A Review on SIFT and SURF for Underwater Image Feature Detection and Matching

Sadaf Ansari
CSIR-National Institute of Oceanography
Goa, India
ansari@nio.org

Abstract—**Fast and robust image recognition and matching is an important task in the underwater domain. Autonomous Underwater Vehicles (AUV) use images for detecting and recognising object or scene of interest. Interest points are extracted from images that are captured by the cameraon the AUV. In this paper, feature detectors SIFT and SURF are investigated as applicable to the underwater environment. They are reviewed and compared in terms of robustness, accuracy, and speed with the aim of identifying a method which is computationally-efficient in finding good corresponding points between two underwater images.**

Keywords—SIFT, SURF, computer vision, underwater imaging, image feature

I. INTRODUCTION

Computer Vision is appliedin many areas of robotics, such as smart robot control, object detection and tracking, autonomous navigation [1, 2]. Visual image information extraction and processing is the core of machine vision research. Image features play an important role in the performance of any image matching algorithm. Therefore, selection, extraction and the subsequent analysis of image features have a practical significance in accomplishing robotic vision-based tasks. The applications of image feature detection and matching continue to grow in the field of ocean research.

Underwater vehicles and underwater photography systems collect image data which contain a substantial amount of information. The data is collected using on-board optic or acoustic sensors. Acoustic sensors are generally used for low-resolution data collection, and optic sensors are used for high-resolution data collection [3]. Underwater image quality is degraded due to the way light propagates in water. For underwater applications, an ideal feature detection technique is one which is invariant and robust to rotation, scale, illumination, and affine transformations [4].

Image feature detectors(or feature descriptors) are the first step of computer vision image analysis. Carefully selected feature descriptors lead to successful image matching algorithms. For this purpose, several techniques have been employed to extractfeatures from images. Out of the many available techniques, two commonly used techniques are Scale Invariant Feature Transform (SIFT) and Speeded Up Robust Features (SURF). The rest of the paper is organised as follows: Section II explains image feature extraction and characteristics of an ideal feature detector. Section III and Section IV elaborate SIFT and SURF respectively. The review is concluded in Section V.

II. FEATURE EXTRACTION

Image recognition and matching is done using local image features (also known as interest points or key points).It is a specific pattern of pixels distinct from the neighbouring pixels, which is generally associated with image properties [5, 6].The properties areedges, corners, regions and so on. A local image feature is invariant to change in scale, rotation, andillumination of the image. Local image features represent essential anchor points that summarise the image content whiledetecting and matching an image or a video [7].Fig. 1 shows a summary of local features.

The algorithm or technique used for detecting (or extracting) features from images is called detector (also known as extractor). The detector algorithm processes image features for the next stage that describes their content, that is, a feature descriptor algorithm.Feature extraction is the intermediate step between different computer vision algorithms. A summary of state of the art feature detectors is given in Table 1.

Feature detection algorithms work bychoosing and extracting features on the basis of intensity patterns in the input image. The selected image features are the backbone of computer vision algorithms and have a big impact on their performance [6]. Ideal feature detectors are typically required to have the following characteristics [5, 7]:

(1) **Quantity**: A large number of features should be extracted so that even small-sized objects have sufficient number of features detected on them.

(2) **Locality**: If the features are local, it reduces the chances of occlusion and effects of clutter. This also allowsgeometric and photometric deformations to be mathematically modelled in images taken under different viewing conditions. The locality of features is extremely important when dealing with underwater images due to the difference in illumination caused by sunlight or artificial lights.

(3) **Distinctiveness**: Distinctiveness refers to the variation in the intensity patterns of the detected features. Informativeness is essential when dealing with underwater images because of blurring effects due to turbid water and presence of bluish or greenish hue in images.

978-1-5386-8158-9/19/$31.00©2019IEEE

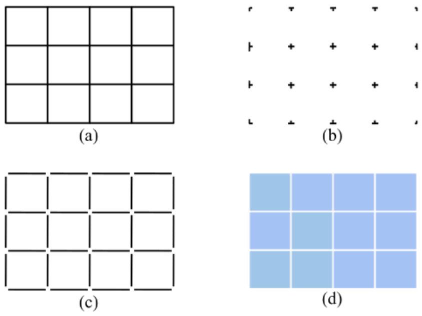

Fig. 1. Figurative representationof image features (a) input image, (b) corners, (c) edges, and (d) regions [7]

Feature Detector	Category			Invariance			Qualities			
	Corner	***Blob***	***Region***	***Rotation***	***Scale***	***Affine***	***Repeatability***	***Localisation***	***Robustness***	***Efficiency***
Harris	Yes			Yes			+++	+++	+++	++
Hessian		Yes		Yes			++	++	++	+
SUSAN	Yes			Yes			++	++	++	+++
Harris Laplace	Yes			Yes	Yes		+++	+++	++	+
Hessian Laplace		Yes		Yes	Yes		+++	+++	+++	+
DoG		Yes		Yes	Yes		++	++	++	++
Salient Regions		Yes		Yes	Yes	Yes	+	+	++	+
SIFT		Yes		Yes	Yes		++	+++	+++	++
SURF		Yes		Yes	Yes		++	+++	++	+++

TABLE I. A SUMMARY OF THE PERFORMANCE OFFEATURE DETECTION ALGORITHMS [5, 7]

(4) **Efficiency**: The detection of features should be fast enough to allow for practical time-critical applications.

(5) **Accuracy**: Accurate localisation of detected features with respect to scale (also shape, if possible) is an important quality of a feature detector.

(6) **Repeatability**: If two images of a scene or object are taken under different viewing conditions, a large number of features detected on the visible part of the scene or object should be found in both images.

When large distortions and deformations are expected, as in underwater images due to the effects of fouling and so on, it is preferred to model the deformations mathematically and then develop methods for feature extraction that are not affected by the mathematical transformations. In this method, repeatability is achieved by invariance. If deformations are relatively small, making the feature detection methods less sensitive to those deformations is generally sufficient. In this case, repeatability is achieved by robustness. Doing this may decrease the accuracy of detection, but not drastically. Robustness tackles image noise, blur, compression artefacts, geometric and photometric deviations from the mathematical models [5].

One quality may be favoured over another depending on the application. However, repeatability is the common most important quality in all applications [5]. It is directly dependent on quantity, locality, distinctiveness, efficiency, and accuracy. Improving one or more qulaities will improve repeatability, where which quality to improve is decided based on the application [7]. A high percentage of features isexpected to improve detection and matching tasks but this also negatively impacts the computation time and is not suitable for underwater systems that have limited hardware capacity and need to be power-efficient.

The most common and also the most importantimage features areedges, corners, and regions. Edgesare features where there is no change in the intensityalong the direction of the edge. Corner shows change in intensity in all directions andit is the intersection point of two (or more) edges. Regions have uniformity in the intensity valuein all directions [7].

Scale Invariant Feature Transform (SIFT) and Speeded Up Robust Features (SURF) have been shown to produce excellent results in a vast number of applications and have performed better than other descriptors in comparison studies [8]. The following sections of this paper discuss SIFT and SURF.

III. Scale Invariant Feature Transform (SIFT)

SIFT features are invariant to rotation, scale, intensity, and viewpoint change [9]. They have been shown to be robust against changes in lighting and viewpoint [10]. A histogram-based descriptor with 128 values is computed for each point in the detected set of interest points [7]. SIFT has four basic steps: (1) determine scale-space extrema using the Difference of Gaussian (DoG), (2) keypoint localisation, (3) keypoint orientation assignment based on local image gradient, (4) descriptor generator for computing the local image descriptor for each keypoint which depends on orientation and magnitude of image gradient [3, 9]. SIFT assembles a high-dimensional vector that represents image gradients within a local region of the image which leads to distinctive keypoints. Due to the distinctiveness, selecting a correct match for a keypoint from a database of other keypoints is possible [9]. High percentage of keypoints can be extracted from typical images, which leads to robustness in detecting small objects among clutter which is particularly useful for underwater images [9].

IV. Speeded Up Robust Features (SURF)

SURF is a local feature detector, partly inspired by SIFT. It is based on the Hessian matrix but uses a very basic approximation, just as DoG [1, 2, 11]. Gaussian filters used by SIFT require more computation as compared to box filters employed by SURF[12]. SURF uses the addition of 2D Haar wavelet responses and depends on the Hessian determinant for location and scale rather than using a different measure for both like in Hessian-Laplace detector. It is invariant to changes in scale and rotation. SURF has been tested for underwater feature-based image matching, but the approach was not successful for blurred underwater images [13]. However, with less murkiness and customised pre-processing, SURF has proved useful and used for feature extraction for underwater visual Simultaneous Localisation and Mapping (SLAM) [15].

V. Conclusion

This paper discusses image feature extraction for underwater application in detail. Two interest point detection schemes, SIFT and SURF, are reviewed, and their applicability in complex underwater environments is explored. The algorithms have been compared in aspects such as repeatability, invariance, robustness, efficiency, and accuracy. SIFT and SURF have proved to be robust methods of image feature extraction. Both the techniques are efficient at extracting distinctive features from underwater images required for numerous applications, though their performance trails off quickly as the water gets more turbid [10].

SIFT and SURF have demonstrated good performance in feature extraction and matching on side-scan sonar images alsowhich were used for underwater vision-based navigation [12]. However, the performance deteriorates as the region of hard sediments and rocks is encountered since there are less texture and isolated point-like variations [12]. It conclusively shows that SIFT and SURF work well on optic as well as acoustic images if texture and unique variations are present. Overall SIFT has been the more dominant algorithm providing more successful matches. SIFT is also the more popular algorithm that can match under different scales, rotations and lighting, but it was found to be significantly slower, thus not suitable for applications with real-time needs [14]. SURF is much faster and computationally-efficient; hence better suited for time-critical real-time applications. It reduces computation requirements of SIFT.

SIFT and SURF have potential applications in view matching for 3D scene reconstruction, object recognition, moving target tracking, underwater SLAM,to get better performances in complex underwater environments. Future work will aim towards an experimental evaluation of image feature detectors and descriptors as well as exploring new feature detectors for underwater applications.

References

[1] Ohya A, Kosaka A, Kak A,"Vision-based navigation by a mobile robot with obstacle avoidance using single-camera vision and ultrasonic sensing,"*IEEE Transactions on Robotics and Automation*, 14(6): 969-978, 1998.

[2] Bonin-Font F, Ortiz A, Oliver G,"Visual navigation for mobile robots: A survey,"*Journal of intelligent and robotic systems*, 53(3): 263- 296, 2008.

[3] A. Mahiddine, J. Seinturier, J.-M. Boï, P. Drap, and D. Merad, "Performances Analysis of Underwater Image Preprocessing Techniques on the Repeatability of SIFT and SURF Descriptors,"*WSCG*, p. 9, 2012.

[4] E. Karami, S. Prasad, and M. Shehata, "Image Matching Using SIFT, SURF, BRIEF and ORB: Performance Comparison for Distorted Images,"*arXiv:1710.02726*, p. 5, 2017.

[5] T. Tuytelaars, K. Mikolajczyk, et al., "Local invariant feature detectors: a survey," *Foundations and trends in computer graphics and vision*, vol. 3, no. 3, pp. 177–280, 2008.

[6] Y. Li, S. Wang, Q. Tian, and X. Ding, "A survey of recent advances in visual feature detection," *Neurocomputing*, vol. 149, pp. 736–751, 2015.

[7] E. Salahat and M. Qasaimeh, "Recent Advances in Features Extraction and Description Algorithms: A Comprehensive Survey," *arXiv:1703.06376 [cs]*, Mar. 2017.

[8] Mikolajczyk, K. and Schmid, C., "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27, 1615–1630, October 2005.

[9] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints', *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[10] K. Oliver, W. Hou, and S. Wang, "Image feature detection and matching in underwater conditions," *SPIE Defense, Security, and Sensing, Orlando*, p. 76780N, Florida, 2010.

[11] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded Up Robust Features," *Computer Vision – ECCV, 2006*.

[12] M. P. King, M. B. Anstey, and D. A. Vardy, "Comparison of Feature Detection Techniques for AUV Navigation Along a Trained Route," *OCEANS*, p. 8, 2013.

[13] D. Lee, G. Kim, D. Kim, H. Myung, and H. T. Choi, "Vision-based object detection and tracking for autonomous navigation of underwater robots,"*Ocean Engineering*, vol. 48, pp. 59–68, Jul. 2012.

[14] E. G. Fidalgo, "Experimental Assessment of Different Image Descriptors for Topological Map-Building and Scene Recognition," MSc Computer Science Thesis, 2011.

[15] J. Aulinas *et al.*, "Feature extraction for underwater visual SLAM," *OCEANS 2011 IEEE Spain*, 2011, pp. 1–7, 2011.