Fingerprint Quality Classification for CSI-based Indoor Positioning Systems

Josyl Mariela Rocamora
Department of Electronic and Information Engineering
The Hong Kong Polytechnic University
josyl.rocamora@connect.polyu.hk

Ivan Wang-Hei Ho
Department of Electronic and Information Engineering
The Hong Kong Polytechnic University
ivanwh.ho@polyu.edu.hk

Man-Wai Mak
Department of Electronic and Information Engineering
The Hong Kong Polytechnic University
enmwmak@polyu.edu.hk

ABSTRACT

Recent indoor positioning systems that utilize channel state information (CSI) consider ideal scenarios to achieve high-accuracy performance in fingerprint matching. However, one essential component in achieving high accuracy is the collection of high-quality fingerprints. The quality of fingerprints may vary due to uncontrollable factors such as environment noise, interference, and hardware instability. In our paper, we propose a method for collecting high-quality fingerprints for indoor positioning. First, we have developed a logistic regression classifier based on gradient descent to evaluate the quality of the collected channel frequency response (CFR) samples. We employ the classifier to sift out poor CFR samples and only retain good ones as input to the positioning system. We discover that our classifier can achieve high classification accuracy from over thousands of CFR samples. We then evaluate the positioning accuracy based on two techniques: Time-Reversal Resonating Strength (TRRS) and Support Vector Machines (SVM). We find that the sifted fingerprints always result in better positioning performance. For example, an average percentage improvement of 114% for TRRS and 22% for SVM compared to that of unsifted fingerprints of the same 40-MHz effective bandwidth.

CCS CONCEPTS

• **Information systems** → **Location based services**; • **Computing methodologies** → **Supervised learning by classification**; **Support vector machines**; • **Hardware** → Wireless devices.

KEYWORDS

Indoor Positioning, Channel State Information, Fingerprint, Logistic Regression

ACM Reference Format:
Josyl Mariela Rocamora, Ivan Wang-Hei Ho, and Man-Wai Mak. 2019. Fingerprint Quality Classification for CSI-based Indoor Positioning Systems. In *PERSIST-IoT Workshop '19: Workshop on Pervasive Systems in the IoT era, July 02, 2019, Catania, Italy.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/--

1 INDOOR POSITIONING USING CSI FINGERPRINTING

Existing indoor positioning systems (IPS) based on radio frequency (RF) technology utilize received signal strength (RSS) or channel state information (CSI) to generate location fingerprints [10]. Unlike RSS, CSI accounts for the presence of multipaths and is not dependent on line-of-sight (LOS) condition [14]. In CSI, the wireless channel fingerprint is either represented in the time domain as channel impulse response (CIR) or in the frequency domain as channel frequency response (CFR). Compared to CIRs, CFRs are easier to extract since many commercially available wireless devices are orthogonal frequency division multiplexing (OFDM) based systems [14].

Using CFR fingerprints, a target can be positioned by employing statistical or matching techniques to identify the best match between the acquired data and the pre-collected training data [10]. Unlike traditional ranging methods, fingerprinting does not depend on multiple anchor points to track a target but relies on a rich database of high-quality CSI.

One of the key challenges for achieving centimeter-level accuracy in CSI-based positioning is the collection of high-quality fingerprint. For instance, in an indoor environment with no moving scatterers, a 280-MHz effective bandwidth generated from high-gain transceivers resulted in near perfect positioning performance with 30-cm resolution [11]. In another CFR-based IPS, a frequency hopping mechanism was used to form fingerprints with up to 1-GHz effective bandwidth, which eliminated the ambiguity issues inherent in wireless channels [2]. By exploiting spatial diversity in MIMO devices, high effective bandwidth can be achieved with antenna arrays embedded on target devices [3]. In general, low effective bandwidth results in poor positioning performance caused by unpredictable environmental dynamics and ambiguous sensitivity of certain locations [2, 3].

IPS in [2, 3, 11] utilized frequency diversity to mask ambiguities and incorporated phase compensation to remove synchronization errors. However, the collected fingerprints may still include unstable phases in certain wireless channels. These IPS also did not include any strategy for detecting channel quality prior to storing fingerprints. In this work, we recognize the imperfections in the wireless channels in practical indoor environment and define a solution to achieve high-quality fingerprint extraction despite channel uncertainties. The major contributions of this paper are as follows.

- We have quantified the concepts of good and bad fingerprints for indoor positioning.
- We have developed a learning algorithm for determining channel quality with high classification accuracy to sift out poor channels in fingerprinting tasks.
- We have enhanced the performance of CSI-based indoor positioning despite the presence of low-quality fingerprints.

This paper is organized as follows. In Section 2, we define good and poor fingerprints through identifying the statistical properties of the subcarrier magnitudes and phases including the channel signal-to-noise ratio (SNR). These definitions are then applied to develop a binary classifier based on logistic regression with gradient descent. In Section 3, we expound on two positioning algorithms, namely, TRRS and SVM. We then evaluate the performance of these positioning systems in Section 4 to demonstrate the impact of poor fingerprints on positioning and that carefully selecting good channels can greatly enhance the positioning accuracy. Finally, we summarize our observations and results in Section 5.

2 FINGERPRINT QUALITY CLASSIFICATION

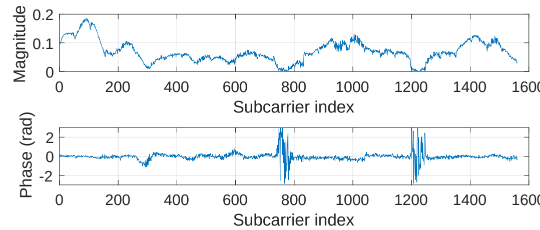
2.1 Definition of good and poor fingerprints

Ideally, a fingerprint is regarded as good if it contains good CFR samples for high positioning accuracy. In a good or high-quality channel, CFR samples can be extracted (i.e., receiver can identify the preamble head) and are stable after phase compensation. On the other hand, a poor or low-quality fingerprint is made up of low-magnitude CFRs with highly erratic phases. The poor sample is due to the collection of CFRs from inherently low-quality channels caused by high environment noise, interference, or unstable hardware. When the channel is noisy, synchronization at the receiver may fail and the CFRs may become more erratic even after phase compensation. The phase compensation method primarily assumes that the magnitude attenuation and phase rotations are only caused by synchronization errors only [2], not by unintended interruptions or low SNR.

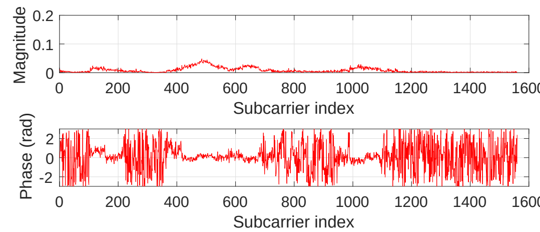
Figure 1 displays good- and poor-quality fingerprints. The good fingerprint has generally high magnitude values and stable phases. On the other hand, the poor fingerprint is quite noisy and erratic. Figure 2 shows the histograms of the CFR subcarrier magnitude, subcarrier phase, and inter-subcarrier phase difference from the good and poor fingerprints. In the good fingerprint, most subcarrier phases have values less than 1 radian, and inter-subcarrier phase differences are mostly between -1 and 1 radian. In the poor fingerprint, most subcarrier magnitudes have a low value of less than 0.01.

To further quantify the concept of good and poor CFRs, we have acquired samples across 1,200 channels, each with 10-MHz. These channels span from 3.42 to 3.91 GHz and are measured at four different time periods with varying antenna gains. One averaged fingerprint is generated from 20 raw CFR samples. Since there are 1,200 channels, then there are a total of 24,000 CFR samples. Each channel is visually checked and labelled as good or poor according to the proposed definitions. If a certain channel is good, then all 20 CFR samples associated to it are good. Else, all the 20 CFR samples are poor. Out of the 24,000 CFR samples, 9,480 are good and 14,520 are poor. Afterwards, statistical properties and the signal-to-noise ratio (SNR) values of the CFR samples are evaluated.

Table 1 shows the statistics of the subcarrier magnitudes and phases of good and poor channels. Each CFR has 52 subcarriers or points. Eight statistical features are extracted by getting the mean, mininum, maximum, and standard deviation of both the magnitude and phase of the 52 points per sample. Good samples have relatively higher subcarrier magnitude and lower subcarrier phase variance compared to that of poor samples.

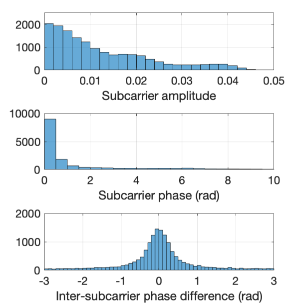


(a) Good-quality fingerprint

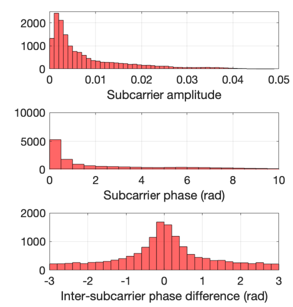


(b) Poor-quality fingerprint

Figure 1: CFR fingerprints with 300-MHz effective bandwidth (30 channels, 52 subcarriers per channel)



(a) CFRs from good-quality fingerprints



(b) CFRs from poor-quality fingerprints

Figure 2: Histograms of CFR samples' subcarrier magnitude, subcarrier phase and inter-subcarrier phase differences

Table 2 lists the SNR statistics of good and poor samples. SNR is measured based on the magnitudes of the desired data against the noise. Note that good channels generally have high SNR compared to poor channels. However, not all high-SNR samples are good. Some poor samples may be collected in LOS channels, where SNR is high, but with erratic phases. Therefore, SNR is not a sufficient metric to determine the channel quality for fingerprinting. In addition, the minimum SNR for poor samples is not known since some of the poor channels are too erratic.

Table 1: Subcarrier magnitude and phase statistics of good and poor CFR samples

Statistical property	Good CFR sample	Poor CFR sample
Mean magnitude	0.0346	0.0076
Min magnitude	0.0230	0.0037
Max magnitude	0.0428	0.0014
Stdev magnitude	0.0051	0.0019
Mean phase (rad)	-0.0510	0.1808
Min phase (rad)	-0.7167	-2.8562
Max phase (rad)	0.6338	3.1662
Stdev phase (rad)	0.3904	1.5884

Table 2: SNR statistics of good and poor CFR samples

Statistical property	Good CFR sample	Poor CFR sample
Mean SNR	5.8102 dB	1.8253 dB
Min SNR	3.7636 dB	N/A
Max SNR	11.2514 dB	7.3461 dB
Stdev SNR	2.7264 dB	2.0065 dB

2.2 How to classify good and poor fingerprints

Given the labels of the 24,000 samples, we can apply a supervised learning algorithm that outputs a binary value, 1 for good and 0 for poor. Logistic regression is a simple cost function for binary classification [9]. It can be trained using 60% of the total data with gradient descent. Another 20% of the total data is used for cross validation, to select which classifier model is the best; while the remaining 20% is utilized as testing data, to calculate the classifier performance. When training the classifier, the gradient descent algorithm is utilized because of its simplicity and scalability to operate even with large feature dimension. Inputs to the learning algorithm can be:

(1) 108 raw features, which include the 52 magnitude values and 52 phase values of the complex CFR sample; or
(2) 8 statistical features, which are the statistical properties of the complex CFR sample

To determine the appropriate model for the logistic regression classifier [8], we look at the cost function (i.e., error) when the regularization parameter λ is varied from 0 (no regularization) to a certain value and when the number of samples vary from 1 to 14,400 (60% of the total). Ideally, the computed value of the cost function should be zero.

When either the 104 raw features or 8 statistical features are used, the cost function has low value of around 0.16 for any regularization parameter. Since the values from training and cross validation are close to each other, the model exhibits neither high variance nor high bias. Similarly, the value of the cost function remains low when the number of training samples increase, and the curve has been converged when the number of training samples is beyond 1,000.

Based on this, a logistic regression classifier using gradient descent and zero regularization is trained using 60% of the training CFR samples. Using the selected model parameters, the testing data is classified as either good or poor and the following parameters

Table 3: Evaluation of the logistic regression classifier

Evaluation Metric	Classifier using 104 raw features	Classifier using 8 statistical features
Accuracy	0.9340	0.9492
Precision	0.9191	0.9415
Recall	0.9117	0.9263
F1 Score	0.9154	0.9338

are calculated: accuracy, precision, recall, and the F1 score. These metrics show the performance of the classifier.

$$\text{Accuracy} = \frac{\text{\# of correctly classified samples}}{\text{total \# of samples}} \tag{1}$$

$$\text{Precision} = \frac{\text{\# of true positives}}{\text{\# of true positives + \# of false positives}} \tag{2}$$

$$\text{Recall} = \frac{\text{\# of true positives}}{\text{\# of true positives + \# of false negatives}} \tag{3}$$

$$\text{F1 score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

Table 3 shows the values of the evaluation metrics of the two sets of input features. The input samples are feature-normalized (i.e., standardized) prior to classification. Based on the metrics, the learning algorithm performs well with both models. The model that utilizes the statistical features shows slightly higher values in all four metrics. This means that the classifier can detect the quality of the CFR samples with high confidence and hence minimize erratic CFR samples in the fingerprinting task.

2.3 Using CFR quality classifier in IPS

The CFR quality classifier based on the 8 statistical features can be used to determine which channel in a fingerprint is good or poor. Poor channels can be sifted out, while good channels can be retained in the fingerprint. The selected channels should be consistent across all the training and testing fingerprints for a particular experiment. The sifted fingerprint can now be used as the input data to the positioning algorithms such as TRRS and SVM.

The sifting function can be done during real-time data collection or offline data processing. In the real-time case, good channels are initially selected based on the classifier and training fingerprints are generated using these channels. When training is done, the same set of channels, no matter the channels remain good or become poor, should be used for testing. For the offline case, a wide range of bandwidth must be scanned and the resulting band of frequencies after sifting may not be contiguous. In this research, we have employed the offline scenario and selected the best 4 or 10 channels to acquire 40-MHz or 100-MHz fingerprints, respectively.

3 POSITIONING ALGORITHMS

3.1 Time-Reversal Resonating Strength (TRRS)

Time-reversal resonating strength (TRRS) is regarded as a measure of similarity between two fingerprints [2, 13], which is the foundation for this kind of fingerprinting technique. TRRS in frequency domain $\gamma[\hat{\mathbf{H}}, \hat{\mathbf{H}}']$ of two CFR fingerprints can be expressed in (5) [2]. $\hat{\mathbf{H}}$ and $\hat{\mathbf{H}}'$ are two CFR fingerprints represented as complex column vectors. $\langle \hat{\mathbf{H}}, \hat{\mathbf{H}} \rangle$ is the inner product between two complex vectors. k is the subcarrier index in the OFDM system.

$$\gamma[\hat{\mathbf{H}}, \hat{\mathbf{H}}'] = \frac{|\sum_{k=1}^{K} \hat{H}_{u_k} \hat{H}'_{u_k}|^2}{\langle \hat{\mathbf{H}}, \hat{\mathbf{H}} \rangle \langle \hat{\mathbf{H}}', \hat{\mathbf{H}}' \rangle} \tag{5}$$

TRRS value ranges between 0 and 1, where 1 occurs if a CFR is the multiple of the other. Looking at the equation, we can see that it is similar to getting the square of the inner product of two normalized vectors. Since the vectors are normalized, the inner product only takes into account the angle separation between the two vectors and not the magnitude. By doing this, the equation returns high value even if the vectors are multiples of each other.

If the training database is already available, TRRS can be used in positioning as shown in (6) and (7) [2]. $\mathbf{H}[l]$ is the training CFR fingerprint from known location l stored in the database. $\mathbf{H}[l']$ is the testing CFR fingerprint from unknown location l'. $\hat{l}'$ is the estimated location of l'. Γ is the localization threshold.

$$l^{\star} = \arg\max_{l=1,2,\ldots L} \gamma[\mathbf{H}[l], \mathbf{H}[l']] \tag{6}$$

$$\hat{l}' = \begin{cases} l^{\star}, & \text{if } \gamma[\mathbf{H}[l], \mathbf{H}[l']] \geq \Gamma \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

We look for the location that yields the highest TRRS and return it as the location estimate, only if the TRRS exceeds the threshold Γ. Otherwise, positioning failed. In this work, we have used $\Gamma = 0$ so that the IPS based on TRRS always outputs a location.

For single training and single testing scenarios, TRRS is calculated based on (5). Since introducing more training data can improve the positioning performance, we utilize a remedy called Continuous Fingerprint Appending (CFA) as proposed in [6]. In this technique, multiple training fingerprints per location are stored in the database. Testing fingerprint is matched to each of the training fingerprint in the database. The TRRS with CFA finds the training fingerprint that yields the highest TRRS and returns its location as the estimate.

3.2 Support Vector Machines (SVM)

SVM has been a solution to various learning problems including classification and regression and is based on the use of Lagrange multipliers in constrained optimization [1]. Since the model parameter training in SVM is a convex optimization problem, the solution is always the global optimum. Unlike TRRS which takes the angle separation of two vectors, SVM considers the Euclidean distance between the vectors. For CSI-based IPS, the linear SVM works well as shown in [11]. For this research, we employ the same linear kernel function.

The output of a linear SVM is shown in (8) [7], which assumes that training data can be separated by a linear hyperplane in the input space. α_k is the Lagrangian multiplier of element k. S is the set containing the indices when $\alpha_k > 0$. $\hat{\mathbf{H}}_k$ is the support vector with corresponding $\alpha_k > 0$. b is the bias parameter. To make our results comparable to TRRS, we modify the SVM to include vector normalization as shown in (9). Both TRRS and SVM then consider the cosine similarity between fingerprints.

$$f(\mathbf{x}) = \sum_{k \in S} \alpha_k y_k \hat{\mathbf{H}}_k \cdot \hat{\mathbf{H}} + b \tag{8}$$

$$f(\mathbf{x}) = \sum_{k \in S} \alpha_k y_k \frac{\hat{\mathbf{H}}_k}{|\hat{\mathbf{H}}_k|} \cdot \frac{\hat{\mathbf{H}}}{|\hat{\mathbf{H}}|} + b \tag{9}$$

To use SVM for multiple classes (i.e., K classes), the one-vs-rest approach is used when there are K SVM classifiers [7]. Each location is treated as one class and each SVM classifier separates a certain class among the rest. For instance, the kth SVM determines whether the input is class k or not. The input vector is processed across all K SVM classifiers. The SVM that yields the highest output score determines the class of the input. In Matlab, each classifier in the the multi-class linear SVM outputs the negated average binary loss for a testing data. The class of the classifier that outputs the lowest loss (i.e., closest to zero) is selected as the estimated location of the testing fingerprint. In this paper, we rescale the SVM scores using min-max normalization so that both TRRS and SVM values have the same range from 0 to 1.

4 PERFORMANCE EVALUATION

4.1 Experiment Setup

Figure 3 shows the indoor environment in a building of the Hong Kong Polytechnic University in April 2019 similar to the setup of our prior studies [6, 11]. The two software defined radio (SDR) devices are Universal Software Radio Peripheral (USRP) N210 devices from Ettus Research [4, 5]. One USRP acts as the access point (AP) or the receiver (RX) and another USRP is the mobile unit (MU) or the transmitter (TX). The USRP RX is fixed inside the room while the USRP TX is stationed in the corridor and is moved during the experiments. The separation between TX and RX is approximately 3 meters. MU transmits an OFDM symbol based on the IEEE 802.11a standard then AP estimates the channel based on the transmitted symbol. Ubuntu Linux machines are connected to the USRP devices and are running python scripts of the GNU Radio Companion (GRC), an open-source software that allows development of wireless systems and signal processing blocks [12].

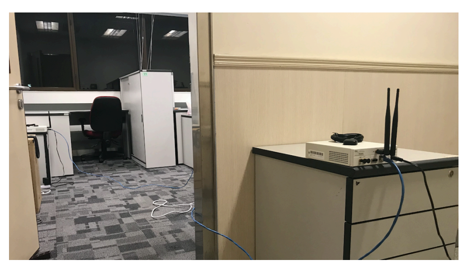


Figure 3: Experiment setup using two USRP devices

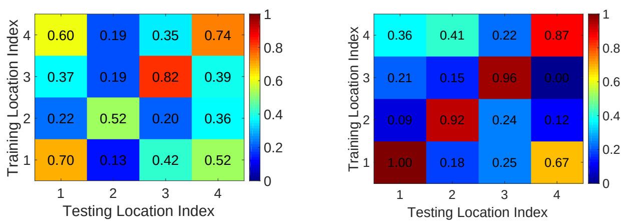
Table 4 lists four NLOS scenarios in our experiment. Both scenarios A and B consider closed-door environment, while scenarios C and D consider open-door environment. For the 4-location scenarios, each location is the corner of a 30 cm x 30 cm square on the table. For the 16-location scenarios, each location is a point in the 4x4 grid with 10-cm grid resolution. For data collection, we utilized center frequencies 3.21:0.01:3.50 GHz. The center frequency

Table 4: List of scenarios

Scenario	Environment	# of locations	Grid resolution
A	NLOS, closed door	4 locations	30 cm (1x1 grid)
B	NLOS, closed door	16 locations	10 cm (4x4 grid)
C	NLOS, open door	4 locations	30 cm (1x1 grid)
D	NLOS, open door	16 locations	10 cm (4x4 grid)

was changed sequentially based on the frequency hopping mechanism introduced in [2]. There were three data collection periods per scenario. The interval between collection period is roughly 1 hour. For multiple training and testing data, we performed leave-one-period-out cross validation to avoid selection bias. Since there are three collection periods, three cross validation tests are conducted and the average accuracy of the three tests is reported.

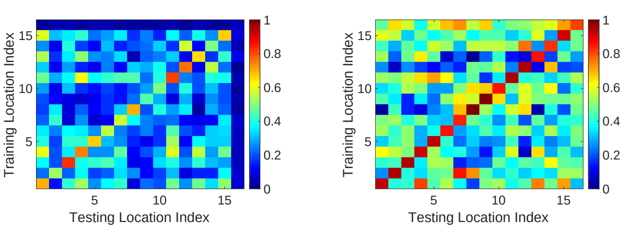
4.2 Positioning Accuracy with and without Fingerprint Sifting



(a) Positioning scores using TRRS with CFA

(b) Positioning scores using Multi-class SVM classifier (scaled range)

Figure 4: IPS output scores ranging from 0 to 1 based on scenario A (4 locations, closed door) using 300-MHz unsifted fingerprints



(a) Positioning scores using TRRS with CFA

(b) Positioning scores using Multi-class SVM classifier (scaled range)

Figure 5: IPS output scores ranging from 0 to 1 based on scenario B (16 locations, closed door) using 300-MHz unsifted fingerprints

Figure 4 shows the output scores of the two positioning systems based on scenario A, which considers 4 locations in a closed-door setting. The y-axis of the matrices represent the training location index for TRRS with CFA and multi-class SVM. The x-axis pertains to the testing fingerprints of four locations. The training data for these matrices are the fingerprints in the first two periods of scenario A, which are composed of 80 unique location fingerprints. On the other hand, the testing data are four location fingerprints from the 3rd period of scenario A.

By looking at the matrix in a column-wise manner, each testing fingerprint yields the highest score in the training fingerprint that matches its actual location. This directly translates to correctly positioning of the testing data. Although the positioning scores for TRRS are not high (i.e., not close to 1), the highest values are still from the fingerprints of the same location. These relatively low TRRS values indicate that the fingerprints are inherently poor. Comparing the two algorithms, SVM performs better than TRRS since the former find the optimal hyperplane that separates one class from the rest, instead of simply finding the similarity between training and testing fingerprints.

Figure 5 shows the output scores of the two positioning systems based on scenario B, which considers 16 locations in a closed-door setting. Since the separation between adjacent positions is much smaller compared to that of scenario A, the off-diagonal scores are higher on the average. Comparing the two algorithms, SVM still performs better than TRRS.

Table 5 summarizes the positioning accuracy of TRRS and SVM when different fingerprints and environments are employed. The first three rows consider unsifted fingerprints while the last two rows consider sifted fingerprints. Looking at the values, higher bandwidth fingerprints always yield better positioning performance since the fingerprints contain more information to better discriminate among classes. When the distance separation between positions decreases, the accuracy is also reduced since fingerprints from close positions become more similar. Closed- and open-door conditions have different responses since the multipaths between the TX and RX behave differently as there is a significant change in the signal scatterers.

Since the quality of this set of data is inherently poor, it follows that the accuracy is not high as compared to our previous work which showed near-perfect accurcy by considering 280-MHz effective bandwidth in a highly controlled environment [11]. Because of this, we incorporate the sifting technique by using the logistic regression classifier introduced in Section 2.

The last two rows in Table 5 show that the performance of the sifted fingerprints is almost always better than that of the unsifted case, since the sifting process ensures the best channels to be used. With 40-MHz, the performance improvements of sifting in TRRS and in SVM are 114% and 22%, respectively. With 100-MHz, TRRS and SVM are improved by 27% and 7%, respectively. In some cases, the 40-MHz sifted fingerprints even achieve higher accuracy than the 100-MHz unsifted case. This is especially prominent for the TRRS method. In a few instances, 40-MHz sifted fingerprints have better performance than that of the 100-MHz sifted fingerprints. This is because the sifting algorithm only selects the best channels out of a total of 30 channels across all locations and collection periods; it is not guaranteed that the retained good channels have the same level of quality.

Table 5: Positioning accuracy (0 to 1) of TRRS with CFA and multi-class linear SVM under varying fingerprint effective bandwidths and environment scenarios

Bandwidth	Scenario A (4 locations)		Scenario B (16 locations)		Scenario C (4 locations)		Scenario D (16 locations)	
	TRRS	SVM	TRRS	SVM	TRRS	SVM	TRRS	SVM
300 MHz	0.992	1.000	0.900	0.965	0.583	0.750	0.738	0.863
100 MHz	0.700	0.792	0.608	0.727	0.550	0.683	0.352	0.560
40 MHz	0.567	0.742	0.329	0.479	0.350	0.425	0.190	0.298
100 MHz (sifted)	0.708	0.883	0.658	0.750	0.775	0.700	0.550	0.617
40 MHz (sifted)	0.822	0.675	0.678	0.610	0.492	0.533	0.690	0.429

5 CONCLUSION AND FUTURE WORK

In this research, we have developed a method for collecting high-quality fingerprints for indoor positioning by utilizing a logistic regression classifier to sift out poor channel frequency response (CFR) samples. Good CFR samples usually have high subcarrier magnitudes (often greater than 0.03) and low subcarrier phase deviation (less than 0.5 rad). Meanwhile, poor CFR samples usually have low subcarrier magnitudes (less than 0.01) and high subcarrier phase deviation (more than 1.5 rad). Since SNR is not a sufficient indicator of CFR quality, the statistical features of the CFR magnitudes and phases are used to train the logistic regression classifier that achieves 94% classification accuracy over 4,800 testing samples. Using the CFR quality classifier in inherently poor fingerprints, the performance improvement of CFR sifting at the same 40-MHz effective fingerprint bandwidth is 114% for TRRS and 22% for SVM. While at 100-MHz, TRRS and SVM are improved by 27% and 7%, respectively. In most TRRS cases, fingerprints with smaller sifted bandwidth can even provide better positioning performance than those with larger unsifted bandwidth.

The CFR quality classifier can be used in real-time indoor positioning to ensure that only high-quality fingerprints are generated and stored in the access point (AP). This aids in developing an indoor positioning system (IPS) that is more robust to noise, interference, as well as hardware instability. Another potential application of the classifier is to detect poor CFR samples and introduce concealment or replacement techniques to maintain the fingerprint bandwidth.

ACKNOWLEDGMENTS

This work is supported in part by the General Research Fund (Project No. 15201118) established under the University Grant Committee (UGC) of the Hong Kong Special Administrative Region (HKSAR), China; and by The Hong Kong Polytechnic University (Project No. G-YBXJ).

REFERENCES

[1] Christopher M. Bishop. 2006. Sparse Kernel Machines. In *Pattern Recognition and Machine Learning*, M. Jordan, J. Kleinberg, and B. Schölkopf (Eds.). Springer, New York, USA, Chapter 7, 325–358.

[2] C. Chen, Y. Chen, Y. Han, H. Lai, and K. J. R. Liu. 2017. Achieving Centimeter-Accuracy Indoor Localization on WiFi Platforms: A Frequency Hopping Approach. *IEEE Internet of Things Journal* 4, 1 (Feb 2017), 111–121. https://doi.org/10.1109/JIOT.2016.2628701

[3] C. Chen, Y. Chen, Y. Han, H. Lai, F. Zhang, and K. J. R. Liu. 2017. Achieving Centimeter-Accuracy Indoor Localization on WiFi Platforms: A Multi-Antenna Approach. *IEEE Internet of Things Journal* 4, 1 (Feb 2017), 122–134. https://doi.org/10.1109/JIOT.2016.2628713

[4] Ettus Research. 2017. UHD and USRP Manual. https://files.ettus.com/manual/index.html

[5] Ettus Research. 2017. USRP N210. https://www.ettus.com/product/details/UN210-KIT

[6] E. R. Magsino, I. W. H. Ho, and Z. Situ. 2017. The effects of dynamic environment on channel frequency response-based indoor positioning. In *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*. 1–6. https://doi.org/10.1109/PIMRC.2017.8292442

[7] Man-Wai Mak. 2018. Lecture 1 Constrained Optimization and Support Vector Machines. http://www.eie.polyu.edu.hk/~mwmak/.

[8] Andrew Ng. 2018. Lecture 10 Advice for Applying Machine Learning. https://www.coursera.org/learn/machine-learning/home/welcome.

[9] Andrew Ng. 2018. Lecture 6 Logistic Regression. https://www.coursera.org/learn/machine-learning/home/welcome.

[10] George Oguntala, Raed Abd-Alhameed, Stephen Jones, James Noras, Mohammad Patwary, and Jonathan Rodriguez. 2018. Indoor location identification technologies for real-time IoT-based applications: An inclusive survey. *Computer Science Review* 30 (2018), 55 – 79. https://doi.org/10.1016/j.cosrev.2018.09.001

[11] J. M. Rocamora, I. W. Ho, and M. Mak. 2018. The Application of Machine Learning Techniques on Channel Frequency Response Based Indoor Positioning in Dynamic Environments. In *2018 IEEE International Conference on Sensing, Communication and Networking (SECON Workshops)*. 1–4. https://doi.org/10.1109/SECONW.2018.8396358

[12] The GNU Radio Foundation. 2017. GNU Radio: The Free and Open Software Radio Ecosystem. https://www.gnuradio.org/

[13] Z. Wu, Y. Han, Y. Chen, and K. J. R. Liu. 2015. A Time-Reversal Paradigm for Indoor Positioning System. *IEEE Transactions on Vehicular Technology* 64, 4 (April 2015), 1331–1339. https://doi.org/10.1109/TVT.2015.2397437

[14] Zheng Yang, Zimu Zhou, and Yunhao Liu. 2013. From RSSI to CSI: Indoor Localization via Channel Response. *ACM Comput. Surv.* 46, 2, Article 25 (Dec. 2013), 32 pages. https://doi.org/10.1145/2543581.2543592