Anonymizing motion sensor data through time-frequency domain

Pierre Rougé, Ali Moukadem, Alain Dieterlen, Antoine Boutet, Carole Frindel

Anonymizing motion sensor data through time-frequency domain

Pierre Rougé
Univ Lyon, INSA Lyon, CREATIS,
Inserm, Lyon, France
rouge@creatis.insa-lyon.fr

Ali Moukadem
UniversitÃľ Haute-Alsace, IRIMAS,
Mulhouse, France
ali.moukadem@uha.fr

Alain Dieterlen
UniversitÃľ Haute-Alsace, IRIMAS,
Mulhouse, France
alain.dieterlen@uha.fr

Antoine Boutet
Univ Lyon, INSA Lyon, Inria, CITI,
Lyon, France
antoine.boutet@insa-lyon.fr

Carole Frindel
Univ Lyon, INSA Lyon, CREATIS,
Inserm, Lyon, France
carole.frindel@creatis.insa-lyon.fr

ABSTRACT

The recent development of Internet of Things (IoT) has democratized activity monitoring. Even if the data collected can be useful for healthcare, sharing this sensitive information exposes users to privacy threats and re-identification. This paper presents two approaches to anonymize the motion sensor data. The first is an extension of an earlier work based on filtering in the time-frequency plane and convolutional neural network; and the second is based on handcrafted features extracted from the zeros distribution of the time-frequency representation. The two approaches are evaluated on a public dataset to assess the accuracy of activity recognition and user re-identification. With the first approach we obtained an accuracy rate in activity recognition of 73% while limiting the identity recognition to an accuracy rate of 30% which corresponds to an activity identity ratio of 2.4. With the second approach we succeeded in improving the activity and identity ratio to 2.67 by attaining an accuracy rate in activity recognition of 80% while maintaining the re-identification rate at 30%.

KEYWORDS

Activity, Privacy, Time-Frequency, Convolutional Neural Networks, Random Forest

1 INTRODUCTION

The wide adoption of Internet of Things (IoT) devices have democratized quantified self applications and revolutionized patient monitoring in healthcare domain [1]. This monitoring relies on sensors that measure motion signals (e.g., accelerometer, gyroscope and magnetometer). These signals are further sent to a cloud server to be analysed and processed through advanced signal processing and machine learning pipeline [7] to compute and present multiple estimators to users or practitioners (such as the number of steps or the activity performed during the day, the quality of the sleep, or the burned calories). Although this information is very useful for self-assessment or remote monitoring, it is closely related to the health status of the associated user and consequently sensitive. Sharing this sensitive information to third party applications exposes users to privacy threats (e.g., attribute inference or re-identification) and discrimination [9]. Moreover, health related data attract much attention nowadays. For instance, an increasing number of health insurers are seeking access to this data to better predict rates and encourage their members to wear fitness trackers [12].

To mitigate the risks of privacy leakage, several approaches have been proposed providing different privacy and utility trade-off. While some of them rely on collaborative learning (federated learning) to avoid sharing data with the server [11], others sanitize raw data to avoid unwanted inferences or re-identification [3]. Other approaches try to minimize the data sent to the server. For instance, [8] extracts locally on the device temporal and frequency features, and sends only the features the most important for the activity detection task while normalizing features leading to re-identification. Instead of processing features from the temporal and frequency domain separately, another approach [4] transforms the signal to a time-frequency representation before filtering high coefficients to limit re-identification. The resulting time-frequency representation is then directly processed by a convolutional neural network (CNN) to predict activity recognition. While this data minimization process (i.e., directly based on classifying time-frequency representation) is simple and attractive, the identification by the CNN of useful information in this representation is complex. In addition, the filtering scheme can be improved to provide a better utility and privacy trade-off (i.e., maintaining an accurate activity detection while preventing re-identification).

In this paper we extend the work initiated in [4] by investigating feature extraction from the time-frequency transform to improve the privacy and utility trade-off. Specifically, motivated by the recent link between Gaussian analytic functions (GAFs) and time-frequency transforms of white Gaussian noise [2], we apply this theoretical work by leveraging the zeros in the time-frequency plane from Short Time Fourier Transform (STFT) and the length of the Delaunay triangles formed by these zeros [5] to identify the most important features in the both classification tasks (i.e., activity detection and re-identification). In addition, we leverage gyroscopic signals and propose a Random Forest-based classifier to better discriminate physical activities of users while anonymizing the motion data.

The pipeline of the solution is illustrated in Figure 1. We evaluate the utility and privacy trade-off provided by our approach based on the zeros of the spectrogram against an optimized version of the state-of-the-art approach [4] using one reference datasets. Results show that the spectrogram's zeros approach offers a better utility-privacy trade-off (80% and 30% in activity and identity recognition

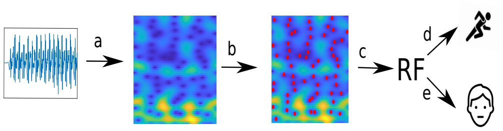

Figure 1: Outline of the approach: a. Transformation from time to time-frequency domain, b. Detection of spectrogram zeros (shown in red), c. Random Forest classifier, d. Activity Recognition, e. Identity Recognition

rate, respectively) than the state-of-the-art approach (73% and 30% in activity and identity recognition rate, respectively).

The rest of the paper is organized as follows. Section 2 gives more details on background and the optimized version of the state-of-the-art approach [4]. Section 3 details our SpectrogramâĂŹs zeros approach while Section 4 presents the results of the evaluation. Finally, Section 5 discusses conclusions and future work.

2 BACKGROUND AND STATE-OF-THE-ART

2.1 STFT and Bargmann connection

Time-frequency domain allows to study the frequency evolution of a signal during time, it is particularly useful to analyse non-stationary signal. The most common transform from the time domain to the time-frequency domain is the Short Time Fourier Transform (STFT). The STFT for a given signal $x(t)$ and a window function $w(t)$ is given by:

$$S_x^{\omega}(t,f) = \int_{-\infty}^{+\infty} x(\tau) w^*(\tau - t) e^{-2j\pi f\tau} d\tau, \qquad (1)$$

In the case of Gaussian window $\omega(t) = g(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-t^2}{2\sigma^2}}$, the STFT can be written as Bargmann transform as follows [2]:

$$S_x^{g}(t,-f) \propto e^{-i\pi tf} e^{-\frac{\pi}{2}|z|^2} B_x(z), \qquad (2)$$

where $z = t + if \in \mathbb{C}$ and $B_x(z)$ is the Bargmann transform defined as follows :

$$B_x(z) = 2^{1/4} \int_{-\infty}^{+\infty} x(t) e^{2\pi tz - \pi t^2 - \frac{\pi}{2} z^2} dt \qquad (3)$$

This link means that the STFT can be completely characterized by its zeros. Equation 2 shows that the zeros of the STFT are the zeros of the Bargmann transform which are also the zeros of Gaussian analytic functions (GAFs). This will ensure some regularity for the zeros distribution of the STFT for white Gaussian noise. The mathematical details and properties of these connections are detailed in [2].

2.2 CNN-based filtering approach

In this section, we present the state-of-the-art approach proposed in [4] and the optimization of the CNN we applied.

Filtering: It was observed in [4] that in the time-frequency representation the difference between activities is encoded through texture and that on the other hand, the difference between subjects is encoded by contrast.

According to these observations, it was proposed in [4] to filter high coefficients of the spectograms to remove user's information to prevent re-identification. In this study we used this filtering method with different percentage of high coefficient removed going from 0% to 90% with steps of 10%.

CNN classifier: In this study, we used six different CNN classifiers that can be broken down into 3 categories: those taking as input accelerometer data only, those taking as input gyroscope data only and those using both accelerometer and gyroscope data as input. In each of these categories, a CNN has been constructed to classify activities into 4 classes and another to classify the user's identity into 24 classes. The results of these models were compared to assess the utility of adding the gyroscope data into the framework.

For all CNN, we considered a model with four convolutional layers followed each by one maxpooling layer and at the end a final softmax dense layer for classification. The number of filters on the first layer was set to a power of two and for each layer the number of filter was set to the next power of two.

Optimisation of CNN using accelerometer or gyroscopic data only: The two tasks (activity recognition and identification) are different by nature, it is therefore necessary to optimize the architecture and the hyperparameters of the CNN independently for these 2 tasks. Fine tuning require to test a lot of hyperparameters combinations. In this work, optimization process focused on certain amount of hyperparameters and possible ranges of values. These hyperparameters were the number of filters on the first convolutional layer, the batch size and the learning rate. Also, during the optimisation process, if an hyperparameter value was obviously not adapted we chose to withdraw this value from the process, so the test of the combinations is not exhaustive.

Two different fusion schemes for the three axes of the sensors were also evaluated and compared: late and early fusion. The early fusion strategy consist in combining images from the 3 axes at the entry of the network. The late fusion strategy consist in using three independent convolutional branches to process each input independently then combining the features map from the three branches just before the dense layer.

The optimisation was made using the acceloremeter data without any filtering and the selected model was later also used on gyroscopic data only.

Model using both accelerometer and gyroscope data: Few tests were done to evaluate the values of the hyperparameters around those found for the model on the accelerometer data to ensure its transferability. Three fusion strategies were also assessed: (i) early fusion on axes, (ii) early fusion on sensors and (iii) late fusion. The early fusion axes consists in merging the images of the two sensors (accelerometer and gyroscope) corresponding to the same axis at the input of the CNN, then three independent convolutional branches process the three axes and finally the characteristic maps of the three branches are merged just before the final dense layer. Early fusion on sensors is the same idea where images of the three axes (x, y, z) are merged at the input of the CNN corresponding to the same sensor. And the late fusion strategy is the same idea as previously except that in this case we have six different convolutional branches each corresponding to a distinct sensor and axis. From our comparisons, we chose the late fusion strategy with a batch size of 512 and a learning rate of 0.0025, which was the set of parameters giving the best results in term of accuracy on the validation set for both activity and identity recognition tasks.

Training implementation: The dataset was split into a training and test sets according to the trials created during the acquisition phase: thus trials 1 to 9 were used for the training phase and trials 11 to 16 for the test phase. More precisely, during the training phase, 90% of the whole set was used for training and 10% for validation. We used categorical cross entropy as the loss function and Adam as the optimizer. The maximum number of epochs was set to 200 and regulated with an early stopping criteria.

3 STFT ZEROS APPROACH

The link established between the STFT and the GAFs guarantees a regular and well known distribution of the STFT zeros in case of white Gaussian noise [2, 5]. In this sense, the distribution of zeros can provide information on the presence of noise or signal for the development of filtering schemes. In this article, we exploit the distribution of zeros to extract handcrafted features to classify activities while preserving privacy. The intuition behind this idea is that the presence of a signal will modify the distribution of zeros in the time-frequency domain and mark this distribution by the signal signature.The first step in this process is to detect the zeros from STFT representation. We use a Gaussian window for the STFT to ensure the link between STFT and GAFs. The value of σ for the Gaussian window is set empirically to 0.05. Since we are working on discrete STFT, the zeros are not perfect and the energy spreading around instantaneous frequencies of the signal's components will affects the intensity of the zeros. To detect them we used a 3x3 mask sliding through STFT: if the value in the center of the mask is the minimum and the maximum value covered by the mask exceeds a certain threshold we consider a zero in the center of the mask. Here the threshold was set to $\max(|S_x^g|)/10^4$, with $|S_x^g|$ the modulus of the STFT representation (Equation 1).

3.1 Features associated with spectrogram zeros

Once the zeros are detected, features associated with their distribution must be extracted for use in a standard classifier. To this purpose we connect them with a Delaunay triangulation and create a graph. Examples of obtained graphs for different subjects and activities are given in Figure 2.

Since numerical zeros are not perfect zeros and their intensity is influenced by the energy in their area, we chose to order them according to their intensity. The advantage of ordering the zeros is that we can easily construct features attached to a zero (for example its intensity) and above all compare them between two different graphs. It suffices then to compare the features associated with the zeros of the same rank in the ordering. Two types of feautres can be constructed: those which characterize the global graph of zeros and those which are attached to a particular zero. In our dataset, the minimum number of zeros detected was 48, so that for each STFT we extracted 48 zeros: the 24 zeros of minimum intensity and the 24 zeros of maximum intensity.

Global Features: To analyse the distribution of the zeros we studied the distance between them. We constructed three distributions: euclidean distance, distance on time-axis and distance on frequency-axis between each zero. For each of these distributions, four statistical moments were used as global features: mean, standard deviation, skewness and kurtosis. From the graph, mean and max edge length were also extracted as features.

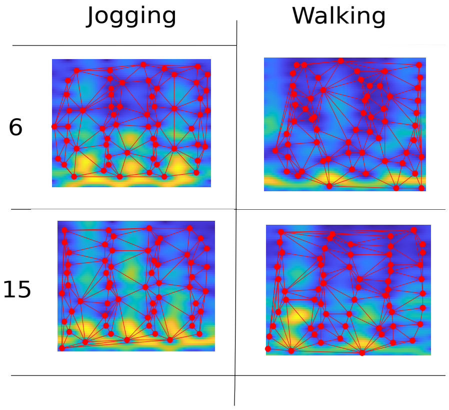

Figure 2: Examples of STFT representations superposed with the associated graph formed from STFT zeros for different subjects in lines and different activities in columns

.

Local Features: To characterize the zero itself, we used as a feature its intensity and its coordinates in the time-frequency plane. From the graph, we also considered the zeros belonging to a neighborhood of order 1 and used their mean intensity as a feature. We also computed the mean energy crossed by each edge to reach the neighbors and used the average of these energies as a feature.

To investigate patterns in the region surrounding the zeros we used Haralick features [6]. Fourteen features were calculated from the gray level co-occurence matrix (GLCM). We considered two GLCM with respectively an offset of (0,1) and (1,0). The idea is to consider the co-occurence along the time and the frequency direction independently. We used a window of size 30x30 to capture the region surrounding the zero. To characterize the edges, we also used the average angle of the edge with respect to the x-axis and the area of the triangles connected to the zero. In the end, a given STFT image results in total of 1694 features.

3.2 Random Forest Classifier

To identify both activity and subjects from the features extracted from the STFT's zeros, we used a Random Forest (RF) classifier. Two parameters were more particularly studied: the number of trees used in the forest and the maximum depth of trees. The values tested were respectively in the following ranges [400, 500, 600] and [25, 50, 75, 100]. To investigate these values, we made a 5-fold cross-validation over each of possible combination and selected the one which gave the best result in accuracy.

As with the CNN-based approach, at a given time there are 6 STFT images from two sensors each with 3 axes. For each of the STFT images, a feature matrix – as described in section 3.1 – has been calculated. To build the RF model we choose to concatenate all the feature matrices associated with the 6 STFTs.

3.3 Feature selection

Once the RF models have been constructed, it is possible to observe the importance of each of the features in the two classification tasks. Our goal is to remove features useful for identity recognition but not for activity recognition. For this, the average importance of the features resulting from each sensor/axis pair was calculated. The sensor/axis pairs contributing most to the activity recognition task have been retained. After that, the correlation between each features was calculated and if this correlation exceeded a certain threshold, we removed it from the model. These operations make it possible to improve the utility/privacy trade-off. In our experiments, the correlation threshold was set at 0.5.

3.4 Dense neural network

In order to compare the performance of our RF, we also used a dense network on the raw data. This network is five unit combining a dense layer and a dropout layer (set to 20%) plus a final dense layer for classification. The results of this network are made with a 5-fold cross-validation.

4 EVALUATION AND DISCUSSION

4.1 Experimental Settings

Dataset: For this study, we used the public dataset Motion-Sense [10]. This dataset was collected with a Iphone 6S placed in the participant's trousers front pocket at a frequency rate of 50 Hz. The dataset provides time-series data from 3-axis accelerometer and gyroscope and includes recordings (15 trials) of 24 different participants for six activities: downstairs, upstairs, walking, jogging, standing and sitting. The time-series are split in sliding windows such as each window corresponds to an activity and a participant. The window length was fixed to 2.5 seconds with an overlap of 50%, the average cadence range of walking is about 1.5 steps by second so about 3 walking steps were captured by window. In this study, only the four dynamic activities were considered: downstairs, upstairs, walking and jogging.

Time-Frequency images: The images formed by the STFT module are considered to train the CNN or to extract features from the zeros of the STFT. The size of the generated images is 65 × 128 which corresponds to 25 Hz × 2.56 sec.

Accuracy: To assess the performance of the CNNs and the RF classifiers, we computed an accuracy score defined as:

$$Accuracy = \frac{1}{n_{\text{samples}}} \sum_{i=1}^{n_{\text{samples}}} 1(y_i, \widehat{y}_i), \tag{4}$$

where n_{samples} is the number of samples and $1(x, y)$ the indicator function which gives 1 if $x = y$ and 0 otherwise. For the CNNs classifiers the given accuracies were averaged over ten experiments and for the RF over a 5-fold cross validation.

Privacy measure: To assess privacy, the ratio between accuracy in activity and in identity was investigated. If this ratio is greater than 1 its mean that we detect more efficiently the activity than the identity of the subject and inversely. In a privacy-preserving framework the goal is to have the highest ratio possible. This ratio was computed for each level of filtering in the case of CNNs. Also, to compare the different CNNs the Area under the utility-privacy Curve (AUC) is studied.

4.2 Optimisation of the CNN

Tables 1 and 2 present the results of the four more efficient architectures in respectively the identity and activity recognition tasks. In these tables, column "Filter" refers to the number of filters in CNN first layer, "Lr" to the learning rate and "Acc Val" to the average accuracy on the validation set over ten experiments.

Modele	Filter	Batch Size	Lr	Acc Val
Late Fusion	**16**	**256**	**0.005**	**0.77**
Late fusion	8	256	0.005	0.75
Early Fusion	16	256	0.005	0.76
Early fusion	8	256	0.01	0.69

Table 1: Accuracy on validation set for the four most efficient architectures in the identification task. In bold the selected model.

Modele	Filter	Batch Size	Lr	Acc Val
Late fusion	**16**	**256**	**0.005**	**0.93**
Early Fusion	16	256	0.005	0.91
Late fusion	8	256	0.005	0.92
Early Fusion	8	256	0.005	0.91

Table 2: Accuracy on validation set for the four most efficient architectures in the activity recognition task. In bold the selected model.

For the identification task, we selected the model with late Fusion and batch size=256, number of filters=16 and learning rate=0.005, because it was the combination showing the best accuracy on validation set (Table 1). For the activity recognition task the results gave similar performances for the different models. We first chose the model with early Fusion, batch size=256, number of filters=16 and learning rate=0.005 because it has fewer parameters than the late Fusion one (103 588 versus 309 892 weights). But during the experiences presented in section 4.3, the early fusion model turned out to be difficult to transfer in the case of filtering, so we decided to return to the model using late fusion but with the same parameters.

4.3 Role of gyroscopic sensors data

Figures 4 and 5 respectively represent the activity and identity accuracy results for different filtering levels for three different CNNs. On the one hand, Figure 5 shows that the CNNs learned from accelerometer or gyroscope data result in degraded accuracy at the same pace for the identity recognition tasks. On the other hand, Figure 4 indicates that, for the activity recognition task, the filtering affects less the performance of the CNN learned from accelerometer data than that learned from gyroscope data. Furthermore, it exhibits that a CNN combining accelerometer and gyroscope data allows to boost the performance in identity recognition but not in activity recognition (to be compared with Figure 5). This observation is also confirmed by the table 3 presenting Activity/Identity ratio considering the level of filtering for the three CNNs. Filtering does not really affect this ratio for the CNN learned gyroscope data unlike that learned from accelerometer data. In terms of normalized AUC, the CNN learned from accelerometer data also outperforms the one learned from gyroscope data as shown in Table 4. Even though the CNN combining accelerometer and gyroscope data presents a slightly better AUC than the one based on accelerometer data, the Activity/Identity ratio (see Table 3) demonstrates that learning only from accelerometer data offers a better utility and privacy trade-off.

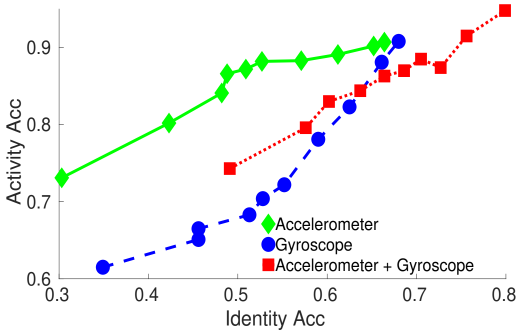

Figure 3: Activity versus identity accuracy for CNN using respectively accelerometer data, gyroscope data and both.

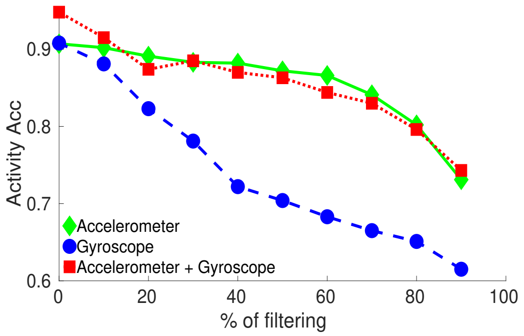

Figure 4: Activity accuracy versus % of filtering for CNN using respectively accelerometer data, gyroscope data and both

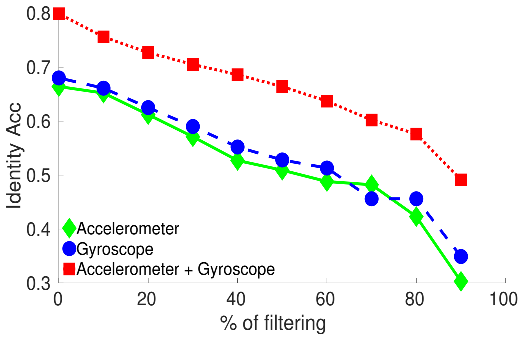

Figure 5: Identity accuracy versus % of filtering for CNN using respectively accelerometer data, gyroscope data and both

% filtering	Gyro	Acc + Gyro	Acc
0	1.3	1.2	1.4
20	1.3	1.2	1.5
40	1.3	1.3	1.7
60	1.3	1.3	1.8
80	1.4	1.4	1.9
90	**1.8**	**1.5**	**2.4**

Table 3: Privacy measure for different levels of filtering and for the three CNNs. In bold the filtering level giving the best activity/identity ratio.

Gyro	Acc + Gyro	Acc
0.719	0.844	0.835

Table 4: Normalized AUC for the CNNs using respectively only accelerometer, gyroscope and both data.

4.4 The zeros of the STFT

With the RF model detailed in 3.2 we first obtained an accuracy score of 85% for the activity recognition task and 72% for the identity recognition task as shown in table 5-model A. The results of the dense network are of the same order of magnitude (+4% in activity and +3% in identity recognition), attesting that our RF does not overfit the dataset. The observations of features importance by pair of sensor/axis demonstrated that for the identity recognition task no pair was preponderant in the decision, whereas for the activity the acceleration on y-axis and the rotation speed around x-axis were decisive. We have therefore decided to limit ourselves to these two channels allowing a first improvement in the measure of privacy as shown in the table 5-model B. Next, we applied our method to remove the correlated features. We managed to reduce the number of features to 409 and significantly improve the utility-privacy trade-off, as shown in the table 5-model C.

Model	Activity Acc	Identity Acc	Ratio	Features
A	0.85	0.72	1.18	10164
B	0.81	0.52	1.56	3388
C	**0.80**	**0.30**	**2.67**	**409**

Table 5: Results for both activity and identity recognition tasks depending on the features used; Model A : all features, Model B: features from acceleration on y-axis and rotation speed on x-axis, Model C : same features as in B with deletion of correlated features. In bold the model giving the best activity/identity ratio.

5 CONCLUSION

In this paper, we presented two privacy-preserving approaches in the time-frequency domain. The first approach has already been tested in [4] but in this work we have strengthened the evaluation of this method by ensuring that the CNNs were properly optimized. Our experiments also highlighted the role of gyroscopic sensor data in identity recognition. The second approach proposes a new way to extract features from the zeros of the STFT. We introduced features based on this representation and then proposed a feature selection method to boost the utility-privacy trade-off. This approach makes it possible in particular to better select the information related to activities and do better than the CNN for which a more naive filtering method was set up before the learning. To extend this work, it may be interesting to study, in addition to the graph of the STFT zeros, the graph of local maxima and also to include the phase of the time-frequency representation. Finally, for more genericity, our approach should now be tested on another dataset.

REFERENCES

[1] B. Ajana. 2017. Digital health and the biopolitics of the Quantified Self. *Digital Health* 3 (2017), 2055207616689509.

[2] R. Bardenet *et al.* 2020. On the zeros of the spectrogram of white noise. *Applied and Computational Harmonic Analysis* 48, 2 (2020), 682–705.

[3] A. Boutet *et al.* 2021. DYSAN: Dynamically sanitizing motion sensor data against sensitive inferences through adversarial networks. In *AsiaCCS*.

[4] N. Debs *et al.* 2021. Motion sensor data anonymization by time-frequency filtering. In *28th EUSIPCO*.

[5] P. Flandrin. 2015. TimeâĂŞFrequency Filtering Based on Spectrogram Zeros. *Signal Processing Letters, IEEE* 22 (03 2015). https://doi.org/10.1109/LSP.2015.2463093

[6] R. Haralick *et al.* 1973. Textural Features for Image Classification. *IEEE Trans Syst Man Cybern* SMC-3 (01 1973), 610–621.

[7] M. Hoogendoorn *et al.* 2018. Machine learning for the quantified self. *On the art of learning from sensory data* (2018).

[8] T. Jourdan *et al.* 2018. Toward Privacy in IoT Mobile Devices for Activity Recognition. In *MobiQuitous*.

[9] D. Leibenger *et al.* 2016. Privacy challenges in the quantified self movement-an EU perspective. *Proceedings on Privacy Enhancing Technologies* 2016, 4 (2016).

[10] M. Malekzadeh *et al.* 2019. Mobile Sensor Data Anonymization. In *IoTDI*.

[11] H. B. McMahan *et al.* 2016. Federated learning of deep networks using model averaging. *arXiv preprint arXiv:1602.05629* (2016).

[12] P. Olson. 2014. Wearable Tech is Plugging Into Health Insurance. *Forbes* (2014).