# An Efficient Indoor Localization Method Based on Visual Vocabulary

Ruolin Guo, Danyang Qin(✉), Min Zhao, and Guangchao Xu

Key Lab of Electronic and Communication Engineering,
Heilongjiang University, Harbin 150080, People's Republic of China
qindanyang@hlju.edu.cn

**Abstract.** This paper proposes a new efficient indoor localization method based on visual vocabulary. The special feature of this method is that no additional components are needed, but only mobile devices equipped with cameras. By matching the query image with a visual vocabulary constructed by a Bag of Self-Optimized-Ordered Visual Vocabulary (BoSOV), the user's position can be accurately determined. In addition, the efficiency of our scheme is compared with that of other schemes, and simulation results reveal that our method has higher indoor positioning efficiency, especially when the amount of image data is large. Simulation results show that our method can well achieve efficient visual indoor positioning when the data volume is relatively large.

**Keywords:** Self-optimization · Visual vocabulary · Feature selection · AP clustering · Indoor localization

## 1 Introduction

Nowadays, there is a growing demand for Location-based Services (LBS) in smartphones, which are provided by external positioning methods such as the Global Positioning System (GPS) or radio communication networks such as the GSM network. Therefore, most LBSs are only suitable for outdoor environments, and the research on indoor localization has received wide attention.

The typical indoor positioning solutions mainly use Wi-Fi technology [1], which performs positioning tasks in indoor environments through Wireless Local Area Networks (WLAN), but it depends on the location of wireless access points. If not taking surrounding Wi-Fi signal into account, errors such as floor positioning wrong are likely to occur. Moreover, the coverage of the wireless access point is limited, the signal is also unstable and it is susceptible to interference. Another solution is to apply Simultaneous Localization and Mapping (SLAM) [2] to locate in an indoor environment. Its process can be described as the machine moving in the environment to be

This work is supported by the National Natural Science Foundation of China (61771186), University Nursing Program for Young Scholars with Creative Talents in Heilongjiang Province (UNPYSCT-2017125), Distinguished Young Scholars Fund of Heilongjiang University, and postdoctoral Research Foundation of Heilongjiang Province (LBH-Q15121).


S. Han et al. (Eds.): AICON 2019, LNICST 286, pp. 398–406, 2019.
https://doi.org/10.1007/978-3-030-22968-9_36

located, and creating a map according to the data recorded by the sensors. Its shortcoming is that it requires complex external facilities to record the relative position of the machine and the ground, and the requirements for the landmarks are very high, but it is difficult to extract the landmarks that meet the requirements by the camera of the mobile phone alone.

This paper proposes a novel vision-based indoor localization scheme that constructs a visual vocabulary through a visual vocabulary bag called BoSOV (Bag of Self-Optimized-Ordered Visual Vocabulary). The difference in this scheme is that the resulting visual vocabulary does not have to be recorded by an additional machines, it has location information itself and can automatically optimize the constructed vocabulary. The method we proposed only requires a smartphone with a camera, which is more cost effective than the previous solution. Moreover, our method is more efficient and accurate in indoor positioning than in Wi-Fi based or SLAM based methods.

The rest of the paper will be organized as follows. Some of the main processes of our proposed indoor positioning solution will be introduced in Sect. 2. The three main processes of our program in the implementation phase will be described in Sect. 3. Section 4 will compare our plan with other programs and evaluate our plan and the full text will be summarized in Sect. 5.

## 2 Methods

This section introduces some main processing procedures of our proposed indoor localization scheme, which are image processing and clustering methods. The most important step in the image processing is introduced, which is the feature selection and extraction of images. We are also introduced two clustering methods.

### 2.1 Image Processing

At present, in image processing, feature selection [3] and extraction of the target is a more appropriate method. Because under the BoVW framework, there are many features that are useless for image recognition, so the selection of features with statistical significance from a large number of original features can reduce the amount of calculation and improve the efficiency of image recognition.

One popular algorithm is SURF [4], which consists of three steps: feature point selection, feature point description, and descriptor pairing. Since the selected feature points have the property of rotation invariance in the process of image recognition, a descriptor is generally given to the feature points in order to maintain this property and make them easy to be distinguished. SURF algorithm has the problem of too much computation, because all feature points extracted by the algorithm need to be stored to describe the image. In addition, when descriptors are paired, a large amount of computation will be generated, and there is also a great demand for storage space.

BoVW (Bag of Visual Word), however, can just solve this problem. It only needs to retain variable descriptors and can represent images with fixed length feature vectors, thus reducing the storage pressure. Therefore, the process of our scheme is to extract the SURF descriptor from the original feature set under the BoVW framework, then

cluster and quantize it to generate visual vocabulary, thus forming a fixed-length visual vocabulary for the vocabulary. The number of occurrences of each word in the statistic is counted to represent the image as a numerical vector.

## 2.2 Clustering Method

For the image representation process, the most important step is to cluster the SURF descriptors into clusters. The commonly used clustering method is to use the K-means algorithm [5], which uses the idea of iteration to aggregate the descriptors into their own specifications. The K cluster makes all the data in the cluster have higher similarity, and the similarity between the cluster and the cluster is low.

Initially, $K$ values are randomly selected as the center of the cluster, the distance from each descriptor to the $K$ centers is calculated, and it is divided into the clusters where the nearest center is located. The iteration is repeated until the center value is unchanged or reached the maximum number of iterations, the clustering process is shown in Fig. 1. Eventually, each graph becomes a numerical vector corresponding to the visual vocabulary. But one of the biggest problems with the K-means algorithm is that the $K$ value is difficult to estimate. Since the number of clusters needs to be specified in advance, the user does not know at first that these descriptors should be classified into several categories.

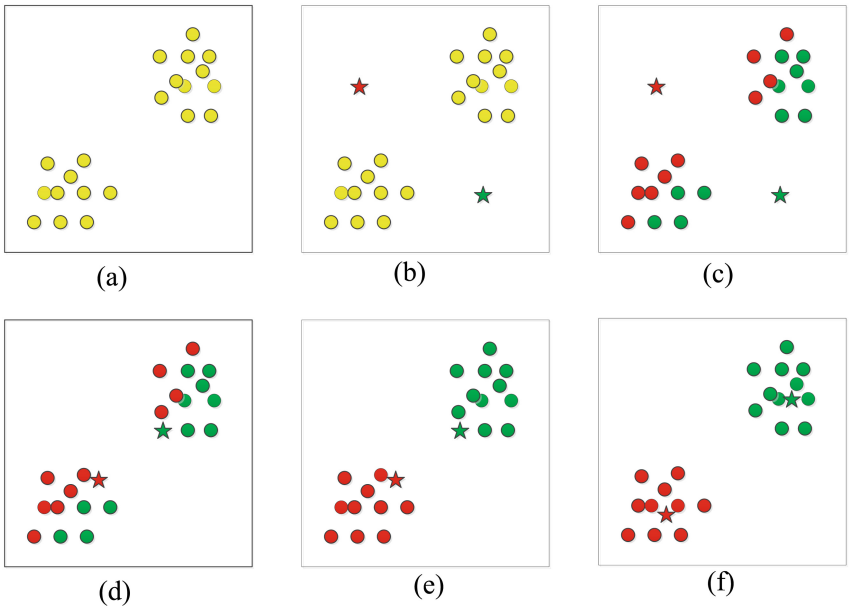


**Fig. 1.** K-means clustering process

The clustering method used in our scheme is affinity propagation (AP) clustering [6]. Its advantage is that it does not need to determine the number of clustering clusters

by itself like the K-means algorithm, nor does it need to take newly generated data as the center of the cluster. It can be selected from existing data points, and the result has small squared error. The basic idea of AP clustering [7] is to treat all descriptors as nodes of a network, and each cluster center is calculated by the information transmitted by each edge of the network. In the clustering process, the attribution degree and the attraction degree are transmitted as the main information between the nodes, and the two values are iteratively updated until a high-quality cluster center is generated. Then the remaining descriptors are allocated to the corresponding clusters.

## 3 Proposed BoVW and Localization

In the execution phase, three main processes in our scheme are introduced: feature selection and extraction, clustering and visual vocabulary generation and matching of reference images.

### 3.1 Feature Selection and Extraction

SURF algorithm is adopted for feature extraction with some preparatory work before extraction. The original SURF feature is generally 64-dimensional or 128-dimensional. We choose 64-dimensional and then insert two elements describing relative spatial information, which becomes a descriptor containing 66 elements. We also need to unify the standard for extracting SURF features, and set the resolution of the image to 1200 m $\times$ 1600 m. In addition, remove the feature points whose brightness is higher than the threshold, because these points are generally derived from light and are not useful for image recognition. The descriptor is as Eq. (1):

$$V = (v_0, v_1, \ldots, v_n, q(x)q(y)) \tag{1}$$

where $q(x)$ and $q(y)$ are quantization functions.

### 3.2 Clustering and Visual Vocabulary Generation

Affinity propagation (AP) clustering is applied to cluster SURF descriptors into clusters. The algorithm flow is as follows:

**Attraction Information Update.** The attraction information of the similar matrix should be updated first, as shown in Fig. 2(a). The updating process can be written as Eq. (2).

$$r(i,k) \leftarrow s(i,k) - \max\{a(i,k') + s(i,k')\} \tag{2}$$

where $a(i,k')$ indicates the attribution value of points other than $k$ for point $i$; $s(i,k')$ represents the attraction of other points except $k$ to $i$; $s(i,k)$ is a similarity matrix (Euclidean distance) between the descriptors $i$ and $k$; and $r(i,k)$ represents the extent to which the descriptor $k$ is suitable as the clustering center for the descriptor $i$, and describes the message from $i$ to $k$.

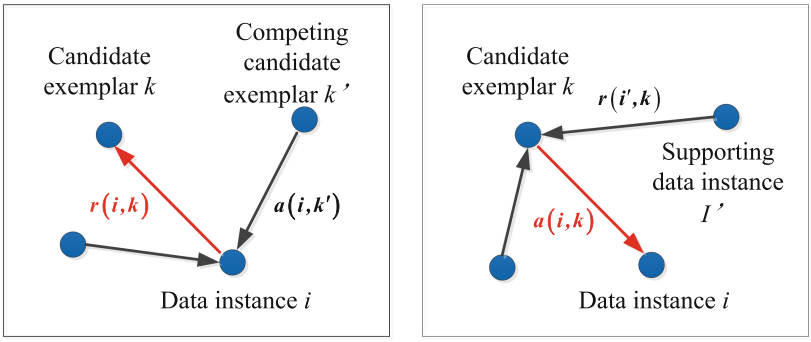


**Fig. 2.** Updating process for (a) the attraction information and (b) attribution information

**Attribution Information Update.** After the attraction information of the similar matrix updated, it turns to attribution information to perform the similar process, as in Fig. 2(b). Such process can be modelled by Eq. (3):

$$\begin{cases} a(i,k) \leftarrow \min\left\{0, r(k,k) + \sum_{i' \subsetneq \{i,k\}} \max[0, r(i',k)]\right\}, & i \neq k \\ a(k,k) \leftarrow \sum_{i' \subsetneq \{i,k\}} \max[0, r(i',k)], & i = k \end{cases} \tag{3}$$

where $r(i',k)$ represents the similarity value of point $k$ as the clustering center of other points except $i$; $a(i,k)$ indicates the degree to which descriptor $i$ selects descriptor $k$ as its clustering center, and describes the messages from $k$ to $i$.

**Summation and Detection.** Now it should sum the attraction information [8, 9] and the attribution [10] information. Moreover, the selected cluster center should be detected. If the cluster center remains unchanged or reaches the maximum number of iterations after several iterations, the algorithm ends. The overall process of Affinity Propagation (AP) clustering is shown in Fig. 3.

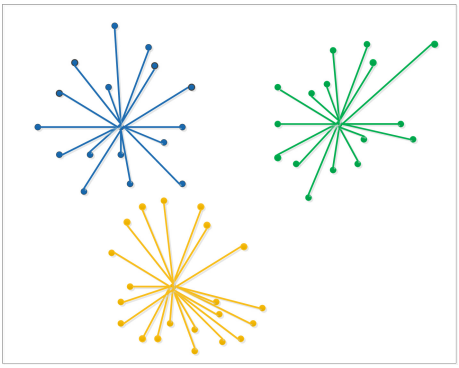

**Fig. 3.** Affinity propagation clustering process

Suppose there are $n$ segment $\{Seg_1, Seg_2, \ldots, Seg_n\}$s, for the $k$th fragment, it contains $m$ images as $\{Img_1, Img_2, \ldots, Img_m\}$. For the $i$th image, a total of $p(i)$ individual feature descriptors are extracted as $\{D_{i1}, D_{i2}, \ldots, D_{ip(i)}\}$. Then, the affinity propagation of the $k$th segment is clustered using the feature descriptors as the combination of $\{(D_{11}, \ldots, D_{1p(1)})(D_{21}, \ldots, D_{2p(2)}) \ldots (D_{n1}, \ldots, D_{np(n)})\}$.

After clustering all the elements into $m$ clusters, a visual vocabulary can be generated from each cluster as $\xi$. Assume that all $p$ descriptors are assigned to set $i(i \leq n)$ and are represented as $\boldsymbol{x} = (x_1, x_2, \ldots, x_p)$. Then perform 1-mean clustering on $x$ to get the corresponding visual word. We find the visual vocabulary $W$ such as the follows Eq. (4):

$$\xi = arg\ \min_W \sum_{i=1}^{p} \|W - x_i\| \tag{4}$$

After applying 1-means clustering in the clustering of each segment, all the visual words sorted by segment number can be added to construct an intermediate visual vocabulary. The segment and visual vocabulary with high mutual information content constitute the final visual vocabulary. The process is shown in Fig. 4. For $i$th segment, there have $m(i)$ visual vocabularies.

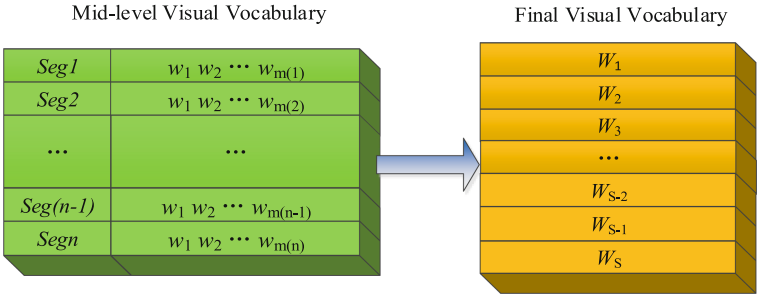


**Fig. 4.** Generation of the visual vocabulary

### 3.3 Reference Image Matching

After matching the query image with all the segments and finding the segment $S$ that is most similar to it, we then searched all the images in it to find the image of the visual vector $T_{query}$ closest to the $T$ distance. Suppose the $S$ segment image is $\{Img_1, Img_2, \ldots, Img_m\}$ and its corresponding visual vector is $\{T_1, T_2, \ldots, T_m\}$, the goal is to find $Img_i$ for any $i$:

$$i = arg\ \min_{k \in \{1,2,\ldots,m\}} \left\| T_{query} - T_k \right\| \tag{5}$$

## 4 Performance Evaluation and Analysis

This section studies and analyzes the efficiency at the construction stage and the efficiency with or without feature selection respectively, and compares our scheme with the scheme using K-means clustering.

### 4.1 Image Matching Efficiency

The higher efficiency and accuracy is the main focus in the research, in which the large amount of image data matching is the most important foundation. Simulations are performed on the indoor image mating efficiency with large amount of images being sampled at offline phase. Time consumptions are evaluated and recorded cost by comparing the image database and the collecting images in online phase. The comparison results between the proposed algorithm and the typical K-means are shown in Fig. 5.

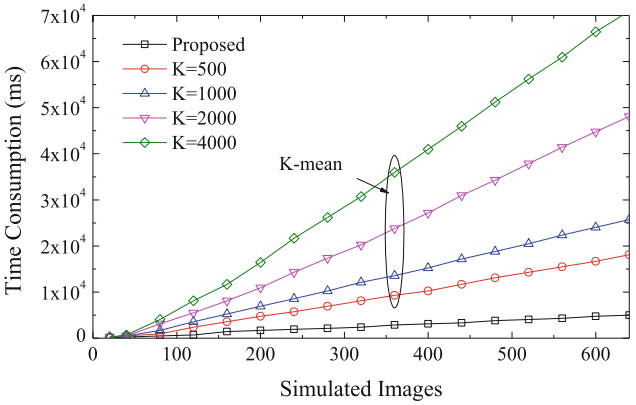


**Fig. 5.** Comparison of matching efficiency in the constructing phase

From Fig. 5, we can clearly see that no matter what the size of image data is, the scheme using K-means clustering consumes much more time in the execution stage than our scheme. When the number of images is 640, the time consumption of our scheme is 5.059 s, while when there are 640 collecting images, the time required by K-means is 19.778 s at least, which is almost four times more than that of the BoVW.

To achieve high execution efficiency will make the practical application available, especially when the image data volume is large. The relationship between the data volume and the execution speed should be analyzed deeply. We take $k = 2000$ as an example to study. When the number of images is 160, it will the K-mean clustering scheme 7.927 times as long as the proposed scheme. And when the collecting images increase to 640, it will take K-mean clustering scheme 9.469 times more duration as long as the proposed scheme.

### 4.2 Performance of Feature Selection

In the above introduction, we mentioned that under the BoVW framework, nearly half of the features that are useless for image recognition increase the computational load of execution. Therefore, selecting the features with statistical significance from the original features can reduce the computational load and improve the efficiency of image recognition. Moreover, feature selection is time-consuming and efficient. Figure 6 shows the time consumption of feature selection and non-feature selection in the proposed scheme, and is compared with the scheme using K-means clustering.

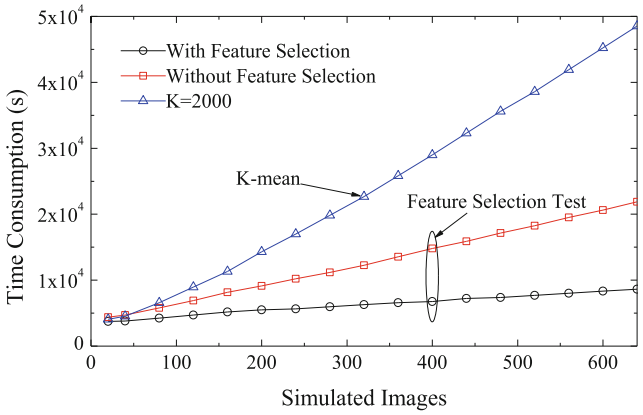


**Fig. 6.** Comparison of efficiency with/without feature selection

It can be seen from Fig. 6 that no matter how large the image data volume is, without feature selection always takes longer than use feature selection, and it becomes more and more obvious with the increase of data volume. In comparison with the K-means clustering scheme, we find that the efficiency of our scheme is much higher than that of the K-means clustering scheme, no matter whether the feature selection is adopted or not.

## 5 Conclusions

In this paper, a new high-precision indoor positioning method based on affinity clustering is proposed, which is compared with the K-means clustering method. According to the performance analysis, our method has higher indoor positioning efficiency, especially when the amount of image data is large. When the amount of image data is greater than 40, the query time consumption of K-means clustering scheme is greater than 4 times that of our method, and the proportion is inversely proportional to the increase of the size of image database. Moreover, our method does not require additional components to record the location, but only requires a smartphone with a camera, which is more cost-effective than previous solutions.

## References

1. Xiao, C., Zou, S.: Improved Wi-Fi indoor positioning based on particle swarm optimization. IEEE Sens. J. **99**, 1–10 (2017)
2. Wang, X., Zhang, C., Liu, F.: Exponentially weighted particle filter for simultaneous localization and mapping based on magnetic field measurements. IEEE Trans. Instrum. Meas. **66**(7), 1658–1667 (2017)
3. Li, J., Cheng, K., Wang, S.: Feature selection: a data perspective. ACM Comput. Surv. **50** (6), 89–99 (2016)
4. Pan, J., Hao, J., Zhao, J.: Improve algorithm based on SURF for image registration. Remote Sens. Land Resour. **40**(6), 60–74 (2017)
5. Dalmiya, S., Dasgupta, A., Kanti, Datta S.: Application of wavelet based K-means algorithm in mammogram segmentation. Int. J. Comput. Appl. **52**(15), 15–19 (2016)
6. He, S., Lin, W., Chan, S.H.G.: Indoor localization and automatic fingerprint update with altered AP signals. IEEE Trans. Mob. Comput. **16**(7), 1897–1910 (2017)
7. Wei, Z., Wang, Y., He, S.: A novel intelligent method for bearing fault diagnosis based on affinity propagation clustering and adaptive feature selection. Knowl.-Based Syst. **116**(1), 1–12 (2017)
8. Jiang, J., Huang, J., Wang, X.R.: Investigating key genes associated with ovarian cancer by integrating affinity propagation clustering and mutual information network analysis. Eur. Rev. Med. Pharmacol. Sci. **20**(12), 2532–2540 (2016)
9. Sun, L., Guo, C., Liu, C.: Fast affinity propagation clustering based on incomplete similarity matrix. Knowl. Inf. Syst. **51**(3), 1–23 (2016)
10. Chen, Q.S., Dan, W., Liu, B.L.: Combining affinity propagation clustering and mutual information network to investigate key genes in fibroid. Exp. Ther. Med. **14**(1), 251–259 (2017)