HAL
open science

Topological Localization using Wi-Fi and Vision merged into FABMAP framework

Mathieu Nowakowski, Cyril Joly, Sébastien Dalibard, Nicolas Garcia, Fabien Moutarde

To cite this version:

Mathieu Nowakowski, Cyril Joly, Sébastien Dalibard, Nicolas Garcia, Fabien Moutarde. Topological Localization using Wi-Fi and Vision merged into FABMAP framework. 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Sep 2017, Vancouver, France. 10.1109/IROS.2017.8206171 . hal-01691920

HAL Id: hal-01691920

https://minesparis-psl.hal.science/hal-01691920v1

Submitted on 24 Jan 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Topological Localization using Wi-Fi and Vision merged into FABMAP framework.

Mathieu Nowakowski[1,2], Cyril Joly[1], Sébastien Dalibard[2], Nicolas Garcia[2], Fabien Moutarde[1]

***Abstract*— This paper introduces a topological localization algorithm that uses visual and Wi-Fi data. Its main contribution is a novel way of merging data from these sensors. By making Wi-Fi signature suited to FABMAP algorithm, it develops an early-fusion framework that solves global localization and kidnapped robot problem. The resulting algorithm is tested and compared to FABMAP visual localization, over data acquired by a Pepper robot in an office building. Several constraints were applied during acquisition to make the experiment fitted to real-life scenarios. Without any tuning, early-fusion surpasses the performances of visual localization by a significant margin: 94% of estimated localizations are less than 5m away from ground truth compared to 81% with visual localization.**

Keywords: topological localization, kidnapped robot, low-cost sensors, data fusion.

I. INTRODUCTION

A. Problem statement

This paper addresses the problem of indoor localization for mobile service robotics. The current market trend consists in a mass deployment of affordable mobile robots interacting with humans. This raises the need for low-cost solutions enabling those robots to: map their environment, and constantly know where they are when they move in it. Numerous projects have been proposed to solve the problem of localization. However, most of these solutions are based on the use of expensive sensors, such as laser range finders, and are designed for specific platforms.

The need for low-cost localization solutions has focused some research on the use of visual sensors. To achieve a long-term and robust global localization, some algorithms describe locations by their appearances [1]. We base our work on the Fast Apperance-Based Mapping algorithm (FABMAP) [2] that uses visual appearance to detect loop closures. This algorithm achieves robust localization with a low rate of false loop closure detection and can manage big maps by employing an inverted-index [3].

However, localization algorithms using visual appearance face the well-known problem of *perceptual aliasing*. Perceptual aliasing happens when two different locations share similar visual appearance (see example on Fig. 1). This problem is inherent in repetitive environment. A solution is the use of a multi-sensor localization for disambiguating such cases.

*This work was supported by SoftBank Robotics Europe.

[1]Authors are with Centre of Robotics, MINES ParisTech, PSL Research University, 60 Bd St Michel, 75006 Paris, France. mathieu.nowakowski@mines-paristech.fr

[2]Authors are with SoftBank Robotics Europe, 43 Rue du Colonel Pierre Avia, 75015 Paris, France.

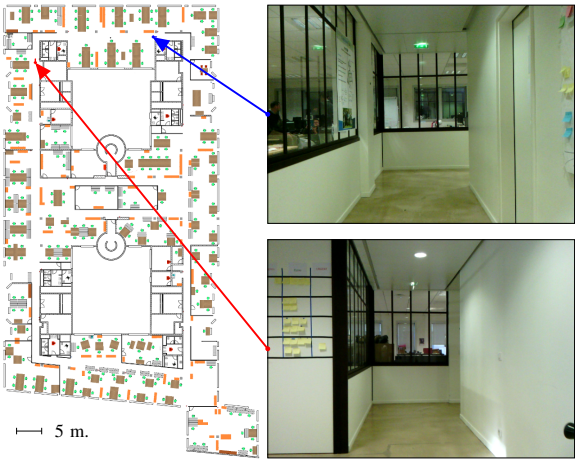

Fig. 1: Perceptual aliasing. Two distant locations in a building share similar visual appearance. But the Wi-Fi signatures received in these locations are different enough to distinguish them.

Using Wi-Fi helps to disambiguate cases where many locations have similar visual appearances. For example, corridors on opposite sides of a building, or at different floors, have different Wi-Fi signatures but can share comparable appearances. Recent work has introduced a way of including Wi-Fi data in the FABMAP algorithm [4] (preprint), but it does not benefit from advantages of both sensors.

In this paper, we introduce a novel way of merging visual and Wi-Fi data in order to solve the global localization and the kidnapped robot problems. We tested our algorithm on Pepper robots, visible on Fig. 2. Because of real-life uses, our approach aims to achieve localization tasks without stopping other actions of the robot.

B. Related Work

In recent years, several indoor localization algorithms based on Wi-Fi sensors have been introduced [6], [7], [8]. This popularity can be explained by two reasons. First, the Wi-Fi coverage in urban environment is dense enough for being used in localization task. Second, it is easy to equip mobile robot with Wi-Fi sensor.

Visual and Wi-Fi data accuracy are complementary. Even if Wi-Fi localization is less accurate [9], it does not suffer from perceptual aliasing, visually dynamic or repetitive environments. Several approaches take advantage of this synergy, and use visual and Wi-Fi sensors to create a low-cost localization. Most of work focusing on solving the localization

Fig. 2: Pepper robot: designed for making his interaction with human being as natural and intuitive as possible. It has been first presented in [5].

problem from these sensors uses particle filters for fusion [10], [11]. However, the hypothesis strongly converge only with motion. Other approaches are sequential, and usually Wi-Fi guided. They define a set of possible locations from Wi-Fi data, and perform visual localization over it [12]. Finally, some methods consist in choosing which sensor is the most reliable for current estimation [13]. These two last approaches are both likely to suffer from one misled sensor.

C. Contribution

Our main contribution is a novel way of merging Wi-Fi and vision for localization tasks. We propose an *early-fusion* process for combining visual and Wi-Fi data, that takes no more inputs than classical FABMAP. In comparison with related work, our approach looks for a compromise on the current estimation by considering data from both sensors together.

D. Paper organization

To this end, we use the FABMAP localization algorithm briefly presented in section II. Section III defines a way to make Wi-Fi data suited to FABMAP. Early-fusion process is introduced in section IV, with other merging styles. Finally, our experimental acquisitions and localization results are presented and discussed in section V.

II. Fast Apperance-Based Mapping

FABMAP [2], [3] discretizes the environment into a succession of topological nodes. Each node constitutes a location L_i, and is associated with one or several visual observations. Given a query image, the goal of FABMAP is to compute the following value for each place L_i in a topological map:

$$p(L_i|\mathcal{Z}^k) = \frac{p(Z_k|L_i, \mathcal{Z}^{k-1})p(L_i|\mathcal{Z}^{k-1})}{p(Z_k|\mathcal{Z}^{k-1})}. \quad (1)$$

Where Z_i is the ith observation and $\mathcal{Z}^i$ is the set of all observations, up to i. Three terms can be identified in (1): the likelihood $p(Z_k|L_i, \mathcal{Z}^{k-1})$, the normalization term $p(Z_k|\mathcal{Z}^{k-1})$, and a prior knowledge on the current pose $p(L_i|\mathcal{Z}^{k-1})$. Note that in our work, we do not use this last term because we focus on the global localization problem. The three next sub-sections respectively introduce the computations of the observation Z_k, the likelihood and the normalization term.

A. Visual Appearance Description

The first step of FABMAP is to transform a query image into a compact image description that is suited to the localization context. This compact description is called observation vector, or appearance vector, and is noted Z in (1). To do this, FABMAP uses the bag-of-words approach introduced in computer vision in [14]. Keypoints are extracted on the image, and their descriptors are then associated with words of a vocabulary. In FABMAP, the observation Z indicates which words are present on the query image.

For a vocabulary of N words, Z thus contains N binary values indicating the presence or absence of the corresponding word in the query. The vocabulary used comes from an offline learning. It is usually built thanks to a clustering method like the k-means, on keypoint descriptors extracted from a lot of learning images.

B. Observation Likelihood

The second step constitutes the core of the algorithm and computes the likelihood term $p(Z_k|L_i, \mathcal{Z}^{k-1})$, simplified into $p(Z_k|L_i)$, assuming independence between current and past observations, conditioned on the location [2]. Approaches using the bag-of-words framework can compute similarity scores between queries and references thanks to methods like the *Term Frequency - Inverse Document Frequency* [15], or hierarchical vocabularies [16]. The main contribution of FABMAP is the use of a Chow-Liu tree [17] that captures correlations between the different words of the vocabulary. This approach is motivated by the fact that certain words are often detected on specific objects and thus, tend to co-occur. The authors show that learning these correlations helps to avoid false associations due to perceptual aliasing [2]. It also helps to achieve correct associations between images, even if they share few words in common.

C. Normalization

In localization tasks, the normalization step allows to detect unknown locations. In [2], the authors split $p(Z_k|\mathcal{Z}^{k-1})$ into two sums, one representing the visited locations M, the other the unknown world $\bar{M}$:

$$\begin{aligned} p(Z_k|\mathcal{Z}^{k-1}) &= \sum_{m \in M} p(Z_k|L_m)p(L_m|\mathcal{Z}^{k-1}) \\ &\quad + \sum_{u \in \bar{M}} p(Z_k|L_u)p(L_u|\mathcal{Z}^{k-1}). \end{aligned} \quad (2)$$

The second summation cannot be evaluated directly. The authors of FABMAP propose to approximate (2) by:

$$\begin{aligned} p(Z_k|\mathcal{Z}^{k-1}) &\approx \sum_{m \in M} p(Z_k|L_m)p(L_m|\mathcal{Z}^{k-1}) \\ &\quad + p(L_{new}|\mathcal{Z}^{k-1}) \sum_{u=1}^{n_s} \frac{p(Z_k|L_u)}{n_s}. \end{aligned} \quad (3)$$

where, $p(L_{new}|\mathcal{Z}^{k-1})$ corresponds to the probability of being in a new location, and is a user-specified input of the algorithm (set to 0.9 in [2]). The second sum of equation (3) consists then in sampling an observation Z to create a place model associated to unknown location. The sampling of Z is realized from training set of n_s images.

In addition to the vocabulary, the Chow Liu tree and the n_s samples, the authors of [2] list some user-specified inputs. In our work, these parameters are set to the values specified in [2]. Fig. 3 summarizes the successive steps of the algorithm.

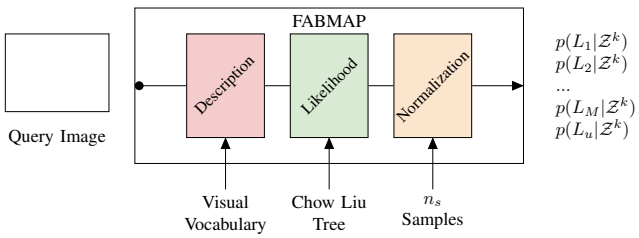

Fig. 3: Inputs of each steps of the FABMAP algorithm.

III. Including Wi-Fi data in FABMAP

In the related literature [6], [9], a Wi-Fi signature - sometimes referred to as a *fingerprint* - consists of a list of visible Access Points (APs), each one being characterized by its MAC address and its signal strength (Received Signal Strength Indication - RSSI). Most Wi-Fi localization algorithms collect Wi-Fi signatures during an exploration stage then generate a map modeling the distribution of Wi-Fi signals in the environment. These approaches particularly suit the topological localization of FABMAP, with the definition of a correct framework. This section introduces the way Wi-Fi information is integrated into FABMAP, following the steps of Fig. 3.

A. Defining a Wi-Fi vocabulary

Each MAC address defines one Wi-Fi word. This choice allows to keep the appearance description style of FABMAP, that uses binary values to indicate the presence or absence of visual words in query image. Here, values of Wi-Fi observation vector Z_{Wi-Fi} indicate which APs are visible in a Wi-Fi signature.

At a given location, Wi-Fi signal strength varies significantly. To make their inputs more reliable, most Wi-Fi localization schemes stay motionless for a while, and compute mean and standard deviation of the RSSI coming from each AP [6], [8]. In this paper, we choose to ignore the strength information because of Pepper's motion behaviours and for simplicity. Potential usage of this information is discussed in section VI and constitutes future work.

Finally, the main difference between visual and Wi-Fi vocabularies is that it is not possible to know the APs a robot is going to encounter in an environment before exploring it. Therefore, it is not possible to define a global and complete Wi-Fi vocabulary in a pre-exploration offline stage. The Wi-Fi dictionary is computed after a first exploration and can be extended when new APs become visible.

B. Tree Structure

Considering now that our robot has explored its environment and defined a vocabulary, we can build a Chow Liu tree that catches correlations between Wi-Fi words. We learn the correlations on the collected Wi-Fi signatures. However, in order to avoid the learning of redundant correlations, we have to ensure that Wi-Fi observation vectors come from different places. Timestamps and odometry data can be used to check this.

C. Normalization and virtual Wi-Fi locations

In the visual world, sampling an observation for normalization is easy. To do this, training images coming from offline learning are used. In the Wi-Fi world, employing this trick is not so simple. One Wi-Fi vocabulary is specific to one environment. Therefore, using real Wi-Fi signatures collected in training environments does not make a lot of sense since the computed observation vectors would only be composed of 0. A solution is to simulate virtual Wi-Fi signatures according to the collected data. To this end, we use a random generator of Wi-Fi observation vectors.

In our measurements, the number of APs in a Wi-Fi signature follows a normal distribution for a specific environment. We thus identify the mean μ and the standard deviation σ on the number of APs perceived in the Wi-Fi signatures collected during the exploration. To randomly generate n_s virtual locations, we repeat the two following steps:

1) randomly select a number of perceived APs following the normal distribution $\mathcal{N}(\mu, \sigma)$,
2) randomly choose Wi-Fi words in known and unknown vocabularies of the same size.

With this formalism, FABMAP localization results based on Wi-Fi data can be computed. However, using Wi-Fi data alone shows poorer accuracy than visual-based localization. Next section introduces how to take advantage of both visual and Wi-Fi sensors.

IV. Merging Visual and Wi-Fi data

This section introduces the early-fusion process. To our knowledge, this approach has never been studied for solving the global localization problem using Wi-Fi and visual data. Our methodology is shown on Fig. 4. The following subsections discuss the interest of this approach, our choices concerning the inputs of the algorithm, and present more classical merging frameworks.

A. Late-fusion and Early-fusion

An intuitive way of merging localization results coming from multiple sensors can be called the *late-fusion*. Each sensor s provides a probability $p_s(L_i|\mathcal{Z}_s^k)$ of being in a location L_i knowing its observations $\mathcal{Z}_s^k$. For a multi-sensor platform composed of two sensors s_1 and s_2, the result p_{lf} from the late-fusion can be written as:

$$p_{lf}(L_i|\mathcal{Z}_{s_1,s_2}^k) = \alpha \times \frac{p_{s_1}(L_i|\mathcal{Z}_{s_1}^k)\, p_{s_2}(L_i|\mathcal{Z}_{s_2}^k)}{p(L_i)}. \quad (4)$$

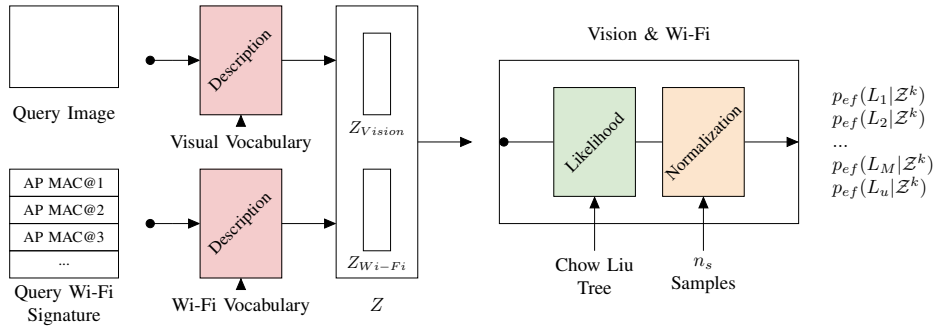

Fig. 4: Early-fusion framework, using visual and Wi-Fi data.

Where α ensures that $\sum_i p_{lf}(L_i|\mathcal{Z}^k_{s_1,s_2}) = 1$. Focusing on the global localization problem and considering that n_M locations have been mapped, plus one location, associated to the unknown world, note that for every i, $p(L_i) = \frac{1}{n_M+1}$.

With the late fusion, each localizer gives a result according to the perceptual area of its sensor. Then, a sensor misled by perceptual aliasing can clearly pervert the system. In this work, we propose to merge the data before giving it to the localization algorithm. We realize thus an early-fusion. The output probabilities of the algorithm are noted $p_{ef}(L_i|\mathcal{Z}^k)$. The idea is to concatenate the observation vectors obtained from visual and Wi-Fi sensors (Fig. 4). The early-fusion asset is the computation of a compromise between the different sensors.

B. Which correlations to learn for early fusion?

When merging Wi-Fi and visual data by early-fusion, the question of which correlations to learn remains. We choose to split the correlations learning. Instead of learning one tree, two trees capture words that respectively co-occur in the visual vocabulary and in the Wi-Fi vocabulary. This choice can be explained by two reasons.

First, learning new correlations between all Wi-Fi and visual words is not obvious. For instance, to learn new visual correlations from nodes collected during exploration, we have to ensure that all visual words of the vocabulary would have been seen at least once during the exploration. In practice, this is not possible.

Second, the normalization step encourages this choice. Indeed the visual and Wi-Fi unknown world is too difficult to simulate whereas using split learning allows to simply concatenate unknown visual and Wi-Fi samples.

Even if the correlations learning is split, we show in section V that early and late fusions generate different results. Both fusions are also compared to more classical Wi-Fi and vision merging style: the sequential fusion.

C. Sequential fusion

Two methodologies can be identified when using sequential fusion with Wi-Fi and visual data:

1) Wi-Fi-guided fusion, in which a visual localization is realized over possible locations determined from Wi-Fi data;
2) Wi-Fi check fusion, where the result from visual localization must be confirmed by Wi-Fi data.

These approaches use the fact that Wi-Fi localization is less accurate than the visual one, but never produces aberrant results.

However, FABMAP normalization style enables the algorithm to detect loop closures with the assumption that no one has been missed. For each query, FABMAP detects at most one loop-closure. To make these sequential fusions work, we need Wi-Fi localization to furnish a set of several possible locations. The FABMAP normalization term introduced in [18] has been used then for Wi-Fi localization in our work.

V. Evaluation

A. Experimental conditions

We evaluated our algorithm on data acquired by a Pepper robot. Pepper is a social robot designed by SoftBank Robotics. It can be found in Japanese shops, where it holds receptionist or demonstrator functions. One of Pepper's key strengths comes from the interaction it has with users. Therefore, practical uses of Pepper result in special constraints applied during acquisitions.

The environment used is the office floors of SoftBank Robotics Europe, that are mainly composed of open spaces (see map and example images on Fig. 1). This facet is significant considering the propagation of Wi-Fi signals. In such environments, Wi-Fi signatures are more difficult to distinguish because there are no obstacles creating important changes. Note that possible uses of Pepper in large indoor environments, like shops, malls or airports, have motivated our choice of this testing conditions.

Acquisitions were done by driving Pepper thanks to a remote control. During the acquisitions, the robot autonomously acquires images every 2s, and Wi-Fi signatures every 10s. Several constraints come from the fact that we want the localization to be natural. For instance, blurry images result from the fact that we do not want the robot to stop for image acquisitions. Moreover, motion behaviours of the robot are kept. When navigating, Pepper looks in the direction of its motion. So in straight lines, images are not taken in discriminative perpendicular directions (left or right), but in the direction of movement.

The paths we run are collected over two months and multiple floor levels. They constitute a set of 4136 images and 556 Wi-Fi signatures, composed of 553 MAC addresses. The total covered distance is 1.2km long. We also pay attention to the diversity and reality of our acquisition scenarios: occlusions, dynamic environment, realistic velocity, user interactions, blur, various times of day, etc.

Finally, the visual vocabulary used was learnt from 3000 images of indoor scenes, extracted from database presented in [19].

B. Annotations: initial exploration and localization

Our formalism defines topological nodes by an associated couple (image ; Wi-Fi signature). In practice, images are collected faster than Wi-Fi signatures. Therefore, we choose to associate acquired images with Wi-Fi signatures under two scenarios: mapping and localization. During the initial exploration phase, we associate images to the estimated spatially closest Wi-Fi signature. During the localization tests, images are associated to the last Wi-Fi signature acquired.

We manually annotate the positions of all collected images, resulting in ground truth of node positions. The different acquisitions are split in two: 40% are used for map creation and 60% constitute our queries.

Note that all Wi-Fi correlations are learnt over examples used for the mapping.

C. Evaluation metrics

For each query, we consider the highest score of the algorithm as the current localization in the map:

$$L_{max} = \underset{L_i, \forall i}{\operatorname{argmax}} \; p(L_i | \mathcal{Z}^k) \tag{5}$$

To evaluate accuracy, the Euclidean distance between the annotated positions (x, y, z) of the query and the associated mapped place L_{max} is computed. When $L_{max} = L_u$, corresponding to unknown location, our evaluation considers the result as rejected.

Next sub-sections discuss the produced results. We use the FABMAP2.0 algorithm [3], and adapt the open source implementation of [20] to our use. For the set of queries, localizations are computed using visual data only, Wi-Fi data only, and different merging styles: early and late fusions, and the two sequential fusions presented in Sect. IV-C. Sub-section V-D presents the obtained results for global localization. The long-term performances of the different algorithms are also tested in sub-section V-E.

D. Global localization results

In this experiment, the robot has entirely mapped a floor of an office building. The robot is then placed in it for global localization.

Results plot on the following figures show:

- *rate of correct localizations*: measured as the cumulative distribution of distances between estimated localizations and the ground truth (Fig. 5),
- *rate of misplaced localizations*: measured as the percentage of queries leading to estimated localizations farther away from ground truth than a given distance (Fig. 6).

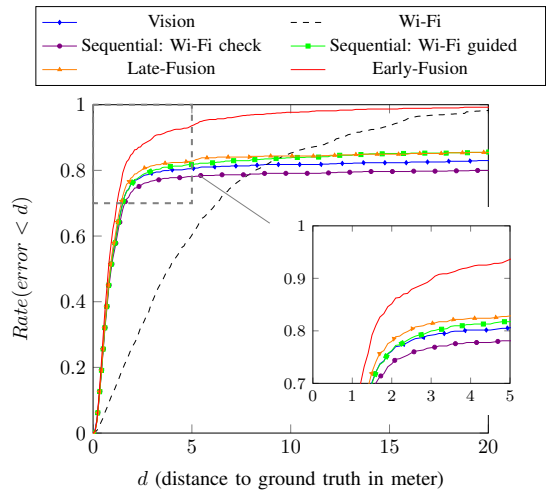

Fig. 5: Rate of correct localizations.

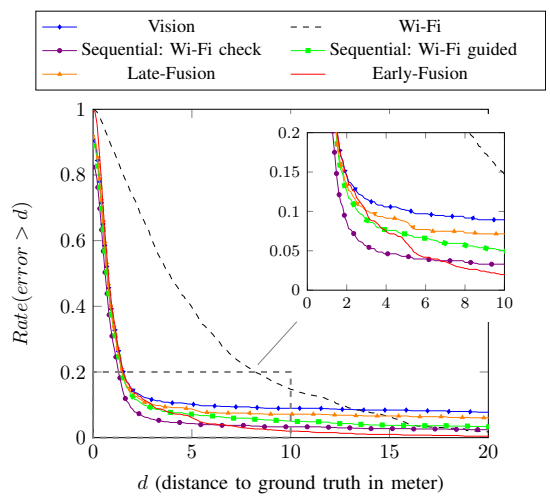

Fig. 6: Rate of misplaced localizations.

Results on Fig. 5 and Fig. 6 highlight the interest of merging Wi-Fi and vision for localization tasks. On Fig. 5, the early-fusion clearly outperforms classical FABMAP with e.g. 94% of queries localized within 5m away from true position, compared to only 81% for vision-only (see left column of Tab. I). On Fig. 6, the rejection rate of sequential fusion with Wi-Fi check reduces the error of this merging style, but the price to pay is high: "unknown localization" answer is provided for nearly 1/5 of queries (18%).

However, early-fusion correctly localizes all the queries within 23m and outperforms other introduced localizations for $d > 7m$. Visual and sequential fusion with Wi-Fi check localizations result in aberrant estimations localized farther than 23m from ground truth. We extract in Tab. I resulting values from early-fusion and Wi-Fi check localizations for

comparison with classical visual FABMAP.

	Correct localization rate (%) within				Misplaced localization rate (%) away from			
	2m	5m	10m	20m	2m	5m	10m	20m
Vision	76	81	82	83	15	10	9	8
Wi-Fi check	74	78	79	80	**9**	**4**	3	2
Early-Fusion	**85**	**94**	**98**	**99**	15	6	**2**	**0.4**

TABLE I: Comparison of classical FABMAP with sequential and early-fusion localizations.

E. Long-Term localization

We tested the long-term robustness of these algorithms by spacing initial exploration phase and localization tests one month apart. The obtained results reveal that visual localization is more deteriorated than the Wi-Fi localization. Small changes occurring in Wi-Fi signals, like new or missing smart-phones mobiles APs, are not significant enough for degrading the outcomes of Wi-Fi localization. However changes in visual appearance, for example caused by furniture moved, deteriorate the results of visual and sequential localizations. Results in Tab. II show that our proposed early-fusion with Wi-Fi enhanced FABMAP gives the best compromise between visual and Wi-Fi data.

	Correct localization rate (%) within				Misplaced localization rate (%) away from			
	2m	5m	10m	20m	2m	5m	10m	20m
Vision	52	60	63	66	34	26	23	20
Wi-Fi check	50	58	61	63	**21**	**13**	10	8
Early-Fusion	**67**	**85**	**94**	**99**	33	15	**6**	**1**

TABLE II: Long-term localization results: initial exploration phase and localization tests one month apart.

VI. Conclusion and discussion

In this paper, we have introduced an early-fusion process that consists in merging visual and Wi-Fi data for localization tasks. Based on the FABMAP algorithm, a low-cost solution for solving global localization and the kidnapped robot problem has been created. We have tested our algorithm on a Pepper robot, in realistic situations and practical uses. Without any tuning, our approach has significantly improved the localization accuracy in repetitive indoor environment in comparison with classical visual FABMAP: 94% of estimated localizations are less than 5m away from ground truth, compared to 81% with visual FABMAP.

An interesting extension would be the integration of Wi-Fi signal strengths into the FABMAP formalism. A solution for this would be the discretization of intensity on binary values to fit within the FABMAP appearance description style. This approach could thus expend the Wi-Fi vocabulary, and improve the localization accuracy.

VII. Acknowledgements

These results have received funding from Bpifrance through the PSPC Project ROMEO 2 and from European Unions Horizon 2020 through the MuMMER project under grant agreement No 688147.

References

[1] S. Lowry, N. Sünderhauf, P. Newman, J. J. Leonard, D. Cox, P. Corke, and M. J. Milford, "Visual place recognition: A survey," *IEEE Transactions on Robotics*, vol. 32, no. 1, pp. 1–19, 2016.

[2] M. Cummins and P. Newman, "Fab-map: Probabilistic localization and mapping in the space of appearance," *The International Journal of Robotics Research*, vol. 27, no. 6, pp. 647–665, 2008.

[3] ——, "Appearance-only slam at large scale with fab-map 2.0," *The International Journal of Robotics Research*, vol. 30, no. 9, pp. 1100–1123, 2011.

[4] J. Wietrzykowski, M. Nowicki, and P. Skrzypczyński, "Adopting the fab-map algorithm for indoor localization with wifi fingerprints," *arXiv preprint arXiv:1611.02054*, 2016.

[5] J. Lafaye, D. Gouaillier, and P.-B. Wieber, "Linear model predictive control of the locomotion of pepper, a humanoid robot with omnidirectional wheels," in *Humanoid Robots (Humanoids), 2014 14th IEEE-RAS International Conference on*. IEEE, 2014, pp. 336–341.

[6] J. Biswas and M. Veloso, "Wifi localization and navigation for autonomous indoor mobile robots," in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*. IEEE, 2010, pp. 4379–4384.

[7] S. Boonsriwai and A. Apavatjrut, "Indoor wifi localization on mobile devices," in *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2013 10th International Conference on*. IEEE, 2013, pp. 1–5.

[8] M. Jirku, V. Kubelka, and M. Reinstein, "Wifi localization in 3d," in *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*. IEEE, 2016, pp. 4551–4557.

[9] H. Liu, Y. Gan, J. Yang, S. Sidhom, Y. Wang, Y. Chen, and F. Ye, "Push the limit of wifi based localization for smartphones," in *Proceedings of the 18th annual international conference on Mobile computing and networking*. ACM, 2012, pp. 305–316.

[10] F. Schwiegelshohn, T. Nick, and J. Götze, "Localization based on fusion of rfid and stereo image data," in *Positioning Navigation and Communication (WPNC), 2013 10th Workshop on*. IEEE, 2013, pp. 1–6.

[11] M. Quigley, D. Stavens, A. Coates, and S. Thrun, "Sub-meter indoor localization in unmodified environments with inexpensive sensors," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*. IEEE, 2010, pp. 2039–2046.

[12] M. Nowicki, "Wifi-guided visual loop closure for indoor navigation using mobile devices," *Journal of Automation Mobile Robotics and Intelligent Systems*, vol. 8, no. 3, pp. 10–18, 2014.

[13] J. Biswas and M. Veloso, "Multi-sensor mobile robot localization for diverse environments," in *Robot Soccer World Cup*. Springer, 2013, pp. 468–479.

[14] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*. IEEE, 2003, pp. 1470–1477.

[15] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Information Processing & Management*, vol. 24, no. 5, pp. 513–523, Jan. 1988.

[16] D. Nistér and H. Stewénius, "Scalable Recognition with a Vocabulary Tree," in *IN CVPR*, 2006, pp. 2161–2168.

[17] C. Chow and C. Liu, "Approximating discrete probability distributions with dependence trees," *IEEE transactions on Information Theory*, vol. 14, no. 3, pp. 462–467, 1968.

[18] E. Stumm, C. Mei, and S. Lacroix, "Probabilistic place recognition with covisibility maps," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 4158–4163.

[19] A. Quattoni and A. Torralba, "Recognizing indoor scenes," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 413–420.

[20] A. Glover, W. Maddern, M. Warren, S. Reid, M. Milford, and G. Wyeth, "Openfabmap: An open source toolbox for appearance-based loop closure detection," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, 2012, pp. 4730–4735.