High-Level Visual Features for Underwater Place Recognition

Jie Li, Ryan M. Eustice, and Matthew Johnson-Roberson

***Abstract*— This paper reports on a method to perform robust visual relocalization between temporally separated sets of underwater images gathered by a robot. The place recognition and relocalization problem is more challenging in the underwater environment mainly due to three factors: 1) changes in illumination; 2) long-term changes in the visual appearance of features because of phenomena like biofouling on man-made structures and growth or movement in natural features; and 3) low density of visually salient features for image matching. To address these challenges, a patch-based feature matching approach is proposed, which uses image segmentation and local intensity contrast to locate salient patches and HOG description to make correspondences between patches. Compared to traditional point-based features that are sensitive to dramatic appearance changes underwater, patch-based features are able to encode higher level information such as shape or structure which tends to persist across years in underwater environments. The algorithm is evaluated on real data, from multiple years, collected by a Hovering Autonomous Underwater Vehicle for ship hull inspection. Results in relocalization performance across missions from different years are compared to other traditional methods.**

I. Introduction

Recognizing a place that has been previously viewed is an important challenge in both robotic navigation and computer vision. It is a prerequisite for visual-based navigation systems, and also important for long-term robot operation using perception. Several robust approaches have been proposed in the last decade for place recognition in the terrestrial domain. However, place recognition in the underwater environment is a more challenging problem. Dramatic changes in scene appearance occur due to viewpoint dependent illumination (underwater robots often carry their own light source) or biofouling, both of which can cause poor performance when executing point-based feature matching. The low density of salient visual features is another factor that also makes underwater image matching hard.

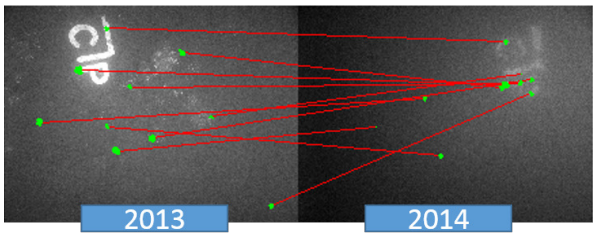
In this work, we address the problem of underwater visual place recognition across years using high-level feature matching Fig. 1(b). The proposed solution has applications for temporally periodic ship hull inspection and long term monitoring of benthic marine habitats.

*This work was supported in part by the American Bureau of Shipping under award number N016970-UM-RCMOP, and in part by the Office of Naval Research under award N00014-12-1-0092.

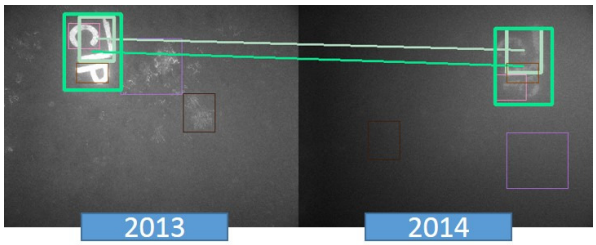
J. Li is with the Department of Electrical Engineering & Computer Science, University of Michigan, Ann Arbor, MI 48109, USA ljlijie@umich.edu

R. Eustice is with the Department of Naval Architecture & Marine Engineering, University of Michigan, Ann Arbor, MI 48109, USA eustice@umich.edu

M. Johnson-Roberson is with the Department of Naval Architecture & Marine Engineering, University of Michigan, Ann Arbor, MI 48109, USA mattjr@umich.edu



(a) Using standard point-based features, we are not able to make a valid match between two corresponding images collected from different years



(b) Using the proposed method, we are able to match two corresponding images collected from different years

Fig. 1. Depiction of the proposed high-level feature matching method with real data. The two corresponding images are collected from the years 2013 and 2014. Obvious texture-fading can be seen, which limits the performance of point-based features. The proposed method, on the other hand, is able to select a salient patch that contains high-level information that tends to last longer, resulting in higher robustness against appearance change.

To deal with challenges in the underwater environment, we propose to match images using high-level features that persist longer when compared to small feature points. Some examples from the tested data include ship identification numbers, portholes, and other man-made structures. Additionally, since individual image matching is unreliable due to insufficient features, neighboring images are also considered in the proposed method to provide more visual evidence for a robust place recognition. The main contributions of this work include:

- A visually-salient patch-based feature proposal method, that enables us to localize the salient regions in underwater images;
- A patch-feature description that is robust to dramatic changes in appearance;
- A robust outlier rejection for patch matching using a geometric constraint.

While the approach can be generalized to many different applications, in this paper we focus on the relocalization task of underwater ship hull inspection across years.

The paper is laid out as follows: Section II gives a brief

introduction about related works in visual-based navigation and visual feature matching; In Section III presents the key steps in the algorithm of high-level feature matching; and In Section IV, we evaluate the approach through the use of ship hull inspection data and compare the approach with some state-of-art algorithms and finally, Section V concludes with a summary and future work discussion.

II. Background

Many mobile robotic navigation systems are formulated as a Simultaneous Localization and Mapping (SLAM) problem, in which a robot is tasked with navigating through a prior unknown environment using a combination of odometry and perceptual sensors. Refer to [1], [2] for a comprehensive survey of solutions to the SLAM problem. Cameras have commonly been used as the primary sensor modality for mobile robots and the ability for visual place recognition has improved dramatically in recent years. Representative works in visual-based place recognition include the Fast Appearance-based Mapping algorithm (FAB-MAP) [3], MonoSLAM [4], and FrameSLAM [5]. These benchmark systems focus on searching for a location with the largest probability given matching results from standard point-based features.

More recently, some representative works that focus on dramatic scene changes within the place recognition problem have been developed. SeqSLAM [6], proposed by Milford et al., estimates the topological location of the vehicle by matching two segments of the robot's trajectory instead of matching two individual images. The robustness of the system is improved when neighboring images are involved in the final matching results. A strict assumption is made that the two matching segments share the same trajectory route. This assumption holds for ground robots like autonomous cars, which drive on the road, however, for underwater robots, it is almost impossible for two individual segments of the path to share the same trajectory, since the vehicle can move freely in space. Naseer et al. [7] proposed a framework that builds upon SeqSLAM by using Flow Network techniques to relax the overlapping route assumption and allow for partial matching. Additionally, they also proposed to use grid HOG (Histogram of Oriented Gradient) as the descriptor of an image, and match two images based upon the dot product of two HOG vectors. Though the HOG representation dramatically improves the performance of image matching in a dataset taken over multiple seasons, the orientation of the vehicle relative to the scene is almost static for an autonomous car. McManus and his colleagues [8] also explored the possibility of matching through high level features, which is similar to our method. They proposed to train Support Vector Machine (SVM) classifiers of signature HOG patch features for a set of neighboring images using an unsupervised method. However, the restriction of requiring small or no pose change relative to the scene limits their approach. Their unsupervised signature patch searching depends on the fact that the image is feature-rich. Neither of these two basic assumptions hold in the underwater environment.

Some effort has gone into addressing the image registration problem for underwater images. Eustice et al. [9] proposed to use pose-constraints from a SLAM graph to pre-constrain the searching area of point-based feature matching within one SLAM graph. Their work was extended in [10] by Carlevaris-Bianco and Eustice to search over a set of neighboring images to provide a larger field of view and more visual features. Both works assume that a pose prior is available to provide a constraint in feature matching; however, no such prior exists when attempting to register two independently navigated missions. Ozog and Eustice [11] proposed a registration between different SLAM graphs that selects candidate images for visual matching. It does this by comparing the image poses with respect to the ship hull using planar features estimated from Doppler Velocity Log (DVL) data. Most of the above work is based on point-based features that are not robust in the case of dramatic appearance change.

III. Approach

A. System Overview

An outline of the proposed image matching method with high-level features is shown in Fig. 2. Given a current image in the SLAM graph of a new mission, the goal of our system is to find the image taken at the nearest location from a set of images collected in previous missions. Given a current image, a set of neighboring images are collected and salient patch features are proposed. For each patch, a binary SVM classifier is trained to detect similar patches in the dataset. Finally, all possible matching features are fed into a geometry verification method to reject outliers and achieve a robust matching result.

B. Salient Patch Proposal

As shown in Fig. 1, most images in the underwater environment, especially those involved in ship hull inspection, are of low feature density. To address this problem, we propose the use of segmentation techniques to select visually salient regions for feature matching, and make use of neighboring image sets to provide more visual evidence to achieve a robust matching result.

Single Image Patch Proposal: Based on human perception principles, two main characteristics are shared by visually salient regions: i) obvious and complete boundaries; and ii) unique color compared to the surrounding area. These principles are widely used in most state-of-the-art salient object extraction algorithms such as [12] [13].

Based upon these characteristics, a salient patch selection algorithm is proposed as outlined in Algorithm 1. First, a graph-based segmentation method [14] is used to decompose the image into a set of components with strong boundaries between each other. The approach makes use of a graph-based representation of an image: $G = (V, E)$, where each pixel is a vertex, and each vertex is connected to its 8-neighbors. The edges are undirected and come with a weight according to pixel intensity similarity $w(e)$. The

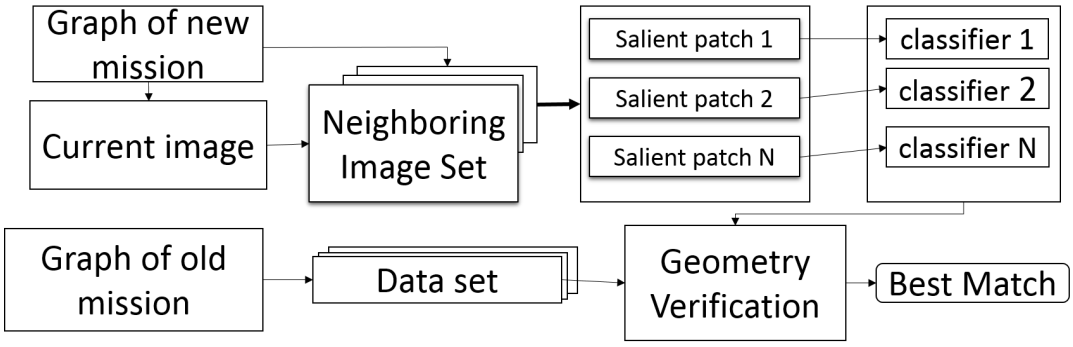


Fig. 2. Flowchart of our high-level feature image matching approach. First, a neighboring image set is gathered for a current image. Then a set of patch features are detected and described with a set of SVM classifiers based on HOG features. Finally, patch feature matching is carried out on the dataset followed by a geometric validation method to reject outliers.

Algorithm 1 Salient Patch Proposal

Initialization: a graph of the image $G = (V, E)$; Initial segments $S^0 = (C_1^0, ..., C_N^0)$. $\mathrm{Int}(C_i^0) = \infty$.
return Final segments $S^* = (C_1, ..., C_r)$

Sort E into $E = (e_1, ..., e_M)$ by non-decreasing edge weight.
for $i = 1$ to M **do**
  $S^i = S^{i-1}$
  Consider $e_i = (v_j, v_k)$ with weight $w(e_i)$, $v_j \in C_{v_j}^i, v_k \in C_{v_k}^i$
  if $C_{v_j}^i != C_{v_k}^i$ **and** $w(e_i) < \mathrm{MInt}(C_{v_j}^i, C_{v_k}^i)$ **then**
    merge $C_{v_j}^i$ and $C_{v_k}^i$
  end if
end for
$S^* = S^M$
for $i = 1$ to $|S^*|$ **do**
  $\mathrm{Sal}(C_i^*) = \Sigma_{B_j \in B} |\mathrm{Mean}(C_i^*) - \mathrm{Mean}(B_j / C_i^*)|)$
  if $\mathrm{Sal}(C_i^*) < \beta$ **then**
    remove C_i^*
  end if
end for

segmentation is initialized as a set of single pixel components $S_0 = (C_i, ..., C_N)$. Then, neighboring components are merged by comparing the minimum weight connecting them and a measure of internal consistency, which is defined as:

$$\mathrm{Int}(C_i) = \max_{e \in \mathrm{MST}(C_i, E)} w(e)$$

$\mathrm{MST}(C_i, E)$ is the *minimum span tree* of C_i constructed upon the edge set E. A minimum internal difference is also defined between neighboring components as a merging threshold.

$$\mathrm{MInt}(C_1, C_2) = \min\left(\mathrm{Int}(C_1) + \tau(C_1), \mathrm{Int}(C_2) + \tau(C_2)\right)$$

The threshold function $\tau(C)$ controls the tolerance level, which is a function of component size. $\tau(C_i) = k/|C_i|$, where k is a manually selected parameter. For a smaller component, merging is encouraged, while for larger components it becomes more difficult. The threshold function also ensures that a larger weight is needed to indicate the existence of a true boundary.

After the segmentation stage, a saliency evaluation is carried out on each component, resulting in candidate patches. The saliency score is based on local contrast and is defined as:

$$\mathrm{Sal}(C_j) = \Sigma_{B_i \in B} |\mathrm{Mean}(C_j) - \mathrm{Mean}(B_i / C_j)|$$

where $\mathrm{Mean}(C_j)$ is the mean pixel value in component C_j, and B is a set of bounding boxes containing C_j in different scales. $\mathrm{Mean}(B_i/C_j)$ refers to the mean value of pixels in B_i not belonging to C_j. Under this definition, a component is compared to its surrounding area at different scales to determine its contrast level. A similar saliency score is defined in some recent salient region/object detection algorithms [15], [16], [17]. The patches are dropped if the saliency score is lower than a manually set threshold β.

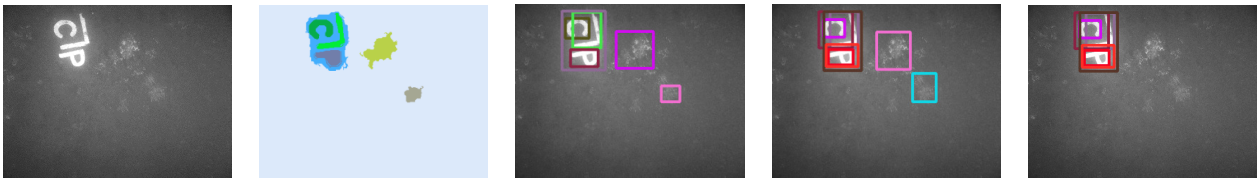
Patch Pooling in Neighboring Image Set: Given a current image I, a set of neighboring images R, including I, is considered. Salient patch proposal is performed on each image in R. All of the proposed patches are included in a patch pool P and considered in the next stage. As shown in Fig. 3(d), the neighboring image helps to propose more combinations of patches. For this step, SIFT matching is used since the neighboring images are collected during the same mission.

C. SVM Classifier for Patch Features

Given the location and scale of a salient patch, a description of the patch and a distance metric is used for patch feature matching. A binary classifier is trained to discriminate matches.

As shown in Fig. 4, HOG features [18] are used to provide a description for the patch, capturing the outline and shape information of a patch, and a linear SVM classifier is trained to detect corresponding patches in a current image. The whole procedure can be summarized in the following steps:

1) For each patch feature P_i, search for corresponding patches in the neighboring images using the propagation method mentioned in §III-B. Extract HOG features for



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

Fig. 3. The intermediate steps in salient patch proposal where a pool of candidate patches are extracted. (a) Original current image. (b) Segmentation results of the current image. (c) Patch proposal of single current image without saliency judgment. (d) Patch pool from all neighboring images projected onto current image frame. (e) Selected salient patches after saliency judgment. Random colors are assigned to every bounding box. Similar colors are not indicative of correspondences between (c) and (d).

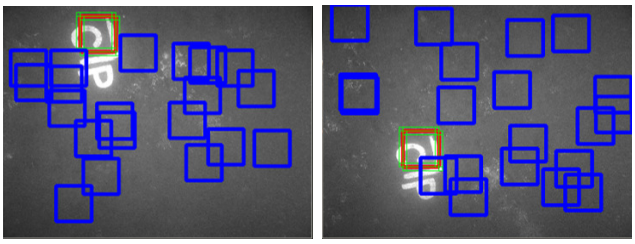


Fig. 4. Training samples are generated in current image and the neighboring images. Red: positive samples; Green: positive samples generated by shifting; Blue: negative samples

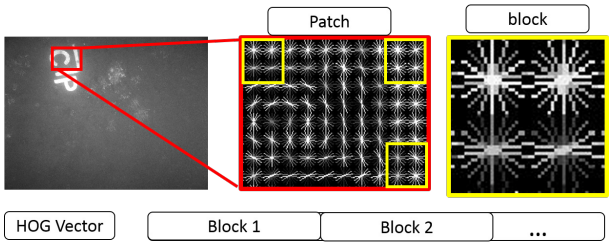


Fig. 5. HOG Structure: A patch is divided into a set of *blocks* by a sliding window. a *block* is divided by 4 *cells*. Within each *cell*, a 9-bin histogram of oriented gradient is accumulated. The *block feature* is a vector of 4 9-bins histograms. The *HOG feature* is a vector of all the *block features*.

all these corresponding patches as positive samples for SVM training.

2) In every neighboring image, generate motion compensated positive samples by slightly shifting the corresponding patches, as shown in the green boxes in Fig. 4.
3) In every neighboring image, randomly sample N patches that are off the location of positive patches and treat them as negative samples, as shown in the blue boxes in Fig. 4.
4) Train an SVM for the patch feature given all positive and negative samples. Discard the patch features whose samples cannot be separated successfully by linear SVM, which means they are not salient enough to be distinguished from other neighboring patches.

The downside of the SVM classifier are the false alarms that occur due to insufficient training samples, as shown in Fig. 6(a) and Fig. 6(b). A method addressing this problem is described in the following section.

D. Geometry Verification

In order to reject false positive SVM matches and achieve a more reliable result, a geometric constraint is used to perform hypothesis testing.

The fundamental matrix F in epipolar geometry [19] is widely used for point-based feature matching and structure estimation in two-view or multi-view systems. It defines a relationship: $x_1^T F x_2 = 0$ where x_1, x_2 are homogeneous coordinates of corresponding image points. F is defined up to scale with one free dimension, so at least 8 pairs of points are needed for estimating F using the standard linear algorithm. However, 8 pairs of patch correspondences are hard to satisfy given the low density of features in underwater images. To deal with this problem, the different parts of the structure of HOG features are used to propose hypothetical fundamental matrices F and perform validation to find the best F matrix as well as the inliers in patch matching.

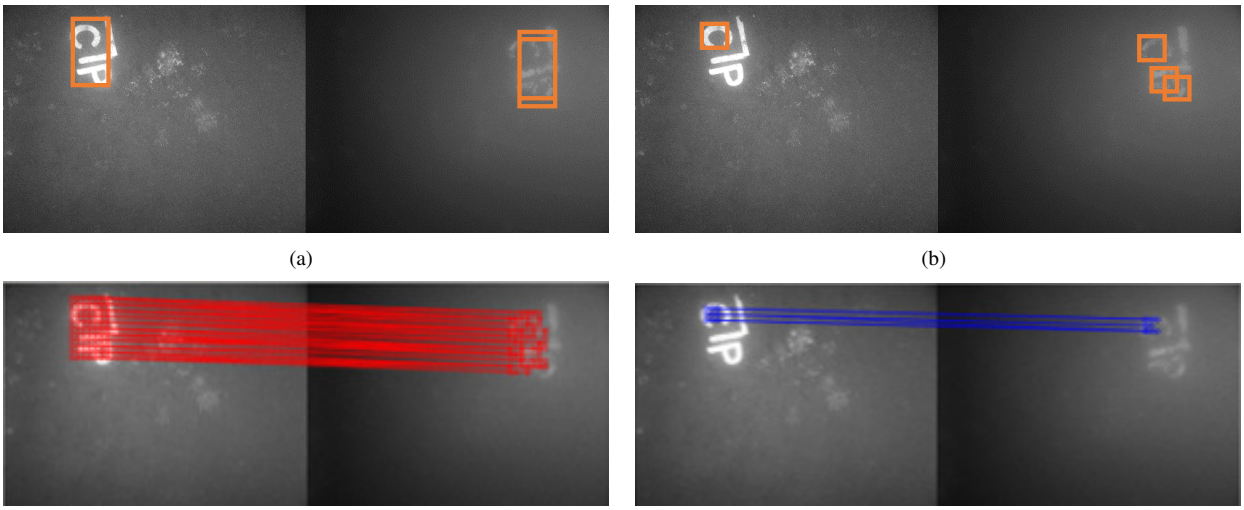
The calculated HOG feature is shown in Fig. 5. The HOG feature is constructed from a set of sub-regions (blocks) in a larger bounding box. Each block in the bounding box contains four cells. A 9-dimension gradient histogram will be extracted for each cell. The block itself is a more localized feature describing a small area relative to the whole HOG feature. By matching block features between two matched patches, a fundamental matrix can be estimated. A hypothesis test is performed as follows:

1) For each patch feature p_i with a positive SVM response (classification) $\{q_{ij}\}$, block feature matching is done for every patch pair $< q_{ij}, p_i >$.
2) For p_i, the best SVM response (the response with largest distance to the boundary) $< q_{ij^*}, p_i >$ is used to estimate a fundamental matrix F_i.
3) Each F_i is tested on other patch features $p_{i'}$, $i' \neq i$. A score is accumulated:

$$G(F_i) = \Sigma_{i'} \max_j (\text{number of inlier in } < q_{i'j}, p_{i'} >)$$

the best response patch in the current image is selected to generate a fundamental matrix hypothesis F_i.
4) $F^* = \text{argmax}_i\, G(F_i)$ is selected to be the optimal geometry model between the two images given the feature matching results, and all matching pairs that satisfy F^* are selected as inliers.

The final matching confidence score of the current and



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

Fig. 6. This figure illustrates how the fundamental matrix is estimated and tested for multiple positive responses returned by SVM classifiers. (a) Left: Patch 1. Right: SVM positive responses. (b) Left: Patch 2. Right: SVM positive responses. (c) Fundamental candidate F is estimated by the best match of Patch 1. (d) The candidate fundamental matrix estimated from Patch 1 is tested on Patch 2 by searching for the best response of Patch 2 that satisfy F.

matched image is given by:

$$M(I_{in}, I_{re}) = \frac{\Sigma_{j \in MP}(Sal(p_j)\lambda_j)}{\Sigma_i Sal(p_i)}$$

where MP is the set of patches that have inlier match pairs, and λ_j refers to the ratio of inlier block matches within the HOG pair.

As shown in Fig. 6, the geometric constraint is able to improve the system by filtering out false alarms in the SVM response. A verification of the effectiveness of the geometry constraint is provided in the experimental results section by comparing performance of the system with and without it.

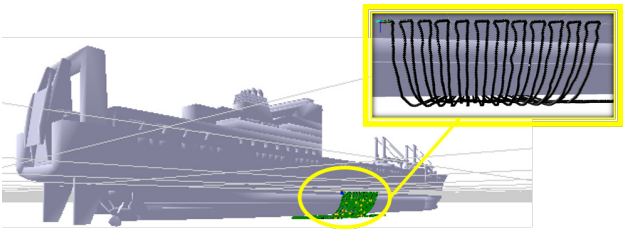
IV. Experimental Results and Discussion

To evaluate the robustness of the proposed place recognition approach, a set of experiments and analysis were performed using a multi-year underwater ship hull inspection dataset. We first introduce the data and ground truth used in the experiments in §IV-A. Experimental results with accuracy analysis compared to other state-of-art algorithms are given in §IV-B. Detailed analysis of the performance of the intermediate steps of the framework are discussed in §IV-B.

A. Experimental Setting

Experiments are performed on data collected by a Hovering Autonomous Underwater Vehicle (HAUV) over three years (2011, 2013, 2014) on the SS Curtiss [20], [11]. External ground truth for underwater data is extremely challenging to gather, so we elected to use hand-labeled data for place recognition ground-truth. Seven places, which included 358 images, were manually labeled from the dataset, where correspondences could be identified by a human. We used the data from the year 2013 (118 images) as the *query* dataset, and the data from the years 2011 and 2014 (240 in total) for matching (what we will refer to as the set of *current* images).

(a) SS Curtiss



(b) Ship hull inspection using HAUV

Fig. 7. Images depicting the ship used in the experiments and the ship hull inspection process. The image of depicts a robots trajectory over the hull when performing an inspection.

One hundred unlabeled images from the mission data in 2013 were randomly sampled (images in unknown places) to be included as noise images to ensure the matching is robust enough to distinguish true matches from non-salient noise. Some examples of correspondences in the manually labeled dataset are shown in Fig. 8. It can be seen that the dataset is quite challenging due to decay and biofouling.

B. Experimental Results

Comparing to standard place recognition methods: Performance is compared between the proposed algorithms and two representative point-feature based place recognition methods. SIFT feature matching with random sample consensus (RANSAC) for geometric constraint estimation

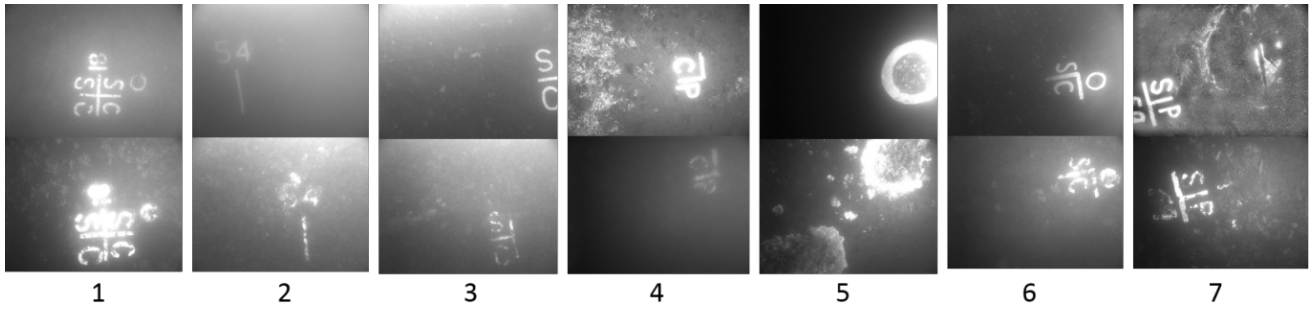


Fig. 8. Examples of manually labeled corresponding images from 7 places on the SS Curtiss across different years. The images in the first row come from the dataset we are searching in. The images in the second row are current images. It can be seen that the data is quite challenging, due to the dramatic decay of patterns and changing light conditions.

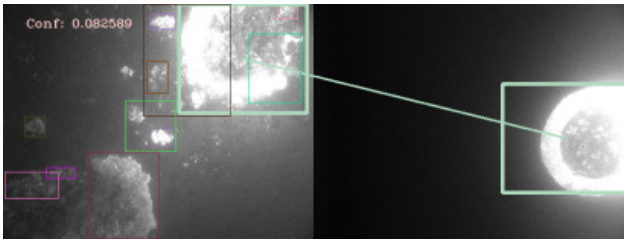
and outliers rejection is included as a standard approach for image matching. A Bag-of-words (BoW) model on SIFT is also compared in the experiment. BoW trains a vocabulary of SIFT features from the query dataset and describes images by the BoW vocabulary. The matching is done based on the image description.

The confusion matrix of matching results for different places is shown in Fig. 11. The proposed method is able to provide reasonable matching results in such a challenging dataset, while the other two standard methods under perform. This result is expected since point-based features are often unable to generate any meaningful correspondences between images with dramatic appearance changes, however, these results can be improved. To better illustrate the experiment results, examples of true matches and false matches are shown in Fig. 12 and Fig. 9. For some places, the proposed method is able to provide almost 80% accuracy given the dramatic appearance changes. In some cases, the decay is too strong for our image segmentation approach to extract a meaningful patch, and the proposed method fails to provide a robust match. For some other cases, when several redundant patches are detected, the system is not able to properly determine which ones are of greater importance within one image. This indicates that a frame-related patch importance weight should be developed to improve the robustness of our approach.

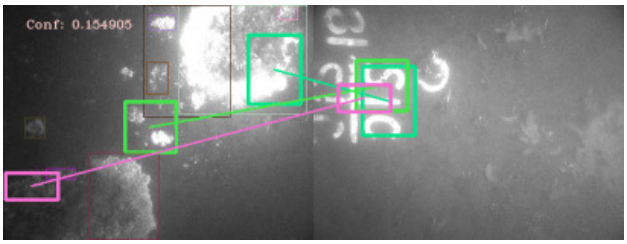
Some matching examples of the proposed method used in above-water images are also provided in Fig. 10 to show that the approach is able to be generalized to non-underwater images. The average matching time for one pair of images is 2.6 seconds on a 2.8GHz CPU. Parallelization should be done in future work to improve the efficiency.

System component analysis: To get a better idea of how the proposed approach is able to make reasonable correspondences between images across different years, we also analyzed the performance with and without certain components of the framework. The results are shown in Fig. 13.

We replaced the SVM classifier of HOG descriptors with Normalized Cross Correlation (NCC) in searching for matched patches. The result is displayed in Fig. 13(a). Comparing that to the original proposed method (Fig. 11), it can



(a) false negative: Although the structure of the round areas are matched, it's considered weak evidence



(b) false positive: many small areas are matched, and the system determines it's strong evidence.

Fig. 9. Example of false positive and false negative matches using the proposed system.

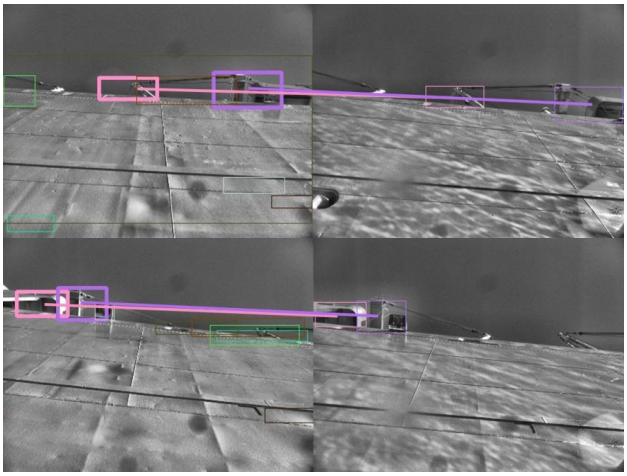
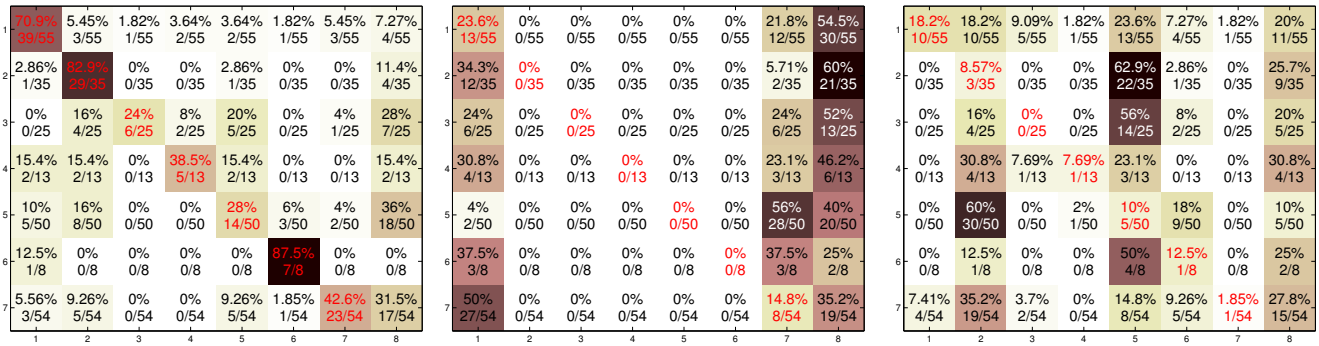


Fig. 10. Examples of image matching using the proposed method in a different domain, above water images.



(a) Proposed Method (b) SIFT+RANSAC (c) Bag-of-Words

Fig. 11. Results from our proposed technique and from point-based approaches from the literature, presented as confusion matrices where the rows and columns refer to places or clusters of co-located images of which samples are shown in Fig. 8. Note the last column has no matched row as it refers to non-labeled random images introduced to the dataset to make matching more challenging/realistic. The red font indicates the accuracy rate of matching result for each place, while the rest of the row indicates how the false matching is distributed in the query dataset.

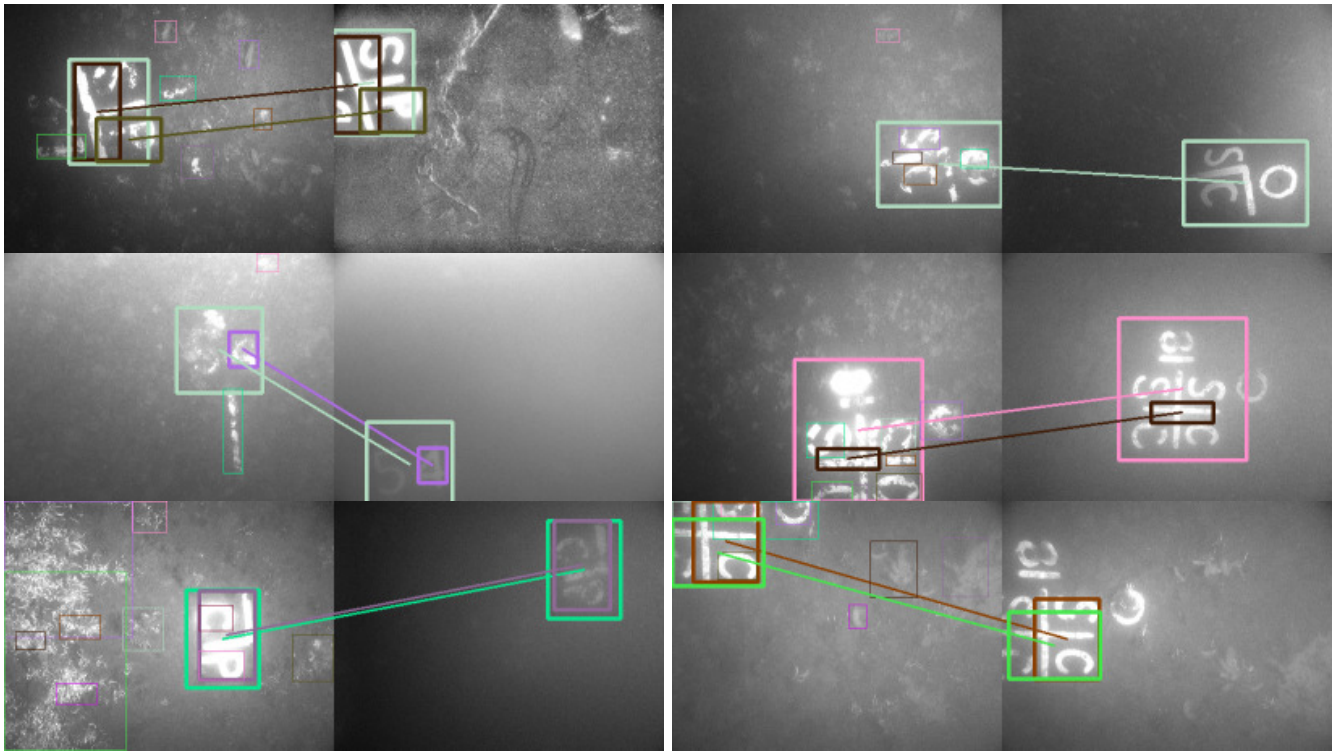
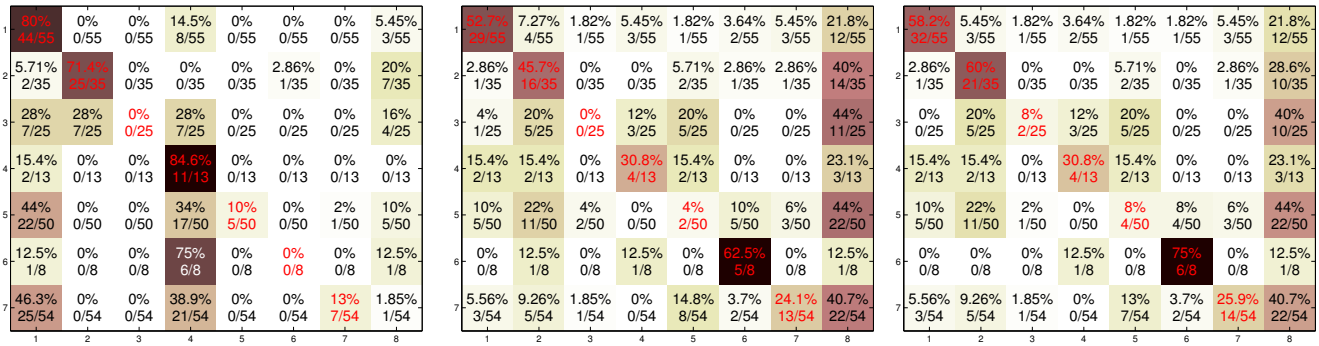


Fig. 12. Examples of successfully matched pairs. The system is able to make valid correspondences between images with dramatic appearance changes.

be seen that SVM with HOG features has greater robustness with respect to providing valid matches, compared to NCC. The main reason is that HOG features have a relatively higher tolerance to small rotation and appearance changes than the image patch intensity comparison. In the case when appearance changes caused by corrosion or so is not obvious, the NCC method tends to perform similar to SVM, such as place 4 in the dataset Fig. 8.

The contribution of geometric validation on false positive rejection is also demonstrated by the comparison experiment with and without it, as shown in Fig. 11 and Fig. 13(b), respectively. It can be seen that the geometric constraint improves the system's robustness significantly by rejecting false positive responses returned by the SVM classifiers. Without the geometric constraint, matches tend to happen between two images with several salient regions, in which case many false positive matches are returned and are included in the final matching score.

The use of neighboring image sets in providing more visual evidence in patch matching is also illustrated in Fig. 13(c), in which no neighboring images are considered. The proposed method supports the use of a default maximum number of neighboring images ($N = 4$), which improves the final performance of the proposed method.



(a) NCC instead of SVM on HOG (b) Without geometry validation (c) Without neighboring image information

Fig. 13. System component analysis. Identification of the contribution of each system component by comparison experiment without certain components or with those components replaced by other simpler alternatives. Results are presented as confusion matrices where rows and columns refer to places or clusters of co-located images of which samples are shown in Fig. 8. Note the last column has no matched row as it refers to non-labeled random images introduced to the dataset to make matching more challenging/realistic.

V. Conclusion and Future Work

In this work, a high-level visual feature image matching approach is presented to address the challenge of underwater image matching under dramatic appearance changes. A salient region identification method is proposed to locate visually salient regions and find high level features. SVM classifiers based on HOG features are used for feature description and matching. Geometric constraints are employed to reject false positive matches provided by SVM. The approach is evaluated on real data collected in multi-year ship hull inspection missions and the performance is compared with other standard place recognition methods. The proposed method strongly outperforms traditional point-based feature matching techniques.

To further improve and generalize the proposed place recognition system, several future directions should be considered: 1) a more thorough system analysis by replacing different system components with other alternatives, such as different segmentation methods (mean-shift, etc); 2) more experiments on datasets from other applications; 3) oriented patch descriptor to increase robustness to viewpoint variation. Furthermore, given the capability to robustly register multi-year datasets, updating the patch classifiers to include the information from different years could increase the stability of the system.

References

[1] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: Part I," *IEEE Robot. Autom. Mag.*, vol. 13, no. 2, pp. 99–110, 2006.

[2] T. Bailey and H. Durrant-Whyte, "Simultaneous localization and mapping (SLAM): Part II," *IEEE Robot. Autom. Mag.*, vol. 13, no. 3, pp. 108–117, 2006.

[3] M. Cummins and P. Newman, "Fab-map: Probabilistic localization and mapping in the space of appearance," *Int. J. Robot. Res.*, vol. 27, no. 6, pp. 647–665, 2008.

[4] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "MonoSLAM: Real-time single camera SLAM," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 1052–1067, 2007.

[5] K. Konolige and M. Agrawal, "FrameSLAM: From bundle adjustment to real-time visual mapping," *IEEE Trans. Robot.*, vol. 24, no. 5, pp. 1066–1077, 2008.

[6] M. J. Milford and G. F. Wyeth, "SeqSLAM: Visual route-based navigation for sunny summer days and stormy winter nights," in *Proc. IEEE Int. Conf. Robot. and Automation*, (Saint Paul, MN, USA), pp. 1643–1649, May 2012.

[7] T. Naseer, L. Spinello, W. Burgard, and C. Stachniss, "Robust visual robot localization across seasons using network flows," in *Proc. AAAI Nat. Conf. Artif. Intell.*, (Québec City, Québec, Canada), pp. 2564–2570, July 2014.

[8] C. McManus, B. Upcroft, and P. Newman, "Scene signatures: Localised and point-less features for localisation," in *Proc. Robot.: Sci. & Syst. Conf.*, (Berkley, CA, USA), July 2014.

[9] R. M. Eustice, O. Pizarro, and H. Singh, "Visually augmented navigation for autonomous underwater vehicles," *IEEE J. Ocean. Eng.*, vol. 33, no. 2, pp. 103–122, 2008.

[10] N. Carlevaris-Bianco and R. M. Eustice, "Multi-view registration for feature-poor underwater imagery," in *Proc. IEEE Int. Conf. Robot. and Automation*, (Shanghai, China), pp. 423–430, May 2011.

[11] P. Ozog and R. M. Eustice, "Toward long-term, automated ship hull inspection with visual SLAM, explicit surface optimization, and generic graph-sparsification," in *Proc. IEEE Int. Conf. Robot. and Automation*, (Hong Kong, China), pp. 3832–3839, June 2014.

[12] H. Jiang, J. Wang, Z. Yuan, T. Liu, N. Zheng, and S. Li, "Automatic salient object segmentation based on context and shape prior.," in *Proc. British Mach. Vis. Conf.*, vol. 3, (Dundee, U.K.), p. 7, Aug. 2011.

[13] B. Alexe, T. Deselaers, and V. Ferrari, "What is an object?," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, (San Francisco, CA, USA), pp. 73–80, June 2010.

[14] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vis.*, vol. 59, no. 2, pp. 167–181, 2004.

[15] R. Achanta, F. Estrada, P. Wils, and S. Süsstrunk, "Salient region detection and segmentation," in *Computer Vision Systems*, pp. 66–75, Springer, 2008.

[16] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, (Providence, RI, USA), pp. 733–740, June 2012.

[17] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Advances in neural information processing systems*, pp. 545–552, 2006.

[18] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, vol. 1, (San Diego, CA, USA), pp. 886–893, June 2005.

[19] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge University Press, 2003.

[20] A. Kim and R. M. Eustice, "Real-time visual SLAM for autonomous underwater hull inspection using visual saliency," *IEEE Trans. Robot.*, vol. 29, no. 3, pp. 719–733, 2013.