Underwater Robot Visual Place Recognition in the Presence of Dramatic Appearance Change

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

***Abstract*— This paper reports on an algorithm for underwater visual place recognition in the presence of dramatic appearance change. Long-term visual place recognition is challenging underwater due to biofouling, corrosion, and other effects that lead to dramatic visual appearance change, which often causes traditional point-feature-based methods to perform poorly. Building upon the authors' earlier work, this paper presents an algorithm for underwater vehicle place recognition and relocalization that enables an underwater autonomous vehicle to relocalize itself to a previously-built simultaneous localization and mapping (SLAM) graph. High-level structural features are learned using a supervised learning framework that retains features that have a high potential to persist in the underwater environment. Combined with a particle filtering framework, these features are used to provide a probabilistic representation of localization confidence. The algorithm is evaluated on real data, from multiple years, collected by a Hovering Autonomous Underwater Vehicle (HAUV) for ship hull inspection.**

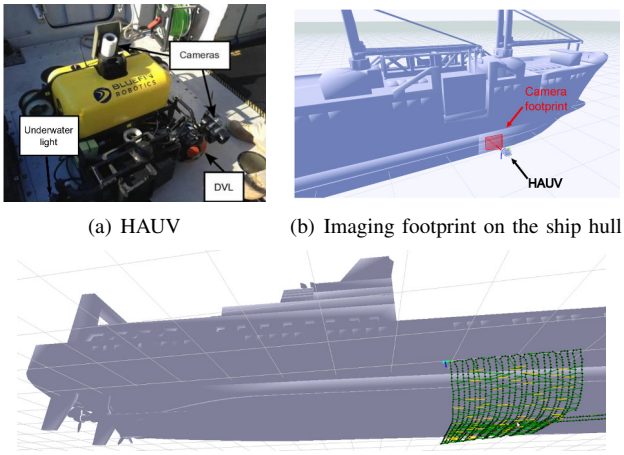
I. Introduction

The localization problem has been an important topic in robot navigation for many years. Being able to localize with respect to a previously seen place or map is a prerequisite in navigation systems for applications like long-term regular surveying. However, the place recognition and relocalization problem is challenging underwater due to contributing factors such as biofouling, corrosion, and other dramatic visual appearance changes. Phenomenon like these break the basic appearance consistency assumption of many popular visual-based algorithms. A low density of visually salient features also make it more challenging to make putative visual correspondences.

To address these challenges in making visual correspondences for underwater images, in the authors' prior work, a high-level structural feature image matching approach was proposed, which obtained a promising matching result between corresponding images collected across years [1]. This paper extends the previous method by combining the high-level structural feature matching approach with a probabilistic framework that considers a set of sequential images observed along the vehicle trajectory as well as measurements from other modalities on the vehicle. The algorithm is evaluated on real data collected in a multi-year ship hull inspection mission carried out with an Hovering Autonomous Underwater Vehicle (HAUV) [2], as depicted in Fig. 1. The contributions of this paper include:

- Development of a high-level feature detector using segmentation and machine learning for images with a low density of visual texture.
- Development of a SLAM-graph global relocalization algorithm using high-level feature matching within a particle filter framework.

The paper is laid out as follows: Section II gives a brief introduction about related work in visual-based place recognition and localization; Section III presents the key steps in the algorithm starting with an overview of the whole framework; Section IV evaluates the approach with localization between ship hull inspection missions across years and finally, Section V concludes with a summary and future work discussion.



(a) HAUV (b) Imaging footprint on the ship hull

(c) Typical robot trajectory

Fig. 1. Depiction of the ship hull inspection application using the HAUV.

II. Related Work

Place recognition and localization against dramatic scene change has been explored for years in within the robotics community. Representative works include [3]–[5]. SeqSLAM [3] localizes the robot using a topological strategy that jointly considers visual feature comparisons along trajectory segments to increase the robustness to dramatic scene changes, as compared to individual image comparison alone. However, such a topological approach make assumptions about temporal trajectory similarly, for example, along a road network, which are often incompatible with unstructured

*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 and M. Johnson-Roberson are with the Department of Naval Architecture & Marine Engineering, University of Michigan, Ann Arbor, MI 48109, USA {eustice,mattjr}@umich.edu

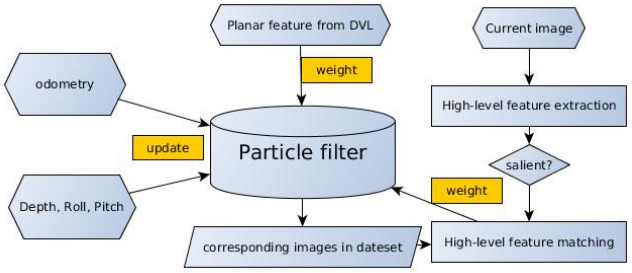
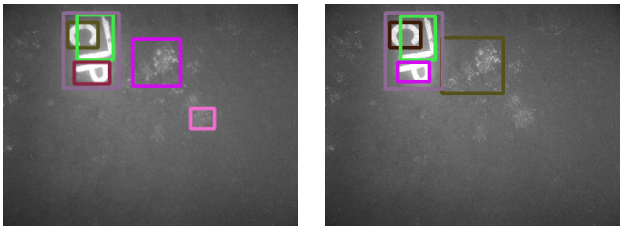


Fig. 2. System Flowchart. A particle filter is initialized on an previous SLAM graph that we will localize against. Onboard sensors including depth and IMU are used to update the particles when the vehicle is moving. Planar features estimated from DVL [6] as well as high-level features extracted from visual images are used to do measurement correction by updating the importance weight of each particle.



(a) Feature patches from raw segmentation (b) Salient feature patch after SVM classification

Fig. 3. High-level feature extraction.

underwater robot surveys. Methods focusing on using higher-level features to make visual correspondences robust against appearance changes are also developed, such as [4] and [5]. Again, assumptions about sensor viewpoint exploited in the ground vehicle domain are not feasible for an underwater navigation system.

In the underwater domain, Ozog et al. proposed a registration algorithm that mitigates visual localization challenges by limiting the search area using measurements from other navigation sensors like the Doppler velocity log (DVL) [6]. A winner-take-all matching is carried out within a pruned candidate region using a traditional point feature the Scale Invariant Feature Transform (SIFT). This approach works robustly in registering data taken within a short time interval, but is not applicable to large time intervals with dramatic appearance changes under less reliable image matching. Our approach shares the same spirit of Ozog et al. in using different sensor measurements to localize the vehicle, however, instead of using a winner-take-all localization decision on image matching, we use the visual matching results within a particle filter framework, where more visual images along the trajectory will be jointly considered in vehicle localization.

III. Methodology

A flowchart of the proposed localization approach is shown in Fig. 2. Here, a particle filter is used to solve the localization problem. An onboard depth sensor and inertial measurement unit (IMU) are used to update the particles, and measurements of the local planar-like shape of the ship's

Algorithm 1 High-level feature extraction and description

initialization

$E = (e_1, ..., e_M)$: Edges connecting neighboring pixels.
$e_i = (v_i^1, v_i^2)$, $w(e_i) = |Img(v_i^1) - Img(v_i^2)|$
$S = (s_1, ..., s_N)$: The set of salient segments is initialed as all individual pixels.
$D = \phi$: A set of SVM as feature descriptors.
$C_{salient}$: A pre-train SVM classifier for salient segments.

Sort E into $E = \{e_i\}$, so $w(e_i) \leq w(e_i + 1)$
for $i = 1$ to M **do**
  Find $s_{v_i^j}$ that $v_i^j \in s_{v_i^j}$
  $MergeThresh = \min(\max_{e_k \in s_{v_i^1}} w(e_k), \max_{e_k \in s_{v_i^2}} w(e_k))$
  if $s_{v_i^2} \neq s_{v_i^2}$ **and** $w(e_i) < MergeThresh$ **then**
    $s_{new} = s_{v_i^1} \cap s_{v_i^2}$, $\text{Insert}(S, s_{new})$,
    $\text{Delete}(S, s_{v_i^1})$, $\text{Delete}(S, s_{v_i^2})$.
  end if
end for
for $j = 1$ to $|S|$ **do**
  if $!\text{SVMTest}(C_{salient}, s_j)$ **then**
    $\text{Delete}(S, s_j)$
  end if
end for
for $k = 1$ to $|S|$ **do**
  $H_{pos} = HOG(S_i)$
  for $u = 1$ to 20 **do**
    $x = \text{Random}(0, Width(Img) - 1)$
    $y = \text{Random}(0, Height(Img) - 1)$
    $P = Img(x, y, Size(s_i))$
    $H_{neg} = H_{neg} \cap HOG(P)$
  end for
  $d_k = \text{SVMTrain}(H_{pos}, H_{neg})$
  $\text{Insert}(D, d_k)$
end for

return S, D

hull and visual features are used to weight the particles at each location. The planar measurements are estimated using the DVL and the visual feature measurements are provided by the high-level visual feature matching approach. When the probability covariance converges to a limited area, the vehicle localized.

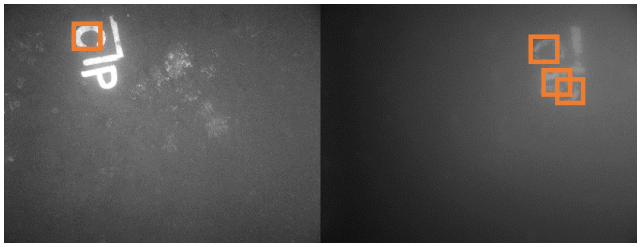
A. Particle Filter

In the proposed method, we use a particle filter to provide an estimate of the vehicle's pose distribution, which enables us to incorporate measurements from different modalities and control inputs from other onboard sensors in a convenient way:

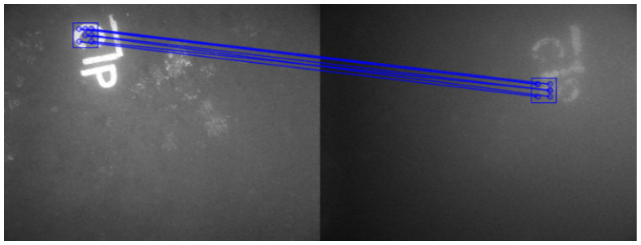
$$p(x_t|Z_{1:t}, U_{1:t}) \propto$$

$$p(Z_t|x_t)p(x_t|x_{t-1}, U_t)p(x_{t-1}|U_{1:t-1}, Z_{1:t-1}).$$

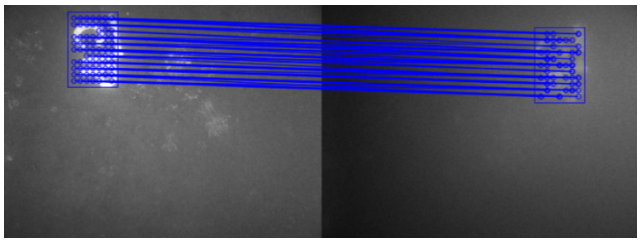
Vehicle pose consists of position and orientation, $x_t = x, y, z, roll, pitch, yaw$. U is the set of control inputs used to propagate particles and Z is the set of observations (i.e., measurements).



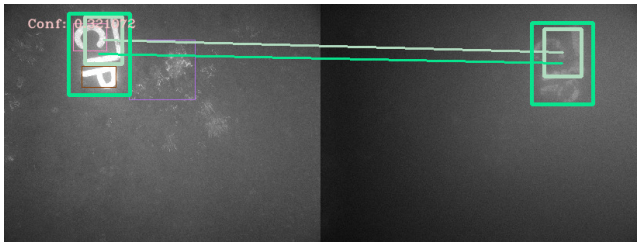
(a) Sample SVM positive response



(b) Sample of geometry constraint defined by a pair of matching features, which is incorrect in this case.



(c) Sample of geometry constraint defined by a pair of matching features, which is correct in this case.



(d) The best geometry relationship model is selected as well as its supported matching pairs.

Fig. 4. Image matching based on epipolar constraint.

1) Updating: We incorporate two different control inputs for particle propagation. An onboard depth sensor and IMU provide direct measurement of depth, roll and pitch. The odometry input estimated from DVL provides a standard odometry update: $p(x_t|u_t^{odom}, x_{t-1}) \sim N(x_{t-1} \oplus u_t^{odom}, \Sigma_{U_z^{odom}})$.
2) Weighting: Two measurements are considered in the system: planar measurements (z_t^{dvl}) and visual measurements (z_t^{cam}); $Z_t = \{z_t^{dvl}, z_t^{cam}\}$ are assumed to be independent given a vehicle pose. We use the planar feature proposed in [6] to evaluate DVL measurements. Given the planar-like shape of the ship hull in the camera field of view, [6] estimates a planar feature $z_{\pi t}$ using Principle Component Analysis (PCA) on DVL point-based outputs. We model the measurement model as follows: $w_p^f = p(z_t^{dvl}|x_t^p) \sim ||z_{\pi t} - \hat{z}_{x_p}||$, where $\hat{z}_{x_p}$ is the expected planar factor estimated from planar features in the old graph neighboring to particle x_p. For visual measurements, $w_p^{dvl} = p(z_t^{dvl}|x_p)$ is assigned according to the high-level feature matching result between current observed image and the nearest image in the old graph corresponding to state x_p. More details of visual feature weighting will be discussed in III-C

B. High-level Feature

An improved version of high-level feature matching approach proposed in the author's previous work [1] is used to provide visual measurements by making correspondences between the current observation and previous observations from the old missions.

An algorithm outline to detect and describe the feature is presented in Algorithm 1. An input image is segmented into a set of segments based on pixel intensity similarity using graph-based segmentation strategy [7]. Then we em-

features	Definition
Contrast	$\lvert \text{Mean}(s_i) - \text{Mean}(b_i) \rvert$
Size	$\lvert s_i \rvert$
SizeRatio	W_{s_i}/H_{s_i}
Shape	$\lvert \lvert s_i \rvert - \lvert b_i \rvert \rvert$

Fig. 5. Features used in salient segment classification. s_i is a set of pixels in the segment. b_i is a set of pixels in the neighboring area of s_i defined by a bounding-box.

ploy a pre-trained Support vector machine (SVM) classifier $(C_{salient})$ to select salient segments as the support region for high-level features, differing from the a hard coded threshold cutting method previously employed in the earlier paper [1]. The vector space that $C_{salient}$ is defined on, is comprised on the feature described in Figure 5. The features are extracted considering the pixels in the segments s_i and the pixels in the bounding box of the segments b_i. These features contribute to how salient the segment is, compared to it's neighboring area. In function $\text{SVMTest}(s_i, C_{salient})$, these features of the s_i are extracted and the SVM response is returned.

For each segment that is classified as salient, a SVM similarity classifier over the Histogram of oriented gradients (HOG) feature is trained($\text{SVMTrain}()$). The positive samples H_{pos} are the HOG features extracted from the image patch defined by the segments bounding box as well as image patches extracted by slightly shifting the bounding box. The negative samples for the training H_{neg} are the HOG features extracted from random sampled image patches with the same size as the positive training examples.

To increase the computation efficiency, only the images over a predetermined threshold of salient segments are used to weight the particles as visual measurements.

Algorithm 2 Image matching using model selection

input

S: Salient segments extracted in Algorithm 1

D : SVM classifiers extracted in Algorithm 1

img2: image from old graph needs to match

initialization

$P = p_i, ..., p_{|D|}$, $p_i = \phi$. Matching pairs in img2.

$F = \phi$: the set of geometry models between two images

for $i = 1$ to $|D|$ **do**
  $[p_i, best_{response}] = \text{SVMTest}(d_i, img2)$
  if $p_i ! = \phi$ **then**
    $\text{Insert}(F, \text{ModelExtract}(s_i, best_{response}))$
  end if
end for
$Best\_F = \phi$
$Max\_Num\_Support = 0$
for $j = 1$ to $|F|$ **do**
  $Num\_Support = \text{ModelTest}(F_i, P)$
  if $Num\_Support > Max\_Num\_Support$ **then**
    $Best\_F = F_j$
  end if
end for

return $Best\_F$

C. Particle Weighting Visual Measurement

When a salient image with its high-level feature descriptors is passed into the particle filter, the image will be matched against all the images in the previous graph consistent with particle positions. A matching strategy using high-level features and a two-view epipolar constraint is given in Algorithm 2.

A set of SVM positive responses for each feature is found by a sliding window search as shown in Figure 4(a). The best positive response of a feature, ranked by its distance to the SVM boundary, is then used to estimate a fundamental matrix between the two images. In the function $\text{ModelExtract}()$, we estimate the fundamental matrix by utilizing correspondences from sub-blocks from within each feature match. Multiple fundamental matrix will be defined by different features, which might indicate different geometry relationships between the two images, as shown in Figures 4(b)and 4(c). To figure out the best fundamental matrix among them, a model selection is carried out to search for the geometry model that is supported by the maximum number of matching features in the previous step. Once the optimal model is selected, an image correspondence with geometric consistency is found, as shown in Figure 4(d). Finally, a matching score between the two image is calculated:

$$S_m = \Sigma_{j=1}^{N} Area(P_{Best\_F})/Area(Img)$$

$$N = Best\_Num\_Support$$

To maintain the diversity in the particle filter representation, the number of particles is often larger than the number of nodes in the old graph. Thus, multiple particles are associated to the same candidate image from the old graph, that is to say, they share the same S_m as visual measurement confidence.

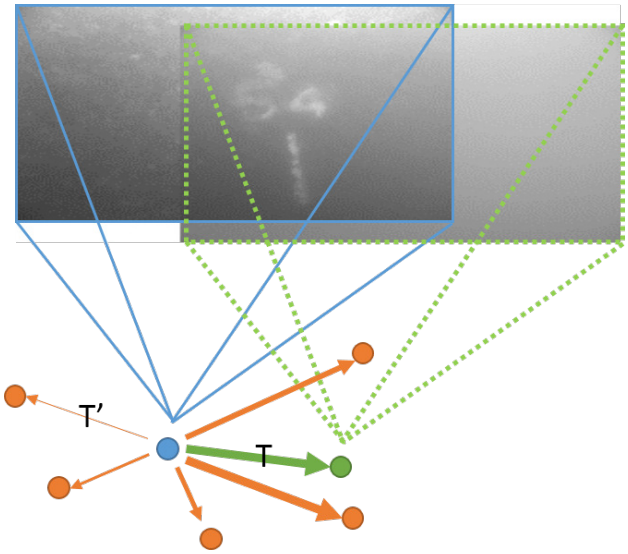


Fig. 6. This figure gives an example when a set of particles (orange dot) are associated with a single image node in previous graph (blue dot). The transform vector T from the old graph node to the new one (green dot) is estimated in image matching. The transform vector T' from old graph node to particles are compared with T. The line weights of T' in the figure indicate the relative value of s_g in this example.

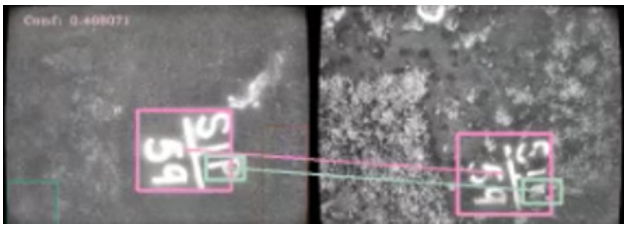
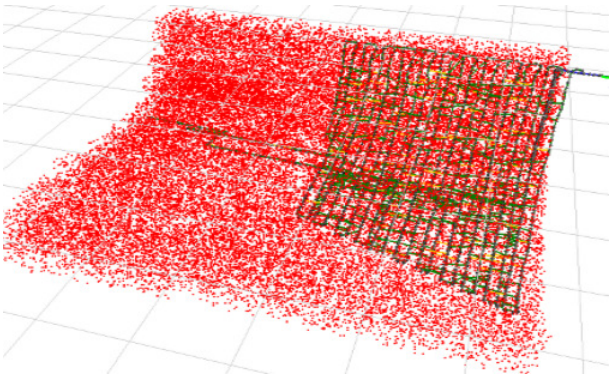


Fig. 8. The best match pair at the convergence location. Left: current image. Right: the best matched candidate image in previous graph.

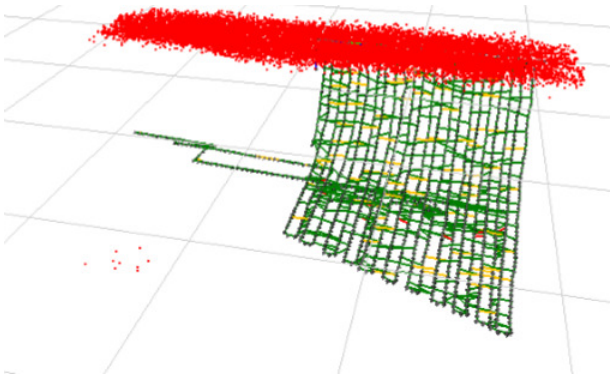
Given the geometric relationship we estimated from the image matching procedure, a particle weight is calculated based on the image matching score as well as the geometric consistency: $w_p^{cam} = s_g^p \cdot s_m^p$. $s_g^p = \frac{TT'}{|T'|}$ is the geometry consistency score of the current particle position given the estimated position from the epipolar constraint in image matching. T is the estimated transformation from the position of the current image to the candidate image position, and T' is the putative position calculated between the particle position and the candidate image position. This step also increases the convergence speed of the particle filter in that it makes the best of the output information of image matching and penalizes particles that are inconsistent with the underlying image geometry.

IV. Experiment and Discussion

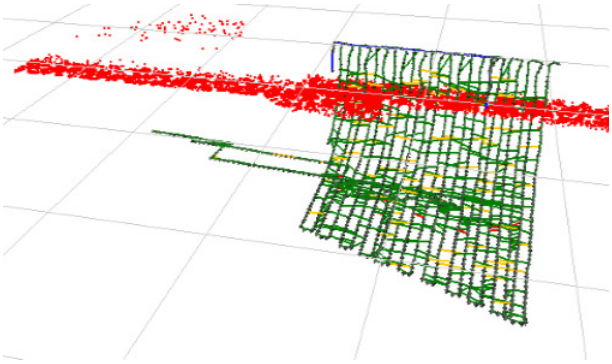
We evaluate the proposed method on data collected for automated ship hull inspection. We derive the trajectory of the HAUV from a graph-based Visual SLAM system [6].



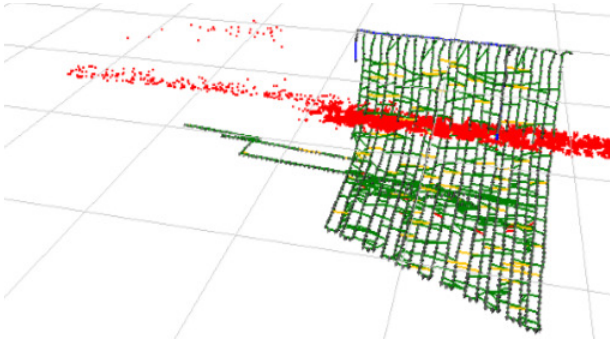
(a) Initial distribution



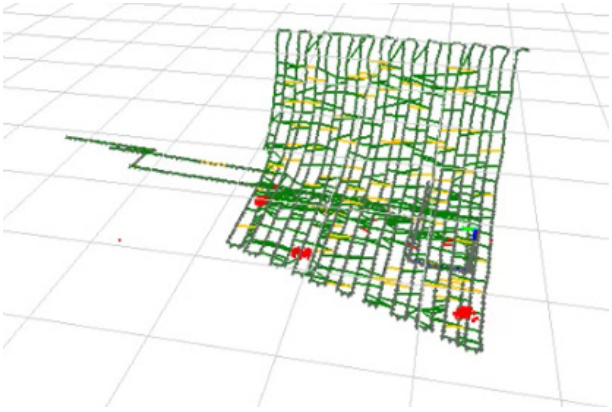
(b) Planar feature: Particle distribution after re-sampling with importance weight provided by planar features.



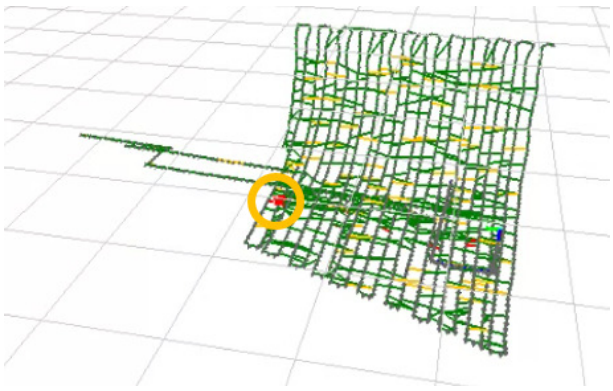
(c) Depth update: When the vehicle is moving, the depth is updated by depth sensor



(d) Non-distinguishable match: A salient image is observed and matched against the previous images, but no distinguish matching is found. The distribution will not change significantly because all the S_m are evenly low



(e) Distinguishable match: A salient image is observed and matched against the previous image corresponding to particle positions. However, more than one previous images are considered to be similar to the current observation, the particles converge to some candidate regions



(f) Particles converge. For the particles around the correct location, when more salient images are consider, their importance weight increase until the particle distribution converge to a small area

Fig. 7. Typical particle filter converging procedure.

We wish to localize the HAUV in two different surveys runs, from different years, into the same reference frame to measure the effectiveness of the proposed algorithm. The proposed method is able to localized the vehicle in a previous built SLAM graph given dramatic appearance changes.

One localization mission that capture a lot of properties of the system is shown in Fig. 7. As shown in Fig. 7(a), the particles are initialized with uniform distribution and updated using estimates from the depth sensor. When a salient image is observed, but which is not capable of making a strong match, the matching score with all candidate images will be uniformly low and the distribution of particles remains almost unaffected, as shown in Figure 7(d). When a salient image produces a strong match, the particles start to converge to the locations where the matching scores are high, as shown in Fig. 7(e). However, the visual evidence provided by a single salient image is insufficient for localization. The particles converge to some candidate regions where the matching scores are outstanding. As the mission goes on, more salient images are used to correct the particle distribution and the probability of the correct position finally stands out. When the determinant of the covariance passes a threshold it is considered localized. Fig. 8 gives the best matching image pair right before the vehicle localized, which indicates the validness of the localization. The highly structured visual features are identified in the boxes and correctly matched across years despite the significant ship hull biofouling.

V. Conclusion

We proposed an algorithm for underwater localization with respect to a SLAM graph using a high-level feature matching approach in a particle filter framework. The approach was evaluated on real data collected by an HAUV in a ship hull inspection mission. Experimental results show that the method is able to make use of high-level visual features from a low feature density environment and perform robust localization by incorporating observations in different locations in a probabilistic framework. Future work will concentrate on involving other onboard modalities, such as imaging sonar, to improve the performance of long-term localization against dramatic environment changes that could take place underwater.

References

[1] J. Li, R. Eustice, and M. Johnson-Roberson, "High-level visual features for underwater place recognition," in *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, pp. 3652–3659, May 2015.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

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