Robust Semantic Place Recognition with Vocabulary Tree and Landmark Detection

Lin Yuan, Kai Chi Chan and C.S. George Lee

***Abstract*—Semantic place recognition problem has attracted growing interests in autonomous robots to expand their application domain. Due to the large in-class variance in semantic place recognition, the recognition performance has been lackluster. In this paper, we hypothesize that the large in-class variance is due to the fact that connections between places cannot be suitably assigned a label. We verified this hypothesis on the COLD localization database. We then propose a robust method that can effectively detect these connections (landmarks), thus improving the accuracy of semantic place recognition systems. The proposed method uses image sequences for landmark detection instead of a single image, thus providing robust results which can be used for topological mapping for mobile robots under different lighting conditions.**

***Index Terms*—Semantic Place Recognition, Bag-of-Words, Visual Vocabulary, Dynamic Time Warping**

I. Introduction

The semantic place recognition of an environment that a robot is traveling will be helpful in autonomous navigation and various human-robot interaction tasks. Efforts in semantic place recognition or classification have emerged since 2005. The semantic place classification problem refers to distinguishing differences between different environmental locations (*e.g.* distinguishing a kitchen from an office). The semantic place recognition problem refers to differentiating different locations when they even may be of the same type (*e.g.* distinguishing office A from office B). Researchers first employed range sensors to solve the semantic classification problem. The distance measurements from range sensors provide a nature information about how cluttered the environment is. These measurements form well distinguishable features for different type of environments. Mozos *et al.* [1] proposed using AdaBoost algorithm for classifying different type of semantic environments (*e.g.* rooms, hallways, doorways, etc.) from range sensor readings. Various geometric measurements calculated from laser range data are then used as weak features for AdaBoost.

As robust feature extraction methods are developed in computer vision [3], [4], vision-based localization methods become a popular research topic and experiments with visual sensors have been carried out to improve the recognition performance [5], [6]. Our hypothesis (misclassification happens mostly at connection between semantic places) is inspired by research [7], [8] in the vision-based localization context. For simplifying naming conventions, we define "landmarks" to be an area on a 2D map where two semantic place units join, which is similar to Ranganathan *et al.* [9]. From now on, we will use the term "landmark" to stand for connections between two adjacent semantic places.

Lin Yuan, Kai Chi Chan and C.S. George Lee are with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, U.S.A. Email: {yuanl, chan56, csglee}@purdue.edu

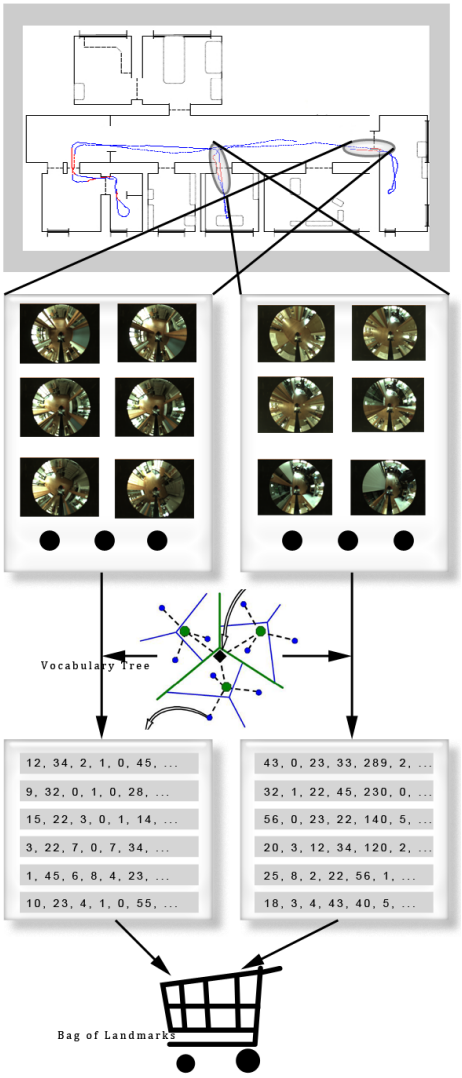


Fig. 1. Landmarks represented with time series of visual words. A part of the figure was adapted from Figure 1 of Nister and Stewenius [2]

II. Related Work

Semantic place recognition and topological mapping share a close relationship because each semantic place is usually a node in the graph of a topological map. And the semantic recognition problem appears earlier in the context of topological mapping. Tapus and Siegwart [10] used line and corner features in omni-directional images to generate signatures for each semantic location. Thus the topological map is represented by a collection of signatures. With a

hierarchical SLAM system, Kouzoubov and Austin [11] can locate the robot in a topological map. Friedman *et al*. [12] used Voronoi random fields to extract the topological structure of an indoor metric map. Ranganathan *et al*. [9] proposed a topological mapping algorithm without metric map. In these topological mapping literatures, a node in the topological map is usually a room on the floor plan, which is a semantic place unit.

Later on, researchers employed Nearest Neighborhood (NN) classification method for vision-based localization. In these NN classification systems, observations are made that misclassification happens mostly at "landmark" positions. Using visual cameras, Zivkovic *et al*. [7] employed SIFT features within images to match against the database images in order to locate a robot to a semantic location. The semantic location is represented with a node in the topological map built with graph-partitioning algorithms. They tested their localization performance by taking one image from the database, matching it against all other images in the database and assigning the nearest neighbor's class to the image. The database is simply a collection of images from one run of the robot at a specific time (i.e., no lighting variations). Thus they were able to achieve 90% recognition rate. However, they pointed out that the misclassification happens mostly at the boundaries between different partitions of their map. Knopp *et al*. [8] also made similar observations. They suppressed confusing features from being used for image classification to achieve higher classification rate. Valgren and Lilienthal [13] investigated the impact of different lighting conditions on the SIFT and SURF descriptors for vision-based localization systems. They pointed out that vision-based localization systems cannot achieve good recognition rate with training and testing sets across different lighting conditions based on a single image.

In recent years, the appearance-based, loop-closure problem has gained significant improvement. In the FAB-MAP system proposed by Cummins and Newman [14], they employed a bag-of-words model for images and used the k-means clustering to generate a visual vocabulary. The vocabulary itself carries moderate information about the location where an image was taken, but is pruned to various conditions. They further employed a graphical model, called Chow-Liu tree, to capture the correlation between those visual words. The resulting learned graphical model significantly helped in the perceptual-aliasing problem.

This paper proposes a new framework under which novel methods can be developed to effectively detect "landmarks" to improve the performance of semantic place recognition (*c.f*. Fig. 1) over alternative methods in [5], [6]. The proposed approach is inspired by [7], [8], [14], but uses a different method for generating vocabulary [2]. The proposed method is based on forming a time-series sample of Bag-of-Landmarks from a sequence of images. We call our method as BoLTS for Bag-of-Landmarks using time series. We generate the visual "landmarks" for a small number of images (10-80) within an image sequence where "landmarks" are identified by a human. Given image sequences collected by the robot, a human picks up segments of images where there is a "landmark". By using the visual vocabulary built with [15], a simplified version of [2], we obtained a high recognition rate for "landmarks". And by removing these "landmarks" from the training set for the semantic place recognition task, we improved the recognition rate, thus validating our hypothesis that misclassifications happen mostly at "landmark" positions. BoLTS advances the state-of-art of visual landmark recognition, by broadening the data-format into time-series with image sequences.

III. Robust Semantic Place Recognition

Our robust semantic place recognition approach consists of two components: a vocabulary tree image classifier and a landmark detector based on time-series pattern matching. We use the vocabulary-tree method [2] to generate signatures for images, which is a histogram of visual words from a pre-built visual vocabulary. These signatures are later used to form an image classifier for semantic place recognition. In order to address the specific boundary issue (*c.f*. Fig. 4) faced in semantic place recognition task, we propose to develop a time-series pattern matching approach, called "Bag-of-Landmarks using time series," for detecting "landmarks".

A. Vocabulary Tree

Bag-of-words modeling of images has been introduced by Sivic and Zisserman [16]. Clustering methods are typically employed for building the visual word dictionary. Several clustering methods (*e.g*. k-means, vocabulary tree, etc.) have been incorporated into an appearance-based localization system [14]. In this section, we will demonstrate how a visual word dictionary built with the vocabulary-tree method can be used for semantic place recognition.

The vocabulary-tree method [2] is a hierarchical iterative k-means clustering method with parent nodes being a quantization representation of their children. Even though the k-means clustering proved to be effective in various applications, we find that the visual word extracted by the vocabulary-tree method together with SIFT features is more consistent under moderate change in lighting condition. This is a crucial factor for robotics applications because the more reproducible the word is, the better the place recognition will be. Thus, we choose to use the vocabulary-tree method for our bag-of-words model with SIFT features.

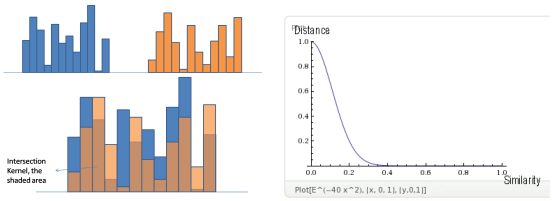
Given the signature of images, typical image categorization methods will build up a nearest neighborhood (NN) classifier from the image database, and then perform image classification. When directly applying this method to semantic place recognition, we face a boundary issue. In these cases, the images collected at the "landmark" positions are difficult for people to assign label for training. These "landmarks", as we mentioned before, are important landmarks in topological mapping [17]. Hence, we propose a landmark detection method based on time-series pattern matching, in which each index of the time series is an image signature. If a "landmark" is detected, we can preclude these images from being used for semantic place recognition, hence improving

the recognition rate. In order to do the time-series matching, the distance between two signatures needs to be matched. Instead of using the normalized difference measure [2], we use a Gaussian histogram intersection kernel measure. Define N as the number of leaves in the vocabulary-tree (dictionary size). Given a query signature $\{s_q : s_q^1 \dots s_q^N\}$ and a database signature $\{s_d : s_d^1 \dots s_d^N\}$, the histogram intersection kernel similarity $sim(s_q, s_d)$ is described in Eq. (1). Note that the similarity ranges from 0 to 1.

$$sim(s_q, s_d) = \frac{\sum_{k=1}^{N} (\min(s_q^k, s_d^k))}{\sum_{k=1}^{N} s_d^k} \quad (1)$$

Next, we need to generate a distance measure based on this similarity measure. Using images found in the database, we find that the similarity measure for two images taken within close proximity to one another can only peak around 0.2. This indicates that there cannot be a linear relationship between the similarity and the distance. Through empirical experience, we find that the Gaussian function e^{-kx^2} in the range of [0,1] is a fit for the data. Equation (2) shows the distance $diff(s_q, s_d)$ generated from the histogram intersection kernel. In practice, we used $k = 40$ for our experiments. Figure 2 further illustrates the histogram intersection kernel and the relationship between distance and similarity.

$$diff(s_q, s_d) = e^{-k sim(s_q, s_d)^2} \quad (2)$$



(a) (b)

Fig. 2. (a) Histogram intersection kernel (b) Relationship between similarity and distance measure

B. Dynamic Time Warping

Dynamic time warping (DTW) is a well-known method for matching time-series data. The advantage of using DTW over other time-series matching algorithms is that the matching between two time series can be of variable length. This advantage is especially suitable for matching image sequences collected by a robot because a robot usually can vary its velocity, thus the number of images collected by the robot over a specific range can be quite different. DTW is also a suitable choice for other distance traveled based schemes [14] since it handles the overlap of images well. Meanwhile, by applying DTW, we assume that the robot traverses through the "landmark" position in a fixed manner. This is generally not a problem if the "landmark" is like a doorway connecting two places, where the only option for the robot is to "go in" or "go out".

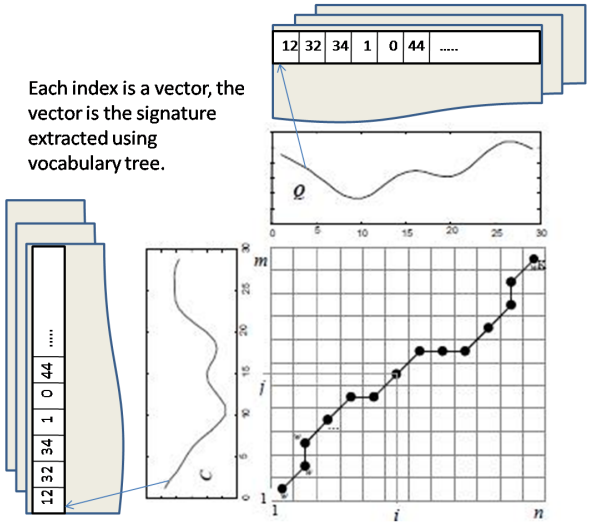


Fig. 3. Iterative Vector Dynamic Time Warping. Adapted from Figure 5 of Chu *et al.* [18]

The DTW algorithm is a dynamic programming algorithm and is described in detail in [18] and [19]. Following Fig. 3, each time series can react with an elastic behavior such that each index of the query time series can find its best match with the index of the reference time series. When matching two time series $[L_q : s_1..s_n]$ and $[L_d : s_1..s_m]$, the cost of matching s_i and s_j, where $1 \leq i \leq n$ and $1 \leq j \leq m$ is described in Eq. (3).

$$DTW(s_i, s_j) = diff(s_i, s_j) + \min(DTW(s_{i-1}, s_j), DTW(s_i, s_{j-1}), DTW(s_i, s_j)) \quad (3)$$

C. Bag-of-Landmarks using Time Series (BoLTS)

In order to apply the DTW algorithm to robot systems, we need to solve the computational issue due to the nature of image sequences collected by the robot. When a new image comes from the visual sensor, the robot will need to concatenate it with variable length of buffered historic images as a query time series L_q. Then the set of variable length query time series is used to match against the "landmarks" of time series trained beforehand, because the "landmarks" of time series in database can have different length. The traditional DTW used here will bring about significant amount of redundant computation. Thus, we save the distance between buffered signatures after each DTW is done in an global buffer. We refer to our modified version of DTW as Iterative Vector Dynamic Time Warping (IV-DTW). Using IV-DTW, whenever a new image is collected, only one more distance needs to be calculated.

By using the vocabulary-tree method with our adapted histogram intersection kernel and modified DTW algorithm, the entire BoLTS for detecting "landmarks" can be described using Algorithm 1. The IV-DTW algorithm is described in Algorithm 2.

Algorithm 1 : Bag-of-Landmarks detection using Time Series

Require: m landmarks $L_1 \dots L_m$ in database. $\forall i$, $L_i = \{s_1 \dots s_{k_i}\}$, k_i is the length of the image sequence used for landmark i.
Initialize $B = 80$, $minlen = 10$
Initialize buffer D, d with size $B \times B$.
for $t = 1$ to T **do**
  Collect new image I_t, generate signature s_t.
  for $j = minlen$ to B **do**
    $dist_j$ = IV-DTW($[s_{t-j}..s_t]$,$\{L_1;..;L_m\}$,D,d)
  end for
  store $Dist_t = \min_j(dist_j)$
end for
Output: The local minimums for $Dist$ below some threshold will be the detected landmarks.

Algorithm 2 : Iterative Vector Dynamic Time Warping (IV-DTW)

Require: m landmarks $L_1 \dots L_m$ in database. $\forall i$, $L_i = \{s_1 \dots s_{k_i}\}$, k_i is the length of the image sequence used for landmark i. Query time series (sequence of signatures): $L_q = \{s_1 \dots s_{k_q}\}$. Buffered distance matrix: d and buffered cumulative distance D.
for $i = 1$ to m **do**
  if d is empty **then**
    d=$[\bigcup_{k_1=1}^{k_i} \bigcup_{k_2=1}^{k_q} diff(s_{k_1}, s_{k_2})]$
  end if
  Initialize $D_c = [d, D]^T$
  for every $d(u, v)$ **do**
    $D_c(u, v) = d(u, v) + \min(D_c(u+1, v), D_c(u+1, v+1), D_c(u, v+1))$
  end for
  Compute $dist_i$ between L_q and L_i by backtracking from the top-right cell of D_c
end for
Output: $\min_{1 \le i \le m} (dist_i)$.

D. Integration

As we have introduced the vocabulary-tree method and the BoLTS method, we will put these two components together to build our semantic place recognition system. The procedure of our system is described in Algorithm 3.

Algorithm 3 : Robust Semantic Place Recognition

Require: m landmarks $L_1 \dots L_m$ in database. $\forall i$, $L_i = \{s_1 \dots s_{k_i}\}$, k_i is the length of the image sequence used for landmark i.
for $t = 1$ to T **do**
  Collect new image, generate signature using vocabulary tree.
  if BoLTS report landmark detection **then**
    Mark the matched time series as landmark.
    Don't do semantic place recognition.
  else
    Do semantic place recognition
  end if
end for

IV. Experimental Work

We validated our semantic place recognition system on the COsy Localization Database (COLD) [20]. The "bag-of-words" and "siftpp" code written by Vedaldi [15] is used for generating signatures using the vocabulary-tree method. We evaluated the performance of our system using the 10 image sequences collected on Path A within the COLD-Freiburg set. First, we demonstrated that the places for misclassification using the vocabulary-tree method are located at "landmark" positions. Next, we use each image sequence as the test set and the "landmarks" in the other 9 image sequences to form the "bag-of-landmarks". The landmark detection rates of BoLTS for all 10 image sequences are discussed. Finally, we compared our system (with and without pruning using BoLTS) to existing methods [5], [6] and demonstrated the robustness of our proposed method.

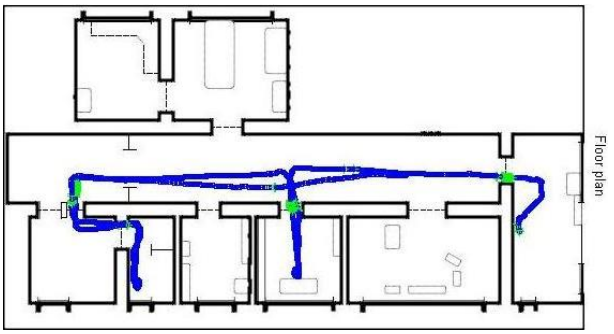


Fig. 4. Misclassified places on the map.

TABLE I
Comparison of Confusion Matrix

(a) Confusion matrix with pruning; (b) Confusion matrix without pruning

(a)

	P_1	P_2	P_3	P_4	P_5
P_1	0.9939	0	0.0061	0	0
P_2	0.0103	0.9897	0	0	0
P_3	0	0	1	0	0
P_4	0	0	0	1	0
P_5	0	0	0	0.0273	0.9727

(b)

	P_1	P_2	P_3	P_4	P_5
P_1	0.9351	0.0227	0.0185	0.0237	0
P_2	0.0163	0.9837	0	0	0
P_3	0.0181	0	0.9819	0	0
P_4	0	0	0	0.9655	0.0345
P_5	0	0	0	0	1

A. Importance of Landmarks

Inspired by Zivkovic *et al.* [7], our hypothesis is that misclassification happens mostly at "landmark" positions. To verify our hypothesis, we used seq1_cloudy1 as a training set and tested the semantic place recognition using only the vocabulary-tree method (VTM). There are 5 semantic places on Path A, and we trained the vocabulary tree with 100 images per place. As shown in Fig. 4, the green marks indicated the images that got misclassified. The obtained confusion matrix for 5 places are compared in Table I. We can see that pruning images at "landmark" positions improved the recognition rate. Furthermore, we trained another vocabulary tree using 3 sequences (seq1_sunny1, seq1_cloudy1, seq1_night1), with 100 randomly sampled images per place. Then we tested the semantic place recognition system on 7 other sequences. The ratio of misclassified images taken at boundaries over all misclassified images is shown in Fig. 5.

B. Landmark Detection Performance

The COLD-Freiburg database [20] was used to test the performance of landmark detection. In the database, there were 10 sequences of images collected at different times

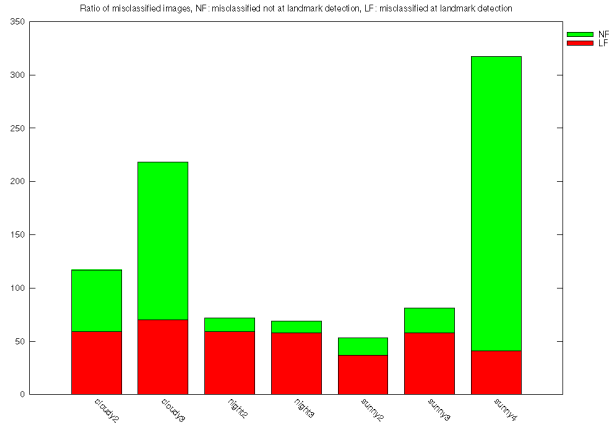


Fig. 5. Misclassification ratio. Red bar represents the misclassified images at boundaries.

under various lighting conditions. Leave-one-out cross-validation was conducted on the 10 image sets to estimate the accuracy of the detection algorithm in practice.

In the training image sets, 8 "landmarks" were trained based on the bag-of-words generated with vocabulary tree. Then, the test images were compared to the "landmarks" using the proposed BoLTS. Figure 6 shows the distance measure for different testing image sets. Each point on the curve represents the minimum distance among distances to the 8 "landmarks" and a sequence of 10 to 80 images ending at that point. The local minima of the curve indicate the ending time of the matching with different "landmarks". Assuming that every landmark is detected at least 60 frames from each other and by sorting the distances of points at different frames, the 8 "landmarks" with at least 60 frames from each other could be found. * The end of time series (each matched landmark) is marked by a red segment in Fig. 6.

For each "landmark" detected, the corresponding sequence of images was retrieved. The accuracy of landmark detection was computed based on the sequence of images. The false positive rates of landmark detection are shown in Table II for all 10 image sequences. Since the ground truth of sequences of images at different "landmarks" were not available, the false positive rates were estimated manually by checking every image in the detected sequence of images. If the image is manually accepted as taken at a "landmark" position, it is treated as a landmark image. Detection results are shown in Fig. 7. In Fig. 7, which shows one of the detection results, the blue line represents the path of a moving robot, and the red line segments indicate the detected "landmarks". All the "landmarks" in the seq1_cloudy3 sequence were detected correctly.

C. Improved Semantic Place Recognition System

We compare our semantic place recognition system with similar visual place classification experiments performed on the same COLD-Freiburg database. We name our methods VTM (vocabulary-tree method) and VTBL(vocabulary tree with bag-of-landmarks). Wang and Lin [6] used a Hull Census Transform (HCT) method to generate features for each omni-directional image and used this HCT feature together

*With a robot collecting images at 30 frames/sec, this means that the robot cannot travel from one "landmark" to another "landmark" within 2 seconds, which is reasonable since a robot usually cannot go from one door to another door within 2 seconds.

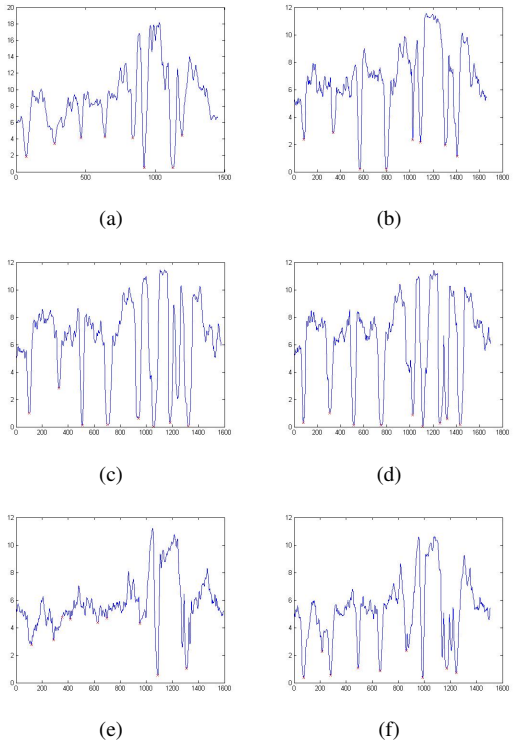


Fig. 6. Extracted IV-DTW distance for example robot run, the red crosses are identified "landmark" positions. The sequence from top-left to bottom-right is (a) cloudy1, (b) cloudy3, (c) night2, (d) night3, (e) sunny1, (f) sunny2. The local minimas show that "landmarks" are properly detected.

TABLE II
RECOGNITION RATES FOR LANDMARK DETECTION

Test	TP%	FP%	TN%	FN%	Total No.
cloudy1	17.2	0	82.8	0	1459
cloudy2	14.0	1.0	84.3	0.7	1632
cloudy3	14.5	0	85.5	0	1672
night1	15.6	0.7	83.7	0	1911
night2	16.1	0.9	81.9	1.0	1582
night3	19.0	0.9	80.1	0	1703
sunny1	11.3	0.9	81.2	6.6	1598
sunny2	17.8	1.0	81.2	0	1514
sunny3	13.8	1.8	82.7	1.7	1451
sunny4	10.4	1.5	83.2	4.9	1777

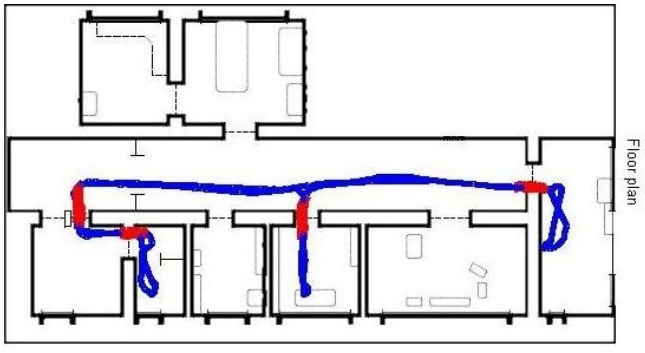


Fig. 7. An example run of landmark detection on the floor plan tested with seq1_cloudy3 image sequence.

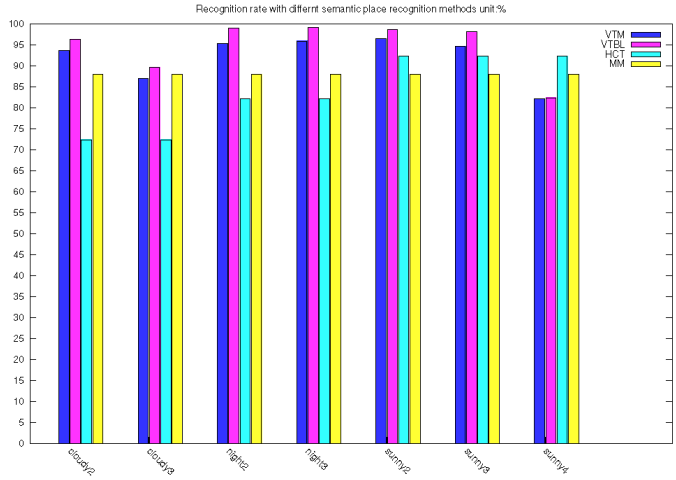


Fig. 8. Comparison of different semantic place recognition algorithms on COLD-Freiburg database. VTM: vocabulary tree without pruning; VTBL: vocabulary tree with bag-of-landmarks pruning; HCT: Hull Census Transform; MM: Multi-model semantic place classification. The result used for HCT is obtained from [6], where their best result for each weather is chosen. The result used for MM is an averaged result of their classification rate on all 3 sequences reported in their paper [5]. It is unknown which 3 sequences of COLD-Freiburg did they use.

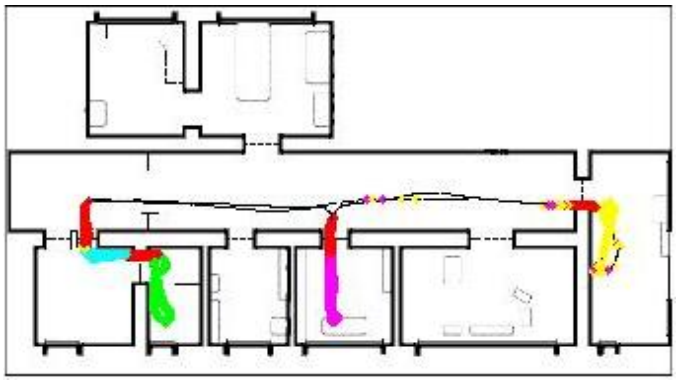


Fig. 9. Final run with our approach for seq1_cloudy3, which has 89% recognition rate.

with SVM for semantic place recognition. Pronobis *et al*. [5] reported the multi-model place classification performance on the COLD-Freiburg database. Figure 8 shows that both of our methods, VTM and VTBL, outperformed existing work. An example run with a final label of the 5 semantic places as well as detected "landmarks" is shown in Fig. 9. More details can be found in the video attachment of this paper.

V. Discussion and Future Work

This paper proposed and developed a semantic place recognition system with vocabulary tree and BoLTS. The proposed system yielded significant improvement over existing methods for the semantic place recognition task. The proposed landmark detection method (BoLTS) is a time-series-based supervised learning approach, which is novel in visual landmark detection context. However, the preparation of training set of transition places may be very dependent on robot trajectory. We will use of salient features other than SIFT and incorporating probabilistic framework like [14] into our landmark detection method to achieve a more robust and general time-series landmark detector, thus requiring no human assistance for marking image segments.

References

[1] O. M. Mozos, C. Stachniss, and W. Burgard, "Supervised learning of places from range data using AdaBoost," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Barcelona, Spain, 2005, pp. 1742–1747.

[2] D. Nister and H. Stewenius, "Scalable recognition with a vocabulary tree," in *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, vol. 2, 2006, pp. 2161 – 2168.

[3] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, 2004.

[4] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Surf: Speeded up robust features," *Computer Vision and Image Understanding (CVIU)*, vol. 110, no. 3, pp. 346–359, 2008.

[5] A. Pronobis, L. Jie, and B. Caputo, "The more you learn, the less you store: Memory-controlled incremental svm for visual place recognition," *Image and Vision Computing*, vol. 28, no. 7, pp. 1080 – 1097, 2010.

[6] M.-L. Wang and H.-Y. Lin, "An extended-hct semantic description for visual place recognition," *The International Journal of Robotics Research*, vol. 30, no. 8, July 2011.

[7] Z. Zivkovic, B. Bakker, and B. Krose, "Hierarchical map building using visual landmarks and geometric constraints," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2005, pp. 2480–2485.

[8] J. Knopp, J. Sivic, and T. Pajdla, "Avoiding confusing features in place recognition," in *European Conference on Computer Vision*. Springer Berlin / Heidelberg, 2010, vol. 6311, pp. 748–761.

[9] A. Ranganathan, E. Menegatti, and F. Dellaert, "Bayesian inference in the space of topological maps," *IEEE Transactions on Robotics*, vol. 22, no. 1, pp. 92 – 107, Feburary 2006.

[10] A. Tapus and R. Siegwart, "Incremental robot mapping with fingerprints of places," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, August 2005, pp. 2429 – 2434.

[11] K. Kouzoubov and D. Austin, "Hybrid topological/metric approach to slam," in *IEEE International Conference on Robotics and Automation*, vol. 1, 2004, pp. 872 – 877 Vol.1.

[12] S. Friedman, H. Pasula, and D. Fox, "Voronoi random fields: extracting the topological structure of indoor environments via place labeling," in *IJCAI'07: Proceedings of the 20th international joint conference on Artifical intelligence*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2007, pp. 2109–2114.

[13] C. Valgren and A. J. Lilienthal, "Sift, surf & seasons: Appearance-based long-term localization in outdoor environments," *Robotics and Autonomous Systems*, vol. 58, no. 2, pp. 149 – 156, 2010.

[14] M. Cummins and P. Newman, "FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance," *The International Journal of Robotics Research*, vol. 27, no. 6, pp. 647–665, June 2008.

[15] A. Vedaldi, "Vlfeat." [Online]. Available: http://www.vlfeat.org/ vedaldi/code/bag/bag.html

[16] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," *IEEE International Conference on Computer Vision*, vol. 2, p. 1470, 2003.

[17] A. Ranganathan and F. Dellaert, "Bayesian surprise and landmark detection," in *IEEE International Conference on Robotics and Automation*, May 2009, pp. 2017–2023.

[18] S. Chu, E. Keogh, D. Hart, and M. Pazzani, "Iterative deepening dynamic time warping for time series," in *Proceedings of 2nd SIAM International Conference on Data Mining*, 2002.

[19] X. Xi, E. Keogh, C. Shelton, L. Wei, and C. A. Ratanamahatana, "Fast time series classification using numerosity reduction," in *Proceedings of the 23rd international conference on Machine learning*. New York, NY, USA: ACM, 2006, pp. 1033–1040.

[20] A. Pronobis and B. Caputo, "COLD: COsy Localization Database," *The International Journal of Robotics Research (IJRR)*, vol. 28, no. 5, May 2009.