WiFi-Visual Data Fusion For Indoor Robot Localization

Yuehua Ding, Jean-françois Dollinger, Vincent Vauchey, Mourad Zghal

WiFi-Visual Data Fusion For Indoor Robot Localization

Yuehua Ding, Jean-François Dollinger, Vincent Vauchey, Mourad Zghal

***Abstract*— In this paper, we propose a WiFi-Visual robot localization method for limiting the unbounded error of image-only localization due to visual environment similarity. The localization problem is modeled as a classification problem based on the WiFi-Visual data collected at labelled positions. The heterogeneous WiFi-Visual data is harmonized by representing the WiFi features in image form to adapt to the strong image processing capacity of the neural network. The WiFi features in image form are fused with the visual features provided by the robot camera. The fused WiFi-Visual features are jointly exploited by a neural network to classify WiFi-Visual features of an unknown position to the most likely class. The labelled position corresponding to the most likely class is taken as the estimated position of the robot. Experiments are carried out on the physical robot platform TIAGO++, which can provide the real-time ground truth reference position. Experiment results show that the proposed WiFi-Visual data fusion method can effectively limit the exceptional unbounded localization errors of image-only localization. The RMSE of the proposed method is less than 2 meters. This value is smaller than that of WiFi localization. The proposed method has more stable performance than WiFi-only localization and image-only localization. Its performance can be further improved by Kalman filtering. During the experiment, a demo video was recorded and it is provided along with this paper [1].**

I. INTRODUCTION

Rapid developments of information technologies are spurring the rise of robot applications in healthcare, industry [1] and so on. To this end, robot localization plays a crucial role for robot's autonomous movements [2]. Various technologies are emerging for robot localization.

Light detection and ranging (LiDAR) [3] can offer accurate mapping and localization, but LiDAR is expensive. In addition, the performance of LiDAR degrades in geometrically ambiguous environments and robot kidnapping situations.

To cope with the limitations of LiDAR, image localization and WiFi localization are two promising ways. Visual methods are based on computer vision, image processing and artificial intelligence (AI) [4]–[7]. Radio methods are based on radio signal processing and the geometric principle or adaptation of radio data. The radio methods are quite rich, plenty of technologies can be used, such as ultra-wide band (UWB) [8], frequency modulated continuous wave (FMCW) [9], WiFi [10]–[14] etc.

Nowadays, the friendly prices of camera and WiFi make the visual method and the radio method using WiFi attractive. Visual localization is usually accurate, however, it can have unbounded localization errors between visually similar points. WiFi localization is less accurate than visual localization because of considerable variations of WiFi signal strength. Fortunately, WiFi signal variation is mainly limited in its coverage range, which can be exploited to avoid catastrophic positioning errors.

Motivated by the complementary natures of WiFi localization and visual localization, the contribution of this paper consists in the proposition of a WiFi-Visual data fusion method, which can limit the unbounded localization error of image-only localization by combining the generally excellent localization accuracy of visual positioning and the bounded localization errors of WiFi localization.

The remaining parts of this paper are organized as follows: part II presents the localization problem, the proposed method is presented in part III, the experiment results are analyzed in part IV, part V concludes this paper.

II. PROBLEM STATEMENT

WiFi-Visual robot localization is based on the information collected by robot WiFi antenna and camera. The positioning process can be modeled as follows:

$$(\hat{x}_n, \hat{y}_n) = g(\mathbf{W}_n, \mathbf{V}_n) \tag{1}$$

where $g(\cdot)$ is a positioning algorithm. $(\hat{x}_n, \hat{y}_n)$ represents the estimate of the n^{th} position (x_n, y_n). $\mathbf{W}_n$ and $\mathbf{V}_n$ represent the WiFi data and image data collected at this position, respectively. Usually, $\mathbf{W}_n$ can be the received signal strength indicator (RSSI). Without loss of generality, $\mathbf{W}_n$ is supposed to have M WiFi samples from K access points, $\mathbf{W}_n$ is written as follows:

$$\mathbf{W}_n = \begin{bmatrix} \mathbf{w}_{n,0} & \mathbf{w}_{n,1} & \cdots & \mathbf{w}_{n,K-1} \end{bmatrix} \tag{2}$$

with

$$\mathbf{w}_{n,k} = \begin{bmatrix} w_{n,0,k} & w_{n,1,k} & \cdots & w_{n,M-1,k} \end{bmatrix}^T \tag{3}$$

where $(\cdot)^T$ represents matrix transpose. $\mathbf{w}_{n,k}$ represents M WiFi samples collected at position (x_n, y_n) from access point k. $w_{n,m,k}$ represents the m^{th} sample received at the n^{th} position from access point k. The image data $\mathbf{V}_n$ includes a sequence of S photos, which are sampled by the robot camera at (x_n, y_n). $\mathbf{V}_n$ can be represented as:

$$\mathbf{V}_n = \begin{bmatrix} \mathbf{V}_{n,0} & \mathbf{V}_{n,1} & \cdots & \mathbf{V}_{n,S-1} \end{bmatrix} \tag{4}$$

where $\mathbf{V}_{n,s}$ represents an image. The localization in (1) is to minimize the error between $(\hat{x}_n, \hat{y}_n)$ and (x_n, y_n).

The problem is usually transformed into a classification problem by uniformly dividing the whole localization surface

Yuehua Ding, Jean-François Dollinger, Vincent Vauchey and Mourad Zghal are with the laboratory CESI LINEACT UR7527, France.

Correspondence: Yuehua Ding, yding@cesi.fr

[1]The demo video available at: https://youtu.be/elVFWTtopoI

into N pieces of small square areas, which are corresponding to classes $L_0, L_1, \cdots, L_{N-1}$. Without loss of generality, the n^{th} position is supposed to be located at square n belonging to class L_n. The problem in (1) is rewritten as:

$$(\hat{x}_n, \hat{y}_n) \leftarrow \hat{L}_n = g(\mathbf{W}_n, \mathbf{V}_n) \tag{5}$$

where $(\hat{x}_n, \hat{y}_n)$ is obtained by taking the center position of the estimated class $\hat{L}_n$.

III. Proposed algorithm

In this section, a novel WiFi-Visual data fusion method is proposed for robot localization. The WiFi features are represented in terms of spectrum matrix and correlation matrix, both of them are in image forms. Before the feature extraction, data preprocessing is performed as follows:

$$\mathbf{w}_{n,k} \leftarrow \frac{\mathbf{w}_{n,k} - \mu_k}{\sigma_k} \tag{6}$$

$$\mathbf{V}_n \leftarrow \frac{\mathbf{V}_n}{255} \tag{7}$$

where μ_k and σ_k represent the mean value and standard deviation of the RSSI of access point k. WiFi localization is heavily influenced by RSSI variations from time to time. To alleviate this effect, the relatively stable WiFi features, such as spectrum and correlation matrix are considered.

Remark: (6) and (7) are normalization processes. To facilitate the representation without influencing the reading, we always use $\mathbf{w}_{n,k}$ and $\mathbf{V}_n$ before and after the normalization processes. In the following steps of the proposed algorithm, $\mathbf{W}_n$ and $\mathbf{V}_n$ refer to the elements normalized by (6) and (7) by default.

A. Spectrum

To obtain the spectrum features of WiFi data, the spectrum of the WiFi data collected at each position can be analyzed. Two-dimensional Fast Fourier Transform (FFT) can be applied on $\mathbf{W}_n$.

$$\tilde{\mathbf{W}}_n = \mathbf{F}_M^T \mathbf{W}_n \mathbf{F}_K \tag{8}$$

where $\mathbf{F}_K$ is the $K \times K$ discrete Fourier transform matrix. $\tilde{\mathbf{W}}_n$ can be considered as a two dimensional image, it is a spectrum matrix of WiFi data received at (x_n, y_n). Each row in $\tilde{\mathbf{W}}_n$ can be considered as the spatial spectrum at the corresponding instant of sampling. Each column stands for the temporal spectrum of the corresponding access point.

B. Correlation matrix

WiFi signals arrive at different positions by experiencing different propagation paths, which can influence the correlation matrix across different access points. At position (x_n, y_n), its access points correlation matrix is given by:

$$\mathbf{R}_n = \frac{\mathbf{W}_n^T \mathbf{W}_n}{M} \tag{9}$$

where $\mathbf{R}_n$ is a matrix of $K \times K$ dimensions. $\mathbf{R}_n$ is symmetric and it can be visualized as an image.

C. WiFi-Visual Feature Fusion

The visual features are the photo sequence $\mathbf{V}_{n,0}, \mathbf{V}_{n,1}, \cdots \mathbf{V}_{n,S-1}$ taken at (x_n, y_n). One notes that the spectrum $\tilde{\mathbf{W}}_n$, the correlation matrix $\mathbf{R}_n$ and the visual features $\mathbf{V}_n$ are in image form, however, $\mathbf{R}_n$, $\mathbf{W}_n$, $\tilde{\mathbf{W}}_n$ and $\mathbf{V}_n$ are of different dimensions. $\mathbf{R}_n$, $\mathbf{W}_n$, $\tilde{\mathbf{W}}_n$ are $K \times K$, $M \times K$, $M \times K$ matrices respectively. $\mathbf{V}_n$ can be considered as a tensor of dimensions $S \times 3 \times height \times width$, it contains S RGB images of dimensions $3 \times height \times width$. They should be further represented in the same dimensions. One notes that $height \times width$ represents the pixels information of R (or G or B) channel for a color image taken by the robot, the values of $height$ and $width$ are at the magnitude level of hundreds or thousands. K and M are much less than these values. For unifying the features, the images are down-sampled as small images of dimensions $K \times K$, which are the dimensions of $\mathbf{R}_n$. If $M < K$, $\mathbf{W}_n$ and $\tilde{\mathbf{W}}_n$ can be filled with zeros to augment their dimensions from $M \times K$ to $K \times K$ (for $M > K$, we have similar processing method). Finally, all these features are unified as $K \times K$, they are input into a neural network $g(\cdot)$ for robot localization:

$$\mathbf{p} = g(\tilde{\mathbf{W}}_n, \mathbf{R}_n, \mathbf{W}_n, \mathbf{V}_n) \tag{10}$$

where $\mathbf{p}$ is a likelihood vector output by $g(\cdot)$. $\mathbf{p}$ can be considered as a probability vector, it is written by

$$\mathbf{p} = [p((x_1, y_1)|\mathbf{W}_n, \mathbf{V}_n), \cdots, p((x_N, y_N)|\mathbf{W}_n, \mathbf{V}_n)]^T \tag{11}$$

Based on Eq. (10), the robot position is estimated as follows:

$$p((\hat{x}_n, \hat{y}_n)|\mathbf{W}_n, \mathbf{V}_n) = max \quad p((x_i, y_i)|\mathbf{W}_n, \mathbf{V}_n) \tag{12}$$

IV. Experiments

A. Experiment platform

In this paper, we use a real physical robot TIAGO++ [15] as the experiment platform. TIAGO++ is a robot platform based on ROS. It has its own WiFi card and camera, which can be used directly to sample the WiFi signal and visual environment. The LiDAR position and mapping system in TIAGO++ makes simultaneous localization and mapping (SLAM) easy, which offers a reference map and real-time positions of high accuracy (centimeter-level). The reference positions given by its LiDAR system are taken as ground-truth values for comparison, the criterion of root mean square error (RMSE) is used. RMSE is calculated as:

$$RMSE = \sqrt{\frac{1}{Q} \sum_{q=0}^{Q-1} [(\hat{x}_q - x_q)^2 + (\hat{y}_q - y_q)^2]} \tag{13}$$

where Q is the number of test points. $(\hat{x}_q, \hat{y}_q)$ is the estimated position of the q^{th} test point. (x_q, y_q) is the reference position given by the robot LIDAR system, which is considered as the ground-truth position.

The parameters K, M and S are 58, 10 and 4 respectively. For a localization, 10 samples of RSSI data are taken from 58 logic access points, and 4 images are sampled on the visual environment.

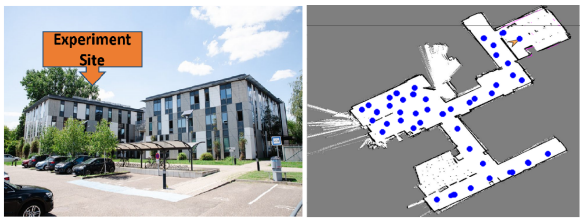

Fig. 1. Mapping in the ground floor of the teaching building.

B. Data collection and training

The experiment environment is shown by Fig.1, the sampled positions are marked in blue color. For each position marked in blue, 1000 samples are taken on WiFi signal by the robot, and then the robot takes 100 images on the visual environment around itself. As shown in Fig.1, there are 43 blue points. Therefore, a database including 43000 WiFi samples and 4300 image samples is constructed. For the selection of neural network, LeNet [16] is chosen as a fundamental structure. The training process is shown by Fig. 2, where the combinations of WiFi features and visual features are input into the network for training.

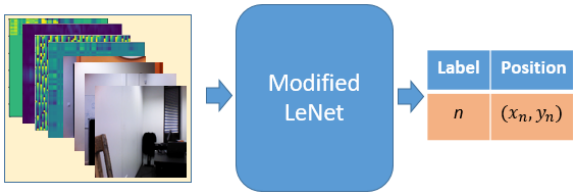

Fig. 2. Training

C. Test results analysis

For test, the robot goes out of the laboratory, and moves in the corridors and the hall. The robot is localized by a PC connecting to WiFi, the PC can send a localization request to the robot, which returns the WiFi-visual data and ground truth position to the PC. The ground truth positions are marked in red color in Fig. 3, which compares the localization results of the proposed method (marked as WiFi-image) with those obtained by separately using WiFi features (marked as WiFi) and photo features (marked as image) in training.

One can note that the image localization trajectory is closest to the true trajectory with an exception of a catastrophic localization error. Fig. 3 visualizes this catastrophic localization error produced by image localization in the environment of corridor similarity. Fig. 4 shows that this error distance is more than 11 meters.

Fig. 4 quantifies the localization errors of these 3 methods. In terms of average precision without counting the catastrophic situation, the image method can reach the precision around 1.1 meters, which is the best performance among the three methods, the WiFi-image method ranks the second with the RMSE precision of 1.59 meters, the WiFi method is the third at the level of 2.56 meters. Unfortunately, the

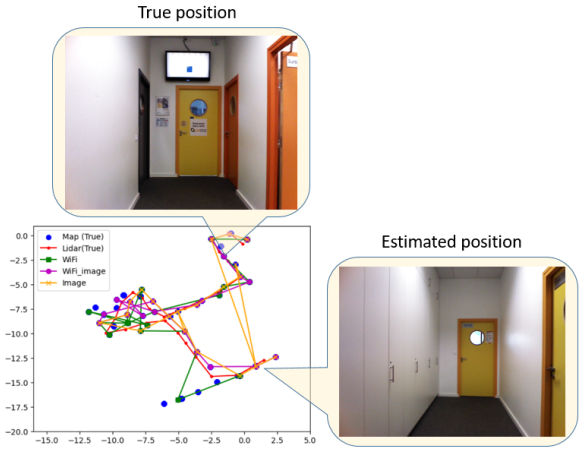

Fig. 3. Trajectory comparison

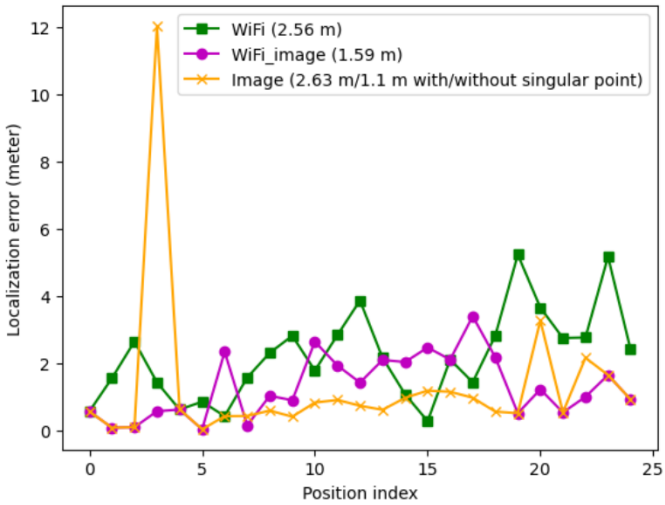

Fig. 4. Localization error comparison

accuracy of the image method can be seriously degraded by homogeneous or unknown environments. In this test, the overall accuracy for the image method is the worst at 2.63 meters.

This catastrophic phenomenon in image localization is illustrated by the histogram in Fig 5, where the statistical data is represented by the bars of different colors. The errors of the image localization are mainly concentrated at 1-meter level with a catastrophic exception at 10-meter level. Fig. 6 illustrates the difference among the performances of the three methods in terms of cumulative density function (CDF), localization using image-only information has no dominant advantage over the other two, it can achieve the best CDF performance for the errors less than 2 meters, however, WiFi-image localization has the best CDF performance in limiting the errors bigger than 2 meters.

Remark: The tests for the trajectory points in Fig. 3 are independent. Changing the order of the test points does not influence the localization performance. Each point can be considered as an initial localization point. Based on the test

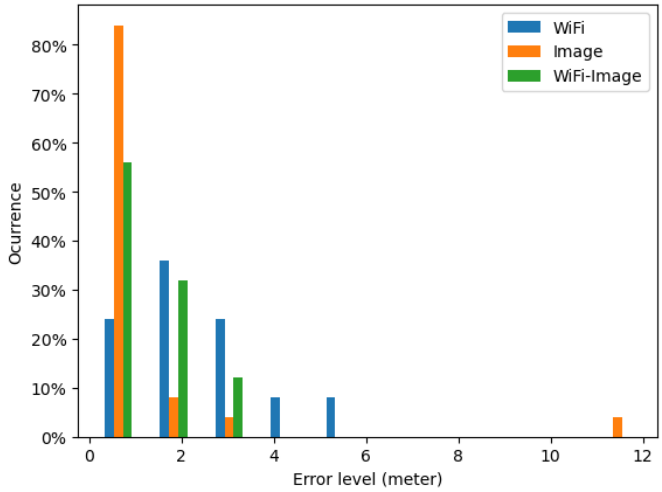

Fig. 5. Histogram of localization errors

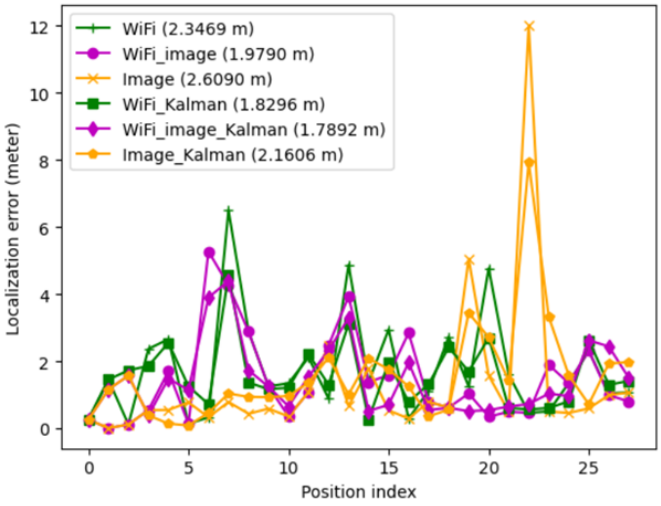

Fig. 7. Localization error comparison (Kalman filtering)

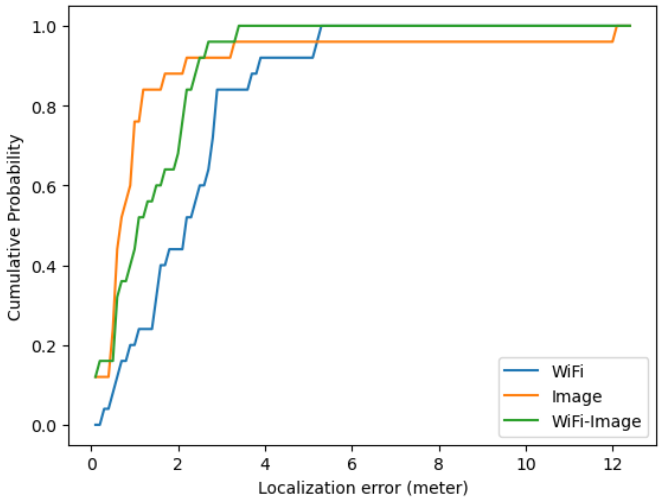

Fig. 6. CDF of localization errors

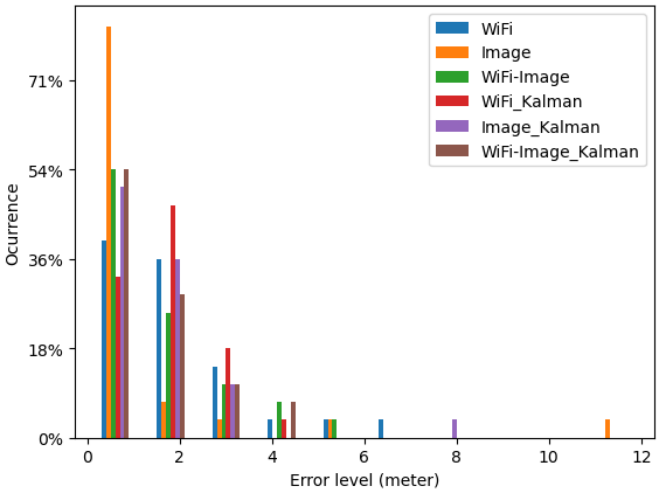

Fig. 8. Histogram comparison (Kalman filtering)

performance and the coverage range of different WiFi access points, the proposed method can effectively resolve the robot kidnapping problem caused by the layout environment change. This is particularly the case when the robot is waken up in a changed environment.

D. Performance discussion with Kalman filtering

To verify the performance of the proposed method under Kalman filtering, another group of test is carried out at the same test site. Fig. 7-9 illustrate the test results. In general, the performances for the three methods without using Kalman filtering are consistent with the results presented above. Fig. 7 shows that the image method usually has the best performance in terms of accuracy, this observation is also illustrated by the statistical tools, such as histogram in Fig. 8 and CDF in Fig. 9. Unfortunately, its average accuracy is not satisfactory due to exceptionally catastrophic case, as shown in Fig. 7, a localization error about 12 meters happens during the image localization.

The improvements brought by Kalman filtering are obvious for WiFi localization and WiFi-image localization. Fig. 7 shows that, in average sense, Kalman filtering can reduce the WiFi localization error by 0.5 meter, and WiFi-image localization error by 0.2 meter, respectively. In the histogram of Fig. 8, the localization errors of WiFi localization and WiFi-image localization are concentrated in an interval from 0 to 3 meters.

One interesting point is that Kalman filtering can not bring improvement to image localization with catastrophic errors. In the CDF of Fig. 9, the image method (without Kalman filtering) is more likely to have localization error less than 1 meter (about 70%), and the probability of its localization error less than 2 meters is about 90 %. However, the image-Kalman method is not better than image localization without Kalman filtering. This is because Kalman filtering can slow down the rapid variation of the estimated robot position by exploiting the historical information. In the same principle, Kalman filtering can also slow down the estimated position

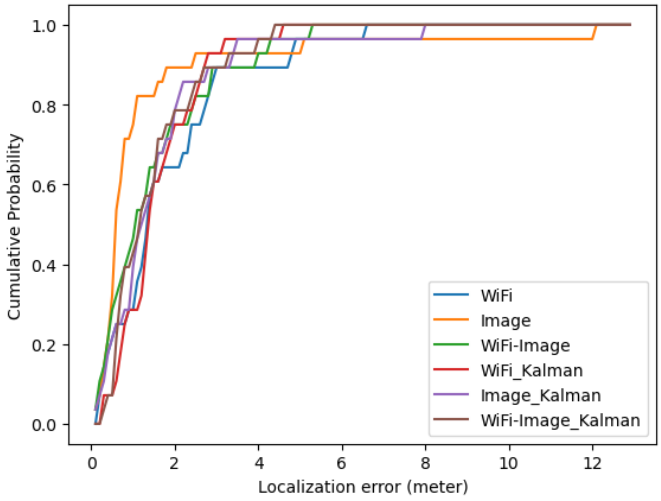

Fig. 9. CDF comparison (Kalman filtering)

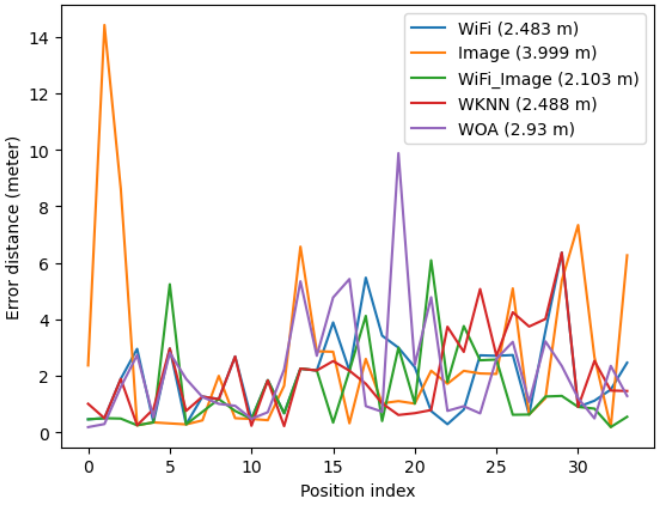

Fig. 10. Localization error comparison with existing methods (In-field test 7 months after the collection of training data)

recovery from the catastrophic state to a normal state, just like it slows down rapid performance degradation.

V. Comparison with other methods in the environment of considerable change

Additional experiment is carried out at the end of September 2024, which is 2 months before this conference and 7 months after the experiments above. During these 7 months, the visual environment is changed in the teaching building. New colorful logos are attached in the walls and doors, some separating boards are removed. The WiFi access points are reorganized. All these changes bring considerable challenges to localization. In this case, the proposed method is also compared with the existing algorithms, such as weighted K-nearest neighbors (WKNN) [17], and whale optimization algorithm (WOA) [18] based on Gaussian process regression (GPR) [19]. Fig. 10 shows the localization performance of different methods in terms of error distance of individual points and RMSE. On can observe that the visual localization degrades significantly due to the visual environment change. The performance of visual localization (RMSE = 3.999 m) is even worse than WiFi localization (RMSE = 2.483 m), which has no significant degradation, thanks to the permanent MAC addresses of the access points despite the reorganization of the WiFi resources. It is also a surprise to find that the proposed WiFi-image localization method can reach the best performance (RMSE = 2.103 m) , which is better than WKNN (RMSE = 2.488 m) and GPR-based-WOA (RMSE = 2.930 m). The CDF performance of WiFi-image localization ranks the first too, according to Fig. 11.

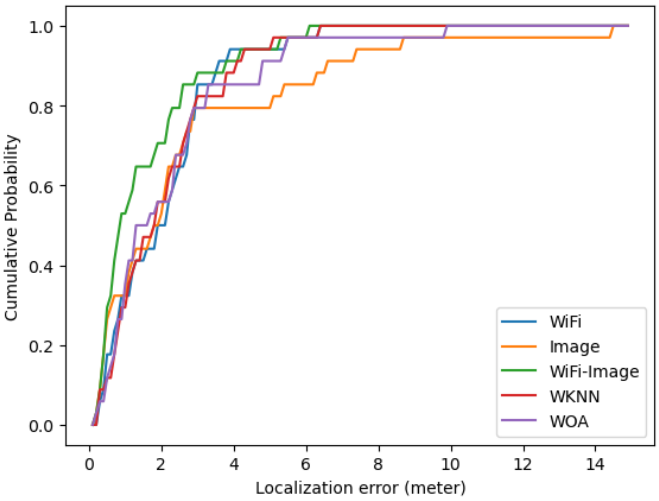

Fig. 11. CDF comparison with existing methods (In-field test 7 months after the collection of training data)

VI. Conclusions

A WiFi-Visual data fusion is proposed for indoor robot localization to limit the unbounded error of image localization in similar visual environments. The localization problem is formulated as a classification problem based on WiFi-Visual features. To jointly exploit the heterogeneous WiFi-Visual data, the WiFi features are represented in image form in order to have the same representation form of visual features. The two features of an unknown position are combined together and input to a neural network, which classifies the input features to the most likely class. The center position of this class is taken as the estimated position. The in-field test is carried out on a true robot platform with ground truth mapping and positioning system. The test results show that the proposed method can effectively limit the exceptional unbounded localization errors of image localization. The RMSE of the proposed method is less than 2 meters, which is smaller than that of WiFi-only localization. In addition, the performance of the proposed method is more stable than those of WiFi-only localization and image-only localization. Kalman filtering can also be used to improve the localization accuracy.

REFERENCES

[1] T. Umetani, Y. Kondo, and T. Tokuda, "Rapid development of a mobile robot for the Nakanoshima Challenge using a robot for intelligent environments," *Journal of Robotics and Mechatronics*, vol. 32, no. 6, pp. 1211-1218, 2020.

[2] T. Lee, C. Kim and D. Cho, "A Monocular Vision Sensor-Based Efficient SLAM Method for Indoor Service Robots," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 1, pp. 318-328, Jan. 2019.

[3] V.Vauchey, Y. Dupuis, P. Merriaux, X. Savatier, "Particle filter meets hybrid octrees: an octree-based ground vehicle localization approach without learning," *Applied Intelligence*, 7 Avril, 2023.

[4] N. Radwan, A. Valada, W. Burgard, "VLocNet++: Deep Multitask Learning For Semantic Visual Localization And Odometry," *IEEE Robotics And Automation Letters (RA-L)*, 3(4):4407-4414, 2018.

[5] A. Valada, N. Radwan, W. Burgard, "Deep Auxiliary Learning For Visual Localization And Odometry," *Proceedings Of The IEEE International Conference On Robotics And Automation*, Brisbane, Australia, 2018.

[6] T. Deng, H. Xie, J. Wang and W. Chen, "Long-Term Visual Simultaneous Localization and Mapping: Using a Bayesian Persistence Filter-Based Global Map Prediction," *IEEE Robotics & Automation Magazine*, vol. 30, no. 1, pp. 36-49, March 2023.

[7] T. Deng, G. Shen, T. Qin, J. Wang, W. Zhao, J. Wang, D. Wang, W. Chen; "PLGSLAM: Progressive Neural Scene Represenation with Local to Global Bundle Adjustment," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 19657-19666

[8] S. -H. Bach, P. -B. Khoi and S. -Y. Yi, "Global UWB System: A High-Accuracy Mobile Robot Localization System With Tightly Coupled Integration," *IEEE Internet of Things Journal*, vol. 11, no. 9, pp. 16618-16626, 1 May, 2024.

[9] A. Venon, Y. Dupuis, P. Vasseur and P. Merriaux, "Millimeter Wave FMCW RADARs for Perception, Recognition and Localization in Automotive Applications: A Survey," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 3, pp. 533-555, Sept. 2022.

[10] J. Jun, L. He, Y. Gu, W. Jiang, G. Kushwaha, A. Vipin, L. Cheng, C. Liu, and T. Zhu, "Low-overhead wifi fingerprinting," *IEEE Transactions on Mobile Computing*, vol. 17, no. 3, pp. 590–603, 2017.

[11] W. Sun, M. Xue, H. Yu, H. Tang, and A. Lin, "Augmentation of fingerprints for indoor wifi localization based on gaussian process regression," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 11, pp. 10 896–10 905, 2018.

[12] R. Ayyalasomayajula, A. Arun et al. Deep learning based wireless localization for indoor navigation. *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, 2020.

[13] A. Arun, R. Ayyalasomayajula, W. Hunter and D. Bharadia, "P2SLAM: Bearing Based WiFi SLAM for Indoor Robots," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 3326-3333, April 2022.

[14] H. Chen, Y. Zhang, W. Li, X. Tao, and P. Zhang, "ConFi: Convolutional neural networks based indoor Wi-Fi localization using channel state information," *IEEE Access*, vol. 5, pp. 18066–18074, Sep. 2017.

[15] https://pal-robotics.com/

[16] Y. Lecun, L. Bottou, Y. Bengio et al. "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, 1998, vol. 86, no 11, p. 2278-2324.

[17] X. Peng, R. Chen, K. Yu, F. Ye, and W. Xue, "An improved weighted k-nearest neighbor algorithm for indoor localization," *Electronics*, vol. 9, no. 12, p. 2117, 2020.

[18] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Advances in engineering software*, vol. 95, pp. 51–67, 2016.

[19] W. Sun, M. Xue, H. Yu, H. Tang, and A. Lin, "Augmentation of fingerprints for indoor wifi localization based on gaussian process regression," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 11, pp. 10 896–10 905, 2018.