samples can be correctly recognised using Mean, Second Central Moment and the normalized Energy. As for static activities, standing, lying and empty room can be recognised with a rate of 98.7% by Mean, Standard Deviation and RMS.

B. Comparison between One Stage and Two Stage Methods

The figure 2 shows the average amplitude of every second of the six activities. The figure illustrates the different volatility of fluctuation of the amplitude of the stationary and movement. Hence, we designed a two stage method, which can roughly distinguish between stationary and dynamic activities and then recognises fine-grained activities within the same coarse-grained activity. To prove the advantage of the two stage method, we compare the accuracy with that obtained by a one staged method which classified all the activities

[7]http://www.ibr.cs.tu-bs.de/users/sigg/CAPS2012/features_shi_CAPS2012.pdf

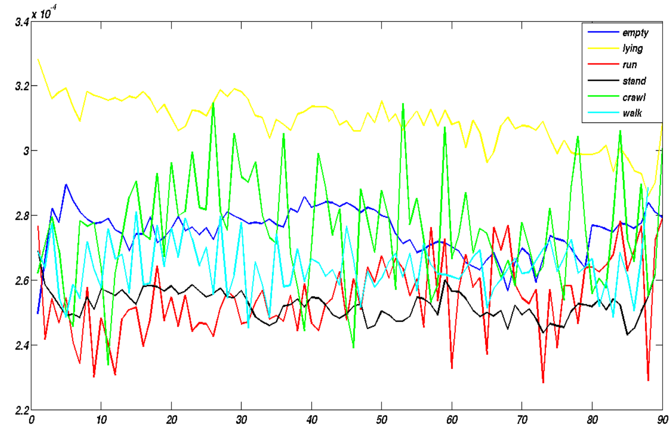

Fig. 2: The average amplitude of the 6 activities per second

simultaneously. All the classification results based on the two staged method are achieved by the feature combination described in section I. Similar to our observation described in section V-A regarding the optimal choice of features, for one stage method, the Mean, RMS and θ are able to classify the activities with the highest precision.

For randomly choosing the samples for training and classification, we ran the 5-fold cross validation algorithm 10 times. The confusion matrices in table II III IV show classification results using SVM, DT and k-NN classification of one and two stage recognition approaches respectively. All of precision are calculated as the mean value of classification results over 10 test cases. Table fields with very low values (i.e. 0.0) are left blank for readability. We observe that, comparing with the one stage approach, the accuracy of all the six experimental activities can be improved by utilising the two stage method.

VI. Conclusion

We implemented activity recognition utilising signals from the RF channel. Three static activities, 'stand', 'lying' and an 'empty room', have been distinguished as well as three dynamic activities 'walk', 'run' and 'crawl'. All these activity classes can be classified with a reasonably high accuracy by properly extracting and interpreting features of RF signals derived from an ambient FM radio station. A USRP SDR device is in charge of sampling FM signals, then, after extracting features, feature samples can be recognised by certain classification algorithms.

To improve the the precision of activity recognition, we mainly considered two ways. One explores a multitude of features in a single recognition step to identify the activities and the other which utilises a two staged method.

References

[1] I. Anderson and H. Muller. Context awareness via gsm signal strength fluctuation. In *4th international conference on pervasive computing, late breaking results*, 2006.

[2] L. Bao and S. Intille. Activity recognition from user-annotated acceleration data. In A. Ferscha and F. Mattern, editors, *Pervasive Computing*, volume 3001 of *Lecture Notes in Computer Science*, pages 1–17. Springer Berlin / Heidelberg, 2004.

true	Classified					
	empty	lying	stand	run	walk	crawl
empty	**1.0**					
lying		**1.0**				
stand			**.939**		.061	
run				**.442**	.176	.382
walk	.03		.061	.152	**.667**	.091
crawl	.05	.1		.075	.15	**.625**

(a) Classification accuracy of the one staged method utilising the SVM classifier

true	Classified					
	empty	lying	stand	run	walk	crawl
empty	**1.0**					
lying		**1.0**				
stand			**.962**		.039	
run				**.924**	.076	
walk	.091		.061	.03	**.696**	.121
crawl				.2	.125	**.675**

(b) Classification accuracy of the two staged method utilising the SVM classifier

TABLE II: Mean accuracy for the classification of the activities 'lying', 'stand', 'empty', 'crawl', 'walk' and 'run' from amplitude based features extracted from an FM-radio signal using SVM classifier in the environment depicted in figure 1

true	Classified					
	empty	lying	stand	run	walk	crawl
empty	**.971**					.028
lying		**1.0**				
stand			**.97**	.03		
run			.029	**.676**		.294
walk				.03	**.658**	.163
crawl		.025		.075	.025	**.875**

(a) Classification accuracy of the one staged method utilising the Decision Tree classifier

true	Classified					
	empty	lying	stand	run	walk	crawl
empty	**.985**					.015
lying		**1.0**				
stand			**1.0**			
run				**.776**	.059	.164
walk	.03		.121	.03	**.7**	.121
crawl				.005	.005	**.9**

(b) Classification accuracy of the two staged method utilising the Decision Tree classifier

TABLE III: Mean accuracy for the classification of the activities 'lying', 'stand', 'empty', 'crawl', 'walk' and 'run' from amplitude based features extracted from an FM-radio signal using the Decision Tree classifier in the environment depicted in figure 1

true	Classified					
	empty	lying	stand	run	walk	crawl
empty	**.993**					.007
lying		**1.0**				
stand			**1.0**			
run				**.824**	.088	.088
walk			.091	.212	**.545**	.151
crawl		.075		.15	.1	**.675**

(a) Classification accuracy of the one staged method utilising the k-Nearest-Neighbor classifier

true	Classified					
	empty	lying	stand	run	walk	crawl
empty	**.996**					.004
lying		**1.0**				
stand			**1.0**			
run				**.726**	.037	.037
walk	.03				**.727**	.121
crawl				.175	.075	**.75**

(b) Classification accuracy of the two staged method utilising the k-Nearest-Neighbor classifier

TABLE IV: Mean accuracy for the classification of the activities 'lying', 'stand', 'empty', 'crawl', 'walk' and 'run' from amplitude based features extracted from an FM-radio signal using k-Nearest-Neighbor classifier in the environment depicted in figure 1

[3] M. Berchtold, M. Budde, H. R. Schmidtke, and M. Beigl. An extensible modular recognition concept that makes activity recognition practical. In *Proceedings of the 33rd annual German conference on Advances in artificial intelligence*, KI'10, pages 400–409, Berlin, Heidelberg, 2010. Springer-Verlag.

[4] D. Figo, P. C. Diniz, D. R. Ferreira, and J. a. M. Cardoso. Preprocessing techniques for context recognition from accelerometer data. *Personal Ubiquitous Comput.*, 14(7):645–662, Oct. 2010.

[5] E. Kim, S. Helal, and D. Cook. Human activity recognition and pattern discovery. *Pervasive Computing, IEEE*, 9(1):48 –53, jan.-march 2010.

[6] J. Lester, T. Choudhury, and G. Borriello. A practical approach to recognizing physical activities. In K. Fishkin, B. Schiele, P. Nixon, and A. Quigley, editors, *Pervasive Computing*, volume 3968 of *Lecture Notes in Computer Science*, pages 1–16. Springer Berlin / Heidelberg, 2006.

[7] K. Muthukrishnan, M. Lijding, N. Meratnia, and P. Havinga. Sensing motion using spectral and spatial analysis of wlan rssi. In *Proceedings of Smart Sensing and Context*, 2007.

[8] G. Ogris, P. Lukowicz, T. Stiefmeier, and G. Tröster. Continuous activity recognition in a maintenance scenario: combining motion sensors and ultrasonic hands tracking. *Pattern Analysis & Applications*, 15:87–111, 2012.

[9] S. N. Patel, T. Robertson, J. A. Kientz, M. S. Reynolds, and G. D. Abowd. At the flick of a switch: Detecting and classifying unique electrical events on the residential power line. In *Proceedings of the 9th International Conference on Ubiquitous Computing (UbiComp 2007)*, pages 271–288, 2007.

[10] M. Scholz, S. Sigg, H. R. Schmidtke, and M. Beigl. Challenges for device-free radio-based activity recognition. In *Proceedings of the 3rd workshop on Context Systems, Design, Evaluation and Optimisation (CoSDEO 2011), in Conjunction with MobiQuitous 2011*, 2011.

[11] M. Scholz, S. Sigg, D. Shihskova, G. von Zengen, G. Bagshik, T. Guenther, M. Beigl, and Y. Ji. Sensewaves: Radiowaves for context recognition. In *Video Proceedings of the 9th International Conference on Pervasive Computing (Pervasive 2011)*, 2011.

[12] M. Seifeldin and M. Youssef. Nuzzer: A large-scale device-free passive localization system for wireless environments. *CoRR*, abs/0908.0893, 2009.

[13] S. Sigg, M. Scholz, Y. Ji, and M. Beigl. Active and passive sensing of activities from amplitude-based rf-features in device-free recognition systems. *IEEE Transactions on Mobile Computing (TMC)*, 2012. submitted.

[14] T. Sohn, A. Varshavsky, A. LaMarca, M. Y. Chen, T. Choudhury, I. Smith, S. Consolvo, J. Hightower, W. G. Grisworld, and E. de Lara. Mobility detection using everyday gsm traces. In *Proceedings of the 8th international conference on Ubiquitous computing*, 2006.

[15] K. Van Laerhoven and H.-W. Gellersen. Spine versus porcupine: a study in distributed wearable activity recognition. In *Wearable Computers, 2004. ISWC 2004. Eighth International Symposium on*, volume 1, pages 142 – 149, 2004.

[16] T. L. Westeyn, G. D. Abowd, T. E. Starner, J. M. Johnson, P. W. Presti, and K. A. Weaver. Monitoring children's developmental progress using augmented toys and activity recognition. *Personal Ubiquitous Comput.*, 16(2):169–191, Feb. 2012.

[17] J. Wilson and N. Patwari. Through-wall tracking using variance-based radio tomography networks. *CoRR*, abs/0909.5417, 2009.

[18] J. Wilson and N. Patwari. Radio tomographic imaging with wireless networks. *IEEE Transactions on Mobile Computing*, 9:621–632, 2010.

[19] K. Woyach, D. Puccinelli, and M. Haenggi. Sensorless sensing in wireless networks: implementation and measurements. In *Proceedings of the Second International Workshop on Wireless Network Measurement (WiNMee)*, 2006.

[20] M. Youssef, M. Mah, and A. Agrawala. Challenges: Device-free passive localizsation for wireless environments. In *Proceedings of the 13th annual ACM international Conference on Mobile Computing and Networking (MobiCom 2007)*, pages 222–229, 2007.

[21] D. Zhang and L. Ni. Dynamic clustering for tracking multiple transceiver-free objects. In *Proceedings of the 7th IEEE International Conference on Pervasive Computing and Communications (PerCom 2009)*, 2009.