

Article

Accurate Indoor Localization Based on CSI and Visibility Graph

Zhefu Wu ¹, Lei Jiang ¹, Zhuangzhuang Jiang ¹, Bin Chen ², Kai Liu ¹, Qi Xuan ¹ and Yun Xiang ^{1,*}

¹ College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China; wuzf@zjut.edu.cn (Z.W.); 2111603003@zjut.edu.cn (L.J.); 201403090110@zjut.edu.cn (Z.J.); lk@zjut.edu.cn (K.L.); xuanqi@zjut.edu.cn (Q.X.)

² School of Design, Zhejiang University of Technology, Hangzhou 310023, China; cjb@zjut.edu.cn

* Correspondence: xiangyun@zjut.edu.cn

Received: 12 June 2018; Accepted: 29 July 2018; Published: 3 August 2018



Abstract: Passive indoor localization techniques can have many important applications. They are nonintrusive and do not require users carrying measuring devices. Therefore, indoor localization techniques are widely used in many critical areas, such as security, logistics, healthcare, etc. However, because of the unpredictable indoor environment dynamics, the existing nonintrusive indoor localization techniques can be quite inaccurate, which greatly limits their real-world applications. To address those problems, in this work, we develop a channel state information (CSI) based indoor localization technique. Unlike the existing methods, we employ both the intra-subcarrier statistics features and the inter-subcarrier network features. Specifically, we make the following contributions: (1) we design a novel passive indoor localization algorithm which combines the statistics and network features; (2) we modify the visibility graph (VG) technique to build complex networks for the indoor localization applications; and (3) we demonstrate the effectiveness of our technique using real-world deployments. The experimental results show that our technique can achieve about 96% accuracy on average and is more than 9% better than the state-of-the-art techniques.

Keywords: indoor localization; CSI; visibility graph

1. Introduction

Indoor localization is an important technique and can be used in numerous real-world applications [1–3]. However, unlike outdoor positioning, indoor localization remains a very challenging and largely unsolved problem. Being able to accurately determine indoor personal positions is a nontrivial task. For example, the indoor environment can be complex, unpredictable, and mingled with walking personals. Therefore, in this work, we propose a novel and accurate indoor localization technique which combines the statistics based and network based CSI features.

Indoor localization is widely used in many wireless applications. The localization accuracy and efficiency demand is ever increasing. The specific application scenarios of indoor location techniques include object tracking, personnel localization, elderly care, somatosensory games [4], etc. Under certain emergent conditions, such as fire rescue, an accurate, efficient and real-time localization scheme can possibly save many lives. However, traditional indoor location techniques employ a received signal strength indicator (RSSI) as the reference physical indicator to determine locations [5,6]. The main problem of RSSI is its lack of stability and sensitivity. It can easily fluctuate over time, resulting in low localization accuracy. Therefore, we use the more stable and accurate CSI as the reference physical indicator [7–9].

However, there are still many challenges for the existing CSI based indoor localization methods. One major problem is that in their models they typically do not consider the inter-subcarrier

correlations. Most existing algorithms only utilize the amplitude and phase information of individual subcarriers. However, the frequency correlations among neighboring subcarriers can also contain important location related information. Therefore, we propose using the VG method [10] to model the inter-subcarrier correlations. The VG technique can transform time or frequency series data into a complex network. The connections of the network nodes are determined by the geometric visual relations of adjacent sequential points. Through the construction of the corresponding complex network, we can derive a CSI based network with subcarriers as the network nodes and the geometric visibility as the edges. It can reflect the frequency correlations between adjacent subcarriers. Based on the network, we extract corresponding network features for indoor localization.

To improve the localization accuracy, we propose a passive indoor location technique based on both statistical and network CSI features. Our technique employs the fingerprint library and can be divided into training and testing stages. During the offline training stage, the CSI information is collected and pre-processed. Then, we construct the VG based complex network from the collected 30 sequential subcarriers. After that, both the network and statistical features are extracted and processed by the appropriate machine learning techniques to form the fingerprint library. During the online testing stage, the testing data are processed similarly. Then, machine learning algorithms are used to classify and estimate indoor locations.

In general, in this work, we have made the following main contributions:

1. We propose a novel passive indoor localization algorithm, which combines both intra-subcarrier statistics features and inter-subcarrier network features. It can greatly improve the localization performance.
2. We develop a modified VG based method to process the frequency-series subcarrier data. Our approach can explore the intra-subcarrier and inter-subcarrier correlations between adjacent subcarriers.
3. We validate our theories and techniques in real-world deployments. The results confirm that our technique can significantly outperform state-of-the-art indoor localization techniques and are more robust as well.

The remaining parts of this article are organized as follows: Section 2 describes the related work; Section 3 introduces the relevant theoretical knowledge; Section 4 presents our technique; Section 5 discusses the experimental setup and the analysis of the results; Section 6 summarizes the full text.

2. Related Work

The related work can be generalized into two categories, which are indoor localization techniques and complex network techniques, respectively.

2.1. Indoor Localization Techniques

Indoor localization is important for many mobile applications. Therefore, accurate and effective indoor positioning has received widespread research interests. Previously, wireless indoor localization techniques mainly rely upon RSSI [11–14], which can be inaccurate, unstable, and vulnerable to multi-path effects. Therefore, CSI based techniques are becoming mainstream. Sen et al. [15,16] propose PinLoc, which is based on the CSI fingerprint library. It can achieve meter level localization accuracy. Wu et al. develop FILA, a fine-grained indoor localization technique [17]. Shaw et al. design the Pilot algorithm [18]. Nasser et al. propose the MonoPHY technique [19]. Wu et al. propose a passive indoor localization technique based on CSI and Naive Bayes classifier [20]. There are problems and challenges for the existing CSI based localization techniques. The traditional CSI based localization methods only utilize the statistical features extracted from the individual subcarrier and do not consider the relationship between subcarriers. However, the relationship may contain important position information. Thus, it is important to quantify the relationship of adjacent subcarriers.

2.2. Complex Network Techniques

Complex network refers to a network with the properties of self organization, self similarity, attractor, small world, and scale free degrees [21]. In real-world applications, many complex systems, such as power grids, aeronautical networks, traffic networks, internet, and social networks, etc., can be modeled using complex networks [22–24]. Meanwhile, the visibility graph is an efficient and thus widely used technique to construct complex networks from complex systems, e.g., CSI. Lacasa et al. proposed the VG based technique to transform time series data into complex networks [10]. Gao et al. improve upon the original VG method and extract the network features for epileptic classification [25,26]. Yan et al. utilize the VG technique to extract network features and distinguish traffic states [27]. Zhu et al. employ multiple VG techniques to transform human sleep data into complex networks and extract network features for sleep classification [28]. Thus, existing research works demonstrate that using VG network to transform time series data into complex networks can reveal the internal relationship of data and improve classification results. Therefore, we propose to transform the time series CSI data into complex network.

3. Preliminaries

In this section, we describe the basic theories and properties of CSI and VG.

3.1. CSI and Localization

Compared with RSSI, CSI can provide more detailed and fine-grain subcarrier information. The channel frequency response (CFR) of a typical wireless channel can be expressed using the following equation:

$$Y = HX + N, \quad (1)$$

where X is the transmitting signal vector, Y is the receiving signal vector, H is the channel state matrix and N is the Gaussian white noises. Thus, the channel state H can be calculated using the following equation:

$$\hat{H} = \frac{Y}{X}, \quad (2)$$

where \hat{H} is the CFR in the frequency domain. The corresponding CSI information can be extracted from a WiFi wireless adapter. For subcarrier k with central frequency f_k , its CSI is defined as $H(f_k)$. The CSI signature contains both amplitude and phase information, which is defined in the following equation:

$$H(a) = |H(a)|e^{j\sin\angle H(a)}, \quad (3)$$

where $H(a)$ is the CSI for the a th subcarrier, $|H(a)|$ is its amplitude, and $\angle H(a)$ is its phase. For multiple in multiple out (MIMO) systems with multiple receiving and transmitting antennas, CSI of each subcarrier can be expressed as a $p \times q$ matrix, where p is the number of transmitting antennas and q is the number of receiving antennas. Thus, for each data packet, we can derive a $p \times q \times N$ matrix, where N is the total number of subcarriers. In this work, the number of N is set to 30 [29].

The underlying reasoning for applying the CSI information in indoor localization applications is that, when the personnel are in different locations, the signal paths and the corresponding multi-path effects vary. Thus, we can observe different reactions and variations on the CSIs. To better identify the human location, the CSI signature should be both stable in the same location and differentiable in the different locations. This characteristics can be readily observed in Figure 1.

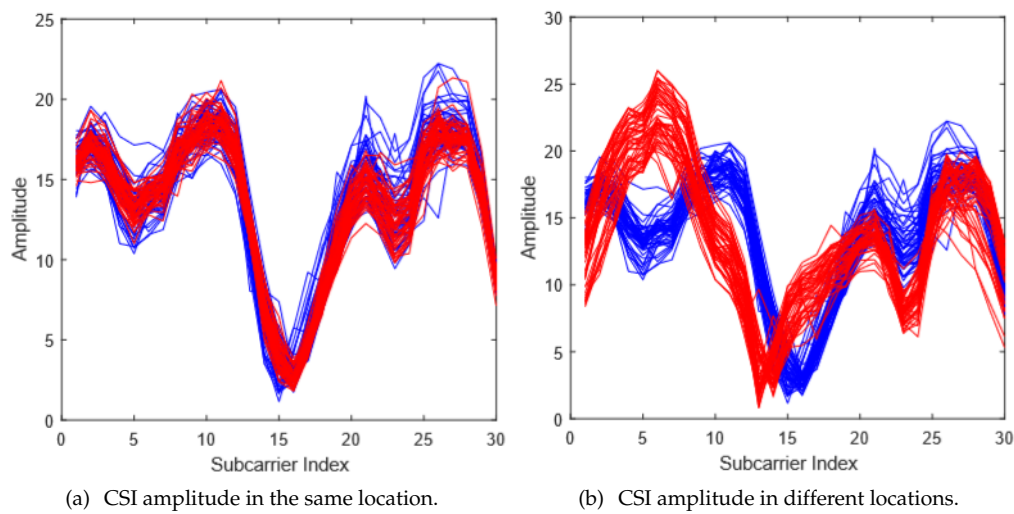


Figure 1. The CSI amplitude comparison.

3.2. VG Introduction

The VG method is widely used to transform time series data into complex networks [10]. Assuming there are three different points in the time series data space, which are denoted as a , b , and c . The network can be constructed using the following equation:

$$y_c < y_b + (y_a - y_b) \frac{t_b - t_c}{t_b - t_a}, \quad (4)$$

where y represents the value on the y -axis and t represents time. Thus, based on Equation (4), we can derive an undirected complex network. A complex network construction example is shown in Figure 2.

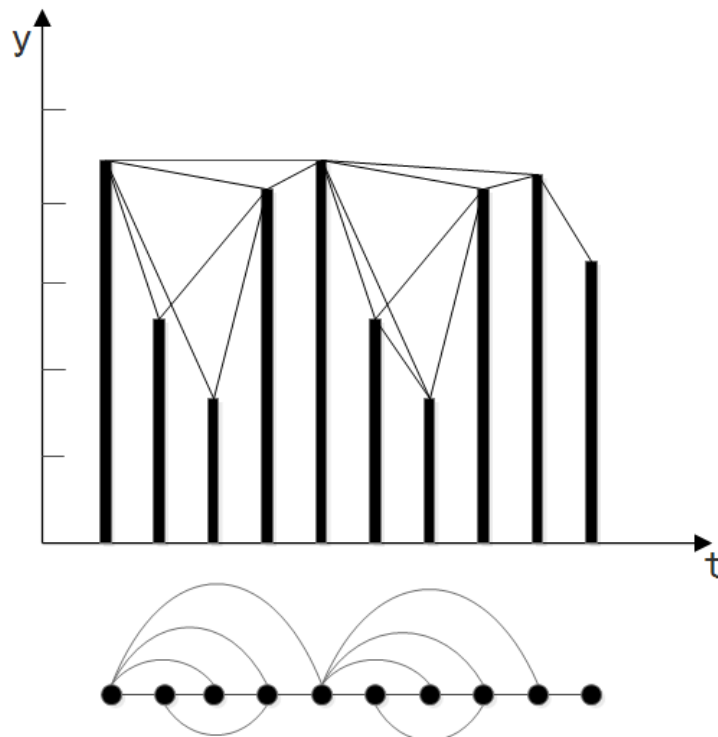


Figure 2. A VG construction example for time series data.

3.3. Machine Learning Algorithms

We introduce several widely used machine learning algorithms and implement them later.

3.3.1. Bayesian Network

Bayesian network (BNet) models the process of human cognitive reasoning. It utilizes a directed acyclic graph (DAG) and its associated conditional probability tables to model the causal inference relation of uncertain events. Each node in DAG represents a random variable, which can be directly observed or hidden. Edges represent the conditional dependency between corresponding variables. Each element in the conditional probability table represents the joint conditional probability between the node and its precedents. By training the Bayesian network, the conditional probability table can be filled and thus used for classification.

3.3.2. Support Vector Machine

Support vector machine (SVM) is a powerful and widely used machine learning technique. The key is to find the boundary hyper plane among different categories, so that different types of samples can be as far away from the boundary hyper plane as possible. Thus, SVM can exhibit an excellent classification ability. Recent SVM techniques utilize various kernel functions, e.g., Gaussian kernel function, to project the plane into a curved surface and greatly improve the performance and application scenarios. SVM is typically resilient to attacks and noises.

3.3.3. Random Forest

Random forest (RF) is an ensemble machine learning algorithm for tasks like classification and regression, etc. Its weak classifiers typically use the classification and regression tree. It first generates a large quantity of decision trees by random selection, and then combines the results of these decision trees to make the final classification decision. Random forest is widely used in real-world applications. Since RF utilizes random sampling, it can minimize modeling variance and have outstanding generalization ability. Therefore, RF is usually resilient to attacks and accurate in general.

4. VG Based Indoor Localization Method

Our proposed technique includes three parts, which are VG network construction, network feature extraction, and CSI fingerprint library. The flow of our method is presented in Figure 3.

4.1. VG Network Construction

The CSI signatures can be considered as frequency series data. Thus, to explore the correlations between adjacent subcarriers, we propose using the VG technique to transfer frequency series CSI signatures to complex networks. The CSI data contain both amplitude and phase information. Specifically, assuming that there are three different subcarriers a , b , and c , we use the following equation to derive the VG network for amplitude:

$$A_c < A_b + (A_a - A_b) \frac{f_b - f_c}{f_b - f_a}, \quad (5)$$

where, for the i th subcarrier, A_i represents its amplitude and f_i represents its frequency.

Similarly, the CSI phase information can also be used to construct the corresponding VG network using the following equation:

$$P_c < P_b + (P_a - P_b) \frac{f_b - f_c}{f_b - f_a}, \quad (6)$$

where P_i is the phase value for the i th subcarrier.

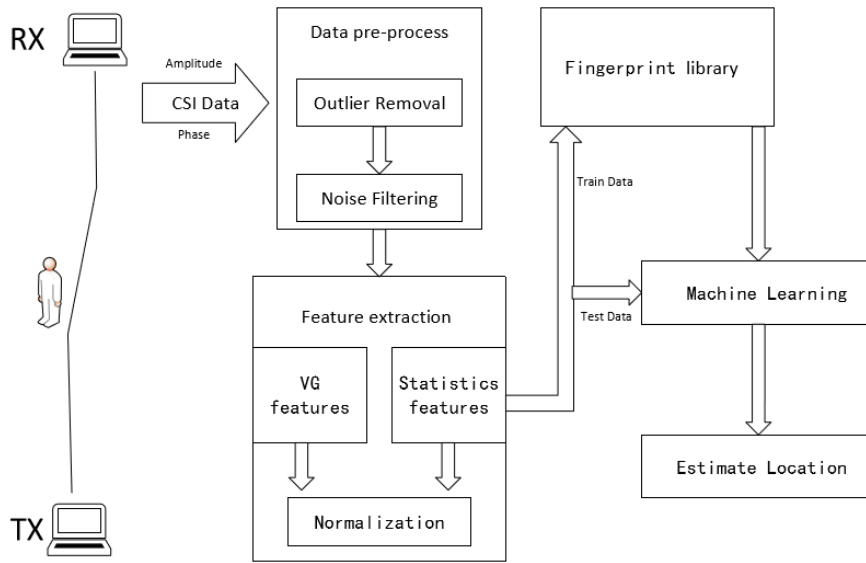


Figure 3. The system flowchart.

Figure 4 shows an example of transforming 30 subcarriers into the VG based complex network. In the figure, the x -axis is the subcarrier frequencies and the y -axis is the amplitude. By applying the VG based network construction rule as presented in Equation (5), we can derive the corresponding complex network.

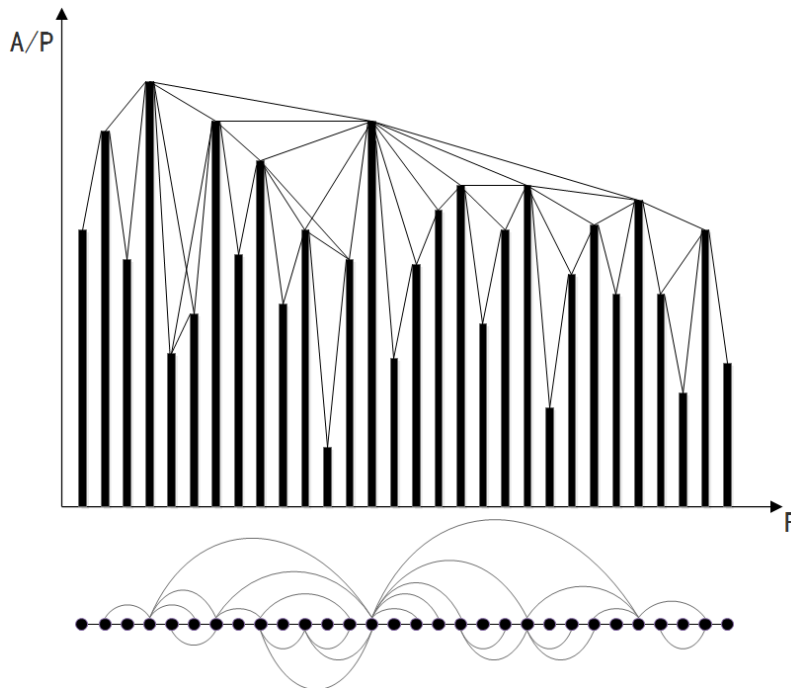


Figure 4. The VG construction process for frequency series CSI data.

In general, we construct an undirected and acyclic network from frequency series CSI signature. We first assign subcarrier indices as network nodes. Then, we apply the VG technique to establish the connections between nodes. Thus, we can use existing network techniques to analyze the CSI data and greatly increasing the number of features for classification.

4.2. Network Feature Extraction

After the construction of CSI complex networks, we extract the network features. In this work, we utilize the following features:

1. Degree deviation [25]

The degree deviation can be calculated using the following equation:

$$k_{std} = \left(\frac{\sum_{i=1}^N (k_i - \bar{k})^2}{N - 1} \right)^{\frac{1}{2}},$$

$$\bar{k} = \frac{1}{N} \sum_{i=1}^N k_i, \quad (7)$$

where k_i represents the number of nodes connecting to node i , k_{std} represents the degree deviation, \bar{k} represents the average degree, and N is the total number of nodes in the network.

2. Degree assortativity coefficient [30]. The Degree assortativity coefficient feature can be extracted using the following equation:

$$r = \frac{M^{-1} \sum_i j_i k_i - [M^{-1} \sum_i \frac{1}{2} (j_i + k_i)]^2}{M^{-1} \sum_i \frac{1}{2} (j_i^2 + k_i^2) - [M^{-1} \sum_i \frac{1}{2} (j_i + k_i)]^2}, \quad (8)$$

where j_i and k_i is the degree of the two nodes connected by edge i and M is the total number of edges.

3. Clustering coefficient entropy [25] The clustering coefficient entropy feature is extracted as follows:

$$C_i = \frac{\tau_{i,\Delta}}{\tau_i}; \quad P_{C,i} = \frac{C_i}{\sum_{i=1}^N C_i},$$

$$Ec = - \sum_{i=1}^N (P_{C,i}) \log(P_{C,i}), \quad (9)$$

where, for node i , C_i is its local clustering coefficient, $\tau_{i,\Delta}$ is the total number of edges connecting to all the neighboring nodes of i , τ_i is the the number of edges connecting to node i , $P_{C,i}$ is the clustering coefficient probability, N is the total number of nodes in the network, and E_C is the clustering coefficient entropy.

4. Average weighted degree [31]. The average weighted degree feature can be extracted as follows:

$$w_{ab} = \frac{1}{M} \sum \left[\arctan \frac{y(x_b) - y(x_a)}{x_b - x_a} \right], \quad (10)$$

where a and b are two separate nodes, x is the corresponding subcarrier frequency, y is amplitude or phase, and M is the total number of edges.

4.3. Fingerprint Library Creation

4.3.1. Statistical Feature Extraction

The raw data of 30 subcarriers contain large amounts of noises and useless redundant signals and, thus, usually cannot be used for classification directly. Otherwise, they can easily cause overfitting and lead to inferior classification performance. Thus, it is common practice to generalize the statistical features from the raw data. In this work, we calculate and utilize four statistical features, which are standard deviation (STD), median absolute deviation (MAD), mean value (MEAN), and median value (MEDIAN), respectively. We do not include the maximal and minimal traits in this work. Experimental

results show that, by adding those two features, the classifiers become more vulnerable to overfitting and tend to derive deteriorated results.

4.3.2. Fingerprint Library

There are significant variations in derived feature magnitudes. Therefore, we normalize the features using the Z-score method as follows:

$$F_{norm} = \frac{F - \mu}{\sigma}, \quad (11)$$

where F_{norm} is the normalized features, F is the original features, μ is the mean value, and σ is the standard deviation. The normalized features, including both statistical and network ones, are combined to create the final fingerprint library.

5. Experimental Results

5.1. Experiment Setup

The experiment platform includes two parts, which are access point (AP) and monitoring point (MP), respectively. The CSI signature is extracted using a CSI-tool [29]. For the experiment, we deploy two notebook equipped with Ubuntu 14.04 (Canonical, London, UK) and Intel 5300 network adapter (Samsung, Suzhou, China). The adapter has three MIMO antennas and thus can form six link pairs in theory. However, in the real-world environment, we can only extract three link pairs stably. Thus, we choose link pair 1-1, 1-2, and 1-3 in this work. The machine learning algorithms are run on a ASUS FH5900V computer (Shanghai, China) equipped with Intel i7-6700HQ CPU and 8 G memory.

During the experiment, the testing participator is standing on the different locations in the room. The AP sends 100 data packets per second and lasts for 10 s. Therefore, at each location, we collect 1000 samples. There are two different testing environments as shown in Figure 5. Figure 5a shows the inside view of testing environment 1, which is a small and noisy conference room. Figure 5b is its vertical view and the room size is approximately $5 \times 6 \text{ m}^2$. Figure 5c shows the inside view of testing environment 2, which is an empty and quiet class room. Figure 5b is its vertical view and the room size is approximately $6 \times 8 \text{ m}^2$.

For both environments, the distance between each data collection point is around 1 m, i.e., the localization resolution is set to be 1 m. In testing environment 1, the MP is placed at 1.2 m high, while the AP is placed at 0.5 m high. Their distance is 2.5 m. In testing environment 2, both AP and MP are placed at 0.5 m high with a distance of 7 m. At each location, we randomly select 600 samples as the training set, and the rest form the testing set.

5.2. Data Pre-Processing

The data pre-processing includes amplitude and phase extraction, abnormality processing, and data smoothing.

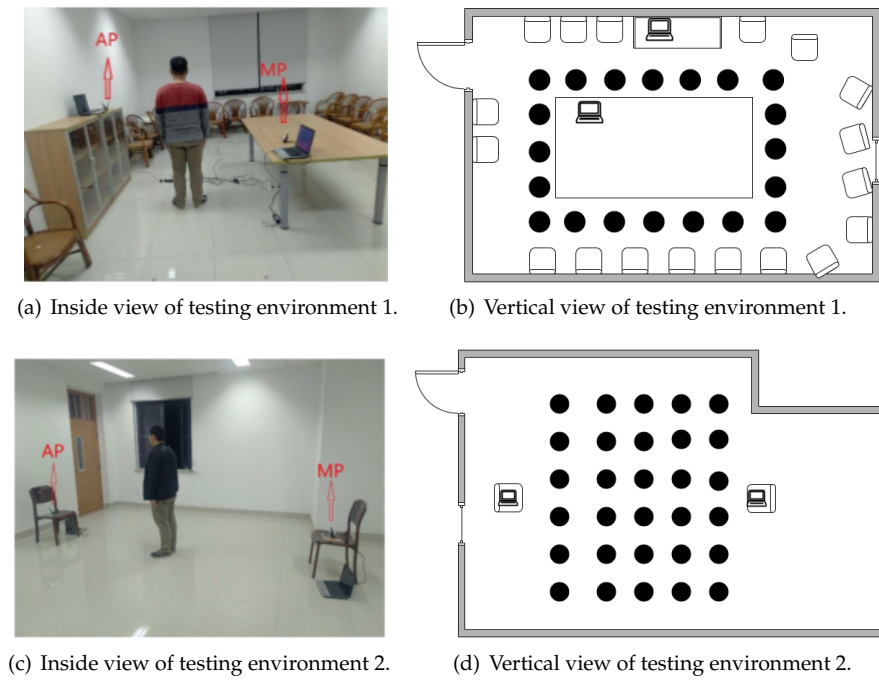


Figure 5. The experiment environments.

5.2.1. Amplitude and Phase Extraction

The CSI amplitude information is stable and easy to extract. Thus, it is the most widely used feature in indoor localization applications. However, the phase information contains significant noise and synchronization problems. Thus, the original phase information must be linearized [32]. The linearization process is shown as follows:

$$\bar{\phi}_i = \tilde{\phi}_i - ak_i - b = \phi_i - \frac{\phi_n - \phi_1}{k_n - k_1} k_i - \frac{1}{n} \sum_{j=1}^n \phi_j, \quad (12)$$

where $\tilde{\phi}$ is the measured phase, ϕ_i is the actual phase, and k_i is the index of the corresponding subcarrier.

5.2.2. Abnormality Processing

Abnormal data have a great impact on training and testing performances. To remove the abnormal data points, we use the Pauta criterion, which is shown in the following equation:

$$V_i^F = \begin{cases} mean_F & |V_i^n - mean_F| > 3std_F, \\ 0, & else, \end{cases} \quad (13)$$

where, for the i th sample and F th subcarrier, V is the CSI value, $mean$ is the mean value and std is the standard deviation.

5.2.3. Data Smoothing

Even after removing the outliers, there are still significant short-term fluctuations among adjacent samples. To remove the unwanted noise fluctuations, we use the moving average filter to process the data [33]. The method is shown in the following equation:

$$CS\bar{I}_n^i = \frac{1}{m} (CSI_n^i + CSI_{n-1}^i + \dots + CSI_{n-m+1}^i), \quad (14)$$

where m is the size of the sliding window, n is the index of the data packets, and i is the index of the subcarriers. There is a trade-off between sliding window size, i.e., data smoothness, and signal integrity. Figure 6 shows an example of data smoothing. The red lines are the original data, while the blue lines are the smoothed one. The data fluctuations are significantly suppressed. After data smoothing, the data are classified using machine learning techniques.

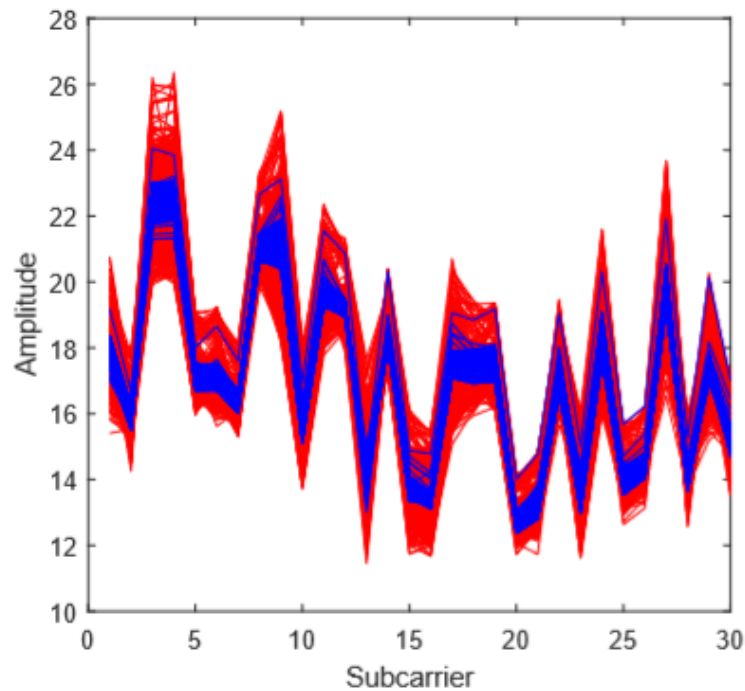


Figure 6. The data smoothing results.

5.3. Result Analysis

In this section, we evaluate the performance of our techniques.

5.3.1. Performance Comparison

In this work, we compare the performance of the following techniques:

1. Confidence: the state-of-the-art CSI based localization method [20], which uses the mean and standard deviation features extracted from CSI amplitude.
2. Statistics: a technique similar to the confidence method but employs four amplitude features and four phase features.
3. VG: a localization technique using only VG network features.
4. Combined: our technique which utilizes both VG and statistics features.

We compare the classification accuracy at two different environments and the results are shown in Figure 7. In the figure, the x -axis is the three machine learning based classification algorithms and the y -axis is the classification accuracy. Among the four compared algorithms, both VG and Combine are our proposed techniques. In both environments, the Statistics method outperforms the Confidence method for all of the three classification methods. This implies that using only the mean and standard deviation features is insufficient to capture the CSI signal spatial variations. Moreover, the combined method constantly gives the best performance, which demonstrates that the network features are truly correlated to the CSI spatial variations. In other words, different indoor locations have an impact on the adjacent subcarrier correlations.

In general, by using the SVM technique, we can achieve the best performance in both environments. For the noisy environment 1, the Confidence method has 80.1% accuracy; the VG only method is 87.2%; the Statistical method is 91.7%; and the Combined method can achieve 92.0% accuracy. Our technique is significantly better in noisy environments, which has 11.9% improvement compared with the state-of-the-art Confidence method. However, the performance improvement is less significant in environment 2. For the best SVM algorithm, our combined method achieves 95.7% accuracy, which has about 8.9% improvement compared with the Confidence method, but only 2.8% compared with the Statistics method. It is possible that the VG network features and statistics features overlap with each other and cause over-fitting.

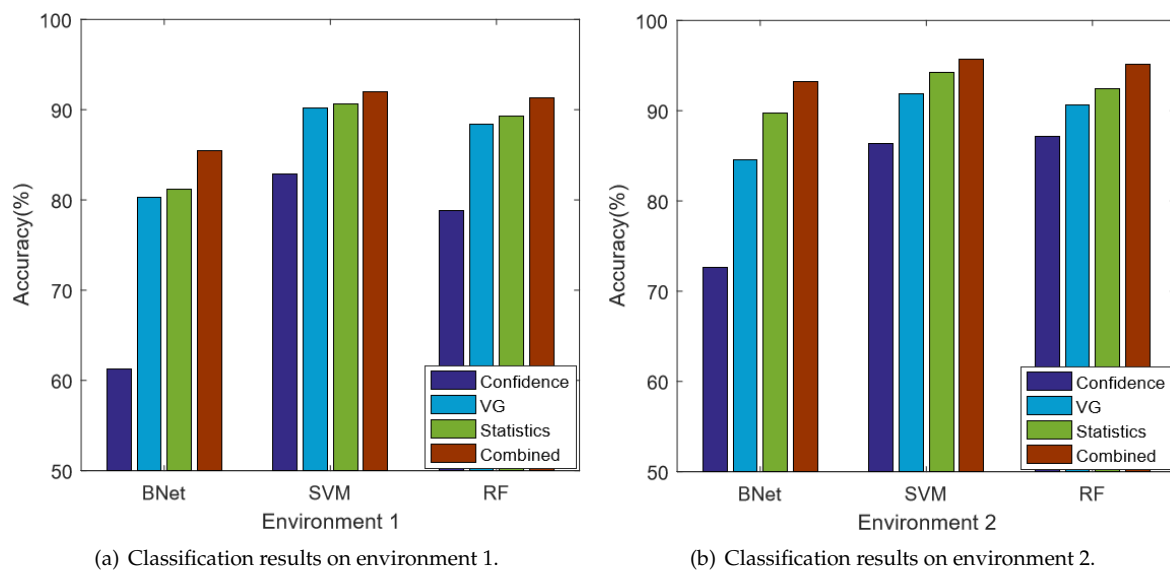


Figure 7. The comparison results.

5.3.2. Performance Analysis

To better illustrate and explain the performance of our combined technique, we plot the confusion matrices of the VG, Statistics, and Combined methods under the complex environment 1, as shown in Figure 8. In the figure, the y -axis is the indices of actual locations and the x -axis is the classification result frequencies. The orthogonal line represents the correctly classified results and the color intensities represent the frequencies.

Figure 8a,b show the corresponding confusion matrices of VG and Statistics methods, respectively. As illustrated in the figure, the misclassification locations for the VG and Statistics methods are mostly different. The annotated areas in Figure 8a,b show the corresponding misclassified locations. They are in different diagonal positions, while implying that, whenever one technique misclassifies, the other one may derive the correct result. Therefore, by combining the two different classes of features, the classification results of the combined technique can be significantly improved. For example, Figure 8a shows that location 20 is frequently misclassified as locations 3 and 6. Figure 8b shows that location 20 is frequently misclassified as locations 6, 11, 12, and 13. However, as shown in Figure 8c, by combining the features, location 20 is rarely misclassified except for few cases. It is clear evidence that the network features, which represent the inter-subcarrier correlations, are different from the statistics features generalized from individual subcarrier and hence can help improve the localization performance.

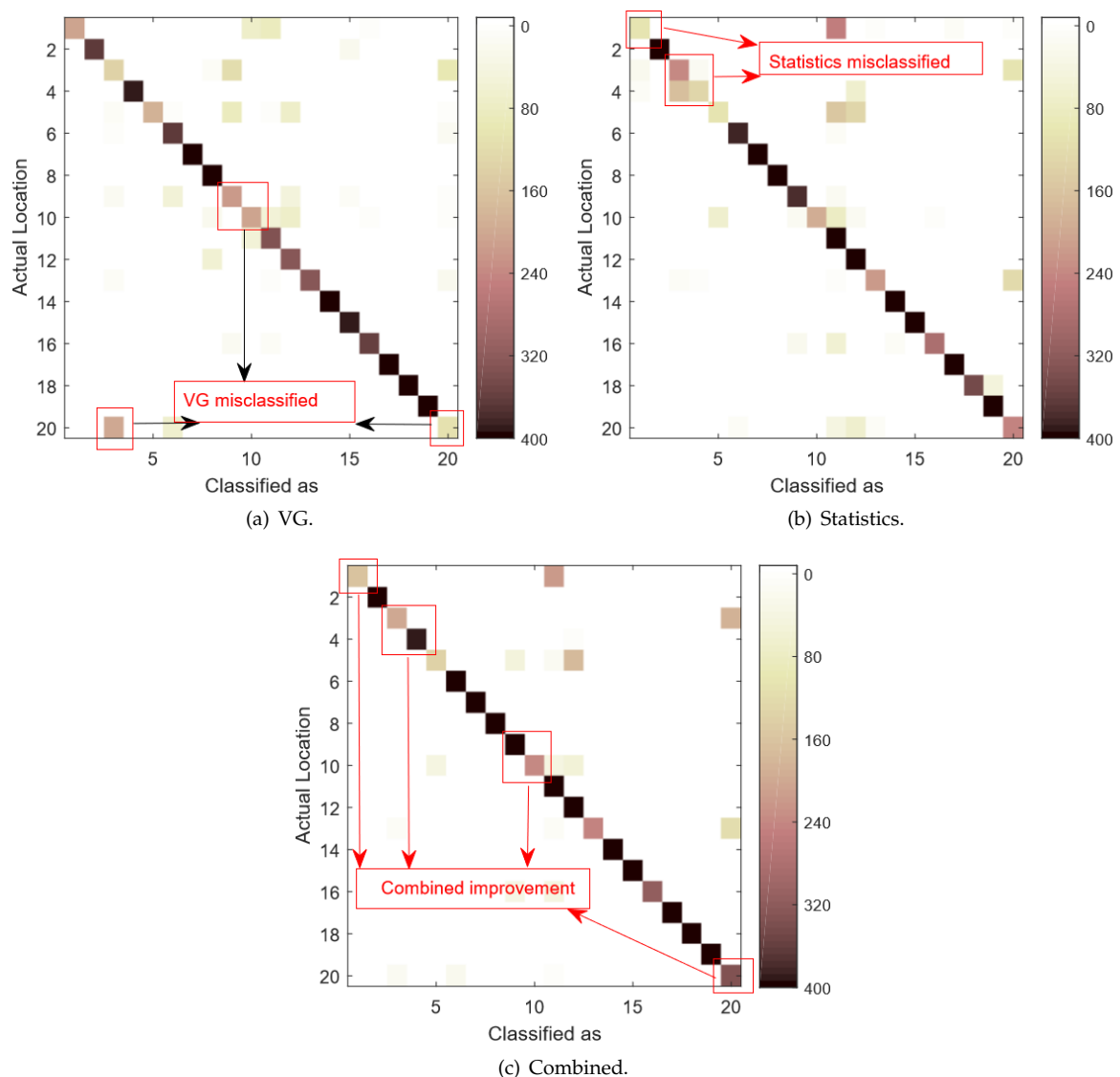


Figure 8. The confusion matrices of different methods.

It should be noted that, although our technique improves significantly, there are still some misclassifications. The reason is as follows.

1. The CSI signatures contain significant noises. Because of environment variations and multi-path effects, the CSI signals are usually unpredictable and fluctuating, which can greatly affect the localization accuracy. For example, in many scenarios, we observe that the CSI signatures of a person standing in certain locations in the middle are very close to the signatures where one stands in the corner. In that case, those two locations are indistinguishable for any CSI based techniques. Therefore, there are always possibilities for misclassifications.
2. The feature selection methods can also cause misclassifications. In this work, we do not use raw CSI data directly. Instead, we extract features from them. Features can filter out the noises and simplify the calculation. However, it is also possible to omit useful information. Our technique is based on two feature sets, which are Statistical and VG features. They stand for the intra and inter correlations of subcarriers, respectively. It is demonstrated in Figure 7 that both feature sets can cause misclassifications. Therefore, for the locations where both feature sets predict incorrectly simultaneously, our technique also misclassifies.

To further analyze the classification results, we compare the false negative (FN) rate and false positive (FP) rate of our combined technique under different environments. Figure 9 shows the comparison results. For the best machine learning algorithm (SVM), the FN rates are constantly below 3% and the FP rates are below 1%. The performances of our algorithms are outstanding. Moreover, for both environments, the FP rate is significantly lower than the FN rate. It is a desirable result since, in most applications, the FP rate is much more important than the FN rate.

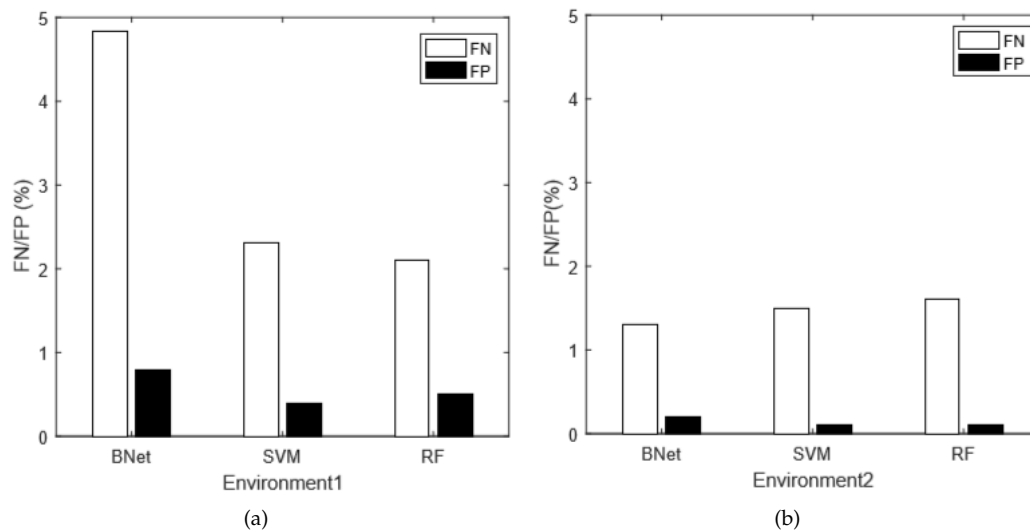


Figure 9. The FN/FP ratios on different environments.

5.3.3. Parameter Selection

We first discuss the selection of amplitude and phase parameter of the CSI signal. We compare the performance of our Combined technique using three different types of CSI data, which are amplitude, phase, and both, respectively. The comparison results are shown in Figure 10, where the x -axis is the three machine learning algorithms and the y -axis is the classification accuracy. The figure demonstrates that, by combining the amplitude and phase information, we can observe a universal improvement. The reason is that the amplitude and phase information represents different aspects of CSI characteristics.

Then, we discuss the impact of training and testing set sizes on algorithm performances. Figure 11 shows the evaluation results at the noisy testing environment 1. Our technique is significantly better than the Confidence method. As the size of the training set grows, the performances improve slightly. Moreover, it is possible that the machine learning based classifiers become overfitted. In that case, the classification performance may even drop slightly, as shown in Figure 11b. Considering the trade-off between training time and classifier performance, we employ 6:4 as the ratio to split the training and testing sets.

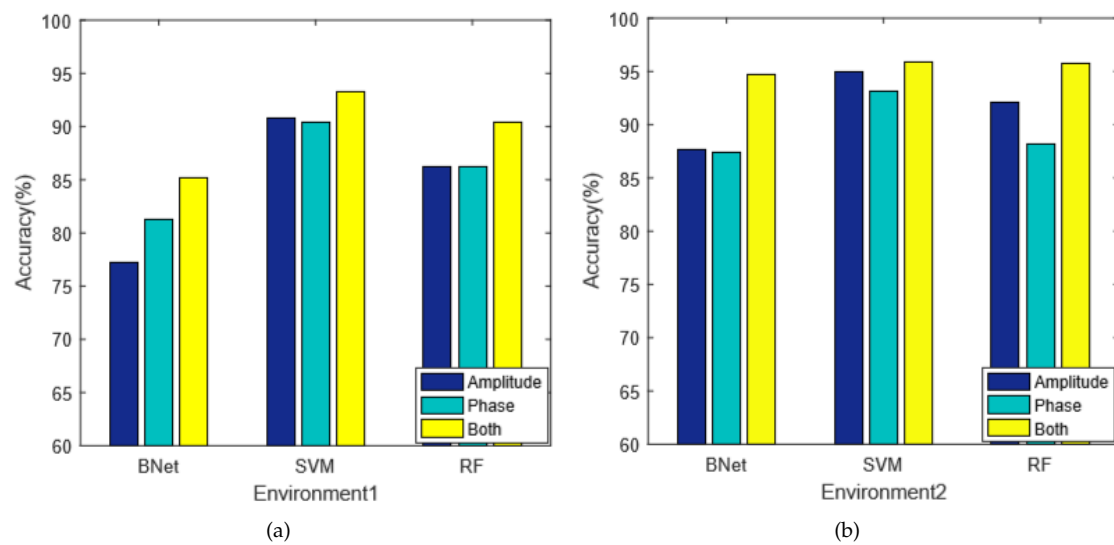


Figure 10. The amplitude/phase selection on classification accuracy.

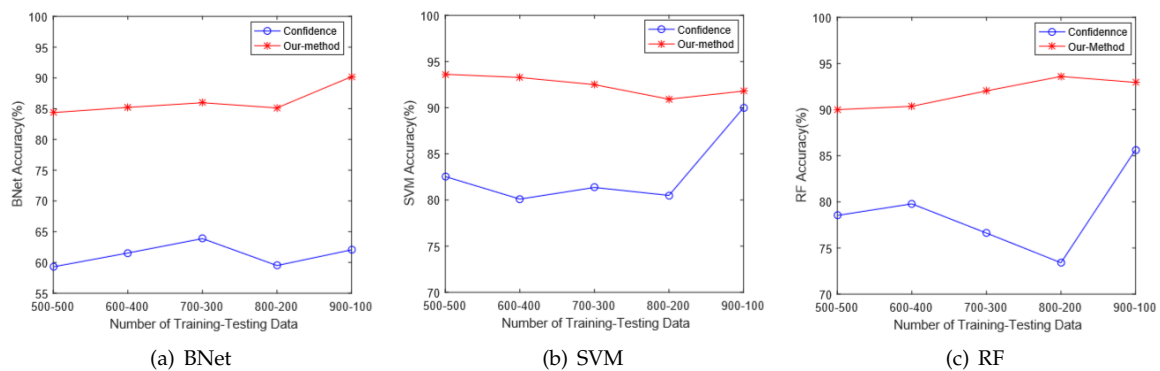


Figure 11. The impact of training and testing sizes.

Table 1 shows the classification accuracy and the corresponding training time. The SVM method has the best performance and moderate training time. Meanwhile, the Bayes net approach requires the least training time. Therefore, there is a trade-off between algorithm performance and the time taken to obtain each model. The specific machine learning technique should be determined by considering all the requirements.

Finally, we evaluate the influences of noise disturbance, e.g., presence of other persons, on algorithm performances. We add two more testing environments, which are environments 3 and 4, respectively. Figure 12 shows the details of these two environments. On testing environment 3, besides the testing personal, an additional person sits idly nearby. On testing environment 4, an additional person walks randomly inside the room. Figure 13 presents the corresponding experimental results. The results show that our technique is very robust compared with the existing methods. Especially in environment 4, the Confidence method is affected significantly by the walking person. Even for the SVM method, the classification accuracy drops below 65%, while, for our combined methods, the accuracy is still about 90%. Therefore, the experimental results demonstrate that the idly seated person does not have a significant impact on CSI based classification techniques. The human activity, on the other hand, can have significant impact. However, by considering both intra-subcarrier statistical features and inter-subcarrier network features, our Combined technique shows great resilience for such disturbance. Figure 14 shows the performance improvements compared with the Confidence method. The noisiest testing environment 4 records the greatest improvements. In general, there is a

trade-off between noises and classification accuracy. However, by considering both intra and inter subcarrier correlations, our algorithm is demonstrated to be very noise resilient.

Table 1. Performance and training time for machine learning algorithms.

Environment	BNet		SVM		RF	
	Acc(%)	Time(s)	Acc(%)	Time(s)	Acc(%)	Time(s)
Env1	85.21	0.56	93.28	1.79	90.70	4.92
Env2	84.74	1.15	95.90	3.21	95.74	8.84

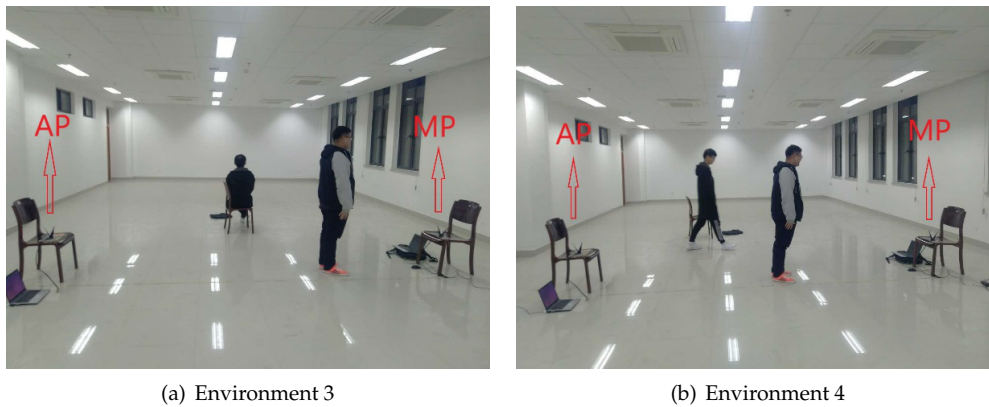


Figure 12. The impact of human interferences.

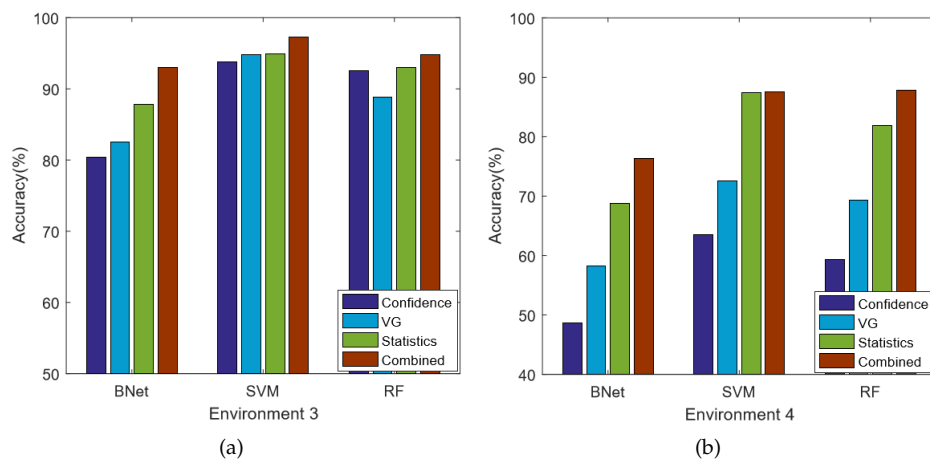


Figure 13. The impact of human interferences.

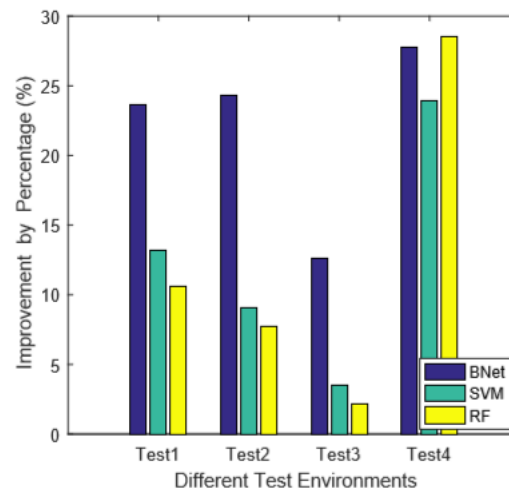


Figure 14. The overall improvements compared with the state-of-the-art technique.

6. Conclusions

In this work, we propose a CSI based indoor localization technique which utilizes both the intra-subcarrier statistics and inter-subcarrier network features. We employ the VG method to transform the CSI amplitude and phase information into complex networks and extract the network features accordingly. Then, we combine the network features with traditional statistics features and use three widely use machine learning algorithms to determine the specific indoor locations. The experimental results show that our technique can achieve 96% classification accuracy, which is about a 9% improvement compared with the state-of-the-art CSI method. Moreover, it is demonstrated that our techniques are more robust to the environment noises and disturbances.

Author Contributions: Conceptualization, Z.W. and Y.X.; Data curation, L.J. and Z.J.; Formal analysis, L.J. and Z.J.; Funding acquisition, Z.W. and Y.X.; Investigation, L.J. and Q.X.; Methodology, Z.W. and Y.X.; Project administration, Z.W.; Resources, B.C.; Supervision, Z.W. and Y.X.; Validation, Q.X.; Visualization, K.L.; Writing—Original draft, L.J.; Writing—Review and editing, Y.X.

Funding: This work is partially supported by the National Natural Science Foundation of China (61502423, 61572439) and the Zhejiang Natural Science Foundation (LY18F030021, LY18F010025, LY14F050004).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kemper, J.; Linde, H. Challenges of passive infrared indoor localization. In Proceedings of the 2008 5th Workshop on IEEE Positioning Navigation and Communication, Hannover, Germany, 27 March 2008; IEEE: Piscataway, NJ, USA, 2008; pp. 63–70.
2. Liu, Y.; Yang, Z. *Location, Localization, and Localizability: Location-Awareness Technology for Wireless Networks*; Springer: Berlin, Germany, 2011.
3. Gonçalves, R.; Reis, J.; Santana, E.; Carvalho, N.B.; Pinho, P.; Roselli, L. Smart floor: Indoor navigation based on RFID. In Proceedings of the 2013 IEEE Wireless Power Transfer (WPT), Perugia, Italy, 15–16 May 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 103–106.
4. Li, X.; Li, S.; Zhang, D.; Xiong, J.; Wang, Y.; Mei, H. Dynamic-MUSIC: Accurate device-free indoor localization. In Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing, Heidelberg, Germany, 12–16 September 2016; pp. 196–207.
5. Wu, Z.; Fu, K.; Jedari, E.; Shuvra, S.R.; Rashidzadeh, R.; Saif, M. A Fast and Resource Efficient Method for Indoor Positioning Using Received Signal Strength. *IEEE Trans. Veh. Technol.* **2016**, *65*, 9747–9758. [[CrossRef](#)]
6. Mazuelas, S.; Bahillo, A.; Lorenzo, R.M.; Fernandez, P.; Lago, F.A.; Garcia, E.; Blas, J.; Abril, E.J. Robust Indoor Positioning Provided by Real-Time RSSI Values in Unmodified WLAN Networks. *IEEE J. Sel. Top. Signal Process.* **2009**, *3*, 821–831. [[CrossRef](#)]

7. Chapre, Y.; Ignjatovic, A.; Seneviratne, A.; Jha, S. CSI-MIMO: Indoor Wi-Fi fingerprinting system. In Proceedings of the 2014 Local Computer Networks, Edmonton, AB, Canada, 8–11 September 2014; pp. 202–209.
8. Wu, K.; Xiao, J.; Yi, Y.; Chen, D.; Luo, X.; Ni, L.M. CSI-Based Indoor Localization. *IEEE Trans. Parallel Distrib. Syst.* **2013**, *24*, 1300–1309. [[CrossRef](#)]
9. Li, F.; Al-Qaness, M.A.A.; Zhang, Y.; Zhao, B.; Luan, X. A Robust and Device-Free System for the Recognition and Classification of Elderly Activities. *Sensors* **2016**, *16*, 2043. [[CrossRef](#)] [[PubMed](#)]
10. Lacasa, L.; Luque, B.; Ballesteros, F. From time series to complex networks: the visibility graph. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 4972–4975. [[CrossRef](#)] [[PubMed](#)]
11. Zanca, G.; Zorzi, F.; Zanella, A.; Zorzi, M. Experimental comparison of RSSI-based localization algorithms for indoor wireless sensor networks. In Proceedings of the 2008 Workshop on Real-World Wireless Sensor Networks, Glasgow, UK, 1–4 April 2008; pp. 1–5.
12. Paul, A.S.; Wan, E.A. RSSI-Based Indoor Localization and Tracking Using Sigma-Point Kalman Smoothers. *IEEE J. Sel. Top. Signal Process.* **2009**, *3*, 860–873. [[CrossRef](#)]
13. Ahn, H.S.; Yu, W. Environmental-Adaptive RSSI-Based Indoor Localization. *IEEE Trans. Autom. Sci. Eng.* **2009**, *6*, 626–633.
14. El-Din, R.A.Z.; Rizk, M. Accurate indoor localization based on RSSI with adaptive environmental parameters in wireless sensor networks. In Proceedings of the First International Conference on Innovative Engineering Systems, Alexandria, Egypt, 7–9 December 2012; pp. 177–183.
15. Sen, S.; Radunovic, B.; Choudhury, R.R.; Minka, T. Precise indoor localization using PHY information. In Proceedings of the 2011 International Conference on Mobile Systems, Applications, and Services, Washington, DC, USA, 28 June–1 July 2011; pp. 413–414.
16. Sen, S.; Choudhury, R.R.; Minka, T. You are facing the Mona Lisa: Spot localization using PHY layer information. In Proceedings of the 2012 International Conference on Mobile Systems, Applications, and Services, Lake District, UK, 26–29 Jun 2012; pp. 183–196.
17. Wu, K.; Xiao, J.; Yi, Y.; Gao, M.; Ni, L. FILA: Fine-grained indoor localization. *Proc. IEEE INFOCOM* **2012**, *131*, 2210–2218.
18. Xiao, J.; Wu, K.; Yi, Y.; Wang, L.; Ni, L.M. Pilot: Passive Device-Free Indoor Localization Using Channel State Information. In Proceedings of the 2013 IEEE 33rd International Conference on Distributed Computing Systems, Philadelphia, PA, USA, 8–11 July 2013; pp. 236–245.
19. Abdel-Nasser, H.; Samir, R.; Sabek, I.; Youssef, M. MonoPHY: Mono-Stream-based Device-free WLAN Localization via Physical Layer Information. In Proceedings of the 2013 IEEE Wireless Communications and Networking Conference (WCNC 2013), Shanghai, China, 7–10 April 2013; pp. 4546–4551.
20. Wu, Z.; Xu, Q.; Li, J.; Fu, C.; Xuan, Q.; Xiang, Y. Passive Indoor Localization Based on CSI and Naive Bayes Classification. *IEEE Trans. Syst. Man Cybern. Syst.* **2017**, *PP*, 1–12. [[CrossRef](#)]
21. Xuan, Q.; Zhou, M.; Zhang, Z.Y.; Fu, C.; Xiang, Y.; Wu, Z.; Filkov, V. Modern Food Foraging Patterns: Geography and Cuisine Choices of Restaurant Patrons on Yelp. *IEEE Trans. Comput. Soc. Syst.* **2018**, *5*, 508–517. [[CrossRef](#)]
22. Fu, C.; Zhao, M.; Fan, L.; Chen, X.; Chen, J.; Wu, Z.; Xia, Y.; Xuan, Q. Link Weight Prediction Using Supervised Learning Methods and Its Application to Yelp Layered Network. *IEEE Trans. Knowl. Data Eng.* **2018**. [[CrossRef](#)]
23. Hu, H.X.; Wen, G.; Yu, W.; Xuan, Q.; Chen, G. Swarming Behavior of Multiple Euler-Lagrange Systems With Cooperation-Competition Interactions: An Auxiliary System Approach. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *PP*, 1–12. [[CrossRef](#)] [[PubMed](#)]
24. Xuan, Q.; Zhang, Z.Y.; Fu, C.; Hu, H.X.; Filkov, V. Social Synchrony on Complex Networks. *IEEE Trans. Cybern.* **2018**, *48*, 1420–1431. [[CrossRef](#)] [[PubMed](#)]
25. Gao, Z.K.; Cai, Q.; Yang, Y.X.; Dong, N.; Zhang, S.S. Visibility Graph from Adaptive Optimal Kernel Time-Frequency Representation for Classification of Epileptiform EEG. *Int. J. Neural Syst.* **2017**, *27*, 1750005. [[CrossRef](#)] [[PubMed](#)]
26. Gao, Z.K.; Cai, Q.; Yang, Y.X.; Dang, W.D. Time-dependent limited penetrable visibility graph analysis of nonstationary time series. *Phys. A Stat. Mech. Appl.* **2017**, *476*, 43–48. [[CrossRef](#)]
27. Yan, Y.; Zhang, S.; Tang, J.; Wang, X. Understanding characteristics in multivariate traffic flow time series from complex network structure. *Phys. A Stat. Mech. Appl.* **2017**, *477*, 149–160. [[CrossRef](#)]

28. Zhu, G.; Li, Y.; Wen, P.P. Analysis and classification of sleep stages based on difference visibility graphs from a single-channel EEG signal. *IEEE J. Biomed. Health Inform.* **2014**, *18*, 1813–1821. [[CrossRef](#)] [[PubMed](#)]
29. Halperin, D.; Hu, W.; Sheth, A.; Wetherall, D. Tool release: Gathering 802.11 n traces with channel state information. In *ACM SIGCOMM Computer Communication Review*; ACM Digital Library: New York, NY, USA, 2011; Volume 41, p. 53.
30. Newman, M.E. Mixing patterns in networks. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* **2003**, *67*, 026126. [[CrossRef](#)] [[PubMed](#)]
31. Supriya, S.; Siuly, S.; Wang, H.; Cao, J.; Zhang, Y. Weighted Visibility Graph With Complex Network Features in the Detection of Epilepsy. *IEEE Access* **2016**, *4*, 6554–6566. [[CrossRef](#)]
32. Wu, C.; Yang, Z.; Zhou, Z.; Qian, K. PhaseU: Real-time LOS identification with WiFi. In Proceedings of the 2015 Computer Communications, Hong Kong, China, 26 April–1 May 2015; pp. 2038–2046.
33. Han, C.; Wu, K.; Wang, Y.; Ni, L.M. WiFall: Device-free fall detection by wireless networks. In Proceedings of the IEEE INFOCOM 2014—IEEE Conference on Computer Communications, Toronto, ON, Canada, 27 April–2 May 2014; pp. 271–279.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).