Received: July 23, 2020. Revised: January 28, 2021. Accepted: February 12, 2021. Published: February 21, 2021.

Image Registration using Local Invariant Relational Gabor Features for Automated Underwater Images Mosaicing

Wan Nural Jawahir Hj Wan Yussof*§, Muhammad Suzuri Hitam*§,
Ezmahamrul Afreen Awalludin**‡ and Zainudin Bachok**§
*Department of Computer Science, Faculty of Science and Technology,
**Institute of Oceanography and Environmental (INOS),
Universiti Malaysia Terengganu, 21030 Kuala Terengganu, Malaysia.

Email: §{wannurwy,suzuri,zainudinb}@umt.edu.my, ‡eafreen@gmail.com

***Abstract*—This paper describes a novel image registration in mosaic framework based on a new type of local invariant feature. Our features are located using Gabor-based features detector which gives spatial distribution of features. We sample a feature descriptor using a newly proposed Relational Gabor Features that adapts relational features onto Gabor magnitude using varying patch sizes obtained during features detection. We tested our proposed features for mosaicing underwater image sequences at ten meters depth taken in Redang Island, Terengganu, Malaysia. Based on the experimental results, the proposed method has produced in highly effective image registration method to obtain a coherent underwater image mosaic.**

***Keywords*—local invariant features, image registration, underwater image**

I. Introduction

A rapid development in underwater robotics [1], [2], [3], [4] makes survey and inspection of deep water available to the science communities e.g., archaeologists, geologists, biologists, among others. Consequently, underwater images are becoming an important tool for studying and acquiring knowledge about the seabed [5], [6], [7]. Object measurement, monitoring and exploration are among the tasks which can be performed using underwater images.

In general, an underwater image is of type video which narrowly limit the photographic view of underwater environment due to significant attenuation of visible light thus leads to lack of image contrast. Therefore, the video processing such as mosaicing is necessary to observe an underwater environment over the wide range. This technique could provide a complete representation of static scenes for a versatile visualization, accessing and analyzing information. The beneficial use of these information is for guiding environmental decisionmakers to maintain compliance with relevant statutes, regulations, and executive orders [8].

In principles, an image mosaicing operation is accomplished through two main steps. The first one is image registration that perform spatial alignment of the images and the second one is image blending that composite two images to form the mosaic. In image registration, the operation is normally solved using feature-based method [9], [10] based on the following steps:

1) Detect local feature points in each image.
2) Extract information at each of the detected feature points.
3) Match the points between two images.

Many of the mosaic algorithms used Harris corner detector [11] to find salient features that correspond between images [12], [13]. The successful of Harris corner detector depends on how the photographer overlapped his shots, and whether there are enough unique and corner-like features in the overlap. In this paper, we propose an alternative method to obtain a spatially diverse set of feature points not necessarily at the corners. Our method is based on Gabor wavelets where a set of local feature points are detected using a non-maximum suppression at every scales of multiorientation response maps. These feature points are then described using the relation with their neighborhoods on the normalized response map generated from the summation of different scales and different orientations divided by the total number of overall response maps. The proposed descriptor is called Relational Gabor features is an alternative to all formerly proposed schemes such as SIFT [14], SURF [15] and MOPs [16], that have considerably greater invariance to image scaling and rotation and robust under change in illumination and 3D camera viewpoint. Although the performance of the proposed scheme is not measured statistically in this paper but mosaicing underwater images is likely to fail if the feature-based method used is not powerful enough for image registration. This is due to the characteristics of the underwater images that often present difficult challenges for processing.

The rest of this paper is organized as follows: Section II is devoted to detail the proposed method to perform image registration process. Then, the next section present the rest of mosaic algorithms used in this work. Some results are presented in Section IV and finally we conclude our work in Section V.

II. Relational Gabor Features

In this section, we begin by describing the proposed feature points detector and next establish the proposed local features

descriptor of Relational Gabor Features (RGF). Our proposed features detector and descriptor are based on Gabor wavelets. For detail explanation of this wavelet, one can refer to [17].

A. Gabor-based Feature Points Detector

A set of Gabor wavelets $\Psi_{v,u}$ of different scale and orientation is convolved with an image $I(z)$ to estimate the magnitude of local frequencies of that approximate scales and orientations. It is defined by:

$$\xi(z; f, \theta) = \Psi(z; f, \theta) * I(z) \tag{1}$$

The common method for reducing the computational cost for the above operation is to perform the convolution in Fourier space. This way, the operation is done based on simple element-wise multiplication with linear time complexity:

$$\xi(z; f, \theta) = \mathcal{F}^{-1}\{\mathcal{F}(\Psi(z; f, \theta))\mathcal{F}(I(z))\} \tag{2}$$

where $\mathcal{F}$ denotes the fast Fourier transform and $\mathcal{F}^{-1}$ is its inverse.

It is important to design a detector that can cope with scaling problem when the point itself lacks information to describe an image. In other words, the detector must have support regions or patches that can be modeled using a scale-space approach. The proposed detector can also be implemented to localize multi-scale regions. The task for detecting multi-scale regions of the proposed method is by generating a series of response maps at every scale and the scale index of each point is known directly while detecting its location. To do this, all the response images in each scale are added together to produce a response map $\hat{\xi}_u(z; f_u)$ that represents an individual scale-space. Thus, $\forall u = 0, \ldots, U-1$,

$$\hat{\xi}_u(z; f_u) = \frac{\sum_{v=0}^{V-1} \psi_{u,v}(z; f_u, \theta_v) * I(z)}{V}. \tag{3}$$

Then, a set of interest points for all scales is obtained by applying non-maximum suppression at all response maps. To do this, the response maps $\hat{\xi}_u$, $\forall u = 1, \ldots, U$ are dilated by performing a grey scale morphological dilation as expressed in the following equation:

$$[\hat{\xi}_u \oplus b](z) = \max_{(s,t)\in b}\{\hat{\xi}_u(x-s, y-t)\} \in [\varepsilon_1, \varepsilon_2] \tag{4}$$

where b is the structuring element with the size $2r+1$ and r is the radii of region considered in non-maximum suppression. The value r is set according to the scale size. The greater the scale size, the smaller r value will be set. This is to ensure that the smaller scale size does not dominate the overall interest points. Local maxima is extracted by finding the points that match the dilated image with values in the range $[\varepsilon_1, \varepsilon_2]$. The threshold value, ε_2 must be greater than threshold value, ε_1 and must be assigned in the interval $\{0; 1]$ whereas ε_1 must be assigned in the interval $[0, 1\}$. For every set of interest points extracted at every scale, the location of points are stored in keypoints list together with their scale index. The process of computing the proposed multi-scale detector is illustrated in Fig. 1.

B. Local Relational Gabor Descriptor

Once a set of feature points is obtained, the next step is to associate features across images. In order to automate features association, we propose a Relational Gabor descriptor that encodes information at each feature point by applying the relational function to the magnitude value difference of the pixels lying on the specific distance and phase to the reference point. The proposed descriptor is motivated from the use of relational kernels introduced in [18].

The relational kernels are based on the Local Binary Pattern (LBP) texture features [19] which threshold the relation between a center pixel and the pixels in its neighborhood into a binary pattern (0 and 1). Applying this to all pixels in a circular neighborhood of the center pixel, a binary pattern is obtained which is then transformed into a unique number as follows:

$$LBP = \sum_{i=0}^{n-1} s(v_i - v_c)2^i \tag{5}$$

where

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0, \end{cases} \tag{6}$$

where v_i and v_c are the grayvalues at a neighboring pixel and the center pixel, respectively. The number of the pixels in the circular neighborhood is denoted by n. Since the signed difference $(v_i - v_c)$ is considered, the effect of grayscale shifts is totally eliminated. Invariants against scaling of the grayscale is achieved by the s operator as the sign of the difference is mapped to 0 or 1.

It is obvious that the discontinuity of the LBP operator (the s function) makes these features sensitive to noise. A small disturbance in the image may cause a big deviation of the feature. To overcome this problem, Schael[18] has introduced an operator which extends the step function in Eq. 6 to a ramp function giving values in the range of $[0, 1]$:

$$rel(x) = \begin{cases} 1, & x < -\varepsilon \\ \frac{\varepsilon - x}{2\varepsilon}, & -\varepsilon < x < \varepsilon, \\ 0, & \varepsilon < x, \end{cases} \tag{7}$$

where ε is a threshold parameter. This way, the features are much more robust against noise. However, to set ε parameter is a non-trivial process. Therefore, we extend a ramp function to a log-sigmoid function for simplicity and to avoid the manual trial and error of what ε parameter that is to be set. Our relational features is defined as:

$$\widehat{rel}(x) = \frac{1}{1 + \exp(-x)} \tag{8}$$

As opposed to LBP, relational features use two circular neighborhoods. Let (x, y) be the coordinate of the central pixel in two-dimension, taking into consideration the phase shift, ϕ, the coordinate of (x_1, y_1) and (x_2, y_2) are given by:

$$(x_1, y_1) = x + r_1\cos(\theta), y + r_1\sin(\theta), \tag{9}$$

and

$$(x_2, y_2) = x + r_2\cos(\theta + \phi), y + r_2\sin(\theta + \phi). \tag{10}$$

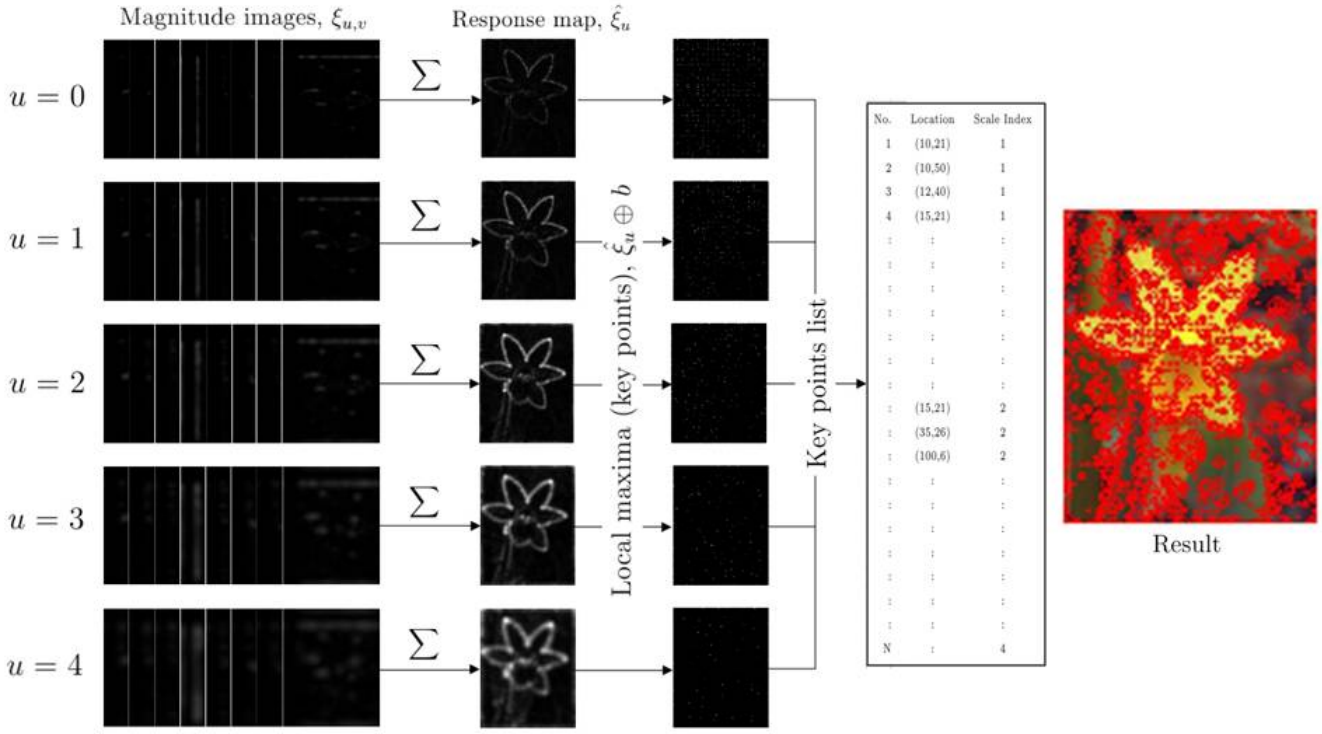


Fig. 1. The proposed multi-scales interest points detector for five scales and eight orientations.

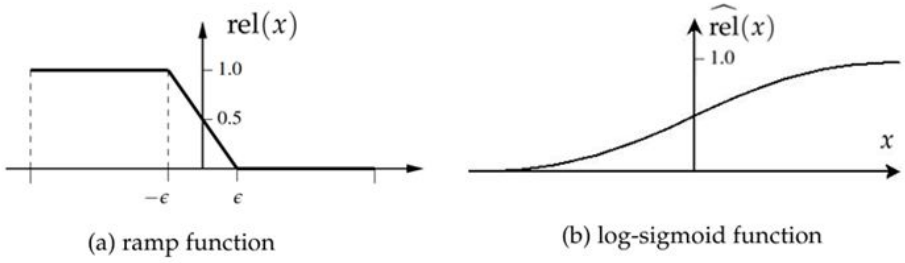


Fig. 2. Graphs of ramp function and log-sigmoid function.

From the equation defined in Eq. (9), r_1 is set based on the scale index obtained during the feature points detection process and r_2 in Eq. 12 is equal to $2r_1$. The θ value is given by $j.2\pi/n,\ \forall j = 1,\ldots,n$ where n is a number of neighborhood points.

Using Eq. 2, we can obtain final Gabor magnitude which we denote as $\grave{\xi}$ as the sum from a series of response maps at all scales and orientations.

$$\grave{\xi} = \frac{\sum_{u=0}^{U-1}\sum_{v=0}^{V-1}\psi_{u,v}(z; f_u, \theta_v) * I(z)}{U * V}. \tag{11}$$

Next, we define our proposed Relational Gabor features, $\mathcal{RG}$ based on relational function defined in Eq. 8 calculated on a keypoint (x, y) of the magnitude image $\grave{\xi}$ as center. The function is given by:

$$\mathcal{RG} = \frac{\sum_{j=1}^{n}\widehat{rel}(\grave{\xi}(x_2, y_2) - \grave{\xi}(x_1, y_1))}{n}. \tag{12}$$

The result of the above equation is a single value representing information of one local point. This kind of feature is not distinctive enough as one point might have the same value to the other points. More features are needed to describe one local point. By using a set of varying ϕ values and concatenating them into one form producing a set of $\mathcal{RG}$ features vector. To do this we systematically set different values of ϕ as follows:

$$\phi_k = \frac{k.(\theta_1 - \theta_0)}{m},\ \forall k = 0,\ldots,m-1. \tag{13}$$

where m is the total number of ϕ we used. Then, the function

$\mathcal{RG}$ with m-tuple can be rewritten as:

$$\mathcal{RG}^m = \left\{ \frac{\sum_{j=1}^{n} \widehat{rel}(\theta_j, \phi_k)}{n} \middle| 0 \leq \theta_j < 2\pi, \theta_0 \leq \phi_k < \theta_1 \quad \forall k = 0, \ldots, m-1 \right\}. \quad (14)$$

The process is illustrated in Fig. 3. As shown in the figure, red node is a keypoint or pixel under consideration which acts as center of the circles. Green nodes and blue nodes are neighborhoods of the first circle and second circle, respectively. The distance for green nodes from the center node are shown with green lines whereas the distance for blue nodes from the center node are shown by blue lines. The feature is formed by integrating all values from the difference between value of the blue node and value of the green node at each θ that put into the $\widehat{rel}$ function. Bilinear interpolation is used for points not lying exactly on image grid. By setting different values of ϕ, we can yield a feature vector of length m. Finally, we sort the features into an ascending order to obtain rotational invariance features.

III. Features Matching

For matching the proposed features descriptor across images, we calculate the squared Euclidean distance, D between two feature vectors. The elements of the first features vector and the second features vector are represented by the notation $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$, respectively where n is the dimensionality of the feature space.

$$D(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} (x_i - y_i)^2 \quad (15)$$

Next, we look for the strongest matches from one image into the other. For this process, we follow Lowe′s method [20] of thresholding the ratio of the best match score to its second best match score. For our implementation, features whose ratios are higher than 0.5 are rejected.

From Fig. 4, some features in the left image are likely to match incorrect correspondences in the right image due to similar image structure they shared. These features are called outliers and need to be removed. We use RANdom SAmple Consensus (RANSAC [21]) to reject outliers that are basically identified as the points that do not follow the dominant image motion estimated by RANSAC. Once we obtain a set of distinctive features or inliers, the next step is to robustly fit homography where a projective transformation is estimated from at least four point correspondences [22]. However, more point correspondences will generate more accurate homography. Using the estimated homography that is computed from the inliers, we can composite the images to form a mosaic.

IV. Application on Underwater Images

Now that all the steps and relevant details of the mosaicing operation have been described, some of the mosaics created for underwater panoramic will be presented. We have tested our method using large-area underwater image sequences of 10 meters depth that were acquired during the experiments in Redang Island, Terengganu, Malaysia in 2011 by marine scientist from Institute of Oceanography (INOS)[1].

TABLE I
Number of Initial Correspondences and Inliers.

Mosaic Images	Initial Correspondences Image(a)-Image(b)	Inliers Image(a)-Image(b)	Initial Correspondences Image(c)-Image(b)	Inliers Image(c)-Image(b)
Set 1	168	146	88	47
Set 2	64	40	54	30

Two sets of image sequences were used for demonstrating the applicability of the proposed features. The results are presented in Fig. 6. In Table I, we show the number of initial correspondences detected by the proposed features and the number of inliers. It should be noted that, the proposed features retained more than 50% of the number of correspondences among all overlapping image pairs from the initial correspondences thus provides a highly effective alignment during image registration process. Obtaining a correct alignment is important to ensure that the quality of mosaic result is at the highest level possible.

V. Conclusion

In this paper, we have presented a new type of local invariant features, which we call Relational Gabor features. The proposed features utilized a Gabor wavelet technique for locating feature points, and relational features were extracted on a magnitude of Gabor response using a log-sigmoid function. These features allow us to recover the spatial alignment between two images by performing image registration process using a method such as RANSAC.

The proposed features have been tested for underwater video images mosaicing, covering large area and results have been presented visually to illustrate its performance. Our results show that the features can be matched not necessarily between adjacent frames. Features in the first frame for example can be matched with the features in the sixth frame as long as some parts of objects in the first image can still be seen in the second one regardless of geometrical transformations happened to the objects.

Based on the demonstration made for the application of underwater images mosaicing, it can be observed that the mosaic images can greatly simplify subsequent processing for underwater image classification. This will be our future work to investigate the usefulness of the proposed features for this process since it is well-known that local invariant features can be generalized for other applications. In addition, we will perform a more detailed, quantitative analysis of feature matching performance.

[1]INOS is a Malaysian 7th Higher Institution Center of Excellent (HiCoE) in Marine Science.

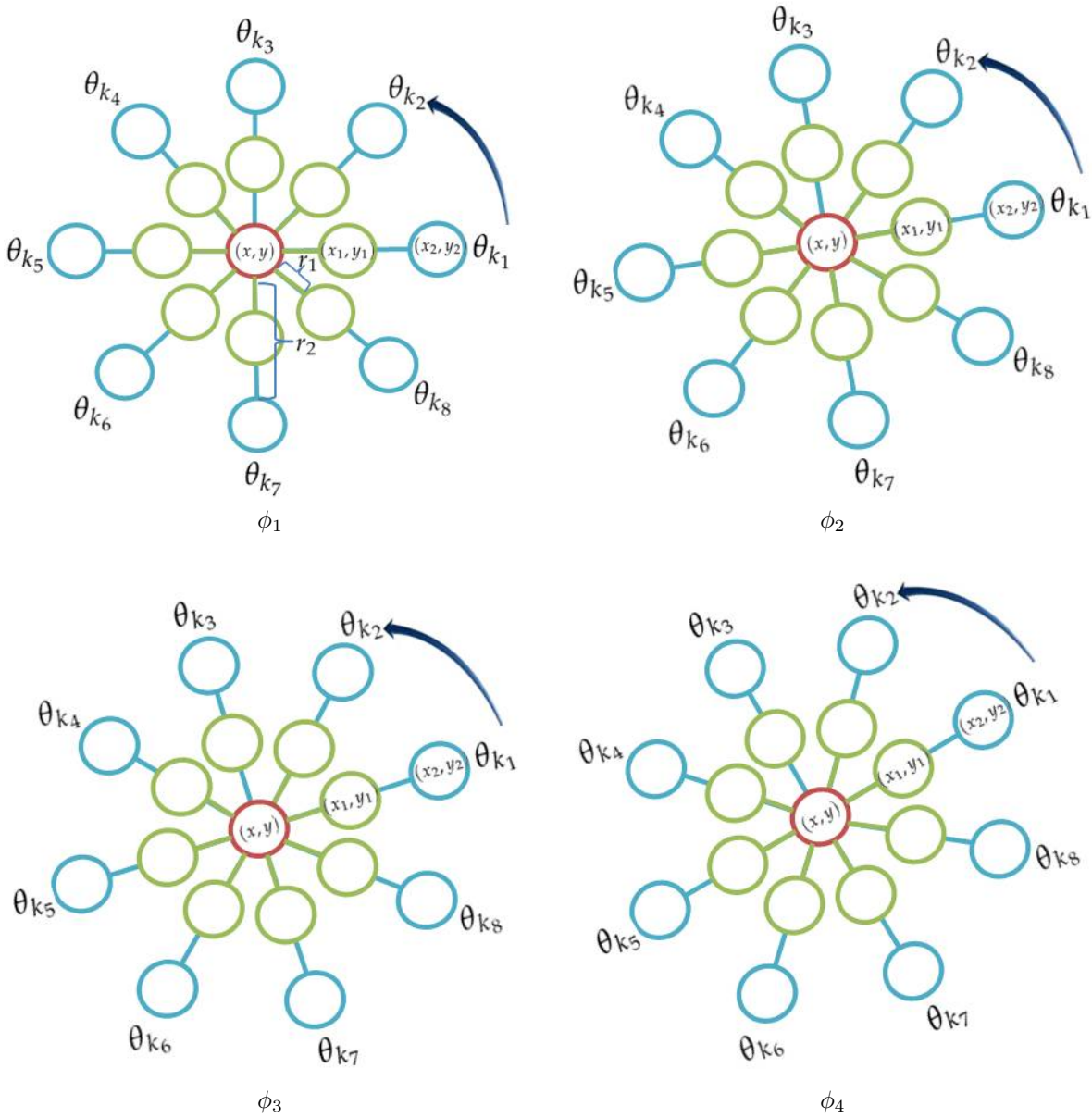


Fig. 3. The illustration for the calculation of the proposed RGF features vector using four-tuple. Note that by changing ϕ value, the locations of the neighbohood are slightly moved at anti-clockwise direction.

ACKNOWLEDGMENT

The authors would like to thank Institute of Oceanography and Environmental (INOS) for providing underwater images for this study.

REFERENCES

[1] C. S. Chin and S. H. Lum, *Rapid modeling and control systems prototyping of a marine robotic vehicle with model uncertainties using xPC Target system*, Journal of Ocean Engineering, vol. 38(17–18), 2011, pp. 2128–2141, 2011.

[2] R. S. Pazmiño, C. E. Garcia Cena, C. A. Arocha and R. A. Santonja, *Experiences and results from designing and developing a 6 DoF underwater parallel robot* Journal of Robotics and Autonomous Systems, vol. 59(2), pp. 101-112, 2011.

[3] P. Kimball and S. Rock, *Sonar-based iceberg-relative navigation for autonomous underwater vehicles*, Deep Sea Research Part II: Topical Studies in Oceanography, vol. 58(11–12), pp. 1301-1310, 2011.

[4] M. Santhakumar and T. Asokan, *Power efficient dynamic station keeping control of a flat-fish type autonomous underwater vehicle through design modifications of thruster configuration*, Journal of Ocean Engineering, vol. 58, pp. 11–21, 2013.

[5] A. Elibol, R. Garcia and N. Gracias, *A new global alignment approach for underwater optical mapping*, Journal of Ocean Engineering, vol. 38, pp. 1207–1219, 2011.

[6] L. Yin, B. Chen, J. Hu and P. Cai, *Research on the Key Technologies for Underwater Image Communication System*, Procedia Engineering, vol. 24, pp. 610-615, 2011.

[7] V. Lucieer, N. A. Hill, N. S. Barrett and S. Nichol,*Do marine substrates look and sound the same? Supervised classification of multibeam acoustic data using autonomous underwater vehicle images*, Estuarine, Coastal and Shelf Science, vol. 117(20) pp. 94-106, 2013.

[8] R. PAmela Reid, D. Lirman, N. Gracias, S. Negahdaripour, A. Gleason and B. Gintert, *Application of Landscape Mosaic Technology to Complement Coral Reef Resource Mapping and Monitoring*, Technical Report: SERDP Project RC-1333, University of Miami, 2010.

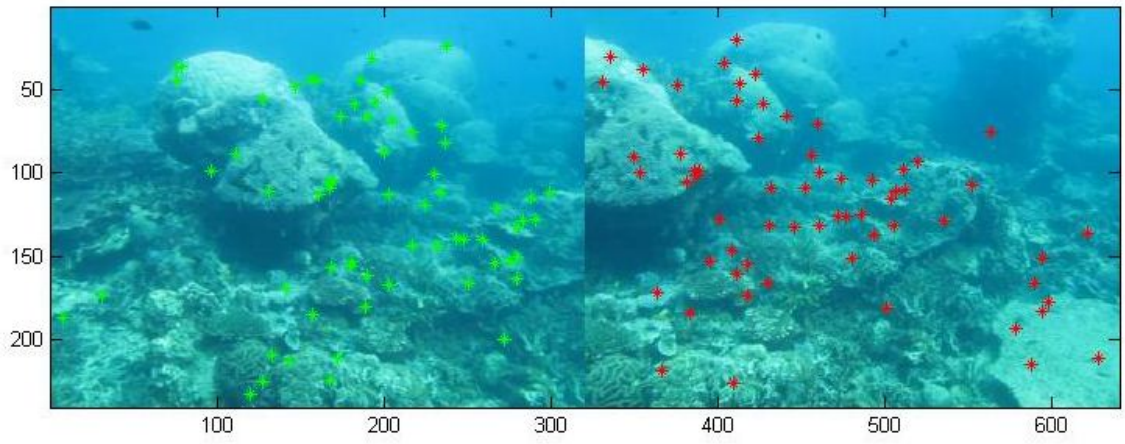


Fig. 4. Initial correspondences of the proposed features.

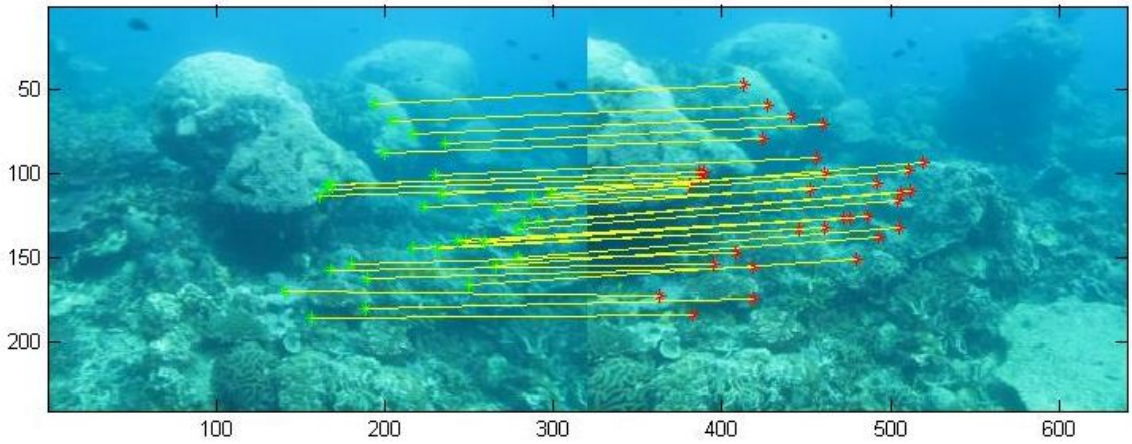


Fig. 5. RANSAC samples the potential inlier features which form homography to composite images.

[9] X. Fang, Z. Pan, G. He, and L. Li, *A new method of feature based image mosaic*. International Journal of Image and Graphics, 6(03), pp. 497-510, 2006.

[10] F. Wu and X. Fang, *An improved RANSAC homography algorithm for feature-based image mosaic*, In Proc. of the 7th WSEAS Int. Conf. on Signal Processing, Computational Geometry and Artificial Vision, pp.202-207, 2007.

[11] C. Harris and M. Stephens, *A combined corner and edge detector*, Alvey Vision Conference, pp. 147–151, 1988.

[12] G. Y. Tian, D. Gledhill and D. Taylor, *Comprehensive interest points based imaging mosaic*, Pattern Recognition Letters, Volume 24(9–10), pp. 1171-1179, 2003.

[13] M. Kise and Q. Zhang, *Creating a panoramic field image using multi-spectral stereovision system*, Computers and Electronics in Agriculture, vol. 60(1), pp. 67-75, 2008.

[14] D. G. Lowe, *Object recognition from local scale-invariant features*, In: Proc. of the International Conference of Computer Vision, pp.1150–1157, 1999.

[15] H. Bay, A. Ess, T. Tuytelaars and L. Van Gool, *SURF: Speeded Up Robust Features*, Computer Vision and Image Understanding (CVIU), vol. 110, no. 3, pp. 346–359, 2008.

[16] M. Brown, R Szeliski and S. Winder, *Multi-image matching using multi-scale oriented patches*, Computer Vision and Pattern Recognition, vol. 1, pp. 510-517, 2005.

[17] J. K. Kämäräinen, *Feature extraction using Gabor filters*, Ph.D Thesis, Lappeenranta University of Technology, 2003

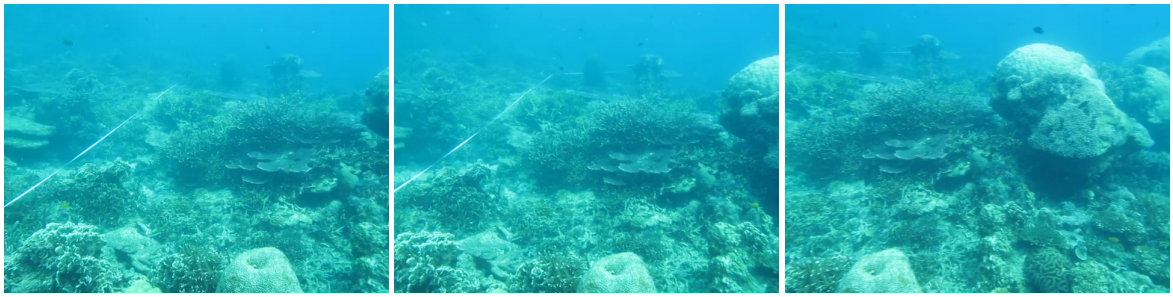
[18] M. Schael, *Invariant grey scale features for texture analysis based on group averaging with relational kernel function*, Internal Report 01/01, University of Freiburg, 2001.

[19] T. Ojala, M. Pietikäinen and T. Mäenpää, *Gray scale and rotation invariant texture classification with local binary patterns*, In: Vernon, D. (ed.) ECCV2000, LNCS, vol. 1842, pp. 404–420, Springer Heidlelberg, 2000.

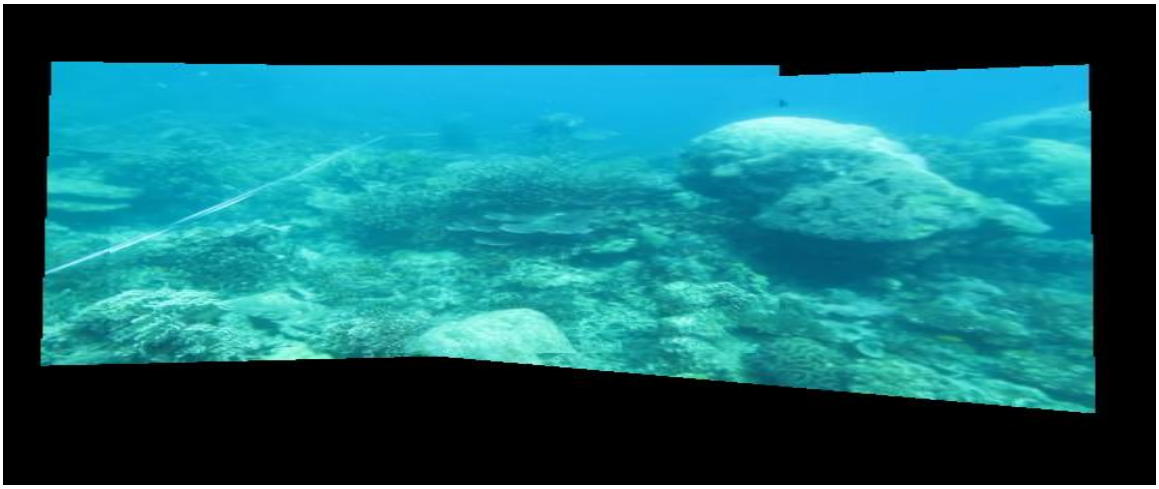
[20] D. G. Lowe, *Distinctive image features from scale-invariant keypoints*, International Journal of Computer Vision, vol. 2, pp. 91–110, 2004.

[21] M. A. Fischler and R. C. Bolles , *Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography*, Comm. of the ACM, vol. 24 (6), pp. 381–395, 1981.

[22] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision (2nd ed.*, Cambridge University Press, 2003.

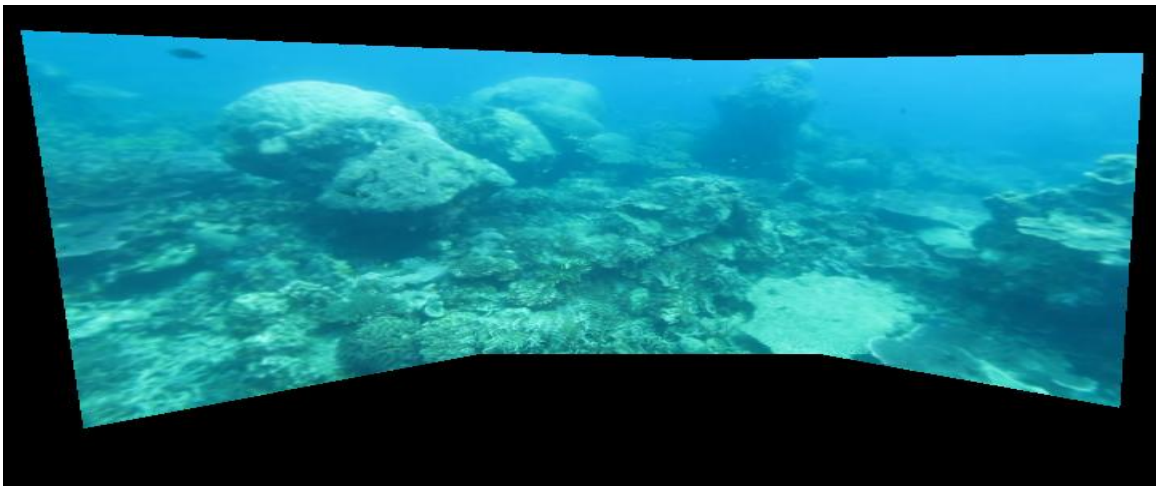


(a) Set of underwater images for mosaicing.



(b) Final mosaic from (a).

(c) Set of underwater images for mosaicing



(d) Final mosaic from (c).

Fig. 6. Mosaicing of underwater images.