Pertanika J. Sci. & Technol. 25 (S): 235 - 242 (2017)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Ensemble of Bayesian Filters for Loop Closure Detection

Mohammed Omar Salameh*, Azizi Abdullah and Shahnorbanun Sahran

Pattern Recognition Research Group Center for Artificial Intelligence Technology,
Faculty of Information Science and Technology University Kebangsaan Malaysia, 43600 Bangi, Malaysia

ABSTRACT

Visual Simultaneous Localization and Mapping (vSLAM) system is widely used by autonomous mobile robots. Most vSLAM systems use cameras to analyze surrounding environment and to build maps for autonomous navigation. For a robot to perform intelligent tasks, the built map should be accurate. Landmark features are crucial elements for mapping and path planning. In the vSLAM literature, loop closure detection is a very important process for enhancing the robustness of the vSLAM algorithms. The most widely used algorithms for loop closure detection use a single descriptor. However, the performance of the single descriptors appears to worsen as the map keeps growing. One possible solution to this problem is to use multiple descriptors and combine them as in Naive and linear combinations. These approaches, however, have weaknesses in recognizing the correct locations due to overfitting and high-bias, which hinder the generalization performance. This paper proposes the usage of ensemble learning to combine the predictions of multiple Bayesian filter models which make more accurate prediction than individual models. The proposed approach is validated on three public datasets; namely, Lip6 Indoor, Lip6 Outdoor and City Centre. The results show that the proposed ensemble algorithm significantly outperforms the single approaches with a recall of 80%, 97% and 87%, with 100% precision on the three datasets, and outperforms the Naive approach and the existing loop closure detection algorithms.

Keywords: Appearance-based localization, Ensemble learning, Loop closure detection

INTRODUCTION

Visual Simultaneous Localization and Mapping (vSLAM) is an unsupervised learning problem. A robot must find its location on a map using input data from cameras and odometry. In the literature, loop closure detection is a crucial process for enhancing the robustness of vSLAM algorithms. Loop closure detection algorithms

ARTICLE INFO

Article history:
Received: 15 August 2016
Accepted: 18 May 2017

E-mail addresses:
m.omar82@siswa.ukm.edu.my (Mohammed Omar Salameh),
azizia@ukm.edu.my (Azizi Abdullah),
shahnorbanun@ukm.edu.my (Shahnorbanun Sahran)
*Corresponding Author

ISSN: 0128-7680 © 2017 Universiti Putra Malaysia Press.

aim to recognize a previously visited location from current visual sensor. Efficient recognition requires accurate feature descriptors to create distinctive landmarks. Although the single descriptor loop closure detection approaches achieve good results, they still rely on single descriptors for describing landmarks; such as, Scale-Invariant Feature Transform (SIFT) used in Fast Appearance-Based Mapping (FAB-MAP) (Cummins & Newman, 2008) and Speeded-Up Robust Features (SURF) used in Real-Time Appearance-Based Mapping (RTAB-Map) (Labbe & Michaud, 2013). In the literature, many loop closure detection approaches have combined multiple visual descriptor to improve their performance. A multiple descriptor approach normally produces different features of visual data for efficient matching between reference images and the observed ones. The main idea of using multiple descriptors to extract repeated features of the same landmark is to increase the probability of relevant features in the landmark. Nuttin and Gool used ten different descriptors and combined the Mahalanobis distance of the matching descriptor vectors (Goedem´e, Nuttin, Tuytelaars, & Van Gool, 2004). Masselli and Zell used the Weighted Gradient Orientation Histogram (WGOH) with the Weighted Grid Integral Invariant (WGII) and combined them by using the normalized histogram intersection (Weis, Masselli, & Zell, 2007). Azouaoui and Achour proposed to combine H with S from Hue Saturation Value (HSV), using color space as input to Fuzzy ART neural network used for loop closure detection (Rebai, Azouaoui, & Achour, 2014). Meanwhile, Ramisa et al. (2009) used the single feature Gradient Location and Orientation Histogram (GLOH) and extracted the features from the three region detectors (Harris-Affne, Hessian-Affne, and MSER) and naive concatenated the features (Ramisa, Tapus, Aldavert, Toledo, & De Mantaras, 2009). Angeli et al. (2009) generated two independent Bag-of-Words (BoW) descriptors using SIFT and HSV features and combined them using the likelihood function part of the Bayes filter (Angeli, Filliat, Doncieux, & Meyer, 2008). However, these algorithms have some limitations in handling high dimensional descriptors as well as numerous images. Besides, they have not tested their algorithms in the standard benchmark datasets.

This paper focuses on an ensemble method and multiple local appearance descriptors for loop closure detection for enhancing vSLAM problems. An ensemble learning method is a technique that combines the predictions from multiple machine learning algorithms to make more accurate prediction than any individual model. One possible problem of the Naive solution in combining multiple local appearance descriptors creates dense features vector. This input vector increases the problem of overfitting and hinder generalization performance. It will eventually lead to inefficient learning. Furthermore, in an unknown environment, landmarks are quite complex with highly dynamic and varying conditions. Therefore, it is hard to thoroughly describe the content of the surrounding environment in one shot due to the existence of redundant features and poor data content, which degrades the matching accuracy. Ensemble learning is used for reducing the lack of single descriptors and managing the strengths of multiple models to improve the accuracy of detection. In this paper, three feature descriptors

(SURF, SIFT and Oriented FAST and Rotated BRIEF (ORB)) are used to generate multiple Bayesian models. After that, the ensemble algorithm is used to combine the output values from the three different models for loop closure detection.

This paper is an extension to the previous work of (Salameh, Abdullah, & Sahran, 2017). However, additional information is added to the feature combination algorithms. In the feature combination, the Naive combination approach is added, and for the experiment two new datasets are used; namely, Lip6 Indoor and Lip6 Outdoor. The results show that the proposed ensemble algorithm significantly outperforms the single model by 80%, 97% and 87% obtained on the Lip6 Indoor, Lip6 Outdoor and City Centre datasets respectively. Compared with the Naive approach, the proposed ensemble loop closure detection increased by 7.08%, 1.49% and 14.48% on the Lip6 Indoor, Lip6 Outdoor and City Centre datasets respectively. The proposed ensemble loop closure detection algorithm outperforms the existing loop closure detection approaches (Angeli et al., 2008; Carrasco, Bonin-Font, & Oliver-Codina, 2016; Cummins & Newman, 2008; Kawewong, Tongprasit, & Hasegawa, 2011; Labbe & Michaud, 2013).

MATERIALS AND METHODS

The aim of loop closure detection in vSLAM is to recognize the the previously visited locations based on the current captured image. The image descriptor is a symbolic representation of the contents of an image, which is faster than using pixel values and abstractions from the image data. In this work, the state-of-the-art feature descriptors in computer vision which are SURF, SIFT and ORB are used in the proposed algorithm.

RTAB-Map (Labbe & Michaud, 2013; 2014) is used to test the proposed algorithm[1]. RTAB-Map uses a topological map with nodes representing the visited locations and contains the location signature which is a set of visual features extracted from the image belonging to the location. Furthermore, the node has an ID and weight. The weight represents the number of visits to this location. There are two types of edge. The first type is linked with the neighbor node and the second type is a loop closure between the two nodes representing the same location. RTAB-Map uses an online incremental BoW with SURF descriptors in a Kd-tree structure. The codebook is built and updated using the Fast Library for Approximate Nearest Neighbours (FLANN) with Nearest Neighbour Distance Ratio (NNDR).

Locations pass through four main stages starting with the Sensory Memory (SM) stage, where the visual features extracted from the current image and generate a visual signature using online BoW to create a new node. The second stage is Short Term Memory (STM) with First-In-First-Out (FIFO) structure and a fixed length. Its main duty is to merge two neighbouring locations in time if it has high similarity matching. When the STM stack is full, the first saved location in the stack will pass to the next stage. The third stage is Working Memory (WM) which is the active part of the memory where a location is identified as a new location or as

[1]Software available on https://github.com/introlab/rtabmap/archive/0.8.3. tar.gz

a loop closure to a previous location. RTAB-Map uses the Bayesian filter to estimate the full posterior probability $p(S_t \mid L^t)$ as Eq.[1], where L_t is the current location and S_t is a set of all loop closure candidates for the location L_t.

$$p(S_t \mid L^t) = \eta p(L_t \mid S_t) \sum_{i=-1}^{t_n} p(S_t \mid S_{t-1} = i) p(S_{t-1} = i \mid L^{t-1}) \quad [1]$$

where η is a normalization term and $L^t=L_{-1},...,L_t$ ffor locations in the search space of the Bayesian filter. The size of the WM changes dynamically according to the real-time operational constraints. If the operation time exceeds the time constraint, RTAB-Map transfers the oldest and the less visited location from WM to the Long Term Memory (LTM), where the rest of the locations are retained. In the final stage, if the current location is registered as a loop closure, WM will retrieve the two neighbours for the current loop closure location where these neighbours have a high probability to be the next loop closure. Ensemble learning has a high potential to solve the weaknesses identified in the previously mentioned approaches (Angeli et al., 2008; Cummins & Newman, 2008; Goedem´e et al., 2004; Labbe & Michaud, 2013; Ramisa et al., 2009; Rebai et al., 2014; Weiss et al., 2007). Ensemble learning can manage the lack of information by utilizing different methods for modeling decisions. The ensemble learning overcomes the problem of overfitting of the large feature vectors and the multiple Bayesian models that support each other to obtain the final decision.

The proposed algorithm generates three Bayesian models using three different feature descriptors SURF, SIFT and ORB and integrates at decision-level using ensemble learning as shown in Figure 1. Three independent codebooks are generated from three different visual features for landmark description.

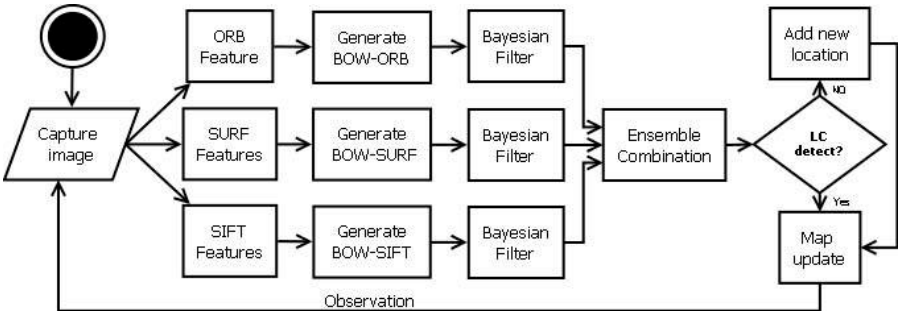

Figure 1. Ensemble approach for loop closure detection

The ensemble learning algorithm is based on the Bayesian filter for making decision for the loop closure detection. A new location is nominated by looking at a number of candidate locations and the Bayesian filter provides a probability value for making the decision. Three ensemble rules are tested to generate the final output; namely, the mean rule, the product rule and the majority voting rule where the product rule and the majority voting give the worst

results which are not reported in this paper. The mean rule is used to combine the outputs from the three Bayesian models using the following equation:

$$Ens(L) = \frac{1}{d}\sum_{i=1}^{d} p\left(S^i\right)_t \mid \left(L^i\right)^t \tag{2}$$

where $(L^i)^t$ is the location associated with the descriptor i=1...d, and d is the number of descriptors. The proportion of matching $p((S^d)_t \mid (L^d)^t)$ is computed using Eq.2. This rule is selected for the proposed algorithm because it can deal with the probability distribution for the Bayesian filter output.

Naive Approach: is an intuitive approach for combining multiple features as a single input used for loop closure detection. In this paper, the Naive approach is implemented and is used as a reference method to compare the proposed algorithm. The Naive approach combines the three feature descriptors (SURF, SIFT and ORB) into a single vector that is used as input for the single Bayesian filter. The Naive approach combines the three descriptors (d_{SURF}, d_{SIFT} and d_{ORB}) as follows:

$$d_{Naive} = d_{SURF} + d_{SIFT} + d_{ORB} \tag{3}$$

where + is the concatenation operator and d_{Naive} is the input for the Bayesian filter.

Public datasets are used to evaluate the proposed algorithm under different conditions. The datasets used are City Centre[2] (Cummins & Newman, 2008) with 1237 images and the captured at rate 0.5Hz. Lip6 Indoor (Angeli et al., 2008) with 388 indoor images and the captured at rate 1Hz. Lip6 Outdoor3 (Angeli et al., 2008) with 1063 outdoor images and the captured at rate 0.5Hz. These datasets test the capabilities of vSLAM in both indoor and outdoor environments. The precision recall curve is a very popular evaluation approach that gives more performance information when highly skewed datasets are used (Labbe & Michaud, 2013; Cummins & Newman, 2008; Abdullah, Veltkamp, & Wiering, 2010). Precision is the number of true positive loop closure detections to the total number of detections, and Recall is the number of true positive loop closure detections to the number of ground truth loop closures. RTAB-Map parameters are setup according to the values reported in (Labbe & Michaud, 2013). The visual features SURF, SIFT and ORB use the default parameters as reported in OpenCV lib. The experimentation is done on an Intel PC, Core i5 2.90GHz, 4Gb RAM and OS Ubuntu 14.04.

RESULTS AND DISCUSSION

Table 1 shows the Recall (%) of the loop closure detection at 100% Precision test on Lip6 Indoor, Lip6 Outdoor and City Centre datasets. Table 1 shows the results of the three single descriptor SURF, SIFT and ORB used with a single model and the combination models.

[2]Dataset available on http://www.robots.ox.ac.uk/~mobile/IJRR_2008_Dataset/
[3]Dataset available on http://cogrob.ensta.fr/loopclosure.html

Table 1
The recall rates quoted are at 100% precision for each method based on the three datasets. NAIVE = combines the three features and uses the single Bayesian model. ENSEMBLE = the proposed algorithm uses the three features to construct multi-model loop closure detection

	Single Approach			Combined Approach	
Dataset	SURF	SIFT	ORB	NAIVE	ENSEMBLE
Lip6 Indoor	74.77	38.49	2.21	73.45	80.53
Lip6 Outdoor	95.3	24.16	59.86	96.3	97.79
City Centre	86.36	85.02	64.39	73.44	87.92

Our experiment shows the comparison among the single descriptor model, the Naive approach and the proposed ensemble algorithm. The result of the experiment shows that the loop closure detection using the SURF descriptor outperforms the SIFT and ORB descriptors based on the three datasets. This was due to the high capability of the SURF in recognizing locations. Furthermore, the loop closure detection using the ORB with Lip6 Indoor shows the worst results because the nature of the image of Lip6 Indoor was collected from the corridor of a building with the least texture. The indoor images contain less information than the outdoor images (Kawewong et al., 2011). Comparing between the single and the combined approaches, the experiments show that the Naive approach outperforms the single descriptor methods and the proposed ensemble Bayesian filter outperforms the Naive and single descriptor methods. The proposed algorithm using ensemble learning with multiple features and multi-model Bayesian filter can handle and improve the loop closure detection in different environments (indoor and outdoor). Moreover, the proposed algorithm is compared with other loop closure detection algorithms, where Table 2 shows this comparison. The proposed algorithm clearly outperforms the single descriptor methods and the other loop closure algorithms. However, RTAB-Map for Lip6 Indoor is higher than the proposed algorithm because RTAB-Map tunes the SURF parameters as reported in (Labbe & Michaud, 2013) and the proposed algorithm uses the default parameters.

Table 2
Comparison among the five different existing approaches. The recall rates quoted are at 100% precision. All algorithms use Bayesian Filter for loop closure detection

	Lip6 Indoor	Lip6 Outdoor	City Center	Features
Proposed Method	80.53	97.79	87.92	SURF, SIFT, ORB
(Carrasco et al., 2015)	74	76	-	SIFT
(Labbe and Michaud, 2013) "RTAB-Map"	98	95	81	SURF
(Kawewong et al., 2011) "PIRF-Nav2.0"	78	-	80	SURF
(Angeli et al., 2008)	80	71	-	SIFT, ColorHist
(Cummins and Newman, 2008) "FAB-MAP"	-	-	37	SIFT

CONCLUSION

The mean rule ensemble of the Bayesian filter model significantly outperforms the individual Bayesian filter models according to the t-test ($p< 0.05$). This is because of the ensemble algorithm being able to improve the independent models if the model predictions are less correlated. The experiments also show that combining the strong and the weak models with the ensemble learning will improve the loop closure detection results because the ensemble algorithm can make a better trade-off between the different performances of the models. Besides the mean rule, we also experimented on other ensemble learning rules; namely, the product rule and the majority voting rule (Polikar, 2006). However, these approaches gave the worst results of 7.6% and 20.1%, respectively, using the City Centre dataset. This is due to: (a) the large probability estimation errors for all nodes of the different feature descriptor models in the working memory of the topological map, and (b) the number of identical or similar nodes of these models which are not similar. These reasons affect the final probability outputs of the ensemble learning algorithm.

The experiments show that each Bayesian filter model nominates different locations for the same input image. By tracking the processes and monitoring the WM of each descriptor, we believe that some locations are missing from each WM, as a result of having been transferred to the LTM. In future work, we plan to focus on recognizing and retrieving the missing locations that support the Bayesian filter model with more completed search space.

ACKNOWLEDGEMENT

The authors would like to extend their appreciation and gratitude to ETP-2013-053 grant for funding this project, and to Mr Omar Hammad for the proofreading.

REFERENCES

Abdullah, A., Veltkamp, R. C., & Wiering, M. A. (2010, December). Ensembles of novel visual keywords descriptors for image categorization. In *11th International Conference in Control Automation Robotics and Vision (ICARCV)* (pp. 1206-1211). IEEE.

Angeli, A., Filliat, D., Doncieux, S., & Meyer, J. A. (2008). Fast and incremental method for loop-closure detection using bags of visual words. *IEEE Transactions on Robotics, 24*(5), 1027-1037.

Carrasco, P. L. N., Bonin-Font, F., & Oliver-Codina, G. (2016). Global image signature for visual loop-closure detection. *Autonomous Robots, 40*(8), 1403-1417.

Cummins, M., & Newman, P. (2008). FAB-MAP: Probabilistic localization and mapping in the space of appearance. *The International Journal of Robotics Research, 27*(6), 647-665.

Goedemé, T., Nuttin, M., Tuytelaars, T., & Van Gool, L. (2004, May). Markerless computer vision based localization using automatically generated topological maps. In *Proceedings of the European Navigation Conference* (Vol. 1, pp. 235-243).

Kawewong, A., Tongprasit, N., & Hasegawa, O. (2011). PIRF-Nav 2.0: Fast and online incremental appearance-based loop-closure detection in an indoor environment. *Robotics and Autonomous Systems, 59*(10), 727-739.

Labbe, M., & Michaud, F. (2013). Appearance-based loop closure detection for online large-scale and long-term operation. *IEEE Transactions on Robotics*, *29*(3), 734-745.

Labbe, M., & Michaud, F. (2014, September). Online global loop closure detection for large-scale multi-session graph-based slam. In *Institute of Electrical and Electronics Engineers / Robotics and Automation Society International Conference in Intelligent Robots and Systems (IROS)* (pp. 2661-2666). IEEE.

Polikar, R. (2006). Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine*, *6*(3), 21-45.

Ramisa, A., Tapus, A., Aldavert, D., Toledo, R., & De Mantaras, R. L. (2009). Robust vision-based robot localization using combinations of local feature region detectors. *Autonomous Robots*, *27*(4), 373.

Rebai, K., Azouaoui, O., & Achour, N. (2014, May). HS combined histogram for visual memory building and scene recognition in outdoor environments. In *ICRA Workshop Visual Place Recog*., Hong Kong.

Salameh, M. O., Abdullah, A., & Sahran, S. (2017). Ensemble of vector and binary descriptor for loop closure detection. In *Robot Intelligence Technology and Applications 4* (pp. 329-340). Springer, Cham.

Weiss, C., Masselli, A., & Zell, A. (2007). Fast vision-based localization for outdoor robots using a combination of global image features. *IFAC Proceedings Volumes*, *40*(15), 119-124.