A Robust Place Recognition Algorithm Based on Omnidirectional Vision for Mobile Robots

Huimin Lu, Kaihong Huang, Dan Xiong, Xun Li, and Zhiqiang Zheng

College of Mechatronics Engineering and Automation,
National University of Defense Technology, Changsha, China
{lhmnew,lixun,zqzheng}@nudt.edu.cn

Abstract. In this paper, bag-of-features, a popular and successful approach in pattern recognition community, is used to realize place recognition based on omnidirectional vision for mobile robots by combining the real-time local visual features proposed by ourselves for omnidirectional vision and support vector machines. The panoramic images from the COLD database were used to perform experiments to determine the best algorithm parameters and the best training condition. The experimental results show that the robot can realize robust place recognition with high classification rate in real-time by using our algorithm.

1 Introduction

In recent years, along with the development of image understanding and pattern recognition, visual place/scene recognition has attracted more and more researchers' interest, and many progresses have been achieved [1][2]. Place recognition can be applied to realize robot topological localization. If the nodes of topological maps are represented by the places like kitchen, corridor, and bathroom, once these places are recognized and classified by robots, topological localization is also realized for the robots. Besides topological localization, place recognition is also important for solving the loop closing in visual odometry, visual SLAM and the kidnapping problem in robot localization.

Pronobis, Caputo, Luo, et al. proposed a robust place recognition algorithm based on SVMs classifier, combined with local visual features computed using a Harris-Laplace detector and the SIFT descriptor in [3]. Because the number of the local visual features in an image is not fixed, the local descriptors are used as the input of SVMs via a match kernel. Then the classifiers can be trained for place classification and recognition. The local visual features are used as the input of SVMs directly, so large memory space is needed to store those features used as support vectors. Therefore, they proposed a memory-controlled incremental SVMs by combining an incremental extension of SVMs with a method reducing the number of support vectors needed to build the decision function without any loss in performance introducing a parameter which permits a user-set trade-off between performance and memory in [2]. They also built up several image databases to provide standard benchmark datasets for the development,

X. Chen et al. (Eds.): RoboCup 2012, LNAI 7500, pp. 286–297, 2013.

evaluation and comparison of different place recognition algorithms: the INDECS [1], the IDOL [1] and the COLD [4] database. All the images in these databases were acquired in indoor environments and with different conditions like different robot platforms, different lighting conditions, and different labs across Europe. Although the COLD database includes panoramic images acquired by the omnidirectional vision system, only perspective images were used to perform experiments to test their place recognition methods in [2][3]. The omnidirectional vision system can provide a 360° view of the robot's surrounding environment in a single image, and it is especially suitable to be a sensor of navigation for mobile robots in large scale environment.

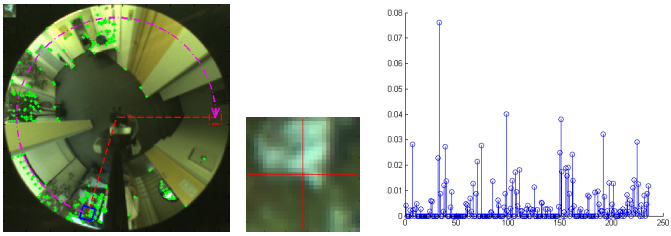
In this paper, bag-of-features [5][6], a popular and successful approach in pattern recognition community, is used to realize robust place recognition based on omnidirectional vision for mobile robots by combining two novel real-time local visual features [7] proposed by ourselves and support vector machines (SVMs) [8]. Some researchers used or extended the bag-of-features method to realize qualitative localization [9], global localization [10], or topological SLAM [11], which are the most similar research with our work in this paper. Only perspective images were used in their work. Furthermore, the local visual features used in their work can not be extracted in real-time, so their algorithms can not be run in real-time actually.

2 Two Real-Time Local Visual Features

Local visual features have become increasingly popular in recent years, and they have been applied very well in many computer/robot vision problems. Although a number of algorithms have been proposed with respect to feature detectors and feature descriptors, a common deficiency for most of the existing algorithms is that their computation costs are usually high. This deficiency limits the actual application of local visual features, especially in those situations with high real-time requirements, such as robot navigation, self-localization. When local visual features are applied to omnidirectional vision, the original algorithms should be modified because of its special imaging character, especially in determining the feature regions [12].

To deal with these problems, we proposed two novel real-time local visual features for omnidirectional vision [7]. Features from Accelerated Segment Test (FAST) [13] is used as the feature detector to detect corner features in the panoramic image. Then we adopted the feature region determining method proposed in [14] to achieve rotation invariance. Rectangular image regions surrounding corner features are firstly determined in the radial direction, and then rotated to a fixed orientation, as shown in Fig. 1(a) and Fig. 2(a). Finally, local binary pattern (LBP) [15] and center-symmetric local binary pattern (CS-LBP) [15] are used as feature descriptors to compute vectors to describe the information of feature regions. So two algorithms named FAST+LBP and FAST+CSLBP were designed. FAST, LBP and CS-LBP are computationally simple, so they can be the basis of our real-time local visual features.

We performed feature matching experiments by using the panoramic images in the COLD database to determine the best algorithm parameters and to compare with SIFT [16]. The final FAST+LBP and FAST+CSLBP with best algorithm parameters are shown in Fig. 1 and Fig. 2 respectively. The descriptor dimension of FAST+LBP and FAST+CSLBP are 236 and 72. The experimental results in [7] show that our algorithms have better performance than SIFT. The computation time needed to extract all the features in an image by FAST+LBP or FAST+CSLBP is from 5ms to 20ms, so our local visual features can be extracted in real-time, and they can be applied to computer/robot vision tasks with high real-time requirements like place recognition for mobile robots in this paper. Their performance will be compared with SIFT and SURF [17] when applied to place recognition in Section 4.2.



(a) (b) (c)

Fig. 1. The final FAST+LBP algorithm. (a) A feature region on the panoramic image. The green points are the detected corner features. (b) The scale-up feature region. (c) The resulting feature descriptor. The descriptor dimension is 236.

3 The Proposed Place Recognition Algorithm Based on Omnidirectional Vision

Bag-of-features [5][6] is a popular approach in computer vision/pattern recognition community, and has been applied successfully to object recognition, video retrieval, scene classification, etc. SVMs [8] is one of the most successful classifier learning methods in pattern recognition. In this section, we try to use bag-of-features to achieve robust and real-time place recognition based on omnidirectional vision for mobile robots by combining the real-time local visual features presented in Section 2 and SVMs. Our place recognition algorithm is divided into two phases: the phase of off-line training and the phase of on-line testing. The diagram of the algorithm is demonstrated in Fig. 3.

In the phase of off-line training, we assume that the number of the panoramic images for training is m, the number of place categories is M, and the corresponding place category of each image is also known. The local visual features f_i are extracted from each training image, where $i = 1...\sum_{j=1}^{m} n_j$, and n_j is the number of the local visual features extracted from the jth training image. After clustering these features with K-means clustering algorithm, we get clustering

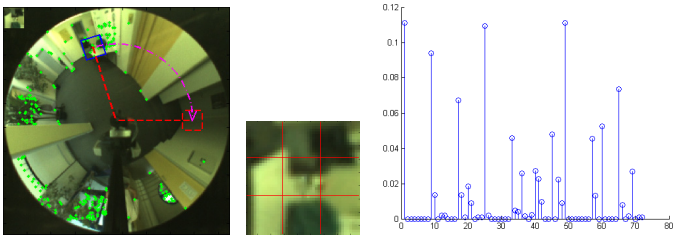


Fig. 2. The final FAST+CSLBP algorithm. (a) A feature region on the panoramic image. The green points are the detected corner features. (b) The scale-up feature region. (c) The resulting feature descriptor. The descriptor dimension is 72.

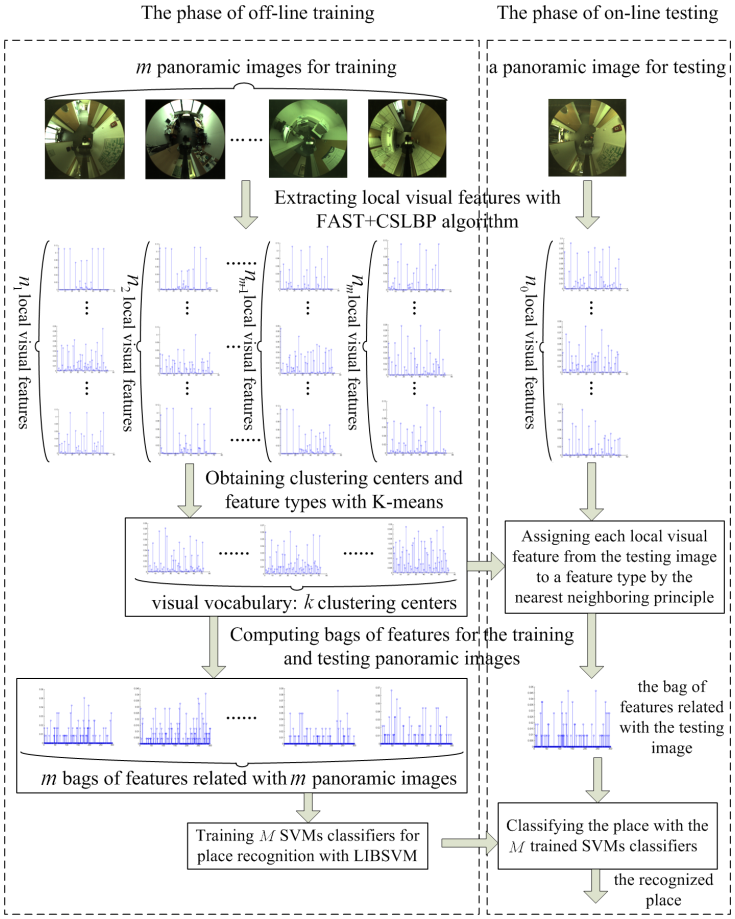


Fig. 3. The diagram of our place recognition algorithm based on omnidirectional vision and local visual features

centers C_i, where $i = 1...k$, and k is the clustering number. These clustering centers form the visual vocabulary which is similar as the word vocabulary for text categorization.

After the clustering is finished, each local visual feature from the training image has also been assigned to the corresponding cluster, which means that the feature type is obtained. A feature vector x_j is computed by normalizing the histogram constructed with the number of occurrences of each type of the features from the visual vocabulary in the jth training image, where $x_j \in R^k$. The feature vector is an effective representation of the image information, and it is named as bag of features. Then the bags of features and the corresponding place categories of all training images are used to learn M classifiers by applying the famous SVMs software - LIBSVM [8] according to the *one-vs-all* strategy.

During the training process mentioned above, the algorithm setup like the different local visual features, the different clustering numbers, the different kernel functions, and the completeness of the visual vocabulary, will affect the performance of our place recognition algorithm. We will determine the best algorithm setup by experiments in the next section.

In the phase of on-line testing, the local visual features are extracted from the testing panoramic image (or the image acquired on-line by the robot's vision system), and then each local visual feature is assigned to a feature type according to its distances to the clustering centers by the nearest neighboring principle. The bag of features of the testing image is computed by normalizing the histogram constructed with the number of occurrences of each type of the features from the visual vocabulary in the testing image. Finally, this bag of features is used as the input of the learned M classifiers, and the outputs are the classification results and the corresponding classification probability. The classification result with the largest classification probability is used as the final place category.

Omnidirectional vision is used in our algorithm, and better performance in place recognition should be achieved than those methods only using perspective images, because omnidirectional vision can provide a 360^o view of the robot's surrounding environment in a single image, which will be verified by the experimental results in the next section.

4 The Experimental Results

In this section, we will introduce the experimental setup firstly, and then test and analyze that how the algorithm performance will be affected by the factors like the choice of the local visual feature, the clustering number, the kernel function, and the training condition. Therefore, the best algorithm parameters and the best training condition can be determined. The performance will be presented in detail when the best parameters and the best training condition are used. Finally the real-time performance will be discussed.

4.1 Experimental Setup

COLD is a freely available database which provides a large-scale, flexible testing environment for vision-based place recognition. COLD contains 76 image sequences acquired in three different indoor environments across Europe. The images are acquired by the same perspective and omnidirectional vision in different rooms and under various illumination conditions. We will use the following six sequences of the panoramic images in COLD-Saarbruecken to perform our experiments: seq3_cloudy1, seq3_cloudy2, seq3_night1, seq3_night2, seq3_sunny1, seq3_sunny2. The "cloudy", "night", and "sunny" indicate the corresponding illumination conditions under which the image sequences are acquired. Four places are included in each of these image sequences: corridor, one-person office, printer area, and bath room. Although there are only four places, the sequences are long-term, and over 700 panoramic images are included in each sequence. More details about COLD can be found in [4].

4.2 The Choice of the Local Visual Feature

In this experiment, we compare the algorithm performance when using different local visual features: FAST+LBP, FAST+CSLBP, SIFT, and SURF. The clustering number was set to be 200, and linear kernel was used in SVMs. During the experiment, we used the image sequence seq3_cloudy1, seq3_night2, seq3_sunny2 for training respectively, and then used seq3_cloudy2, seq3_night1, seq3_sunny1 for testing respectively. Because there is a certain degree of randomness in the clustering results obtained by using K-means clustering algorithm, the training and testing processes were run several times to get the average place classification rate. The experimental results are shown in table 1 when different local visual feature was chosen. We see that the overall performance is much better when using FAST+LBP or FAST+CSLBP than using SIFT or SURF, which also validates that the discriminative power of FAST+LBP and FAST+CSLBP are good. There is not much difference in the overall performance when using FAST+LBP or FAST+CSLBP. However, the descriptor dimension of FAST+CSLBP is 72, and it is much smaller than that of FAST+LBP, which causes the lower computation cost of the place recognition algorithm. So we choose FAST+CSLBP as the local visual feature in the following experiments.

4.3 The Clustering Number

In this experiment, we compare the algorithm performance affected by different clustering numbers. FAST+CSLBP and linear kernel were used in the algorithm. During the experiment, we used the image sequence seq3_cloudy1 for training, and then used seq3_cloudy2, seq3_night1, seq3_sunny1 for testing respectively. The training and testing processes were also performed several times to get the average place classification rate. The experimental results are shown in Fig. 4 when the clustering number was set to be 50, 100, 150, 200, 250, 300, 350 and 400. In the general trend, the algorithm performance increases as the increase of

Table 1. The place classification rates when choosing different local visual feature

		training		
		seq3_cloudy1	seq3_night2	seq3_sunny2
seq3_cloudy2 for testing	FAST+LBP	0.9375	0.9245	0.8216
	FAST+CSLBP	0.9611	0.9364	0.9468
	SIFT	0.8168	0.7282	0.8361
	SURF	0.8313	0.7255	0.7954
seq3_night1 for testing	FAST+LBP	0.9472	0.9621	0.9149
	FAST+CSLBP	0.9254	0.8659	0.9470
	SIFT	0.8495	0.8342	0.8892
	SURF	0.7225	0.7849	0.8150
seq3_sunny1 for testing	FAST+LBP	0.8867	0.7238	0.7493
	FAST+CSLBP	0.8666	0.6802	0.6863
	SIFT	0.8083	0.7362	0.6335
	SURF	0.7760	0.7711	0.7813

the clustering number, which is consistent with the research results [5] in pattern recognition community. But the increase of the clustering number will make the vocabulary size larger and then cause higher computation cost in the testing process, so a compromise should be made between the classification rate and the clustering number. In the following experiments, the clustering number will be set to be 300 as the best parameter.

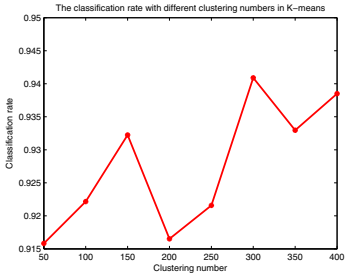


Fig. 4. The place classification rates when using different clustering number in the algorithm

4.4 The Choice of the Kernel Function in SVMs

In this experiment, we compare the algorithm performance affected by using different kernel functions in SVMs: linear kernel, RBF kernel, Sigmoid kernel. FAST+CSLBP was used, and the clustering number was set to be 300 in the algorithm. During the experiment, we used the image sequence seq3_cloudy1 for training, and then used seq3_cloudy2, seq3_night1, seq3_sunny1 for testing respectively. The training and testing processes were also performed several times to get the average place classification rate. The experimental results are shown

in table 2. We see that there is not much difference in the overall performance when using different kernel functions. Because linear kernel is computationally simplest, it will be used as the best kernel function in the following experiments.

Table 2. The place classification rates when using different kernel function in the algorithm

	linear kernel	RBF kernel	Sigmoid kernel
seq3_cloudy2 for testing	0.9521	0.9632	0.9313
seq3_night1 for testing	0.9594	0.9477	0.9516
seq3_sunny1 for testing	0.9336	0.9475	0.9476

4.5 The Completeness of the Visual Vocabulary

In this experiment, the best algorithm parameters determined above were used, which means that FAST+CSLBP and linear kernel were chosen, and the clustering number was set to be 300. During the experiment, we used the image sequence seq3_cloudy1, seq3_night2, seq3_sunny2 for training respectively, and then used seq3_cloudy2, seq3_night1, seq3_sunny1 for testing respectively. The average place classification rates were acquired to compare which image sequence was best for training to achieve the best performance. The experimental results are shown in table 3. We clearly see that the highest classification rate is achieved when using the image sequence acquired under "cloudy" illumination condition for training. The same conclusion can also be obtained from the experimental results in Section 4.2. When the illumination condition is "night" or "sunny", and the robot is located in the position where the illumination is affected greatly by the natural light, the acquired image may be less-exposed or over-exposed. Then some local visual features cannot be extracted, which may cause that the visual vocabulary is incomplete. The incompleteness of the visual vocabulary will lead to the decrease of the place classification rate, which is the same as the situation in text categorization that the incompleteness of the word vocabulary will result in the decrease of the text categorization rate. The "cloudy" illumination is more stable than "night" and "sunny", so the image sequence acquired under "cloudy" illumination condition is best for training.

Furthermore, we used seq3_cloudy1, seq3_night2 and seq3_sunny2 jointly for training, and then used seq3_cloudy2, seq3_night1, seq3_sunny1 for testing respectively. The experimental results are also shown in table 3. The place classification rates are improved when using seq3_cloudy2 and seq3_night1 for testing. But when seq3_sunny1 is used for testing, the performance is still much worse than that when only using seq3_cloudy1 for training. So in the following experiments, seq3_cloudy1 will be used for training.

Table 3. The place classification rates when different image sequences acquired under different illumination conditions were used for training

	training			
	seq3_cloudy1	seq3_night2	seq3_sunny2	all seqs
seq3_cloudy2 for testing	0.9296	0.9366	0.9523	0.9634
seq3_night1 for testing	0.9550	0.7190	0.9516	0.9707
seq3_sunny1 for testing	0.9380	0.5919	0.6379	0.8362

4.6 The Performance with the Best Parameters and Training Condition

Through the experiments mentioned above, we have determined the best algorithm parameters, and the illumination condition under which the best training image sequence is acquired. In this experiment, the best parameters were used in the algorithm. The best image sequence seq3_cloudy1 was used for training, and seq3_cloudy2, seq3_night1, seq3_sunny1 were used for testing respectively, so the algorithm performance can be analyzed in detail.

Because of the randomness of the clustering process, we only demonstrate the best results after training several times. When seq3_cloudy2 was used for testing, the detailed result of place classification is shown in table 4, where the statistics of how many images being correctly and wrongly classified are listed. The place classification rate is 0.9806. Some panoramic images which were wrongly classified are shown in Fig. 5. When seq3_night1 was used for testing, the detailed result of place classification is shown in table 5. The place classification rate is 0.9594. Some panoramic images which were wrongly classified are shown in Fig. 6. When seq3_sunny1 was used for testing, the detailed result of place classification is shown in table 6. The place classification rate is 0.9429. Some panoramic images which were wrongly classified are shown in Fig. 7.

Table 4. The detailed result of place recognition when using seq3_cloudy2 for testing

real places ↓	recognition results			
	corridor	one-person office	printer area	bath room
corridor	277	0	0	0
one-person office	5	106	0	0
printer area	2	0	77	0
bath room	7	0	0	246

From the experimental results, we clearly see that high place classification rate can be achieved by using our algorithm. Most of those panoramic images which were wrongly classified are acquired when the robot is located near the border of two different places. Because the omnidirectional vision system can provide a 360° view of the robot's surrounding environment, when the robot is located near the border, both of the scenes belonging to the two places will be

Table 5. The detailed result of place recognition when using seq3_night1 for testing

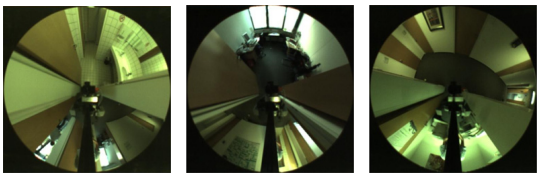
real places ↓	recognition results			
	corridor	one-person office	printer area	bath room
corridor	290	11	0	1
one-person office	6	108	0	0
printer area	6	0	94	0
bath room	7	0	0	241

Table 6. The detailed result of place recognition when using seq3_sunny1 for testing

real places ↓	recognition results			
	corridor	one-person office	printer area	bath room
corridor	253	9	4	2
one-person office	2	95	0	0
printer area	5	0	99	0
bath room	21	0	0	263

included in the panoramic image. Furthermore, the panoramic images are not labeled according to their content but to the position of the robot at the time of acquisition. So in this case, the classification error cannot be completely avoided.

In comparison with the place classification results in [3], where only the perspective images in the COLD database were used, better performance is achieved by our algorithm. This can be explained as follows: our method is based on omnidirectional vision, and the changes of the panoramic image with the different robot's positions are not so rapid as that of the perspective image, so omnidirectional vision is more suitable for place recognition than perspective camera; our FAST+CSLBP feature is discriminative and robust; the bag-of-features method itself is powerful for place recognition.

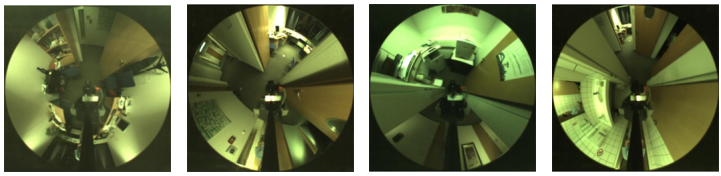


(a) (b) (c)

Fig. 5. Some wrongly classified images when using seq3_cloudy2 for testing. Bath room (a), one-person office (b), printer area (c) were wrongly classified as corridor.

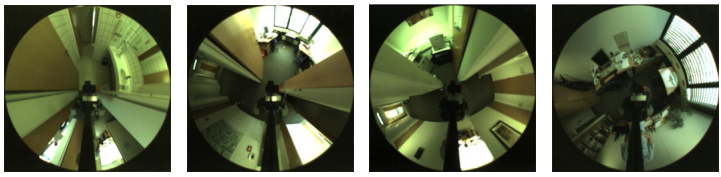
4.7 The Real-Time Performance

The real-time performance is very important in the actual application for mobile robots. Because the training process of our algorithm is off-line, only the on-line testing process should be analyzed. In the testing process, the algorithm



(a) (b) (c) (d)

Fig. 6. Some wrongly classified images when using seq3_night1 for testing. One-person office (a), printer area (c), bath room (d) were wrongly classified as corridor. (b) Corridor was wrongly classified as one-person office.



(a) (b) (c) (d)

Fig. 7. Some wrongly classified images when using seq3_sunny1 for testing. Corridor was wrongly classified as bath room (a), one-person office (b), and printer area (c). (d) One-person office was wrongly classified as corridor.

consists of three parts: the extraction of local visual features, the construction of bag of features, and place classification with SVMs. The computer is equipped with 2.26 GHz Duo CPU and 1.0G memory. According to the experimental results in [7], the computation time needed to extract all the local visual features in a panoramic image by FAST+CSLBP is from 5 to 20 ms. When the best parameters in Section 4.5 are used, the construction of bag of features and place classification with SVMs can be finished in 10 ms. The whole place recognition can be finished in 30 ms, so our algorithm can be run in real-time. We can clearly see that using the real-time local visual features is very important to make our algorithm satisfy the real-time requirement.

5 Conclusion

In this paper, the bag-of-features method is used to solve place recognition based on omnidirectional vision for mobile robots by combining the real-time local visual features proposed by ourselves and SVMs. The panoramic images in the COLD database were used to perform experiments to test the affection on the performance by different algorithm factors like the choice of the local visual feature, the clustering number, the choice of the kernel function in SVMs, and the completeness of the visual vocabulary. So the best algorithm parameters and the illumination condition under which the best training image sequence was acquired were determined, and the performance of place recognition with these best parameters and the best training condition was analyzed in detail.

The real-time performance was discussed finally. The experimental results show that place recognition can be realized in real-time with high classification rate by using the proposed algorithm.

References

1. Pronobis, A., Caputo, B., Jensfelt, P., Christensen, H.I.: A realistic benchmark for visual indoor place recognition. Robotics and Autonomous Systems 58, 81–96 (2010)
2. Pronobis, A., Jie, L., Caputo, B.: The more you learn, the less you store: Memory-controlled incremental SVM for visual place recognition. Image and Vision Computing 28, 1080–1097 (2010)
3. Ullah, M.M., Pronobis, A., Caputo, B., et al.: Towards robust place recognition for robot localization. In: Proceedings of the 2008 IEEE ICRA, pp. 530–537 (2008)
4. Pronobis, A., Caputo, B.: COLD: The Cosy Localization Database. The International Journal of Robotics Research 28(5), 588–594 (2009)
5. Csurka, G., Dance, C.R., Fan, L., et al.: Visual categorization with bags of keypoints. In: Proceedings of ECCV 2004 Workshop on Statistical Learning in Computer Vision, pp. 59–74 (2004)
6. Sivic, J., Zisserman, A.: Video Google: A Text Retrieval Approach to Object Matching in Videos. In: Proceedings of the 9th IEEE ICCV, pp. 1–8 (2003)
7. Lu, H., Zheng, Z.: Two novel real-time local visual features for omnidirectional vision. Pattern Recognition 43, 3938–3949 (2010)
8. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2, 27:1–27:27 (2011), Software available at, `http://www.csie.ntu.edu.tw/~cjlin/libsvm`
9. Filliat, D.: A visual bag of words method for interactive qualitative localization and mapping. In: Proceedings of the 2007 IEEE ICRA, pp. 3921–3926 (2007)
10. Fraundorfer, F., Engels, C., Nistér, D.: Topological mapping, localization and navigation using image collections. In: Proceedings of the 2007 IEEE/RSJ IROS, pp. 3872–3877 (2007)
11. Cummins, M., Newman, P.: FAB-MAP: Probabilistic localization and mapping in the space of appearance. The International Journal of Robotics Research 27, 647–665 (2008)
12. Svoboda, T., Pajdla, T.: Matching in Catadioptric Images with Appropriate Windows, and Outliers Removal. In: Skarbek, W. (ed.) CAIP 2001. LNCS, vol. 2124, pp. 733–740. Springer, Heidelberg (2001)
13. Rosten, E., Drummond, T.: Machine learning for high-speed corner detection. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 430–443. Springer, Heidelberg (2006)
14. Andreasson, H., Treptow, A., Duckett, T.: Self-Localization in non-stationary environments using omni-directional vision. Robotics and Autonomous Systems 55, 541–551 (2007)
15. Heikkilä, M., Pietikäinen, M., Schmid, C.: Description of interest regions with local binary patterns. Pattern Recognition 42, 425–436 (2009)
16. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision 60(2), 91–110 (2004)
17. Bay, H., Ess, A., Tuytelaars, T., et al.: Speeded-Up Robust Features (SURF). Computer Vision and Image Understanding 110, 346–359 (2008)