Ensemble of Bayesian Filters for Loop Closure Detection

Mohammad Omar Salameh, Azizi Abdullah, Shahnorbanun Sahran

Pattern Recognition Research Group
Center for Artificial Intelligence
Faculty of Information Science and Technology
University Kebangsaan Malaysia 43600 Bangi, Selangor
E-mail: omar82@siswa.ukm.edu.my,{azizi, shahnorbanun}@ukm.edu.my

Abstract. Loop closure detection for visual only simultaneous localization and mapping needs effective feature descriptors to obtain good performance results. Currently, the most widely used feature description is the global or local descriptor such as color histogram and Speeded Up Robust Features. The global features can be computed either by considering all points within a region, or only for those points on the boundary of a region. In contrast, the local features are obtained by considering the boundary of an object that represents a distinguishable small part of a region. One possible problem of these approaches is that the number of features become very large when a dense grid is used where the histograms are computed and combined for many different regions or points. The most popular solution for the problem is to use a clustering algorithm to create a visual codebook to create a histogram of visual keywords present in a visual image. In this paper, we designed and implemented an ensemble learning method namely mean rule to combining three different local features: Scale-invariant feature transform (SIFT), Speeded Up Robust Features (SURF) and Oriented FAST and Rotated BRIEF (ORB). The aim of using ensemble learning is to enhance learning speed and final performance of different local visual keywords descriptors for loop closure detection. Furthermore, the Real-Time Appearance-Based Mapping (RTAB-Map) using a Bayes filter is used to evaluate loop closure hypotheses. Experimental results on a public dataset contains 2464 images show that the ensemble algorithm outperform the single bag-of-features approach.

Key words: loop closure detection, ensemble rules, bag-of-features, visual features combination

1 Introduction

Loop closure detection algorithms aim to recognize a previously visited place from current visual sensor measurements. During the last decade researches in machine vision have proposed many effective descriptors for dealing with the complex problem of handling high dimensional pixel representations. These algorithms can be used in robotics such as to solve the loop closure detection problems. In literature, most Simultaneous Localization and Mapping (SLAM) algorithms working with sensory data such as laser to capture surrounding information or odometry which provides the rotational speeds of the robot's wheels for building map and estimate the location. However, it still has limited capabilities in the context of understanding, recognizing and validating surrounding environments. The algorithms used probabilistic approaches in a strictly cartesian space to precisely describe a complex 3-D geometrical model of the environment. However, it is very complicated to produce models, especially to recover topology of the environment from noisy or complex real world information, which can increase problems of maintaining a global map for localization. Thus, the need for more robust for environment description and preserve only its topology became unavoidable. Nowadays, using cameras for capturing visual information for description have become popular. This sensor is not only cheap but also many complex computer vision algorithms can be applied for feature extraction and association in different views of a scene. These features can be applied to vision based SLAM or visual SLAM to build up a map within unknown environment for autonomous mobile robots. In visual SLAM, one of the crucial parts to enhance mapping and metrical SLAM algorithms is loop closure detection. This problem consists in detecting a previously visited place in an environment. Once a loop closure is detected the actual pose can be precisely estimated from the map. Not only the pose but the global localization problems or for recovering from a kidnapping robot can also be improved. As a result, it attracts many computer vision researches to conduct research to detect loop closure using vision sensors.

One of the most earlier works on using visual SLAM for mobile robot localization is from Ulrich and Nourbakhsh [1]. In this work, they have introduced the concept of appearance-based representation for visual place description and a similarity measure to obtain similarities of the location choices. This concept basically is a popular concept in developing computer vision systems such as in image retrieval systems. In the system the the image descriptor is used for description and similarity measure for classification. Thus, it has gained more attention especially computer vision researches to conduct research in visual SLAM especially to detect loop closure using vision sensors. Besides using the raw image features such as color, texture or shape for description, the use of local appearance approach by clustering feature vectors extracted from local points features into similar group patterns is also popular. Thus, Newman et al. [2] have proposed to used the visual appearance and laser ranging sensor for outdoor SLAM. In this system, a sequence of images from the camera are used to detect loop closure using a novel appearance-based retrieval system. They reported that the method is robust to repetitive visual structure and provides a probabilistic measure of confidence. In [3] proposed to use the most common local appearance based approached namely bag-of-features for loop closure detection. In this research, they found that the visual location can be easily identified using a set of unordered cluster features. The features are computed from a local image features to represent patches such as SIFT [4] and SURF [5] for vector descriptor or ORB [6] for binary descriptor. After that, a clustering technique is applied to group similar patterns into a cluster visual codebook. In the loop closure detection, the codebook can be constructed either offline using identified location images for training or online using incremental clustering algorithm [7]. Besides using similarity measures for landmark matching, in [8] proposed to use Bayesian filter to estimate the loop closure from a previously visited place. Similar to previous methods the concept of visual words is used to symbolic representation of visual places.

In this paper, a visual loop closure detection using an ensemble of Bayesian learning technique is proposed. We employ Bayesian filters to learn to recognize the visual locations generated from visual words of local features. After that, the mean rule ensemble learning is used to combine all probability outputs of Bayesian filters. The method can combine the best performing classifier by combining global and local features. However, in our experiments we used only the local descriptors because it preserves more discriminative information over a small number of variations.

2 Methods

In this section we will first describe the RTAB-Map for realtime appearance based mapping [9], since the method presented in this paper use it for loop closure detection. After that we will describe the local image features used and Bayesin learning for recognizing visual places.

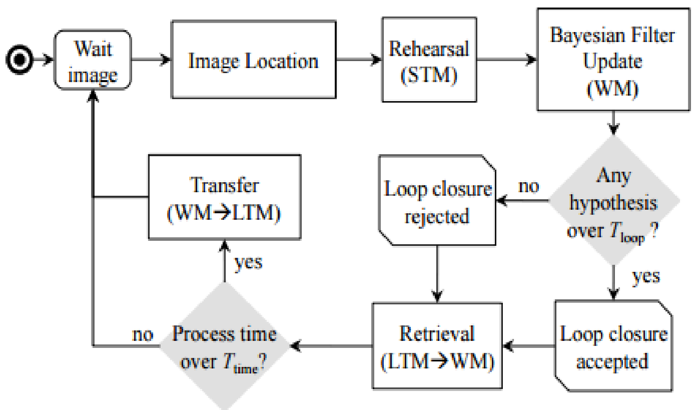

Fig. 1. Architecture of the RTAB-Map memory management loop closure detetcion steps.

2.1 RTAB-Map

RTAB-Map is a memory management approach for real time appearance based loop closure detection. It uses the local appearance descriptor for description and a Bayes filter for evaluating the loop closure hypothesis. The map is constructed by linking new acquire images with previous ones based on the loop closure probability values and it is fully incremental. RTAB-Map uses (a) Working Memory (WM) - to keep the most recent and frequent observed locations (b) Long Term Memory (LTM) - storing other observed locations. When a match is found between the current location and one stored in WM, associated locations stored in LTM can be remembered and updated and (c) Short Term Memory (STM) - storing the poses and retrieve them from the long term memory when loop closing is required. Figure 1 shows the RTAB-Map memory management loop closure detection steps [9].

2.2 Local Image Features

We used the following local image descriptors for image description: (a) **SIFT** - We applied the SIFT descriptor proposed by Lowe [4] which constructs the histograms of gradient orientations computed around the points as the descriptor. SIFT uses an interest points detector to detect salient locations which have certain repeatable properties. Afther that an 128-bin is used for image description. (b) **SURF** - We also applied the SURF proposed by Bay et. al [5] which computes the sum of the Haar wavelet response around the point of interest for image description. Incontrast to SIFT, it uses the integral image for approximating the second-order derivatives for points detection. We used an 64-bin for image description. (c) **ORB** - ORB [6] uses image pyramids for scale invariance and intensity centroid for rotation invariance. However ORB may be an efficient alternative to SURF or SIFT because it works on binary descriptors. After that an 256-bit long binary string is used for image description.

2.3 Bayesian Learning

RTAB-Map uses a discrete Bayesian filter to estimate an unknown probability density function over time for possible loop closure pairs. It keeps track the loop closure by estimating the probability of the current location L_t matches one of an already visited location in the Working Memory at time t.

2.4 Ensemble Learning

When estimator of the different filters contain large errors, it can be more efficient to combine their estimated probability values by mean rule m as follows:

$$P_j^m(x^1, ..., x^n) = \frac{1}{n} \sum_{k=1}^{n} P_j^k(x^k) \tag{1}$$

where x^k is the pattern representation (bag-of-features) of the k^{th} descriptor for location j of n different locations.

3 Findings and Arguments

We have evaluated the ensemble technique on the CityCentre dataset. The dataset contains 2474 images of size 640 x 480 pixels acquired from left and right cameras. In general, the datatset contains dynamic visual images collected along public roads near the city center and has many dynamic objects such as traffic and pedestrians. The images are collected on a windy day with bright sunshine weather that create more challenging situations for recognition. The dataset is available at $http://www.robots.ox.ac.uk/$~mobile$/IJRR_2008_Dataset/$. Table 1 and Table 2 show the results of the different descriptors and combination of the different descriptors respectively.

Table 1. The recall performance observed at 100% precision for the different descriptors

	SURF	SIFT	ORB
%	82.00	81.29	65.24

Table 2. The recall performance observed at 100% precision for the different combination descriptors

	SURF + SIFT	SURF + ORB	SIFT + ORB	SURF + SIFT + ORB
%	84.12	84.85	83.78	84.67

Finally, we have compared our approach with other experimental results using the same the evaluation method. The FAB-MAP [10], PIRF-Nav2.0 [11] and RTAB-Map[9] gave 37%, 80% and 81% respectively. It shows the the ensemble can be used to increase the loop closure detection performance.

4 Conclusions

In this paper, we have introduced an approach for recognizing loop closure using an ensemble technique for combining multiple image descriptors. The results show that the mean rule handles multiple features efficiently and performs significantly better that a standard Bayesian filter.

References

1. Ulrich, I., Nourbakhsh, I.: Appearance-based place recognition for topological localization (2000)

2. Newman, P., Cole, D., Ho, K.L.: Outdoor SLAM using visual appearance and laser ranging. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Orlando Florida USA (May 2006)
3. Nister, D., Stewenius, H.: Scalable recognition with a vocabulary tree. In: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2. CVPR '06 (2006) 2161–2168
4. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision **60**(2) (2004) 91–110
5. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (surf). Comput. Vis. Image Underst. **110**(3) (2008) 346–359
6. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: Orb: An efficient alternative to sift or surf. In: Proceedings of the 2011 International Conference on Computer Vision. ICCV '11, IEEE Computer Society (2011) 2564–2571
7. Angeli, A., Filliat, D., Doncieux, S., Meyer, J.: Fast and incremental method for loop-closure detection using bags of visual words. IEEE Transactions on Robotics **24**(5) (2008) 1027–1037
8. Angeli, A., Doncieux, S., Meyer, J.A., Filliat, D.: Real-time visual loop-closure detection. In: Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on, IEEE (2008) 1842–1847
9. Labbe, M., Michaud, F.: Memory management for real-time appearance-based loop closure detection. In: Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on. (2011) 1271–1276
10. Cummins, M., Newman, P.: Fab-map: Probabilistic localization and mapping in the space of appearance. The International Journal of Robotics Research **27**(6) (2008) 647–665
11. Kawewong, A., Tongprasit, N., Tangruamsub, S., Hasegawa, O.: Online and incremental appearance-based slam in highly dynamic environments. The International Journal of Robotics Research (2010)