

AW-FLoc: An Adaptive Weighted Fusion Localization Algorithm Based on CSI and Multidirectional Image

Wen Liu¹, Mingjie Jia¹, Zhongliang Deng¹, Shude Li¹ and Yingeng Zhang¹

¹*School of Electronics Engineering, Beijing University of Posts and Telecommunications, Beijing 100089, China*

Abstract

Localization algorithms play an important role in various location-based services (LBS), including production, public services, living, entertainment, and social networks. In order to improve the accuracy of the indoor positioning algorithm, this paper proposes an adaptive weight-based channel state information (CSI) and multi-azimuth image fusion positioning method to solve the limitation of single-sensor positioning. In order to reduce the image dimension and weaken the interference of low-texture images on fusion features, an image feature fusion algorithm based on texture degree evaluation (TDE-IFF) is proposed. In TDE-IFF, the texture evaluation algorithm assigns weights to images of different orientations and fuses image features of opposite orientations based on the weights to obtain a vectorized representation of multi-orientation images. On this basis, this paper proposes an adaptive weight-based fusion location framework (AW-FLoc), which measures CSI features and multi-directional image features through a gating network to adaptively enhance the effective signal features of different fingerprint points. This improves the accuracy of the fusion positioning algorithm. Experiments show that the fusion positioning based on AW-FLoc achieves better positioning performance, and the average error in the comprehensive office scene is 0.61m.

Keywords

Indoor localization, adaptive weighted, fusion, localization

1. Introduction

Location algorithms play an important role in a variety of location-based services (LBS), including production, public services, living, entertainment, and social networks. With the intelligent development of the information industry, LBS has seen wider development in rescue and relief, logistics tracking, and high-end manufacturing cite [1] [2]. Therefore, there is an urgent need for highly reliable, high-precision indoor positioning technology to accomplish the task of positioning in complex indoor environments [3].

In terms of how to improve the accuracy of indoor positioning algorithms, it has been shown that multi-source fusion is more advantageous than single-source positioning. The reason for this is that multiple sources provide a richer set of signal characteristics, whereas single sources inevitably suffer from feature loss, signal distortion, and signal-to-noise ratio degradation due to interference, occlusion, and non-visual range caused by the complex indoor environment. Therefore, single-source positioning algorithms can achieve a certain degree of positioning

IPIN 2022 WiP Proceedings, September 5 - 7, 2022, Beijing, China


✉ liuwen@bupt.edu.cn (W. Liu); jiamingjie@bupt.edu.cn (M. Jia); dengzhl@bupt.edu.cn (Z. Deng);

lishude123654@126.com (S. Li); zyg99@126.com (Y. Zhang)

ORCID 0000-0002-6450-1969 (W. Liu); 0000-0002-6862-7905 (M. Jia)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

accuracy, but it is difficult to ensure the stability of the positioning. Multi-source localization methods provide a more adequate signal and can further improve localization accuracy and stability. Therefore, positioning methods based on multi-source signal fusion are gradually becoming a key focus of research in positioning technology.

As the frequency response signal of the channel, CSI can reflect the signal changes in the indoor environment, including signal reflection, diffraction, and power changes, with higher granularity, and compared to the received signal strength, CSI is an essential description of the propagation process of wireless signals in space. Although CSI-based localization has been publicly reported [4] to have worked with positioning accuracy up to the meter level, it still suffers from insufficient stability and mismatching in complex indoor environments. Images, on the other hand, are widely used in positioning and navigation due to their low acquisition cost. Although image-based indoor positioning methods have the advantages of rich location features and high stability, there are repetitive regions in the image signals collected from different locations, resulting in low feature differentiation. In summary, in the field of indoor positioning, both image and CSI signals contain rich location features and can form complementary roles. Therefore, it is theoretically possible to implement a fusion algorithm for stable and highly accurate positioning services using both signals.

Existing fusion multi-source fusion localization methods usually use model integration or feature fusion. The former lacks the interaction between different features and has the problem of insufficient fusion depth. Although the latter increases the interaction between features, the current feature fusion method is relatively simple, and it is difficult to take advantage of the positioning advantages of different signal sources. Moreover, the fused features lack the evaluation of different signals, and it is easy to introduce noise signals, resulting in insufficient positioning accuracy of the algorithm.

The most important contribution of this paper is to design an algorithm framework suitable for multi-source fusion positioning and apply it to the fusion positioning of CSI and image signals. The detailed innovation points are as follows:

1. Given the large dimension of multi-directional image features and the introduction of extra noise in low-texture images. In this paper, a multi-directional image fusion method based on texture degree evaluation (TDE-IFF) is proposed. The method first scores the texture degree of the image and uses the score as the weight of the image features. Afterward, the method performs image feature fusion of opposite orientations based on texture score to enhance the spatial perception ability of the network on each coordinate axis.
2. For the problem of lack of contribution evaluation of multi-source signals, this paper proposes a fusion localization model with adaptive weights (AW-FLoc). The model first analyzes the contribution of different signals to positioning and proposes an adaptive weighting strategy based on the gating mechanism to realize the complementary advantages of different signals, thereby generating more effective location features.
3. In addition, this paper verifies the effectiveness of the above two methods through comparative experiments and ablation experiments.

2. Related works

In the data representation of CSI, we use a representation method based on a multi-head self-attention mechanism and efficient CSI fusion (MHSA-EC) [5]. The method uses a multi-head attention model to increase the aggregation ability of CSI non-adjacent features and combines the extracted features with effective CSI. Effective CSI can add distance constraint information to the CSI signal representation and increase the stability of the positioning algorithm. This paper mainly focuses on the fusion localization of features, so the output of this method is directly applied to the feature extraction of CSI.

Indoor images contain rich location information, so the representation methods of images have a significant impact on the fusion positioning accuracy. In terms of image feature representation, Xiao A et al. [6] used the Faster R-CNN algorithm to identify the static target uploaded by the mobile phone and established a collinear equation according to the control point on the static target and the geographical position of the observer to solve the final position of the photographer. [7] proposed to apply transfer learning to image indoor localization to generate robust image features. The algorithm uses the image data collected by the robot as the training set to train the VGG-16 network parameters, and after the model training is completed, the penultimate layer features of the network are used for matching, and good results have been achieved in the positioning task. At present, image representation methods based on deep learning have been widely used in image localization tasks. However, indoor localization based on multi-directional images still has the problem that high feature dimensions and low-quality images will introduce noise. There is a lack of an efficient unified vector representation for representing multi-directional images. In this paper, TDE-IFF algorithm is proposed, which effectively reduces the interference of low-quality images on fusion features and facilitates subsequent multi-sensor fusion.

Compared with the decision-level fusion method, the feature-level fusion method has more in-depth interaction between features, which is beneficial to the final model decision. This method is widely used in the field of multi-source fusion localization. Localization methods based on feature-level fusion often map the interrelated features of different signal sources into space to form a richer feature representation for the final location decision. Chiou et al. [8] proposed a method of building a graph network to aggregate multi-directional image features to achieve mutual fusion of image features, and verified the improvement of localization performance by this method through experiments; Jiao et al. [9] talked about HOG features and PAC features in images. Perform fusion and verify that the fusion features help improve the stability of the positioning system. The above methods are only for image feature information, and the lack of other information to introduce leads to limited positioning accuracy. Huang et al. [10] proposed a new fingerprint database construction method, which fused geomagnetic signals and CSI features at the feature level to generate a new fingerprint database, and based on this, proposed a multi-scale KNN matching algorithm for matching and localization. (MDS-KNN), achieving a localization accuracy of 1.7m in the test. Although this method introduces multi-source signals, the fusion method is relatively simple, lacks further analysis of different signals, and the fusion features do not make tradeoffs and judgments for each signal, which makes it difficult to improve the discrimination of fusion features.

In addition, some scholars have restricted the fusion features by introducing gated signals in

the fields of time series signal processing, aiming to further improve the effectiveness of the fusion features. The Gated Recurrent Unit (GRU) network [11] performs a weighted fusion of long and short-term information by introducing a gating signal, thereby enhancing the network's feature extraction capability for time series. The MMOE algorithm [12] constrains multiple expert networks through a gating mechanism, and the output features of each expert network are fused according to weights to enhance the performance of the recommendation algorithm. Click-through rate prediction selection enhanced deep network (Gate-Net) [13] by introducing a gating mechanism to the weighted fusion of different feature embedding representations, so as to better improve click-through rate prediction performance. Therefore, the gating mechanism helps to improve the discrimination of fused features and enhance their effectiveness of fused features. Based on the inspiration of the above research, this paper applies the gating mechanism to the multi-source signal localization task and proposes a fusion framework with adaptive weights. By evaluating the contribution of different signals, the framework enhances effective signals and reduces the influence of invalid signals, thereby increasing the positioning accuracy.

3. Fusion and localization based on AW-FLoc

In this paper, a multi-source signal fusion localization framework (AW-FLoc) is proposed to achieve high-precision indoor localization. To ensure the fusion effect, the algorithm first builds a separate feature representation method for the multi-directional image signal and the CSI signal. The TDE-IFF algorithm is proposed to fuse multi-directional image features. The AW-FLoc algorithm fuses and fine-tunes CSI features and image features. In addition, the MHSA-EC network is used for feature extraction for the CSI signal framework. The frame is shown in figure 1.

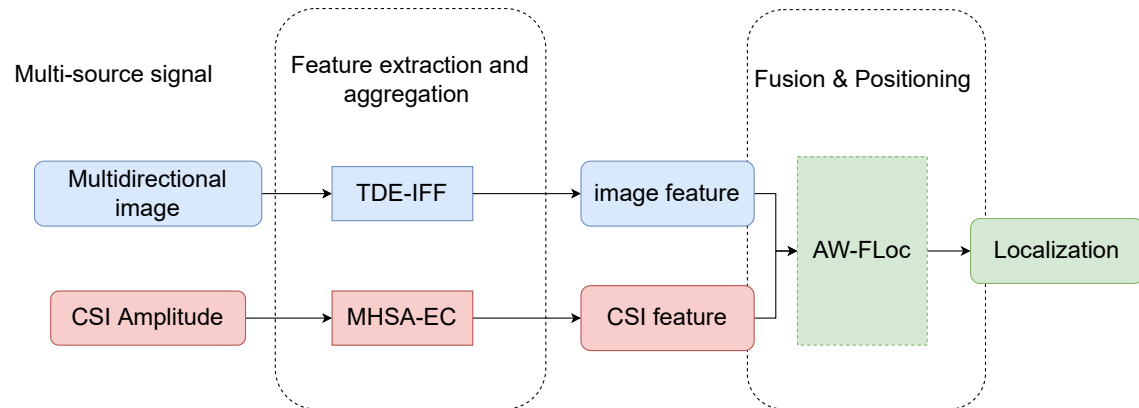


Figure 1: Framework of fusion indoor localization based on AW-FLoc.

3.1. TDE-IFF algorithm

The TDE-IFF algorithm is a fusion representation algorithm for multi-directional image features. In this paper, images from four directions of east, west, north and south are selected as input, and the fused image features are output by TDE-IFF algorithm.

3.1.1. Problem analysis

In order to illustrate the advantages of the TDE-IFF algorithm, we first analyze the problem of multi-directional image feature fusion with two subordinate points.

(1) Low-texture images introduce noise

In image feature representation, low-texture images can introduce noise interference. As shown in figure 2, in an indoor scene, two images in opposite directions are collected from a reference point close to the wall. It can be seen that, due to the proximity to the wall, the image collected in the southern position is only a picture of the wall, lacking texture and position information. The northern position image has a larger viewing angle and is rich enough to capture larger indoor scene texture and location information. If these two image features are directly fused, it will not only be difficult to increase the information but will lead to the introduction of noise into the image features with more texture information. Therefore, an algorithm is needed to score the texture degree of the image, to reduce the interference of low texture signals to the fusion features.

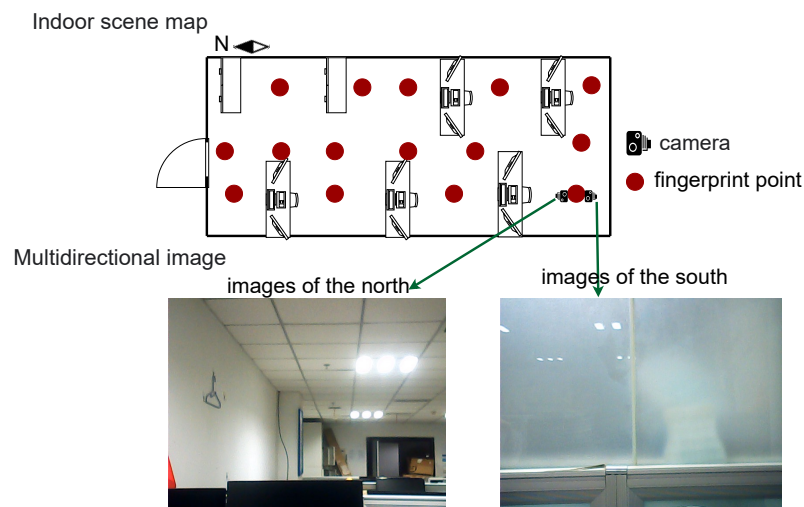


Figure 2: Differences in texture between different images

(2) The position on the coordinate axis is strongly correlated with the image features along that orientation

In indoor scenes, image features oriented along a certain axis (x or y axis) tend to exhibit strong correlations with positions in that axis (x or y axis). The image changes of different fingerprint points in the same direction are shown in figure 3:

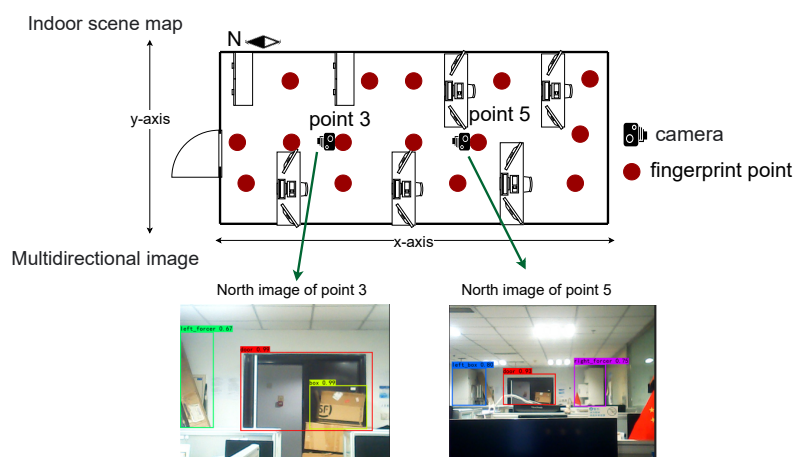


Figure 3: Image variation of different fingerprint points along the x-axis

It can be seen from figure 3 that when the camera is far away from the landmark, the surface objects such as doors and cabinets become significantly smaller in the image. To further describe the relationship between this image feature and physical location, this paper model the problem in figure 4.

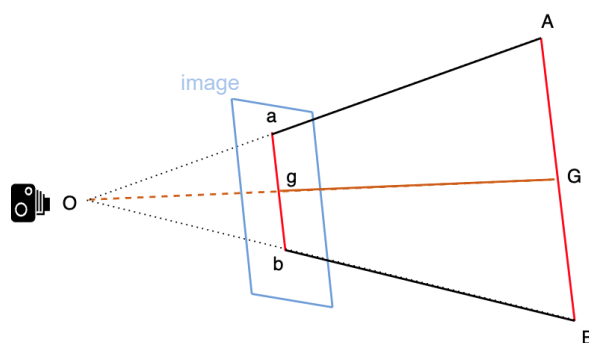


Figure 4: The relationship between the distance of the object in the image and the actual position

In figure 4, O is the focus of the camera; AB is the two endpoints of the actual object, and ab is the position of the two endpoints in the image; G is the midpoint of AB, and g is the midpoint of ab; OG is the distance between the camera and the midpoint. distance. So the following equations can be constructed:

$$OG = \frac{og}{ab} * AB$$

For a given object AB, the distance OG between the camera and the object and the height ab of the object projected in the image satisfy the above formula. For images taken along the x-axis, when the object AB is at the center of the image, OG can represent the position of the camera on the x-axis. Furthermore, it can be confirmed that the image features along the

x-axis are strongly correlated with the distance along the x-axis, and the same is true for the south-oriented images along the x-axis. Since the image features in two opposite directions along the x-axis show a strong correlation with the x-axis. Therefore, fusing the image features of two opposite orientations (north-south orientation) on the x-axis can theoretically further increase the positioning capability of the algorithm on the x-axis. The same is true for the y-axis.

3.1.2. TDE-IFF algorithm

Based on the above analysis, an algorithm is needed that can aggregate image features of opposite orientations on the same coordinate axis (x-axis or y-axis), and at the same time can reduce the interference of low-texture image features on fusion features. For the multi-directional image feature aggregation problem, the block diagram of the model built in this project is shown in figure 5:

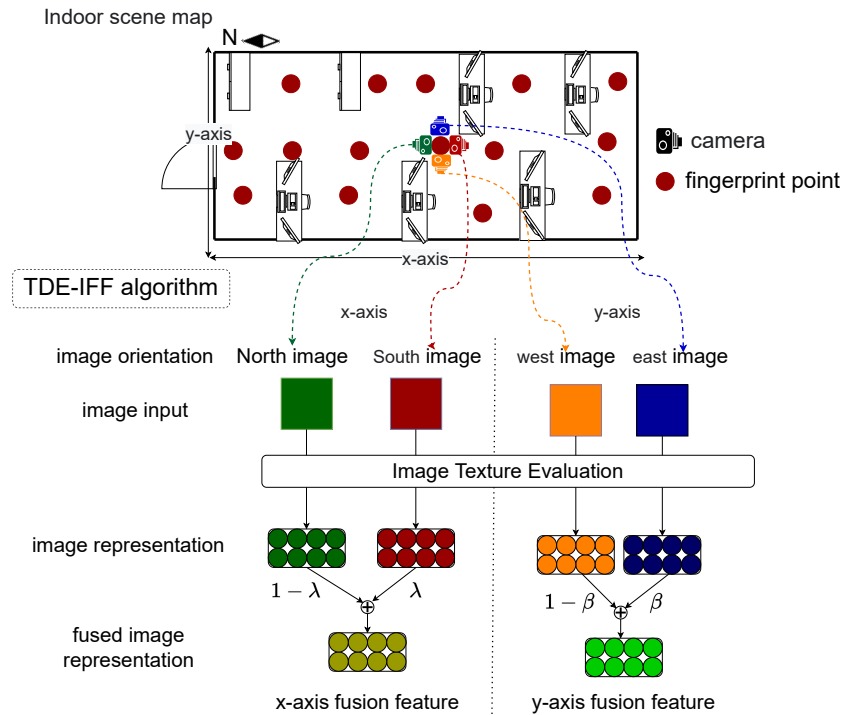


Figure 5: TDE-IFF algorithm

First, in the physical space, the x-axis and the y-axis are grouped image $[image_x, image_x][image_y, image_y]$, x and x' are two opposite directions on the x-coordinate axis, y and y' are the same. The images of opposite orientations are put into a group, and then the image features $[F_x, F_{x'}][F_y, F_{y'}]$ are obtained through the image representation algorithm. Image features in opposite orientations

are progressively aggregated through TDE-IFF:

$$FusionFeat_x = \lambda F_x + (1 - \lambda) F_x,$$

$$FusionFeat_y = \beta F_y + (1 - \beta) F_y.$$

$FusionFeat_x$ is the fused image feature in two directions on the x-axis, and $FusionFeat_y$ is the fused image feature in the two directions on the y-axis. The parameters λ and β are the scores for the image texture features, which are used to evaluate the texture features of images in different orientations, and aim to reduce the feature weight of low-texture images. The evaluation of image texture features is obtained by calculating the variance of the image gray-scale histogram, taking $image_x$ as an example, First, the gray-scale histogram statistics are performed on the input image:

$$p(z_i) = \frac{n_i}{n}$$

where n is the total number of pixels, n_i is the number of pixels of the gray level z_i , and $p(z_i)$ is the frequency corresponding to the gray level z_i . Then calculate the histogram variance:

$$u(image) = \sum_0^{L-1} (z_i - m)^2 p(z_i)$$

$$m = \sum_0^{L-1} p(z_i) z_i$$

where m is the mean frequency of the histogram, L is the total number of gray levels, and $u(image)$ is the histogram variance of the $image$. In this paper, $u(image)$ is used as the evaluation of the image texture, and the parameters λ and β :

$$\lambda = \frac{u(image_x)}{u(image_x) + u(image_x)}$$

$$\beta = \frac{u(image_y)}{u(image_y) + u(image_y)}$$

According to the parameters calculated by this method, the interference of low-texture picture features can be eliminated to a certain extent. The TDE-IFF algorithm can effectively fuse multi-directional map line features and enhance the localization performance of the fusion algorithm.

3.2. AW-FLoc algorithm

In the field of multi-source fusion, previous algorithms often ignore that the contribution of different signals to the location dimension is different. The significance of assigning weights according to different coordinate axis is that the signals strongly correlated with the position of the coordinate axis can be enhanced, and other irrelevant or weakly correlated signals can be reduced at the same time.

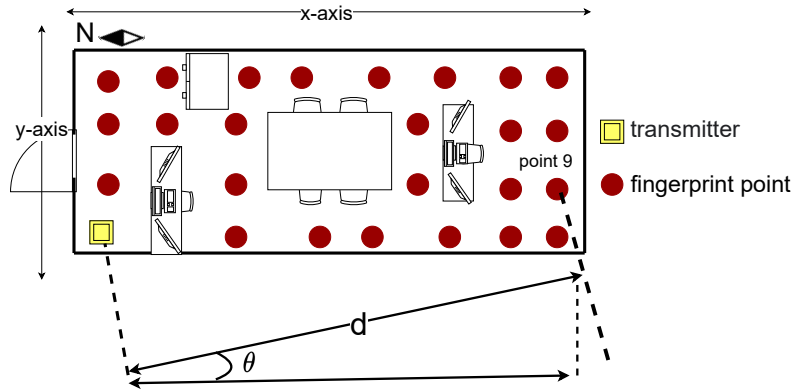


Figure 6: The relationship between distance and coordinate axis in narrow and long environment

In the CSI representation, this paper introduces effective CSI as distance constraint information, and this distance information contributes differently to different coordinate axis in a narrow and long space. Figure 6 is an example:

As shown in Figure 6, the fingerprint point 9 is a point far away from the transmitting end, and the corresponding relationship between this point and the transmitting end is represented by a d vector, where the angle between d and the x-axis is θ . Since the environment is long and narrow, θ is a small angle, assuming that the cosine distance of the included angle is used as the method of correlation evaluation, $\cos\theta$ is larger. In other words, the distance d has a strong correlation with the x-axis distance. And this distance information exists in the fusion representation of CSI, so in this case, the representation has different contributions to different coordinate axis, which also needs to be further measured. For image features of different orientations, the image features aggregated on the x-axis are more closely related to the position of the x-axis because the feature changes of landmarks in this dimension can be observed, and the same is true on the y-axis.

The above example illustrates that the signal has different contribution degrees to the positioning of different coordinate axes, which is not only for the CSI signal but also not only for the distance information. For the fusion positioning framework, what is needed is a method that can adaptively evaluate the contribution of different signals to the positioning and the adaptive evaluation of the contribution of the signals to the positioning of different coordinate axes. This can greatly improve the positioning accuracy and universality of the algorithm.

Based on the above analysis, this paper designs the following multi-source fusion framework network and proposes a block diagram as shown in figure 7.

Taking the fusion representation and CSI features of multi-directional images as an example, the algorithm firstly takes the well-trained features as input and performs further features, dimensionality reduction, and extraction for the input. At the same time, considering that the algorithm locates the positions of the x and y axis respectively, separate gating networks and position determinators are designed for the two coordinate axis. The gating network uses the original feature as the original feature, evaluates the contribution of each feature to the dimension, and outputs the corresponding weight. The feature will be weighted and fused based

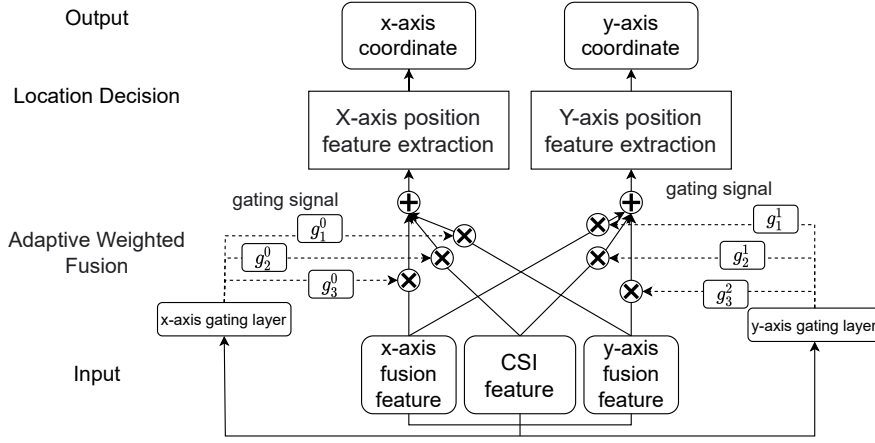


Figure 7: The AW-FLoc algorithm

on the weight, and input to the position determiner corresponding to the given dimension. The position decider is composed of multiple linear layers and activation functions and outputs the position of the corresponding coordinate axis according to the fusion feature.

(1) Input feature processing

The features of the network include x-axis fused image features, CSI features based on MHSA-EC, and y-axis fused image features, denoted as $[FusionFeat_x, Feat_{csi}, FusionFeat_y]$. To ensure the consistency of feature dimensions, further feature extraction and compression are performed on the aggregated image features. The CSI fusion feature considers the uniformity of the scale, and the layer normalization process is performed here, and the processed features are $[spaceFeat_x, \widehat{Feat}_{csi}, spaceFeat_y]$.

(2) Gating network and feature fusion

Considering that each feature has different contributions to different spatial coordinate axis, two gating layers are introduced to evaluate the contribution adaptively. The gating network takes the original features as input, and performs adaptive weight scoring according to the input features:

$$g_k = softmax(W_{gk}[FusionFeat_x, Feat_{csi}, FusionFeat_y])$$

$$softmax(x_i) = \frac{exp(x_i)}{\sum_j exp(x_j)}$$

Where W_{gk} is the linear layer weight of the gated network, k corresponds to the two tasks of x-axis positioning and y-axis positioning (ie, $k = 0$ corresponds to the x-axis positioning task, $k = 1$ corresponds to the y-axis positioning task), g^k will return a three-dimensional vector $[g_1, g_2, g_3]$, the value of each dimension is in the (0,1) interval, corresponding to the contribution value of each feature. The adaptive weighted fusion algorithm fuses the contribution value with the feature:

$$locfeat_k = g_1 * spaceFeat_x + g_2 * \widehat{Feat}_{csi} + g_3 * spaceFeat_y$$

The input of the decider of each axis is $locfeat_k \in [locfeat_0, locfeat_1]$, corresponding to the fused features, and then through multiple linear layers and activation functions to complete the fitting of each position axis. The gradient descent algorithm of the model uses Adam, and the loss function is set to cross-entropy.

4. Experiments

4.1. Experiment environments

This part mainly verifies the fusion localization performance based on AW-FLoc. The experiment includes CSI and image acquisition equipment, and the two types of equipment correspond in figure 8 and figure 9, respectively. The CSI acquisition equipment includes a transmitter and a receiver, both of which are equipped with Intel 5300 network cards, and parse the CSI data by changing the underlying driver. The transmitter uses one antenna, and the receiver is equipped with three antennas. Ubuntu 16.04 system is installed on each device. In the received data packet, signals such as the number of antennas, noise, CSI, and RSSI are included. The image acquisition equipment used four low-cost cameras. Four cameras are fixed in the four directions of the box and connected with the same laptop computer through the data line respectively. Therefore, the multi-directional image data can be collected by a single computer operation.

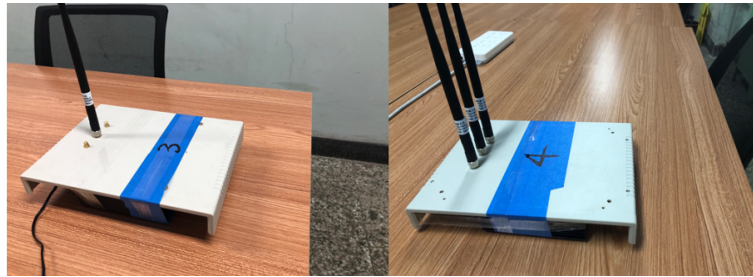


Figure 8: Transmitter (left) and receiver (right)

The experimental scene is a comprehensive office scene, as shown in figure 10. The complex indoor environment and rich multipath effects can be used as a typical comprehensive indoor environment. The scene is composed of an office area, a meeting room, and an external corridor. The three environments are isolated from each other. The office area and the conference room are isolated by the glass, and the corridor is isolated from the other two rooms by walls. There are many desks, chairs and computers in the office area and conference room. The total area of the scene is $152.9m^2$, of which the meeting room area is $8.4*1.8m^2$, and the laboratory area is $16.4*4.4m^2$. 59 reference points and 59 test points were established in the scene, and the distance between the reference points was 1.2m. The data collection included 59 reference position points and 59 test points, 2000 CSI packets and 40 multi-directional images collected at each position.



Figure 9: Four-directional camera: top view (left) side view (right)

4.2. The evaluation of TDE-IFF

To verify the positioning performance of TDE-IFF, this part mainly compares TDE-IFF with the positioning method based on VGG-16 proposed by Wozniak P et al in 2018 from positioning accuracy and stability [14].

It can be seen from figure 11 that in terms of overall positioning accuracy, the algorithm performance of TDE-IFF is better. In the low positioning error interval of 0-4.1m, there is a higher degree of confidence. The maximum error is 8.5m for the TDE-IFF method and 10.6m for the VGG-based positioning method. The samples are in the corners of the scene, which makes them prone to mismatching with each other, which in turn increases the positioning error of long-distance points. The TDE-IFF algorithm will have better control in this regard.

Table 1 compares the average positioning error and standard deviation of the two methods. The average error of TDE-IFF is 0.94m, which is lower than the 1.09m of the VGG method, and the positioning accuracy is improved by 16%. In terms of stability, TDE-IFF is 1.02m, which is 9% less than the 1.11m of the VGG method. Combined with the CDF map, the TDE-IFF algorithm can improve the accuracy and stability of the positioning.

Table 1

Image Localization Errors of Different Methods

Method	TDE-IFF	VGG
mean/m	0.94	1.09
std/m	1.02	1.11

4.3. The evaluation of AW-FLoc

To verify the performance of the AW-FLoc algorithm, this section conducts ablation experiments with or without the gating layer and compares it with the MDS-KNN algorithm [10].

- (1) Adaptive weight comparison experiment

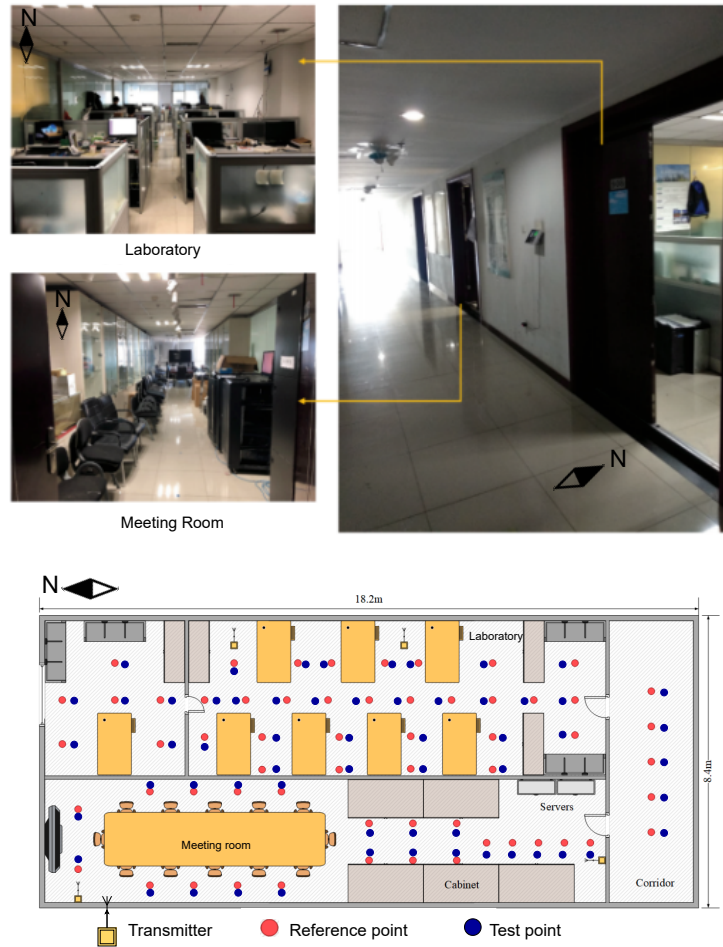


Figure 10: Comprehensive office scene

In order to verify the adaptive weights, this paper compares the AW-FLoc method with the fusion localization algorithm without adaptive weights (w.o.gate AW-FLoc). The weight of the signal is set to 1, and the experimental results are shown in figure 12(a). In the interval of 0-5.3m, the fusion positioning algorithm based on adaptive weight has higher confidence. Although in the interval of 5.3m-6m, the confidence of the network without adaptive weight is slightly ahead, overall, The performance of AW-FLoc is better, and the positioning error is also smaller. This can prove the effectiveness of the adaptive weights.

This section selects the algorithm based on MDS-KNN proposed by Huang et al [10]. as a comparison item to verify the localization ability of the network. The positioning results of the experiment are shown in the figure 12(b). In the low error interval of 0-2.2m, the AW-FLoc algorithm slightly leads the MDS-KNN algorithm. In the range of 2.2m-8.5m, the AW-FLoc algorithm has higher confidence and better positioning performance. When the confidence level reaches 1, the error of the AW-FLoc algorithm is 8.5m, while the error of the MDS-KNN

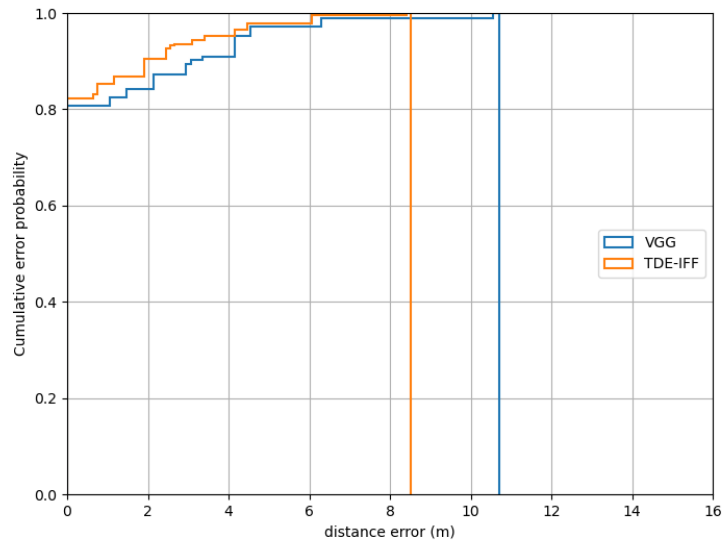


Figure 11: The CDF of TDE-IFF

algorithm is 10.2m. In terms of positioning accuracy, the experiment shows that compared with the simple feature fusion in MDS-KNN, the features of the AW-FLoc algorithm are more effective, and due to the introduction of the adaptive algorithm, the fused features are more prominent and effective information, making different fingerprints point The feature discrimination is greater and the localization performance is better.

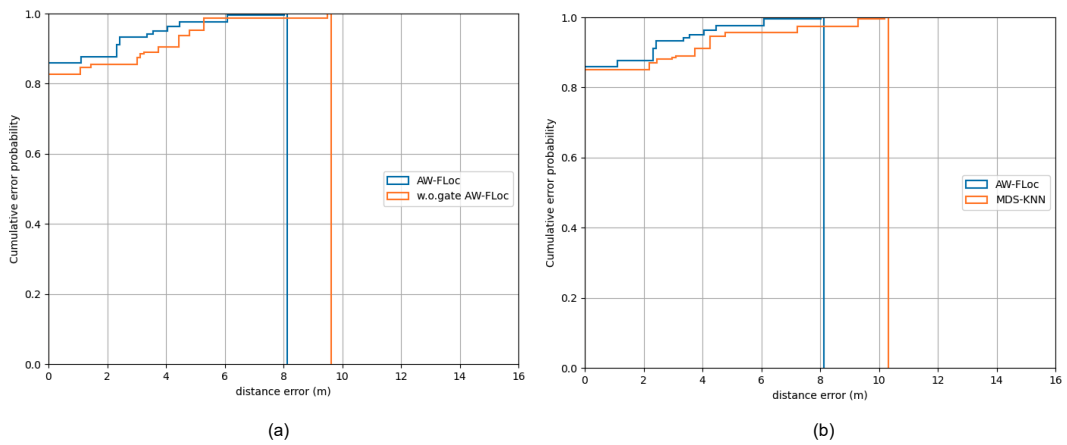


Figure 12: The CDFs of different methods.

(2) Compare with other methods

To further verify the stability of the AW-FLoc algorithm, the average positioning error

and standard deviation indicators are further compared here, as shown in Table 2. AW-FLoc has the best accuracy and positioning stability. Compared with the MDS-KNN system, in a comprehensive office scenario, the accuracy is improved by 20%, and the stability is improved by 17%. Experiments show that the AW-FLoc method outperforms other existing methods in localization accuracy. In addition, the adaptive weight and TDE-IFF algorithm have great gains in the improvement of the overall positioning performance, which verifies the necessity of the two methods for the multi-source heterogeneous signal fusion positioning algorithm.

Table 2
Fusion Localization Errors of Different Methods

Method	AW-FLoc	MDS-KNN
mean/m	0.61	0.73
std/m	0.73	0.85

5. Conclusion

In this paper, we propose a fusion localization framework based on AW-FLoc to improve localization accuracy and robustness. This paper first proposes the TDE-IFF algorithm, which constructs the fusion representation of multi-directional images, reduces the interference of low texture on the fusion representation through the texture degree evaluation algorithm, and provides high-quality image features for subsequent fusion positioning methods. On the basis of TDE-IFF, this paper proposes a fusion localization framework of AW-FLoc. The algorithm adaptively weighs CSI and image features for different fingerprint points, enhances the proportion of effective signals in the fusion representation, improves the signal-to-noise ratio of the fusion representation, and further optimizes the localization capability of the algorithm. Experiments show that in a comprehensive office scenario, TDE-IFF effectively improves the positioning accuracy, with an average error of 0.94m. AW-FLoc is further improved on this basis, with an average error of 0.61m. Compared with the MDS-KNN method, the accuracy is improved by 20%, and the stability is improved by 17%.

6. Acknowledgments

This work was financially supported by the National Natural Science Foundation of China under Grant No.61871054.

References

- [1] H. Huang, G. Georg, Current trends and challenges in location-based services, *International Journal of Geo-Information* 7 (2018) 199.
- [2] H. T. Hoi, V. T. Le, Location-based services, *ABC Journal of Advanced Research* 8 (2019) 89–94.

- [3] D. Lymberopoulos, R. R. Choudhury, J. Liu, S. Sen, X. Yang, V. Handziski, Microsoft indoor localization competition: Experiences and lessons learned, *Getmobile Mobile Computing Communications* 18 (2015) 24–31.
- [4] Z. Deng, Csi amplitude fingerprinting for indoor localization with dictionary learning, *Entropy* 23 (2021).
- [5] W. Liu, M. Jia, Z. Deng, C. Qin, Mhsa-ec: An indoor localization algorithm fusing the multi-head self-attention mechanism and effective csi, *Entropy* 24 (2022). URL: <https://www.mdpi.com/1099-4300/24/5/599>. doi:10.3390/e24050599.
- [6] A. Xiao, R. Chen, D. Li, Y. Chen, D. Wu, An indoor positioning system based on static objects in large indoor scenes by using smartphone cameras, *Sensors* 18 (2018) 2229.
- [7] S. Xu, W. Chou, H. Dong, A robust indoor localization system integrating visual localization aided by cnn-based image retrieval with monte carlo localization, *Sensors* 19 (2019) 249.
- [8] M.-J. Chiou, Z. Liu, Y. Yin, A.-A. Liu, R. Zimmermann, Zero-shot multi-view indoor localization via graph location networks, in: *Proceedings of the 28th ACM International Conference on Multimedia, 2020*, pp. 3431–3440.
- [9] J. Jiao, X. Wang, Z. Deng, Build a robust learning feature descriptor by using a new image visualization method for indoor scenario recognition, *Sensors* 17 (2017) 1569.
- [10] X. Huang, S. Guo, Y. Wu, Y. Yang, A fine-grained indoor fingerprinting localization based on magnetic field strength and channel state information, *Pervasive and mobile computing* 41 (2017) 150–165.
- [11] R. Dey, F. M. Salem, Gate-variants of gated recurrent unit (gru) neural networks, in: *2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS), IEEE, 2017*, pp. 1597–1600.
- [12] J. Ma, Z. Zhao, X. Yi, J. Chen, L. Hong, E. H. Chi, Modeling task relationships in multi-task learning with multi-gate mixture-of-experts, in: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018*, pp. 1930–1939.
- [13] T. Huang, Q. She, Z. Wang, J. Zhang, Gatenet: gating-enhanced deep network for click-through rate prediction, *arXiv preprint arXiv:2007.03519* (2020).
- [14] P. Wozniak, H. Afrisal, R. G. Esparza, B. Kwolek, Scene recognition for indoor localization of mobile robots using deep cnn, in: *International Conference on Computer Vision and Graphics, Springer, 2018*, pp. 137–147.