Tefnut: An Accurate Smartphone Based Rain Detection System in Vehicles

Hansong Guo[1], He Huang[1,2(✉)], Jianxin Wang[1], Shaojie Tang[3], Zhenhua Zhao[1], Zehao Sun[1], Yu-E Sun[1,2], Liusheng Huang[1], and Hengchang Liu[1]

[1] University of Science and Technology of China, Hefei, China
[2] Soochow University, Suzhou, China
huangh@suda.edu.cn
[3] University of Texas at Dallas, Richardson, USA

Abstract. Real-time and fine-grained rain information is crucial not only for climate research, weather prediction, water resources management, agricultural production, urban planning and natural disasters monitoring, but also for applications in our daily lives. However, because of the lack of rain detection systems and the high variable attribute of rain, both in time and space, the rain detection today is still not precise enough. In such context, we propose and implement Tefnut (Tefnut is the rain deity in Ancient Egyptian religion.), a novel system that exploits opportunistically crowdsourced in-vehicle audio clips from an alternative, nowadays omnipresent source, smartphones, to achieve precise detection of rain leveraging a supervised recognizer constructed from a series of refined features. We conduct extensive experiments, and evaluation results demonstrate that Tefnut can detect the rain with 96.0 % true positive rate, when deciding with a one-second-long in-vehicle audio segment only.

Keywords: Rain detection · Supervised classification · Signal processing · Smartphone

1 Introduction

According to the information released by the United Nations Office for Disaster Risk Reduction [1], developing countries lack rain detection systems, making them more vulnerable to natural disasters, such as flood, erosion, waterlogging, landslide and debris flow, caused by extreme rain events. Depending on sophisticated and expensive equipments and infrastructures, developed countries can achieve reliable daily and city-wide rain detection. However, rain has high variability, both in time and space, that is to say, the rain may start suddenly, only last for a very short period of time and then stop unexpectedly or within a very small area, the weather conditions may be totally different, in other words, it may be raining on one side, but it is completely sunny on the other side, which is not far away. But on the one hand, even in developed countries, the

Q. Yang et al. (Eds.): WASA 2016, LNCS 9798, pp. 13–23, 2016.
DOI: 10.1007/978-3-319-42836-9_2

rain detection systems are sparsely deployed and can not cover all areas. On the other hand, for most of the cities all over the world, these rain detection equipments and infrastructures are placed far away from the urban centers to avoid the interferences from human activities and only periodically record the rain data. These two situations mentioned above both aggravate the inaccuracy of rain detection. To illustrate that precise rain information also plays an important role in our daily lives, we just provide the following one scenario here, due to the page limitation.

Mr. A likes running. He plans to run to a park several kilometers away this morning. He wants to know whether it is raining there. Because if it is raining, he can change into waterproof shoes and clothes or will not go there. But the rain informations released by all weather apps nowadays are not accurate enough, both with respect to temporal and spatial resolution. The precise rain informations crowdsourced and shared by people whose vehicles are near the park utilizing Tefnut can help him.

Nowadays, smartphones are becoming more and more ubiquitous in our daily lives, which not only just serve as communication devices, but also are equipped with abundant advanced built-in sensors. The development of smartphones stimulates the blooming of mobile sensing researches, such as healthcare [3,31], localization [8,13,20,29], safety [12,16,26,27], human computer interaction [30] and makes our lives more efficient [21,25,28], more intelligent [4,10,22] and more enjoyable [5,9]. However, little attention has been paid to the field of rain detection. In this paper, we explore this area, propose, implement and evaluate Tefnut, which is, to the best of our knowledge, the first rain detection system exploiting opportunistically crowdsourced audio clips from smartphones in vehicles, in both industrial and academic communities.

2 Related Work

In recent years, several other kinds of rain detection systems are proposed. Allamano et al. [2], Roser et al. [19], Gormer et al. [6] and Nashashibi et al. [15] employ images for rain detection. In some literature, Grimes et al. [7] and Wardah et al. [24] leverage satellites and Leijnse et al. [11], Messer et al. [14], Overeem et al. [17], Zinevich et al. [32] and Rayitsfeld et al. [18] exploit microwave links for rain detection.

3 System Design and Implementation

3.1 System Overview

In this part, we provide the system overview of Tefnut (Fig. 1). At the very beginning, Tefnut recognizes the in-vehicle environment adopting the EEMSS proposed in [23]. The microphone will be turned on if and only if the user is in the vehicle.

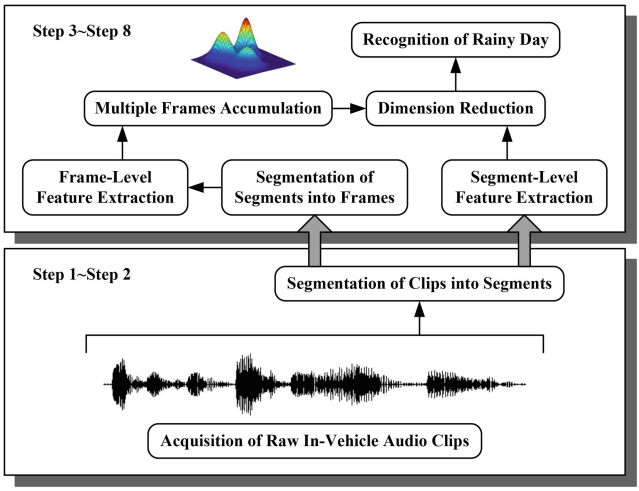

Fig. 1. System overview of Tefnut.

The first step is the acquisition of raw in-vehicle audio clips from smartphone. Then Tefnut divides these audio clips into segments, as presented in Sect. 3.2. In the third step, Tefnut computes several segment-level features in power spectrum, which contain important cues for distinguishing segments generated in rainy and sunny days in vehicles, as reported in Sect. 3.3. In order to study and depict the characteristics of these two kinds of segments more subtly, and then ultimately recognize the segments generated in rainy days in vehicles, Tefnut divides every segment into frames in the fourth step, as given in Sect. 3.4. Based on every frame, Tefnut extracts a series of frame-level features in frequency spectrum and time domain, which are economical but effective, as demonstrated in Sect. 3.5. Then Tefnut accumulates features extracted from different but continuous frames within every segment, as described in Sect. 3.6. In the seventh step, Tefnut conducts the dimension reduction, as outlined in Sect. 3.7. Finally, Tefnut constructs a recognizer, which outputs the recognition results, namely, rainy or sunny day.

3.2 Divide Audio Clips into Segments

The raw audio segments acquisition algorithm we utilized in Tefnut is based on End-Points Detection and Sliding Windows, which are two of the most popular segmentation methodologies. The length of every segment is 1 s in this paper.

3.3 Segment-Level Feature Extraction

In this part, we introduce the features which are selected to detect the rain. These features are all calculated in power spectrum of every segment.

Root Mean Squared Error-Low. This feature measures the smooth degree of power spectrum curve in low frequency part (i.e., less than 7500 Hz). The RMSE-L of segment S_i is calculated as[1]:

$$RMSE\text{-}L_i = \sqrt{\frac{1}{nl_i^p}\sum_{j=1}^{nl_i^p}(a_{ij}^p - \tilde{a}_{ij}^p)^2}$$

where nl_i^p indicates the total number of low frequency components in the power spectrum of segment S_i. a_{ij}^p indicates the amplitude of the j-th frequency component of segment S_i and $\tilde{a}_{ij}^p$ indicates the predicted amplitude of the j-th frequency component calculated by carrying out linear fitting of the power spectrum curve of segment S_i.

Amplitude of Middle Frequency. This feature stands for the amplitude of 7500 Hz in the power spectrum of segment S_i.

Amplitude of Cut-Off Frequency. This feature denotes the amplitude of 15000 Hz in the power spectrum of segment S_i.

Min., Med., Avg. and Var. Amplitude. These features focus on the basic shape of the power spectrum curve.

Energy-Low. This feature pays attention to the signal energy in low frequency part. The E-L of segment S_i is calculated as:

$$E\text{-}L_i = \sum_{j=1}^{nl_i^p} a_{ij}^p$$

Spectral Similarity. This feature describes the similarity degree between power spectrum curves in low and high frequency parts. The SS of segment S_i is calculated as:

$$SS_i = \frac{1}{nl_i^p}\sum_{j=1}^{nl_i^p}|\hat{a}_{ij}^p - \hat{a}_{i(nl_i+j)}^p|$$

where $\hat{a}_{ij}^p$ indicates the modified amplitude of the j-th frequency component calculated by aligning power spectrum curves in low and high frequency parts to x-axis respectively, in other words, by subtracting the average value of all amplitudes in low (high) frequency part from a_{ij}^p if f_{ij}^p is a low (high) frequency component, which corresponds to a_{ij}^p.

3.4 Divide Segments into Frames

In this part, Tefnut divides every segment into frames, whose lengths are 0.032 s in this paper, with an overlap of 50 % between consecutive frames, and then applies Hanning window to every frame to avoid frequency distortion.

[1] In this paper, the superscript p, t or f on a variable indicates that this variable is calculated in power spectrum, time domain or frequency spectrum respectively.

3.5 Frame-Level Feature Extraction

The following are several features employed to recognize the sounds generated in rainy days in vehicles, such as raindrops hitting on windows and windshield wipers pivoting, which are all calculated in time domain or frequency spectrum of every frame.

Spectral Centroid. This feature characterizes the barycenter of the frequency spectrum, which is correlated with the perceptual attribute of timbre, i.e., brightness. The SC of frame F_i is calculated as:

$$SC_i = \frac{\sum_{j=2}^{n_i^f} a_{ij}^f \log_2 \frac{f_{ij}^f}{1000}}{\sum_{j=2}^{n_i^f} a_{ij}^f}$$

where n_i^f indicates the total number of frequency components and a_{ij}^f indicates the amplitude of the j-th frequency component f_{ij}^f, in the frequency spectrum of frame F_i.

Spectral Spread. This feature denotes the shape of the frequency spectrum, that is to say, whether it is concentrated in the vicinity of its centroid, or spread out over the frequency spectrum. The SS of frame F_i is calculated as:

$$SS_i = \sqrt{\frac{\sum_{j=2}^{n_i^f} (\log_2 \frac{f_{ij}^f}{1000} - SC_i)^2 a_{ij}^f}{\sum_{j=2}^{n_i^f} a_{ij}^f}}$$

Spectral Roll-Off. This feature captures the frequency below which 95 % of the signal energy is contained. The SR-O of frame F_i is calculated as:

$$SR\text{-}O_i = \min j', \text{ subject to:}$$

$$\sum_{j=1}^{j'} a_{ij}^f \geq 0.95 \sum_{j=1}^{n_i^f} a_{ij}^f$$

Mel Frequency Cepstral Coefficients. These features collectively represent the shape of the spectrum. To calculate the MFCC of frame F_i, Tefnut firstly employs the Hamming window to minimize the maximum side lobe. Then, Tefnut transforms the time domain of frame F_i into frequency domain by performing the DFT (Discrete Fourier Transform). In the third step, utilizing a set of triangular filters, Tefnut computes the Mel scale from the frequency components obtained

above. Finally, Tefnut takes the logarithms of powers at all Mel frequencies and conducts the DCT (Discrete Cosine Transform) of these logarithm values. The amplitudes of the output spectrum are the Mel Frequency Cepstral Coefficients.

Avg. Zero-Crossing Rate. This feature counts the average number of occurrences that the sampling points of audio signal pass through the zero axis in time domain within a particular frame. The AZ-CR of frame F_i is calculated as:

$$AZ\text{-}CR_i = \frac{1}{n_i^t - 1} \sum_{j=2}^{n_i^t} [s_{ij}^t s_{i(j-1)}^t < 0]$$

where $[P]$ is the Iverson Bracket, a notation whose numerical value is 1 if the proposition P within square brackets is satisfied, and 0 otherwise. n_i^t indicates the total number of sampling points in time domain of frame F_i and s_{ij}^t indicates the j-th sampling point of frame F_i.

3.6 Multiple Frames Accumulation

In this paper, we put forward a novel approach for describing the distributions of features extracted from different but continuous frames within every segment based on estimating a GMM (Gaussian Mixture Model), which can also be considered as a one-state CDHMM (Continuous Density Hidden Markov Model). Then we employ the parameters of the GMM as new features to conduct the recognition procedure.

3.7 Dimension Reduction

The segment-level features are 9 dimensions and the features output by GMM are 72 dimensions. In order to evaluate the sparseness, maximize the synergies between different features and then reduce the dimension of features, we explore LDA (Linear Discriminant Analysis) and PCA (Principal Component Analysis). Ultimately, according to experimental results, we choose the LDA and integrate it into Tefnut.

4 Evaluation

In this section, we present the results of our experiments. This section consists of three parts. In the first part, we compare the differences in recognition performance among six common recognizers. Then in the following part, we report the experimental results of Tefnut employing two kinds of dimension reduction algorithms. Finally, we demonstrate the time consumption of every step in the recognition process for every segment.

Our experimental dataset contains 5400 rainy day and 5400 sunny day audio segments crowdsourced and labeled by our 25 participants in rainy and sunny days in their vehicles respectively.

4.1 Recognition Performance of Different Recognizers

In this part, we construct six recognizers based on different recognition algorithms, which are Decision Tree, Random Forest, Naive Bayes, Multi Layer Perceptron, k-Nearest Neighbors and Support Vector Machine respectively. Then we conduct a series of 10-fold cross-validation experiments on our dataset. Figure 2 presents the confusion matrixes, Fig. 3(a) illustrates the TPR (True Positive Rate), Fig. 3(b) highlights the FPR (False Positive Rate) and Fig. 4 reports the time consumption for recognition on our experimental dataset of these six recognizers.

a=Rainy Day b=Sunny Day

	Classified as		Classified as		Classified as		Classified as		Classified as		Classified as	
	a	b	a	b	a	b	a	b	a	b	a	b
a	5138	262	5317	83	5159	241	5347	53	5299	101	5245	155
b	247	5153	264	5136	509	4891	78	5322	298	5102	384	5016
	(a) DT		(b) RF		(c) NB		(d) MLP		(e) k-NN		(f) SVM	

Fig. 2. Confusion matrixes of different recognizers.

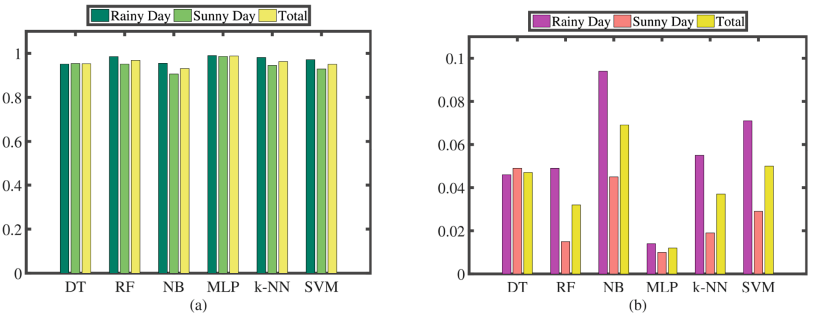

Fig. 3. TPR (a) and FPR (b) of different recognizers. (Color figure online)

We can make two main observations here. Firstly, as presented in Figs. 2 and 3, Multi Layer Perceptron outputs the best recognition performance, namely, 98.8 % TPR, 1.2 % FPR and 131 misrecognition segments, Random Forest achieves the second best recognition performance, that is, 96.8 % TPR, 3.2 % FPR and 347 misrecognition segments and Naive Bayes yields the worst recognition performance, namely, 93.1 % TPR, 6.9 % FPR and 750 misrecognition segments. Secondly, as reported in Fig. 4, we only pay attention to the time

consumption of the test process, because the training process can be accomplished offline. k-Nearest Neighbors is the most time-consuming for testing our experimental dataset, 10800 labeled rainy day or sunny day in-vehicle audio segments, which spends 85.69 s. Support Vector Machine is the second most time-consuming and spends 13.84 s. The Decision Tree and Random Forest can both complete the whole test process within 0.4 s.

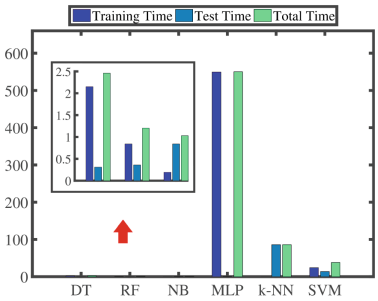

Fig. 4. Total time consumed of different recognizers. (Color figure online)

4.2 Experimental Results

In this part, we construct the recognizer based on Random Forest, in addition, we exploit Principal Component Analysis and Linear Discriminant Analysis to reduce the dimension of features. Then we conduct a series of 10-fold cross-validation experiments on our dataset. Figure 5 illustrates the confusion matrixes, TPR, FPR and feature dimension of these two recognition algorithm combinations.

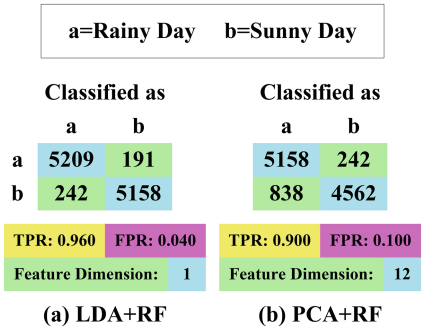

Fig. 5. Confusion matrixes, TPR, FPR and feature dimension of RF along with LDA and PCA.

We can observe that the combination of Linear Discriminant Analysis and Random Forest achieves better recognition performance, namely, 96.0 % TPR, 4.0 % FPR and 433 misrecognition segments, which also reduces the features to 1 dimension and the combination of Principal Component Analysis and Random Forest yields worse recognition performance, that is, 90.0 % TPR, 10.0 % FPR and 1080 misrecognition segments and reduces the features to 12 dimensions.

4.3 Time Consumption of Every Computational Step

In this part, we conduct the recognition experiment on our dataset for ten times and compute the average time consumption for every segment, as shown in Table 1, which is specific to every step. We can observe that Segment-Level Feature Extraction is the most time-consuming, because we need to calculate the power spectrum first, and only then can we extract features. For every segment, the whole recognition process can be accomplished within about 675.93 ms.

Table 1. Time consumption (Avg.±Std.Dev.) of every step in the recognition process for every segment.

Step	Computational Process		Time
1-2	Acquisition of Raw In-Vehicle Audio Clips		(10.40±0.79)ms
	Segmentation of Clips into Segments		
3	Segment-Level Feature Extraction		(578.29±15.85) ms
4-6	Segmentation of Segments into Frames		(46.62±0.74) ms
	Frame-Level Feature Extraction		
	Multiple Frames Accumulation		(40.56±1.17) ms
7-8	Dimension Reduction (LDA)	Train	(0.03±0.01) ms
		Test	0.01 ms
	Recognition of Rainy Day (RF)	Train	0.01 ms
		Test	0 ms
Total			(675.93±16.20) ms

5 Conclusion

Rain detection with high temporal and spatial resolution is significant not only for professional researches and decisions-making, but also for applications in our daily lives. In this paper, we present Tefnut, which is, to the best of our knowledge, the first rain detection system exploiting opportunistically crowdsourced in-vehicle audio clips from smartphones, in both industrial and academic communities. Tefnut utilizes a supervised recognizer constructed from a series of refined features. The evaluation results of extensive experiments demonstrate

that Tefnut can detect the rain with 96.0 % true positive rate, when deciding with a one-second-long in-vehicle audio segment only.

Acknowledgement. This paper was supported by National Natural Science Foundation of China under Grant No. 61572342, 61303206 and 61472384, Natural Science Foundation of Jiangsu Province under Grant No. BK20151240 and BK20140395, China Postdoctoral Science Foundation under Grant No. 2015M580470.

References

1. United Nations Office for Disaster Reduction. https://www.unisdr.org
2. Allamano, P., Croci, A., Laio, F.: Toward the camera rain gauge. Water Resour. Res. **51**(3), 1744–1757 (2015)
3. Aminikhanghahi, S., Wang, W., Shin, S., Son, S.H., Jeon, S.I.: Effective tumor feature extraction for smart phone based microwave tomography breast cancer screening. In: Proceedings of the 29th Annual ACM Symposium on Applied Computing, pp. 674–679. ACM (2014)
4. Dhondge, K., Song, S., Choi, B.Y., Park, H.: WiFiHonk: Smartphone-based beacon stuffed WiFi Car2X-communication system for vulnerable road user safety. In: IEEE 79th Vehicular Technology Conference (VTC Spring), 2014, pp. 1–5. IEEE (2014)
5. Gao, X., Tian, J., Liang, X., Wang, G.: ARPP: an Augmented Reality 3D ping-pong game system on Android mobile platform. In: WOCC, pp. 1–6. IEEE (2014)
6. Görmer, S., Kummert, A., Park, S.B., Egbert, P.: Vision-based rain sensing with an in-vehicle camera. In: Intelligent Vehicles Symposium, 2009 IEEE, pp. 279–284. IEEE (2009)
7. Grimes, D., Diop, M.: Satellite-based rainfall estimation for river flow forecasting in Africa. I: rainfall estimates and hydrological forecasts. Hydrol. Sci. J. **48**(4), 567–584 (2003)
8. Gutierrez, N., Belmonte, C., Hanvey, J., Espejo, R., Dong, Z.: Indoor localization for mobile devices. In: ICNSC, pp. 173–178. IEEE (2014)
9. Jing, T., Cui, X., Cheng, W., Zhu, S., Huo, Y.: Enabling smartphone based HD video chats by cooperative transmissions in CRNs. In: Cai, Z., Wang, C., Cheng, S., Wang, H., Gao, H. (eds.) WASA 2014. LNCS, vol. 8491, pp. 636–647. Springer, Heidelberg (2014)
10. Kim, H., Lee, S.K., Kim, H., Kim, H.: Implementing home energy management system with upnp and mobile applications. Comput. Commun. **36**(1), 51–62 (2012)
11. Leijnse, H., Uijlenhoet, R., Stricker, J.: Rainfall measurement using radio links from cellular communication networks. Water Resour. Res. **43**(3) (2007)
12. Li, F., Yang, Y., Wu, J.: CPMC: an efficient proximity malware coping scheme in smartphone-based mobile networks. In: INFOCOM, pp. 1–9. IEEE (2010)
13. Liu, Z., Chen, Y., Liu, B., Wang, J., Fu, X.: Aerial localization with smartphone. In: Wang, X., Zheng, R., Jing, T., Xing, K. (eds.) WASA 2012. LNCS, vol. 7405, pp. 386–397. Springer, Heidelberg (2012)
14. Messer, H., Zinevich, A., Alpert, P.: Environmental monitoring by wireless communication networks. Science **312**(5774), 713 (2006)
15. Nashashibi, F., de Charrette, R., Lia, A.: Detection of unfocused raindrops on a windscreen using low level image processing. In: ICARCV, pp. 1410–1415. IEEE (2010)

16. Novak, E., Li, Q.: Near-pri: private, proximity based location sharing. In: INFOCOM, pp. 37–45. IEEE (2014)
17. Overeem, A., Leijnse, H., Uijlenhoet, R.: Country-wide rainfall maps from cellular communication networks. Proc. Nat. Acad. Sci. **110**(8), 2741–2745 (2013)
18. Rayitsfeld, A., Samuels, R., Zinevich, A., Hadar, U., Alpert, P.: Comparison of two methodologies for long term rainfall monitoring using a commercial microwave communication system. Atmos. Res. **104**, 119–127 (2012)
19. Roser, M., Geiger, A.: Video-based raindrop detection for improved image registration. In: ICCV Workshops, pp. 570–577. IEEE (2009)
20. Tan, G., Lu, M., Jiang, F., Chen, K., Huang, X., Wu, J.: Bumping: a bump-aided inertial navigation method for indoor vehicles using smartphones. IEEE Trans. Parallel Distrib. Syst. **25**(7), 1670–1680 (2014)
21. Tang, Z., Guo, S., Li, P., Miyazaki, T., Jin, H., Liao, X.: Energy-efficient transmission scheduling in mobile phones using machine learning and participatory sensing. IEEE Trans. Veh. Technol. **64**(7), 3167–3176 (2015)
22. Tian, J., Wang, G., Gao, X., Shi, K.: User behavior based automatical navigation system on Android platform. In: WOCC, pp. 1–6. IEEE (2014)
23. Wang, Y., Lin, J., Annavaram, M., Jacobson, Q.A., Hong, J., Krishnamachari, B., Sadeh, N.: A framework of energy efficient mobile sensing for automatic user state recognition. In: MobiSys, pp. 179–192. ACM (2009)
24. Wardah, T., Bakar, S.A., Bardossy, A., Maznorizan, M.: Use of geostationary meteorological satellite images in convective rain estimation for flash-flood forecasting. J. Hydrol. **356**(3), 283–298 (2008)
25. Wen, Y., Shi, J., Zhang, Q., Tian, X., Huang, Z., Yu, H., Cheng, Y., Shen, X.: Quality-driven auction-based incentive mechanism for mobile crowd sensing. IEEE Trans. Veh. Technol. **64**(9), 4203–4214 (2015)
26. Wu, L., Du, X., Wang, L., Fu, X., Mbouna, R.O., Kong, S.G.: Analyzing mobile phone vulnerabilities caused by camera. In: GLOBECOM, pp. 4126–4130. IEEE (2014)
27. Wu, L., Du, X., Wu, J.: MobiFish: a lightweight anti-phishing scheme for mobile phones. In: ICCCN, pp. 1–8. IEEE (2014)
28. Yang, S., Thormann, J.: Poster: crowdsourcing to smartphones: social network based human collaboration. In: MobiHoc, pp. 439–440. ACM (2014)
29. Yoo, S., Kim, E., Kim, H.: Exploiting user movement direction and hidden access point for smartphone localization. Wireless Pers. Commun. **78**(4), 1863–1878 (2014)
30. Yue, Q., Ling, Z., Fu, X., Liu, B., Ren, K., Zhao, W.: Blind recognition of touched keys on mobile devices. In: Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, pp. 1403–1414. ACM (2014)
31. Zhang, Z., Wang, H., Wang, C., Fang, H.: Cluster-based epidemic control through smartphone-based body area networks. IEEE Trans. Parallel Distrib. Syst. **26**(3), 681–690 (2015)
32. Zinevich, A., Messer, H., Alpert, P.: Frontal rainfall observation by a commercial microwave communication network. J. Appl. Meteorol. Climatol. **48**(7), 1317–1334 (2009)

http://www.springer.com/978-3-319-42835-2

Wireless Algorithms, Systems, and Applications
11th International Conference, WASA 2016, Bozeman, MT, USA, August 8-10, 2016. Proceedings
Yang, Q.; Yu, W.; Challal, Y. (Eds.)
2016, XV, 583 p. 260 illus., Softcover
ISBN: 978-3-319-42835-2