WiFi-guided Visual Loop Closure for Indoor Localization using Mobile Devices

Submitted: 3th June 2014; accepted: 11th June 2014

Michał Nowicki

DOI: 10.14313/JAMRIS_3-2014/23

Abstract:
Mobile, personal devices are getting more capable every year. Equipped with advanced sensors, mobile devices can use them as a viable platform to implement and test more complex algorithms. This paper presents an energy-efficient person localization system allowing to detect already visited places. The presented approach combines two independent information sources: wireless WiFi adapter and camera. The resulting system achieves higher recognition rates than either of the separate approaches used alone. The evaluation of presented system is performed on three datasets recorded in buildings of different structure using a modern Android device.

Keywords: *mobile devices, localization, sensor fusion*

1. Introduction

Mobile devices, like tablets or smartphones, are nowadays equipped with more sensors than few years ago. Those sensors combined with increasing processing capabilities allow to develop more complex, real-time algorithms that can be used for personal navigation or detection of potentially dangerous situations. Those algorithms have not only academic, but also commercial significance due to the popularity of personal mobile devices in the modern world.

One of the sensors available in every, recent Android device is a WiFi adapter. Most users use this adapter to connect to wireless Access Points (APs), but it can be used as a sensor that measures the strength of surrounding wireless networks. The researched approaches utilizing WiFi scans can be divided into two groups: WiFi triangulation or WiFi fingerprinting. The WiFi triangulation uses three or more APs that are visible in line-of-sight and triangulates the user position based on the measured signal strength of each network [1]. This approach is effective if the localization is performed in open-space areas. In a typical building with cluttered environment that is rich in corridors and additional rooms, WiFi triangulation is still applicable, but the number of APs needed to perform successful localization is higher. Therefore, if there exists an additional prerequisite to use only the already existing APs infrastructure, WiFi triangulation can provide misleading localization as the number of signal reflections negatively impacts the measured signal strength. In structured environment, the WiFi information can be used to determine the measured position based on the list of available wireless networks in a single scan. This technique, called WiFi fingerprinting, determines the similarity of current scan to previous scans or to the entries in a recorded database of WiFi scans. The efficient, working solutions utilizing WiFi fingerprinting were presented in [3], [4] and [12]. Other researches focus on using sensors that are equivalent to the equipment present in typical mobile devices, but do not perform the experiments on actual mobile devices [3], [19].

This information might be used to provide an estimate of the user's localization, but the precision of signal measurement depends greatly on the orientation of the measurement with respect to the APs. Holding the mobile device in different way or shadowing the signal with the person's body affects the obtained results and can have a negative impact on the repeatability of the measurements. Therefore, to alleviate this influence, it is beneficial to incorporate information from additional sensors, e.g., an inertial sensor. Modern mobile devices are in most cases equipped with a 3-axis accelerometer, a 3-axis gyroscope, and a 3-axis magnetometer. The information from these sensors can be used to create a system estimating the orientation of a smartphone [10]. The orientation estimate can be later effectively used to enhance the WiFi measurement.

Another sensor that is a standard in mobile devices is a camera. The sight plays significant role in the localization strategy of human beings and therefore image processing is researched in robotics and computer vision communities. Methods estimating the total motion based on consecutive image-image estimates are called Visual Odometry and are especially important for mobile robots [14]. Typically, those methods find a sparse set of features that are matched/tracked in consecutive images. The positions of features in compared images are used to estimate the transformation. Due to the frame-frame estimation, those methods suffer from an estimation drift arising due to error summation over time. This approach provides a continuous estimate of motion, but is also computation-demanding and thus energy-consuming. Energy-efficiency is especially important for small, portable devices, and from user's point of view should not have a significant, negative impact on the battery lifetime.

The WiFi and vision based approaches to indoor localization are usually researched separately, neglecting the possible synergies of both information sources and gains due to data fusion. The known works approaching the problem of multi-sensor fusion for indoor localization on mobile devices are

dominated by the continuous data fusion paradigm, employing a filter-based framework [8]. The results being presented are often achieved with custom experimental setups [19], not actual mobile devices. Thus, these works avoid confronting the problems of limited computing power and energy. Some other approaches focus on enhancing the WiFi-based localization with data from inertial sensors, but do not use cameras [11], [18].

This paper presents a prototype system that determines on demand the position of a person inside a building using data from the WiFi and camera of a mobile device (smartphone). The acquired WiFi scan is used to determine the best fingerprint match to the WiFi scans recorded previously and stored in a database of known locations. Then, the WiFi-based position estimate is confirmed and refined by matching a compact representation of the location's visual appearance to the image-based description of the known locations, also stored in a database. Thus, the proposed system combines data from both sources of localization information available in a typical mobile device, achieving higher recognition rates than eithor of subsystems and is less prone to failures caused by the peculiarities of a particular environment. Moreover, the system is energy-efficient as the loop closure detection procedure is triggered only when needed, as a discrete event. To the best knowledge of the author, a similar idea has not been yet presented in the literature.

In section 2, the structure of the proposed system is presented, as well as the details of the WiFi-based and image-based subsystems. The next section 3 focuses on the experimental evaluation of each subsystem and the integrated solution. Moreover, it describes three datasets recorded in different environments and used for evaluation. The last section 4 concludes the paper and mentions future work.

2. System Structure

2.1. WiFi Fingerprinting

The WiFi fingerprinting approach was firstly described in [1]. As the WiFi fingerprint allows only to localize in a known environment, the system based on WiFi fingerprint operates in two stages:

- data acquisition stage,
- localization stage.

In the data acquisition stage, certain positions are chosen as references, where available WiFi signals are scanned and stored in a database. These positions can be randomly chosen, uniformly chosen or based on the structure of the building. Due to the energy considerations, the proposed system scans only the positions that are important for user navigation, .e.g., doors that have to be crossed, beginning of the long corridor or the entrance to a new part of the building. Due to these limitations, it is assumed that the user is capable of performing local navigation whereas system provides global position information that the user can apply to plan his/her movement. The WiFi fingerprint approach assumes that each position can be uniquely defined by the combination of access points' MAC addresses and RSSI signal strength values. An exemplary situation is represented in Fig. 1, where the user movement is represented by dashed lines, whereas the discrete events, when WiFi scanning is performed, are drawn using circles. Each WiFi found in a single position is marked using a line connecting the AP and user position. The list of WiFi networks available in each position is the list of lines that are pointing towards user's position.

Assuming that the WiFi database of a floor is created, it is essential to efficiently compare the list of scanned WiFis $\mathcal{X}$ to the WiFi scans stored in the database D. The comparison has to be performed using a function that evaluates the difference of two scans: new scan $\mathcal{X}$ and one of the scans $\mathcal{Y}$ in the database D. Typically, the WiFi scans are compared using the Euclidean norm [1]:

$$d(\mathcal{X},\mathcal{Y}) = \frac{1}{N}\sqrt{\sum_{i=1}^{N}(\mathcal{X}_i - \mathcal{Y}_i)^2}, \tag{1}$$

where $\mathcal{X}_i$ and $\mathcal{Y}_i$ represent the strengths of i-th network found in both scans, $\mathcal{X}$ and $\mathcal{Y}$. Number N is the count of networks found in both scans.

Finding the best correspondence in the database can be written as finding a record, which distance function to current scan is minimal:

$$\mathcal{Y}_{min} = \underset{\mathcal{Y}\in D}{\operatorname{argmin}}\, d(\mathcal{X},\mathcal{Y}) \tag{2}$$

The Euclidean distance is usually applied as it allows to precisely position user based on the measured RSSI values. But in the case of sparse position set it is more important to rely on the unique set of found networks than on the strength of these networks. Therefore an evaluation of various distance/similarity functions is performed in section 3.

Moreover, as the system operates, it gathers new data that might be stored as the scans that have been correctly matched to some WiFi fingerprint from the database, or as unclassified cases. This way the system might gather new information, which can be used to detect, when user revisits position previously added to the database. The information about new positions can be also used to provide user with the database containing positions important for particular user, which due to the personal importance might be revisited in the future.

2.2. Visual Loop Closure

Visual loop closure is a technique that tries to determine if the currently observed scene had been previously encountered based on the captured images.

Computer vision algorithms usually try to process only a subset of available image information in order to reduce the processing time. This observation is also valid for visual loop closure, for which the detection of a sparse set of salient features is performed. In most

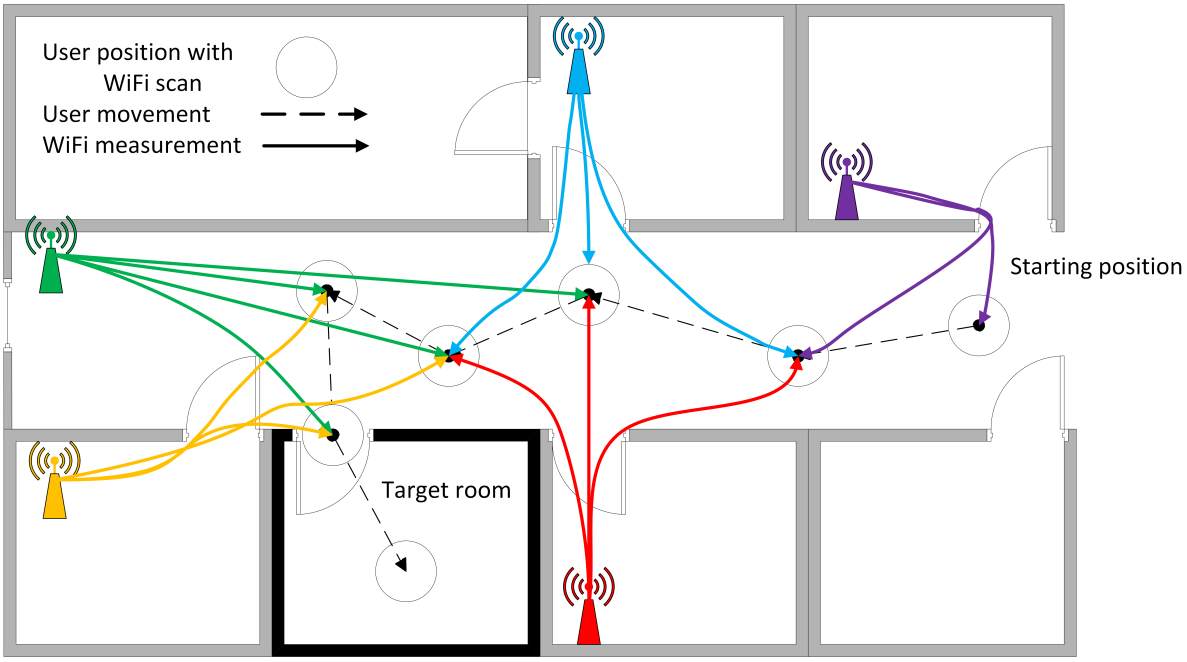

Fig. 1. In WiFi fingerprinting approach, user's position is recognized based on the combination of scanned WiFi networks

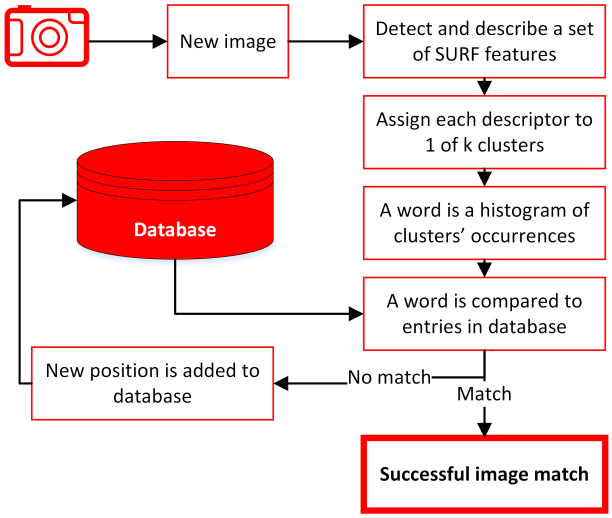

Fig. 2. The processing steps of the Visual Bag of Words approach

cases the SURF [2] detector and descriptor is used. Another possible approach may utilize the HoG [7] information. Each feature is then described by the set of values representing its local neighbourhood. Descriptors for all salient features are then compressed into a single image descriptor called a word. This approach is known as the Bag-of-Words approach [6]. To compress the information into an image's word, the Bag-of-Words approach firstly determines the k clusters of descriptor types using the k-means algorithm and then labels each image descriptor with the number of cluster it has been assigned to. The numbers of descriptors assigned to each cluster is used to create a histogram representing an image in further computations. The process results in reducing the representation of a single image into one vector of floating point values. The processing flow of Bag-of-Words is represented in Fig. 2.

In practical applications, the vision-based loop closure is hard to detect robustly. Even a small difference in the observation's orientation can influence the observed feature set and therefore prevent the system from correctly recognizing that the place was previously visited. What is more, the database of image words takes a lot of memory and may grow with the system's running time, therefore the corresponding image's are not stored.

2.3. WiFi-guided Visual Loop Closure

The main contribution of this paper is the combination of the already known algorithms in creation of a robust, data integrating system. The idea behind the proposed algorithm is simple: try to match WiFi information giving global estimate than can be a good initial estimate for further confirmation from the vision-based loop closure subsystem.

The system starts with gathering WiFi and image information into database during the preparation task to allow further loop closures. Due to the WiFi mechanism, WiFi scanning time takes one to five second depending on the used WiFi adapter drivers. These, relatively long scanning times make WiFi fingerprinting useless in case of a dynamic motion, e.g., person running through a building. What is important, dynamic motion also negatively impacts the vision-based loop closure as the images would contain significant amount of motion blur. Therefore, in the proposed system, dynamic motion is detected using the combination of gyroscope and accelerometer and in that situation new information is not inserted into the database. Assuming that the motion speed is below the chosen threshold, WiFi scan, image and orientation from the Android-based orientation estimation system are stored. Between the starting and ending time of the WiFi scanning, 20–40 images can be cap-

tured. From those images, the images with most distinct orientations are chosen to best represent different points of view. For each scan, the image taken approximately in the half of WiFi scan duration is chosen and will be referred to as the mid-scan image. If there are images with orientation significantly different than the mid-scan image, those images are considered to be used for visual loop closure. Maximally, mid-scan image and two additional images with highest orientation difference are processed per scan in the visual loop closure approach. For each image, it's salient features are detected and described using descriptors. The descriptors are then used to form an image's word using Bag-of-Words approach. The created word is a shorter representation of the image and allows efficient comparison between images.

The processing of the localization mode of the proposed system is presented in Fig. 3. The system gathers the WiFi and image information. From the image, Bag-of-Words technique creates a word representing observed location. Then the WiFi scan is compared to the database entries and in case of successful WiFi fingerprints match, the comparison of words representing the images is performed. If the WiFi match is confirmed by the image match, the mobile device is believed to have been successfully localized. If the position is not recognized, the image and the WiFi scan are stored in the database as a new position used in the recognition process.

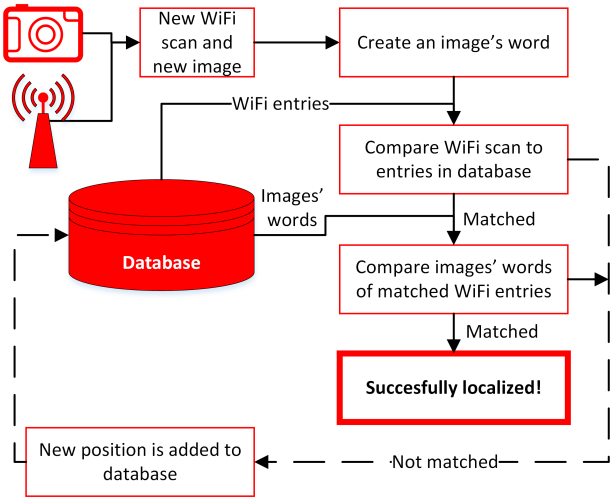

Fig. 3. Processing steps of the proposed loop closure approach

2.4. Implementation Remarks

The proposed approach is application-orientated, therefore it has been tested on the Samsung Galaxy Note 3, which uses the Android 4.4 as an operating system. The information about the WiFi signal strength was captured using Android-available functionality in the Java API. The time of a single WiFi scan time depends on the wireless adapter driver installed on the mobile device and on the Samsung Galaxy Note 3 it takes approx. 4 seconds.

The image processing was done using the commonly used OpenCV library (2.4.8) [5], which is available for x86/x64 and Android platforms [16]. The proposed application consists of a Java-part used for GUI and less demanding computations, and C++ NDK libraries for more demanding taska, e.g., image processing. The structure was proved to be a good trade-off between programming complexity and processing time and has been already proven to work well in another Android-based experiments [15]. A similar project structure is also used in [13].

3. Experimental Evaluation

3.1. Recorded Datasets

The experiments were performed on the dataset recorded in two buildings of the Poznan University of Technology (building of Mechatronics, Biomechanics and Nanoengineering (PUT CM) and the Lecture Center (PUT CW) and a shopping mall located in Poznań (SM). The user equipped with a smartphone was moving around the buildings gathering WiFi scans and corresponding images in places that seemed important for user localization due to the building structure, e.g. short corridor connecting two parts of the building or unique objects in sight. The dataset PUT CM contains 14 places of possible loop closures, where the dataset PUT CW contains 20 places of possible loop closures. The shopping mall dataset SM contains also 20 places of possible loop closure. For each place, several WiFi scans and several images were recorded. For each of those positions, one recording was assumed to be inserted into the database created prior to localization. The remaining samples were used in a testing phase. More information about the datasets is presented in Table 1. Exemplary images from the datasets are presented in Fig. 4.

Tab. 1. Short description of recorded datasets

dataset name	PUT CM	PUT CW	SM
num. of positions	14	20	20
num. of records	140	100	100
avg. num. of WiFis in the scan	14.21	39.21	20.22
avg. RSSI of 5 strongest WiFis	-75.045	-41.642	-32.143
Building structure	corridors	open-space	shopping mall

3.2. Testing the Nature of WiFi Signal

The evaluation starts with an assessing the repeatability of WiFi scans. In a perfect environment with APs in line-of-sight, the measurement should be perfectly the same. In a cluttered environment with possible, multiple reflections and additional disturbances due to moving people, the scans information might be noisy. What is also essential to propose a distance function measuring the similarity of two scans is the probability distribution of measurements. This experiment consist of performing 1439 consecutive scans in a single spot using the Samsung Galaxy Note 3.

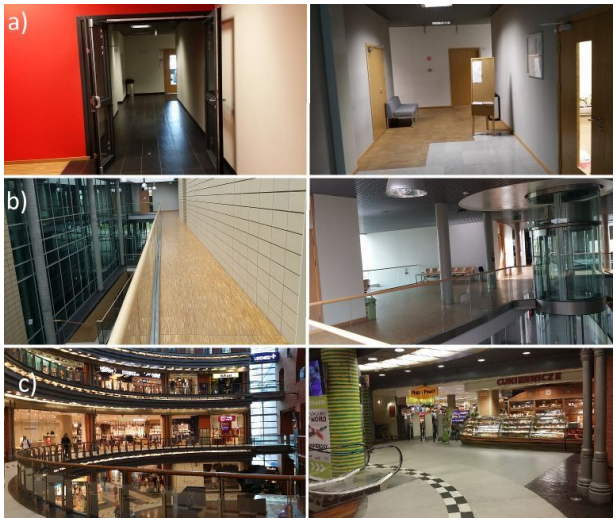

Fig. 4. Exemplary images presenting different building structures for PUT CM *(row a),* PUT CW *(row b) and* SM *(row c) datasets*

In the experiment the average RSSI signal is measured, while looking for the standard deviation of the measurement. Also the repeatability was measured to determine if there is a clear correspondence to the measured signal strength. The results of the experiment are presented in Tab. 2.

Tab. 2. WiFi signal floating example for 1439 measurements taken in a single spot

WiFi id	avg(RSSI)	std(RSSI)	Network detection percent
1	-49.12	5.51	100.00%
2	-74.24	3.05	100.00%
3	-74.57	2.99	100.00%
4	-83.49	3.42	56.12%
5	-83.99	1.83	94.02%
6	-84.15	2.69	82.82%
7	-86.08	3.33	94.65%
8	-86.64	1.78	94.65%
9	-87.45	1.21	95.76%
10	-87.57	1.47	67.39%

The presented results show that in most cases the stronger the signal, the higher is the standard deviation of these measurements. Moreover, with an exception for network 4, the stronger networks are detected with higher repeatability percentage and thus they are a good indicator if the user is in a vicinity of a previously stored WiFi scan. Also, in Fig. 5 the histogram of values for two WiFi networks with the greatest average signal strength is presented. Due to the cluttered environment, the achieved probability distributions are not Gaussian in all cases (like for WiFi with id=1). This observation indicates that when possible, it is better to rely more on the combination of detected networks than trust the measured signal strength, which can differ up to 20 dBm in a single spot.

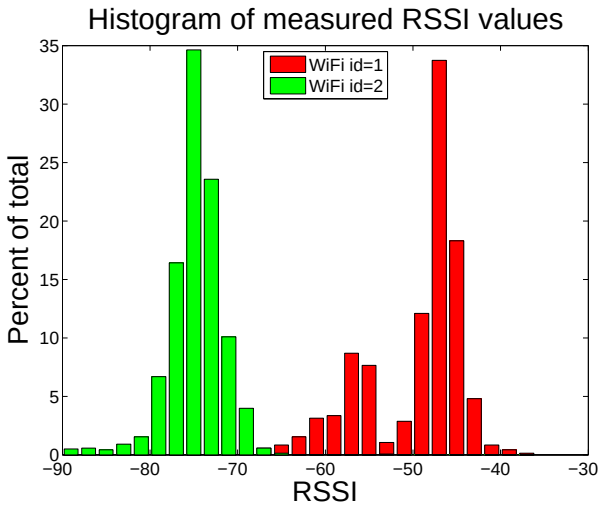

Fig. 5. Experimental distribution of RSSI for series of measurements in a single spot

3.3. Testing the Distance Functions used for WiFi Scans Comparison

The WiFi fingerprinting in the proposed approach is used to localize in the discrete set of positions. Therefore, the WiFi fingerprinting returns information about the most similar pose stored in the database or information about the unsuccessful match. Due to these assumptions, the comparison of WiFi scans using the standard Euclidean distance might not be the best choice as the combination of detected WiFi networks in most cases is sufficient to determine in which position the scan was performed. To determine the correctness of this statement, several definitions of distance/similarity functions are proposed and evaluated on the recorded datasets. For each position, one scan was treated as the database entry, whereas other scans were compared to all available database entries. The results are presented in Tab. 3.

Tab. 3. Comparing WiFi fingerprinting distance functions on the recorded datasets

Used function	PUT CM	PUT CW	SM
Simple similarity	72.86%	67%	75%
Euclidean norm	61.43%	53%	94%
Euclidean norm II	61.43%	53%	94%
Gaussian with $\sigma = 2$	82.14%	99%	94%
Gaussian with $\sigma = 3$	84.29%	99%	96%
Gaussian with $\sigma = 5$	86.43%	100%	97%
Gaussian with $\sigma = 10$	85.71%	100%	95%
Gaussian with $\sigma = 15$	83.57%	98%	89%
Gaussian with $\sigma = 20$	83.57%	96%	84%

The first tested function was a simple similarity function, which for both scans represents the number of WiFi networks that are detected in both scans. This function does not use the RSSI information, but has been chosen as the baseline approach, that can be a reference point for other approaches. This method achieved position recognition rates of 72.86%, 67%

and 75% for PUT CM, PUT CW and SM datasets respectively. The high recognition rates of this simple approach are believed to be task specific. In the performed tests, a sparse set of WiFi measurements is taken in locations that are separated by several meters. As the localization positions are not placed closely to each other, in many cases the combination of network names is sufficient to correctly determine the user's location. The second tested function is the Euclidean distance defined as in the state-of-the-art works [1]. Surprisingly, the Euclidean norm results in lower regonition rate for PUT CM and PUT CW datasets, which is in contrast to better recognition rate for SM. The author believes that those results are caused by different structures of the building. In case of PUT buildings, the APs are usually placed inside rooms and regardless of the corridor type, the WiFi information that reaches the mobile device was probably deflected several times. In case of the shopping mall, the open-spaces result in a WiFi signal propagating directly to the user, thus resulting in lesser number of deflections. Another tested functions was an Euclidean norm with an additional subtracted discount for each correctly matched network (called Euclidean norm II). This approach was based on an observation, that WiFi scans with higher number of matched networks are intuitively more likely to be the same. This modification didn't have any significant impact and resulted in values similar to the Euclidean distance approach. Due to the low recognition rate achieved with the Euclidean distance propositions, the simple similarity idea was expanded to incorporate RSSI values. For simplicity, the RSSI of the same networks are assumed to have Gaussian distribution. Then the similarity of networks found in two scans is defined by Gaussian membership values. The similarity between two scans $\mathcal{X}$ and $\mathcal{Y}$ is measured as a sum of Gaussian membership values for all networks available in both scans. Formally, it can be written as:

$$S_{\text{Gauss}}(\mathcal{X},\mathcal{Y},\sigma)=\sum_{i=1}^{N}\exp\{-\frac{(\mathcal{X}_i-\mathcal{Y}_i)^2}{-2\sigma^2}\}, \qquad (3)$$

where, N is the number of common networks found in both $\mathcal{X}$ and $\mathcal{Y}$ scans. The σ is the standard deviation of the measurement used to define the shape of Gaussian membership function. The choice of σ is arbitrary, but from the experiment measuring the WiFi scans in a single spot, it was assumed that best results should be achieved for a value in the range of 2 to 7. To confirm this assumption, different σ values have been chosen. As expected, the best results were obtained for σ equal to 5. The results using modified similarity values turned up to be better when compared to previous approaches. For PUT CM the recognition rate increased to 86.43%, for PUT CW to 100%, for SM to 97%. In case of PUT CW, the WiFi information is sufficient to precisely localize the mobile device. In the remaining cases, the usage of image information may be useful in finding loop closures in scenarios, where WiFi matching failed.

3.4. Testing the Vision-based Loop Closure

The next tests concern the recognition rate of the vision-based loop closure subsystem. Similarly to the WiFi evaluation, for each distinct position one image was chosen as a reference. The remaining images were then compared against all of the images in the database in order to find a positive match.

The images taken with the Samsung Galaxy Note 3 have a maximum resolution of 1920×1080 pixels (Full HD). Due to the mobile platform processing power, the resolution of 640×480 pixels (VGA) is chosen as the image of reduced size have 7 times less pixels to process. This results in obvious processing speed up. A detailed comparison with VGA and FullHD images is presented in Tab. 4.

The most time consuming part of any system using the SURF detector/descriptor is the detection of keypoints that takes almost 1s on the Samung Galaxy Note 3. The obvious reduction of needed time can be achieved by lowering the number of keypoints used by system and thus described by the descriptor. Unfortunately, the minimal number of keypoints needed to achieve a robust system is application dependent and in the proposed tests 500 strongest keypoints were chosen. Another time reduction strategy is to use different detector/descriptor pairs [15], but tests concerning the choice of detector/descriptor pairs are not a part of presented research.

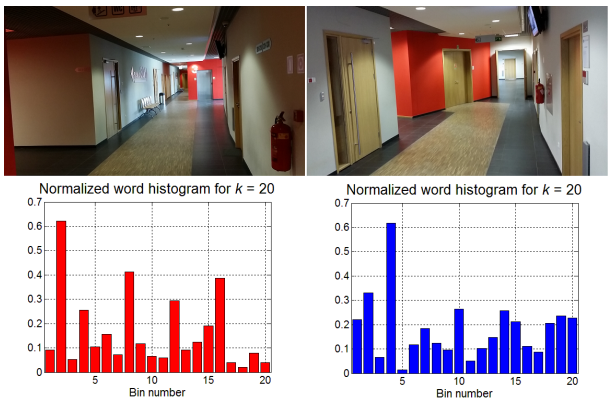

Fig. 6. Similar corridors with corresponding image words observed in two distant localization positions in PUT CM dataset

When the detection and description parts are finished, a k-dimensional word creation is initiated. The process start with the classification of descriptors of keypoints found in the image. The descriptors are classified into clusters, for which the minimal error to the centroid is obtained. The centroids are computed prior to the localization. The centroids are found by performing a k-means algorithm computation on a dataset consisting of every descriptor found in all reference images. After the classification, each image is described by a histogram of length k with number of descriptors classified into each cluster. The construction of the image word finishes with a normalization procedure of the histogram. The exemplary computed words for images from PUT CM dataset are presented

Tab. 4. Processing time of the proposed visual loop closure subsystem

System part	S. Galaxy Note 3 VGA	S. Galaxy Note 3 Full HD	Nexus 7 VGA	Nexus 7 Full HD
Image resizing	18.86 ms	-	34.0 ms	-
Keypoints detection	460.53 ms	2376.32 ms	443.57 ms	2516.21 ms
Keypoints description	486.73 ms	720.86 ms	431.21 ms	648.57 ms
Word creation (K=5)	30.71 ms	30.14 ms	34.00 ms	33.64 ms
Word creation (K=20)	127.21 ms	126.78 ms	116.79 ms	116.07 ms
Word creation (K=50)	351.21 ms	349.43 ms	374.00 ms	300.43 ms
Word creation (K=200)	1417.36 ms	1378.71 ms	1303.36 ms	1133.71 ms
Estimated total time per word creation (K=20)	1093 ms	3224 ms	1026 ms	3281 ms

in Fig. 6. After the normalization, k-dimensional word for an image is successfully computed, the subsystem determines the correct match to the reference frames stored in the database by comparing current image to all entries in the database. In the proposed subsystem the comparison of words is done using the Euclidean distance. If the smallest distance between matches is higher than a preset threshold, the match is considered to be correct.

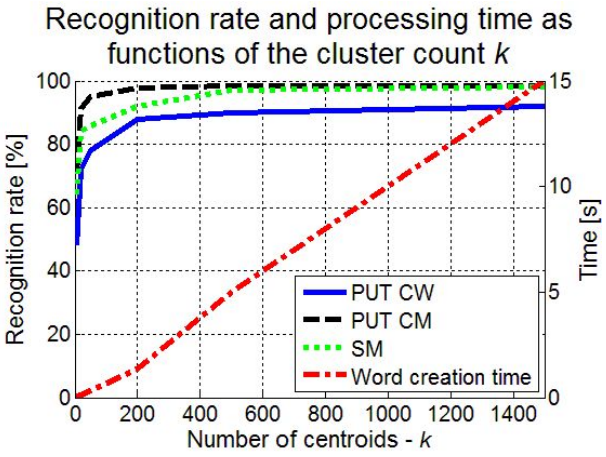

Fig. 7. The recognition rate and time taken for word creation for different number of centroids evaluated at PUT CW

To determine, what number of classes k used by the k-means algorithm results in the highest recognition rate, different number of k values were evaluated. The results are presented in Fig. 7. For the proposed datasets, higher number of classes for the k-means algorithm results in higher recognition ratee. But, in the presented vision-based loop closure approach, each descriptor is assigned to one of k clusters. If the k value is higher, the total time needed to classify descriptors is higher. Therefore, it is necessary to find a k value that results in high recognition rate within reasonable time. From Fig. 7, the value of k equal to 200 is chosen as the best choice and used in described subsystem.

The results obtained by the proposed approach are also presented in Tab. 5. The visual loop closure has the highest recognition rate of visited places in case of PUT CM dataset, which is equal to 97.86% for chosen k equal to 200. The small number of distinct objects poses a great challenge for the visual system as the detected features are in most cases similar for all of the positions in the sequence. The problem of similar places also arises for PUT CW, for which the lowest performance is achieved. The recognition rate of 92% for the SM dataset in most cases is a result of situations, when passing pedestrians are present in a significant part of the image and thus make images from training and testing sets look different.

Tab. 5. Accuracy of the proposed visual loop closure approach

	PUT CM	PUT CW	SM
k = 5	70.71%	48%	64%
k = 20	91.43%	72%	84%
k = 50	95%	78%	86%
k = 200	97.86%	88%	92%
k = 500	98.57%	90%	97%
k = 1500	98.57%	92%	98%

3.5. Results – Testing WiFi Guided Vision Loop Closure

The system that combines information from both subsystems is expected to outperform either of them. In Tab. 6 the best results obtained from both subsystems are presented. The comparison shows, that WiFi fingerprinting provides more reliable estimate for the PUT CW or SM datasets, whereas visual loop closure works better for the PUT CM dataset. In case of application tailored for a specific building, the system designer may decide to use only one source of information. If the single-source solution is inefficient, there exists a need for a system integrating data from both subsystems.

In case of an unknown building structure, it is essential to correctly weight information from both subsystems. In the presented research, three methods are proposed and tested:

1) method I – rank-based,
2) method II – normalize and sum,
3) method III – normalize and multiply.

Method I, called rank-based, for each position to evaluate assigns the ranks based on the similarity of WiFi scans and distance functions for vision-based loop closure to positions stored in the database. For

each position to classify, the most probable estimates from both subsystems are provided. Then, separately for each subsystem, the most probable estimate is assigned rank 1, the second most probable is assigned rank 2 and so forth. At this point of processing, each position to process contains two ranks representing the estimates from both subsystems. Then, for each reference position, a summation of assigned ranks is performed. The position in the database with a lowest sum of ranks is chosen as a combined system estimate.

Method II, normalize and sum, tries to incorporate also information about the distances between position in the estimates of separate subsystems. To include this information into the proposed system, subsystems estimates must have a similar range of values. Therefore, for each position a vector of distances to all database entries is created and then normalized in L2 norm. The normalized vector of estimates for WiFi fingerprinting is denoted by w. The equivalent, normalized vector for vision loop closure is denoted by v. As the WiFi subsystem operates using similarities between classes, whereas vision-based loop closure uses distances, the final estimation is computed as difference of estimates ($w - v$). The finally inferred position is based on finding an index of maximal element in a $w - v$ vector:

$$\text{estimatedID}_{\text{II}} = \underset{i}{\text{argmax}}\{w(i) - v(i)\}, \qquad (4)$$

Method III, normalize and multiple, uses a similar strategy to previously presented method II. In this case, the distances from vision loop closure are recomputed to represent similarities by exchanging each value x in a vector v with $1 - x$. The resulting vector is again normalized. The best position estimated by the integrated system corresponds to an index of maximal value after elementwise multiplication of vectors v and w:

$$\text{estimatedID}_{\text{III}} = \underset{i}{\text{argmax}}\{(1 - v(i)). * w(i)\}, \qquad (5)$$

The proposed system is evaluated in the same way as shown for the subsystems. The results are presented in Tab. 6. It is shown that when concerned about the PUT CM building, all of different functions for the system using WiFi data performed worse than vision-based loop closure. In other cases, the proposed system performs the same or better than either of the subsystems. The best results are obtained for method II and it is the recommended method if the structure of the building is unknown or if the system must operate in changing conditions regarding the uniqueness of images and number of available WiFi networks. In case of a system created for specific building it is recommended to record a training and testing set and perform experiments to correctly weight the input of each subsystem. In some cases it might be also necessary to detect same positions based solely on WiFi, whereas in other completely rely on gathered images. These mentioned remarks are application-specific and cannot be applied to universal system. In case of the proposed system operating in three different buildings, the recognition rate was equal or greater than 90% in each, tested case without usage of additional subsystem weights.

Tab. 6. Localization recognition rate of subsystems and different approaches to the system combining information from WiFi and Vision subsystems

	PUT CM	PUT CW	SM
WiFi fingerprinting	86.43%	100%	97%
Visual loop closure	97.86%	88%	92%
Method I (rank-based)	92.14%	97%	97%
Method II (sum)	90%	100%	98%
Method III (product)	88.57%	100%	98%

4. Conclusion

The proposed event-based, WiFi-guided visual loop closure approach presents a new approach to data integration of mobile platforms' sensor information that results in a system than outperforms each individual approach. The information from camera usually helps in localization in areas with small number of WiFis, e.g., corridors or staircases. What is surprising, the system performed well in the case of corridors that seemed alike. The system works with lesser recognition rate in case of a shopping mall, where sudden pedestrian's occlusions negatively affect the visual localization. Moreover, the achieved results suggest that WiFi and vision information complement each other and provide a data needed to create a more robust localization system.

Contrary to proposed event-based localization, the further works will focus on providing a continuous estimate at the user by estimating the motion through the vision-based monocular visual odometry with additional incorporation of WiFi information.

ACKNOWLEDGEMENTS

The author would like to thank Piotr Skrzypczyński for numerous discussions regarding the presented localization system.

This work is financed by the Polish Ministry of Science and Higher Education in years 2013-2015 under the grant DI2012 004142.

AUTHOR

Michał Nowicki* – Institute of Control and Information Engineering, Poznań University of Technology, Poznań, Poland, e-mail: michal.nowicki@cie.put.poznan.pl.

*Corresponding author

REFERENCES

[1] P. Bahl, V. N. Padmanabhan, "RADAR: An In-Building RF-Based User Location and Tracking System." In: *19^{th} Annual Joint Conf. of the IEEE*

Computer and Communications Societies (INFOCOM), 2000, pp. 775–784. DOI: http://dx.doi.org/10.1109/INFCOM.2000.832252.

[2] H. Bay, A. Ess, T. Tuytelaars, L. Van Gool, "SURF: Speeded up robust features", *Comp. Vis. and Image Underst.*, vol. 110, no. 3, 2008, pp. 346–359. DOI: http://dx.doi.org/10.1016/j.cviu.2007.09.014.

[3] J. Biswas, M. Veloso, "WiFi localization and navigation for autonomous indoor mobile robots." In: *10 IEEE Int. Conf. on Robotics and Automation (ICRA)*, 20, 2010, pp. 4379–4384. DOI: http://dx.doi.org/10.1109/ROBOT.2010.5509842.

[4] S. Boonsriwai, A. Apavatjrut, "Indoor WIFI localization on mobile devices." In: *2013 10th Int. Conf. on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, 2013, pp. 1–5. DOI: http://dx.doi.org/10.1109/ECTICon.2013.6559592.

[5] G. Bradski, "The OpenCV library", Dr. Dobb's Journal of Software Tools, opencv.org, 2000.

[6] G. Csurka et al., "Visual categorization with bags of keypoints." In: *Workshop on Statistical Learning in Computer Vision*, ECCV, 2004, pp. 1–22.

[7] N. Dalal, B. Triggs, "Histograms of oriented gradients for human detection." In: *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2005, vol. 1, pp. 886–893. DOI: 10.1109/CVPR.2005.177.

[8] T. Gallagher et al., "Indoor positioning system based on sensor fusion for the Blind and Visually Impaired." In: *Int. Conf. Inndoor Positioning and Indoor Navigation (IPIN)*, 2012, pp. 1–9. DOI: 10.1109/IPIN.2012.6418882.

[9] A. Glover et al., "OpenFABMAP: An Open Source Toolbox for Appearance-based Loop Closure Detection." In: IEEE Int. Conf. on Robotics and Automation, St Paul, Minnesota, 2011.

[10] J. Gośliński, M. Nowicki, "Performance Comparison of EKF-based Algorithms for Orientation Estimation on Android Platform."

[11] M. Holĉik, "Indoor Navigation for Android," M.S. thesis, Faculy of Informatics, Masaryk Univ., Brno, 2012.

[12] H. Liu et al., "Accurate WiFi Based Localization for Smartphones Using Peer Assistance," *IEEE Transactions on Mobile Computing*, vol. PP, no. 99, 2013, pp. 1. DOI: 10.1109/TMC.2013.140.

[13] K. Muzzammil bin Saipullah, A. Anuar, N. A. binti Ismail, Y. Soo, "Real-time video processing using native programming on Android platform." In: *Proc. IEEE 8th Int. Col. on Signal Proc. and its App.*, 2012, pp. 276–281.

[14] M. Nowicki, P. Skrzypczyński, "Combining photometric and depth data for lightweight and robust visual odometry." In: *European Conference on Mobile Robots (ECMR)*, 2013, pp. 125–130. DOI: 10.1109/ECMR.2013.6698831.

[15] M. Nowicki, P. Skrzypczynski, "Performance Comparison of Point Feature Detectors and Descriptors for Visual Navigation on Android Platform," *Int. Wireless Communications and Mobile Computing Conference (IWCMC)*, 2014.

[16] K. Pulli et al., "Real-time Computer Vision with OpenCV," Commun. ACM, 2012, vol. 55, no. 6, pp. 61–69. DOI: 0.1145/2184319.2184337.

[17] N. Ravi, P. Shankar, A. Frankel, A. Elgammal, L. Iftode, "Indoor localization using camera phones," Proc. 7th IEEE Work. on Mobile Comp. Sys. and App., 2006, pp. 1–7.

[18] U. Shala, A. Rodriguez, "Indoor Positioning using Sensor-fusion in Android Devices," M.S. thesis, Dept. Computer Science, Kristianstad Univ., Kristianstad, 2011. http://hkr.diva-portal.org/smash/record.jsf?pid=diva2:475619

[19] M. Quigley, D. Stavens, A. Coates, S. Thrun, "Sub-meter indoor localization in unmodified environments with inexpensive sensors." *Proc. IEEE/RSJ Int. Conf. on IROS*, Taipei, 2010, pp. 2039–2046. DOI: 10.1109/IROS.2010.5651783.