Original Research Paper

Spatial Representation to Support Visual Impaired People

Abbas Mohamad Ali and Shareef Maulod Shareef

Department of Software Engineering, College of Engineering, Salahaddien University, Erbil, Kurdistan, Iraq

Abstract: The rapid development of Information and Communication Technology (ICT) enhances the government services to its citizen. As a consequence affects the exchange of information between citizen and government. However, some groups can experience some difficulties accessing the government services due to disabilities, such as blind people as well as material and geographical constraints. This paper aims to introduce a novel approach for place recognition to help the blind people to navigate inside a government building based on the correlation degree for the covariance feature vectors. However, this approach faces several challenges. One of the fundamental challenges inaccurate indoor place recognition for the visual impaired people is the presence of similar scene images in different places in the environmental space of the mobile robot system, such as a computer or office table in many rooms. This problem causes bewilderment and confusion among different places. To overcome this, the local features of these image scenes should be represented in more discriminatory and robustly way. However, to perform this, the spatial relation of the local features should be considered. The findings revealed that this approach has a stable manner due to its reliability in the place recognition for the robot localization. Finally, the proposed Covariance approach gives an intelligent way for visual place people localization through the correlation of Covariance feature vectors for the scene images.

Keywords: E-Government, Covariance Features Vectors, SIFT Grid, K-Means

Introduction

The development of new technologies might prove to be a great facilitator for the integration of disabled people provided that these environments are accessible, usable and useful; in other words, that they take into consideration the various characteristics of the activity and the needs and particularities (cognitive, perceptive, or motive) related to the disability of the users (Paciello, 2000). The objective of this paper is to determine the real contributions of accessible E-services for visually disabled persons to identify places. This is achieved by introducing a mobile application that recognizes a place, hence will help blind person to identify the location. However, place recognition is one of the basic issues in mobile robotics based accurate localization through the environmental navigation. One of the fundamental problems in the visual place recognition is the confusion of matching visual scene image with the stored database images. This problem is caused by instability of local feature representation. Machine learning is used to improve the localization process of known or unknown

environments. This led the process to have two modes; supervised mode like (Booij *et al.*, 2009; Miro *et al.*, 2006) and unsupervised mode, like (Abdullah *et al.*, 2010). The most common tools that used in machine learning is the *K*-means clustering technique to cluster all probabilistic features in the scene images in order to construct the codebook. Several works used clustering techniques, where the image local features in a training set are quantized into a "vocabulary" of visual words (Ho and Newman, 2007; Cummins and Newman, 2009; Schindler *et al.*, 2007). Clustering technique may reduce the dimensionality of features and the noise by the quantization of local features into visual words. The process of quantizing the features is quite similar to the BOW model as in (Uijlings *et al.*, 2009). However, these visual words do not possess spatial relations. However, this model employed to make more accurate features for describing the scene image in place recognition.

Cummins and Newman (2009), they used BOW to describe an appearance for Simultaneous Localization And Mapping (SLAM) system, which was used for a large scale rout of images. Schindler *et al.* (2007) an

informative features proposed to add for each location and vocabulary trees (Nister and Stewenius, 2006) for recognizing location in the database. In contrast, (Knopp *et al.*, 2010) measured only the statistics of mismatched features and that required only negative training data in the form of highly ranked mismatched images for a particular location. Artac *et al.* (2002), an incremental Eigen space model was proposed to represent the panoramic scene images, which is taken from different locations, for the sake of incremental learning without the need to store all the input data. The work in (Ulrich and Nourbakhsh, 2000) was based on color histograms for images taken from omnidirectional sensor, these histograms were used for appearance based localization. Recently, most of works in this area have been focusing on large-scale navigation environments. For example, in (Murillo and Košecka, 2009) a global descriptor for portions of panoramic images was used for similar measurements to match images for a large scale outdoor Street View dataset. Kosecka *et al.* (2003) qualitative topological localization established by segmentation of temporally adjacent views relied on similarity measurement for global appearance. Local scale-invariant keypoints were used as in (Košecka *et al.*, 2005) and spatial relation between locations was modeled using Hidden Markov Models (HMM). Sivic and Zisserman (2003), the Support Vector Machines (SVM) was used to evaluate the place recognition in long-term appearance variations. The performance of the covariance proved by (Tuzel *et al.*, 2006) used covariance features with integral images, so the dimensionality is much smaller and getting fast computational time. Most of the implementations need spatial features, which arise when the robot is navigating in some places that are similar to each other. For example, like two offices those have the same type of tables or in corridor navigation. E.g., in feature based Robot navigations to help the visual impaired people, Land Marks are commonly used to find the correspondence between the current scene and the database. Wang and Yang (2010) the covariance is also used with SVM for classification purpose called Locality-constrained Linear Coding. In General, the covariance implementation results of the previous studies indicate that it has promising results for the recognition process.

The main contribution of this work is to introduce an e-service for blind person by using the Covariance features to give a spatial relation to the visual words to decrease the confusion problem for visual places recognition in the large indoor navigation process.

Methodology

Clustering image features is a process of learning visual recognition for some types of structural image contents. Each image I_j contains a set of features $\{f_1, f_2, ..., f_m\}$ and each f_i is a 128 size element. To organize all these features into K clusters $C = (C_1 ... C_k)$, the features that are close to each others will be grouped together (Sivic and Zisserman, 2003), as in (1):

$$KCL(K) = \sum_{i=1}^{n} \min_{1 \le j \le k} \left(f_{i-} x'_j \right)^p \tag{1}$$

where, K is the number of clustering means of features, p is the measurement of the distance between these features; and $x'_1, x'_2, ... x'_k$ are the means. In this study, SIFT grid approach is used to extract the local feature fs for the images of 30×30 grid block. Matlab code used for this purpose (Lazebnik *et al.*, 2006).

The local features for any selected image is represented by distance for these features from the centroid c of the codebook B, which is represented by a distance table containing m of distance vectors of size (128) from each centroid c in B as in Equation (2):

$$Dt(c) = dist(c, x_{1..m}) \tag{2}$$

The covariance (COD) of Dt in Equation (3) gives the covariance distances of all features related to the selected images. Let Sb is the row size of matrix Dt:

$$COD = diag(\frac{1}{Sb} Dt \times Dt) \ sb \ne 0 \tag{3}$$

The Minimum Distance (MDT) for the table Dt in Equation (4), produces a row of minimum value for each column in the table. The size of this row is the number of centroid c in the code book (B), informed it Sb:

$$d = MDT(c) = min_i (Dt_i) \tag{4}$$

The covariance of minimum distance for each image, d will be expressed as:

$$cov(d) = \frac{1}{Sb} d * d' \quad sb \ne 0 \tag{5}$$

The eigenvalues Er and eigenvectors Ev are calculated for the constructed covariance matrix and used in Equation (6) to give the covariance matrix (T). The work study is based on Standard Deviation (STDEV) and mean for calculating the distance of local features from the centroids in the codebook this principle has been exploited to optimize the filtered sums by multiplying it by the upper bound of the STDEV and the mean of the feature vector

```
Im=Read Image
sf=Sift Feature(Im)
d= calculate_ minimum distance (sf, code-book)
sb=size(d);
n=better division (sb)
k=1;
for  i=n:sb step n
part_of_dₖ= dᵢ...ᵢ₊ₙ₋₁
cfₖ =calculate the covariance (part_of_dₖ as in (5,6))
k=k+1;
End for
cf= {cf₁,cf₂, ....cfₘ};
```

Fig. 1. The pseudo code for speedup *ef*

$$cf = d \times Ev \times diag\left(\sqrt{\frac{1}{diag\left(Er\right)+0.1}} \times Ev\right)$$

$$\times \frac{e^{(mean(d)+std(d))}}{e^{(mean(d)*std(d))}} \qquad (6)$$

The size of the covariance feature vector (*cf*) is the same size as *d*. To speed up this calculation for the *Er* and *Ev*, the minimum distance *d* is subdivided into n parts to calculate the covariance for each part separately as in Fig.1.

Classification and Correlation

The recognition process in this study simply uses the calculation of the Covariance of the Minimum Distance (CMD) generated from the query scene image as in Equation (6). To examine the similarity of two images like x and y, the correlation between the two Covariance feature vectors *cf1* and *cf2 is calculated as in* Equation (7):

$$corr\left(cf1,cf2\right) = \frac{cov(cf1,cf2)}{std\left(cf1\right)*std(cf2)} \qquad (7)$$

where, the correlation coefficient is Pearson's coefficient for the two variables *cf1* and *cf2*, varies between -1 and +1.

The results for all correlation values are sorted; then, the maximum values are taken to be the best matching visual places. This approach is also called as a k Nearest Neighbor (k-NN). The average precision can be calculated as in (Bin Abdullah, 2010), where the Precision (*P*) of the first N retrieved images for the query Q is defined as:

$$p(Q,N) = \frac{\left|\left\{Ir \mid Rank(Q,Ir) \leq N \; suchthatIr \in g\left(Q\right)\right\}\right|}{N} \qquad (8)$$

where, *Ir* is the retrieved image and g (*Q*) represents the group category for the query image.

Experiments and Results

Two types of experiments have been conducted to check the accuracy performance of CMD.

The first experiment is to test the accuracy of the proposed approach, through working on the data set of IDOL (Pronobis *et al.*, 2009), Figure 2 shows some groups in the dataset. The SIFT features were extracted using SIFT grid algorithm for each image. The size of each frame image was 230×340. The CMD features vectors extracted using cluster number 260, it was used to express different places of the environmental navigation, namely an office for one person, a corridor, an office for two-people, a kitchen and a printer area. To demonstrate the accuracy performance of CMD, the algorithm implemented on various illumination condition groups (sunny, cloudy and night) for IDOL dataset each group divided into two parts such as train and test images, each parts divided into 16 subgroups. Different running tests were used around 5 times. In addition to the experiments, with mixed of these groups have been used also. Then the performances were reported using the average of the obtained classification results. Table 1 shows the experimental results for HBOF, MDT and CMD approach implemented on one IDOL data set using K-NN and SVM for WEKA software, to classify the images corresponding to their places.

The performance of the proposed approach using k-NN is more accurate than SVM. This will not give an indication that k-NN better than SVM, since the theoretical background for the two methods is known; therefore k-NN is adapted in the second experiments for navigation process. The accuracy, performance under various illumination conditions (sunny, cloudy and night) is about 97%, depending on the specific environmental difficulties. Figure 3, shows random selections of images for testing the retrieval of the best similar 5 images according to the highest correlation values.

B- Indoor Experiment

In this section, an experiment is performed as a simulation of navigation for the whole IDOL dataset using CMD approach used, to check the accuracy performance for the robot navigation. This done by using pre-stored some images as landmarks from the data set with locations and then by giving each places its color to know the error of confusing place recognition. Figure 4, shows the results of simulation. Each color indicates group in the dataset. The wrong correlation leads to confusion of place recognition, which leads to give the wrong color in the topological map.

Table 1. Comparison of some approaches

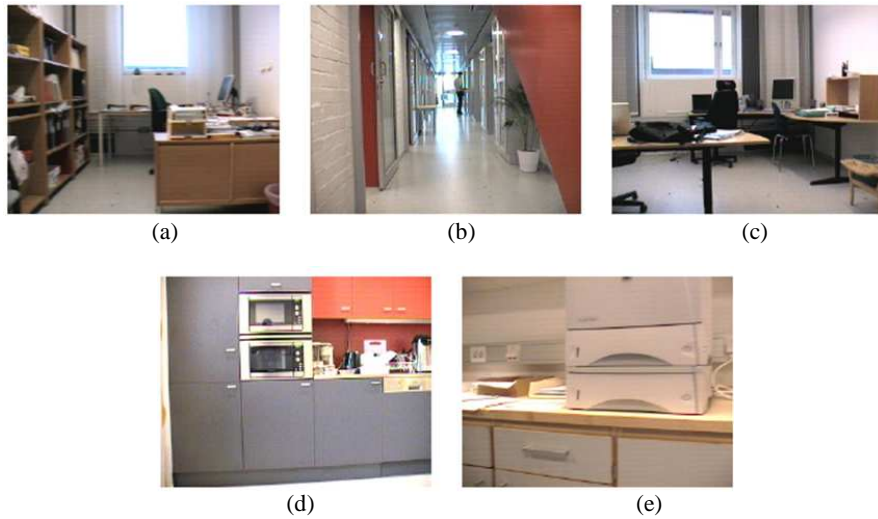| Class	K-NN	SVM
HBOF	84.0784±0.1937	85.2723±0.4176
MDT	92.6356±0.2676	90.8±0.2333
CMD	97.8509±0.2859	92.8078±0.1682

Fig. 2. IDOL dataset (a) office for one person (b) Corridor (c) office for two persons (d) Kitchen (e) Printer area
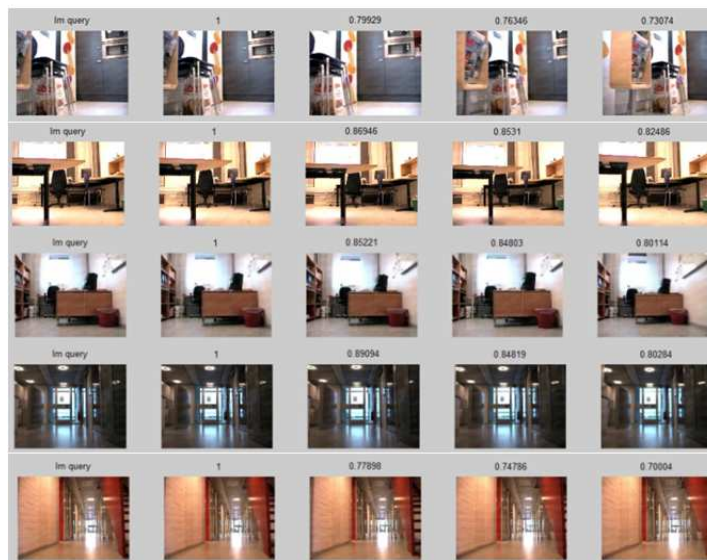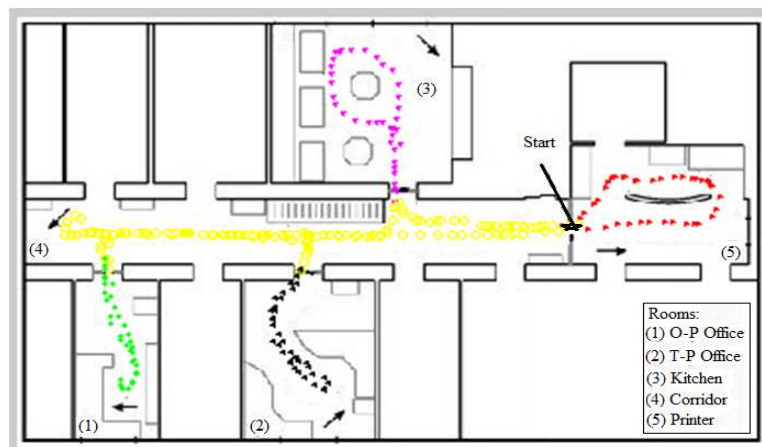

Fig. 3. Random query images and image retrieving

Fig. 4. IDOL dataset groups recognition

Discussion

The place recognition is done through a sequence of image scenes converted to visual words; the CMD for these visual words give some relation between them. The correlation between these CMD features gives indication that how extent the two images related to each other. The decision making for localization is done according to the correlation values related to correlation values between the current image scene and all the stored landmarks. The maximum values of k nearest neighbor used to select the current localization place for the visual impaired people. This algorithm can be used as an auditory advising for the blind person navigating through indoor environment. Place recognition based on CMD gives some reliability and accurate perception for the global localization and it reduces the confusion for place recognition.

The system consistently shows on-line performances more than 97% for environments, recognition, the image for the landmarks stored as covariance features, which was extracted from the quantized SIFT features according to the codebook. The more number of landmarks will give more accurate localization process within the navigated environment. Figure 5 shows a confusion matrix for idol data set to show the confuse place recognition through landmark recognition, which effects on the localization process. The landmarks selected inaccurate way to be discriminated from each other, in such a way that gives accurate localization for the robot within the topological map.

Fig. 5. Confusion matrix for IDOL (a) CMD (b) MDT (c) HBOF

Conclusion

One of the important responsibilities of government in developed countries across throughout the world is to provide services to everyone, particularly disable persons. However, one important issue in the visual impaired people through robot localization is accurate place recognition in the environment to give accurate mapping. The problem of confusion for the similar place recognitions is a challenging issue in computer vision. Accurate spatial representation of the visual words, it may give a good solution for this issue. The work, proposed a novel approach using correlation of Covariance Minimum Distance (CMD) for place recognition. CMD has been compared with some approaches using the same data set to evaluate and measure the accuracy performance. The experimental results show that the proposed method can be outperforming than others. It is an establishment of an algorithm to conceptualize the environment; using spatial relations of clustered SIFT features in navigation and localization techniques.

Acknowledgement

Funding Information

Author's Contributions

Abbas Mohamed Ali: Contribute to the creation of an idea and identify certain issue, finding out the possibilities that seem right for the objective of the research. Gathering and evaluating information. Working out and formulating main argument. Design and creation of the figures and necessary tables. Identify suitable tool that achieves the goal of the article. Drafting the article.

Shareef Maulod Shareef: Contribute to the conception and design of the article. Plan for the structure and organize the content of the article. Interpreting and analyzing the data reviewing and editing the article critically for significant intellectual content. Critically and re-evaluation as well as the accurate documentation of the resources.

Ethics

The data and information are used and have been gathered legally and not plagiarized or claimed the results by others. The data have not been submitted whose accuracy they have reason to question. The resources have been used based on the academic standards, along with the accurate citation. The article has not been written in a way that deliberately makes it difficult for the reader to understand them.

References

Abdullah, A., R.C. Veltkamp and M.A. Wiering, 2010. Fixed partitioning and salient points with MPEG-7 cluster correlograms for image categorization. Patt. Recogn., 43: 650-662. DOI: 10.1016/j.patcog.2009.09.007

Artac, M., M. Jogan and A. Leonardis, 2002. Mobile robot localization using an incremental eigenspace model. Proceedings of the IEEE International Conference on Robotics and Automation, May 11-15, pp: 1025-1030. DOI: 10.1109/ROBOT.2002.1013490

Bin Abdullah, A., 2010, Supervised learning algorithms for visual object categorization, Utrecht University Repository.

Booij, O., Z. Zivkovic and B. Kröse, 2009, Efficient data association for view based SLAM using connected dominating sets. Elsevier Robotics Autonomou Syst. 57: 1225-1234. DOI: 10.1016/j.robot.2009.06.006

Cummins, M. and P. Newman, 2009. Highly scalable appearance-only SLAM-FAB-MAP 2.0. Oxford University Mobile Robotics Group.

Ho, K.L. and P. Newman, 2007. Detecting loop closure with scene sequences. Int. J. Comput. Vision, 74: 261-286. DOI: 10.1007/s11263-006-0020-1

Knopp, J., J. Sivic and T. Pajdla, 2010. Avoiding confusing features in place recognition. Proceedings of the 11th European Conference on Computer Vision, Sep. 5-11, IEEE Xplore Press, Heraklion, Crete, Greece, pp: 748-761. DOI: 10.1007/978-3-642-15549-9_54

Košecka, J., F. Li and X. Yang, 2005. Global localization and relative positioning based on scale-invariant keypoints. Robotics Autonomous Syst., 52: 27-38. DOI: 10.1016/j.robot.2005.03.008

Kosecka, J., L. Zhou, P. Barber and Z. Duric, 2003. Qualitative image based localization in indoors environments. Proceedings of IEEE Computer Society Computer Vision and Pattern Recognition, Jun. 18-20, IEEE Xplore Press, Madison, WI, pp: II-3-II-8. DOI: 10.1109/CVPR.2003.1211445

Lazebnik, S., C. Schmid and J. Ponce, 2006. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Jun. 17-22, IEEE Xplore Press, pp: 2169-2178. DOI: 10.1109/CVPR.2006.68

Miro, J.V., W.Z. Zhou and G. Dissanayake, 2006. Towards vision based navigation in large indoor environments. Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Oct. 9-15, IEEE Xplore Press, Beijing, pp: 2096-2102. DOI: 10.1109/IROS.2006.282487

Murillo, A.C. and J. Kosecka, 2009. Experiments in place recognition using gist panoramas. Proceedings of the 12th International Conference on Computer Vision Workshops, Sept. 27-Oct. 4, IEEE Xplore Press, Kyoto, pp: 2196-2203. DOI: 10.1109/ICCVW.2009.5457552

Nister, D. and H. Stewenius, 2006. Scalable recognition with a vocabulary tree. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Jun. 17-22, IEEE Xplore Press, pp: 2161-2168. DOI: 10.1109/CVPR.2006.264

Paciello, M., 2000. Web accessibility for people with disabilities. Lawrence, KS: CMP Books.

Pronobis, A., B. Caputo, P. Jensfelt and H.I. Christensen, 2009. A realistic benchmark for visual indoor place recognition. Robotics Autonomous Syst., 58: 81-96. DOI: 10.1016/j.robot.2009.07.025

Schindler, G., M. Brown and R. Szeliski, 2007. City-scale location recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Jun. 17-22, IEEE Xplore Press, Minneapolis, MN, pp: 1-7. DOI: 10.1109/CVPR.2007.383150

Sivic, J. and A. Zisserman, 2003. Video Google: A text retrieval approach to object matching in videos. Proceedings of the IEEE International Conference on Computer Vision, Oct. 13-16, IEEE Xplore Press, Nice, France, pp: 1470-1477. DOI: 10.1109/ICCV.2003.1238663

Tuzel, O., F. Porikli and P. Meer, 2006. Region Covariance: A fast Descriptor for Detection and classification. Proceeding of the 9th European Conference on Computer Vision, May 7-13, Graz, Austria, pp: 589-600. DOI: 10.1007/11744047_45

Uijlings, J.R.R., A.W.M. Smeulders and R.J.H. Scha, 2009. Real-time bag of words, approximately. Proceedings of the ACM International Conference on Image and Video, Jul. 08-10, Santorini, Fira, Greece, DOI: 10.1145/1646396.1646405

Ulrich, I. and I. Nourbakhsh, 2000. Appearance-based place recognition for topological localization. Proceedings of the IEEE International Conference on Robotics and Automation, Apr. 24-28 IEEE Xplore Press, San Francisco, CA, pp: 1023-1029. DOI: 10.1109/ROBOT.2000.844734

Wang, J. and J. Yang, 2010. Locality-constrained linear coding for image classification. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Jun. 13-18, IEEE Xplore Press, San Francisco, CA, pp: 3360-3367. DOI: 10.1109/CVPR.2010.5540018