

Extraction of visual descriptors in underwater images

Varbanov Vladimir

Institute of Metal Science, Equipment and Technology with Hydroaerodynamic Center at Bulgarian Academy of Sciences, 67 Shipchenski Prohod Street, 1574 Sofia, Bulgaria
vdvarbanov@abv.bg

Abstract: Visual points of interest have various application in different fields of engineering and applied sciences. In example they are used to estimate position and orientation of the scanning device, whether it is a camera, LIDAR or sonar, which is crucial in environment, lacking inertial navigation system. Also are heavily used in Simultaneous Localization And Mapping (SLAM), as well in scene reconstruction and other applications in robotics. The current paper explore the performance of different image descriptors in underwater environment. The lack of navigation system in such environment requires use of said descriptors to estimate the position and orientation of the camera in respect to not moving object.

Keywords: Image descriptors, Underwater, Orb, Harris, BRISK

1. Introduction

Ocean is becoming more and more lucrative for industrial and scientific endeavors. With increase of the infrastructure build off shore and more renewable energy and agricultural resource being obtained from the ocean, a new set of problems start to emerge. Solving this problems require development and construction of specific underwater equipment. Sometimes such equipment is based physical principles, which are well known like the hydro acoustic. Unfortunately such equipment comes with high cost, because of the niche market and respectively small volume of devices made. One of the possible solutions is to adapt well-known technologies, which were mass produced for the needs of "on shore" industries to work in the harsh and specific maritime environment [1]. As result we can see increased usage of optical devices – lasers, cameras, LIDARs, etc. Often these devices work in conjunction with traditional acoustic devices. Unfortunately operating in different environment require additional research and adaptation. However there is silver lining – many of the algorithms, technologies, and principles in general have already been well studied[2, 3]. The paper focus on the comparison of different algorithms for feature extraction in underwater images. Such visual features are called image descriptors and represent specific distribution of intensity around a point of interest, usually which usually represent a corner in the image.

2. Basic principles of image descriptors

2.1 Harris

A corner is a point surrounded by two dominant edges, pointing in different direction. It usually is represented as rapid change of intensity in the neighboring pixels, as shown on fig. 1 directions. In other words, a corner can be interpreted as the junction of two edges, where an edge is a sudden change in image brightness. Finding a corner point is done by creating small window around every pixel and moving that window in different directions until we find position in which small shift of the window in any direction will result in big change of the squared sum of the intensity of all pixels, under that window[4, 5].

$$E(u, v) = \sum_{x,y} w(x, y) [I(x + u, y + v) - I(x, y)]^2 \quad (1)$$

in equation (1) $E(u, v)$ represents the change of intensity I , by shifting the window u and v pixels horizontally and vertically. We have to maximize this function $E(u, v)$ for corner detection. That means, we have to maximize the second term.

$$E(u, v) \approx [uv] \Sigma \begin{bmatrix} I_x^2 & I_{xy} \\ I_{xy} & I_y^2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = [uv] M \begin{bmatrix} u \\ v \end{bmatrix} \quad (2)$$

The directions for the largest and smallest intensity increases are found by solving the eigenvectors of M , as shown in equation (3), where λ_1 and λ_2 are the Eigen values of M , and R is a score

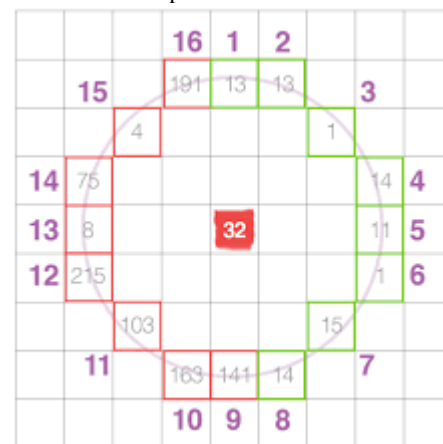
$$\begin{aligned} R &= \det M - k(\text{trace} M)^2 \\ \det M &= \lambda_1 \lambda_2 \\ \text{trace} M &= \lambda_1 + \lambda_2 \end{aligned} \quad (3)$$

Based on equation (3) we decide whether the pixel belong to image corner, flat region or edge based on the R [6]:

- $|R| \rightarrow 0$ means $\lambda_1 \rightarrow 0$ and $\lambda_2 \rightarrow 0$ and the pixel lie on a flat region
- $R < 0$ means $\lambda_1 \gg \lambda_2$ or $\lambda_2 \gg \lambda_1$ and pixel lie on an edge
- $R \gg 0$ means $\lambda_1 \gg 0$ and $\lambda_2 \gg 0$ and pixel lie on a corner

2.2. BRISK

Brisk feature detector finds descriptors which are scale and rotation invariant to a degree. The edges are detected by using FAST edge detector algorithm where we shift a window around every pixel. In every shift we check the values of the pixels that lie on a circle distanced from the center, as shown on figure 1. If more than half the consecutive pixels on that circle have higher/lower



value than the central pixel, then the pixel is considered a corner.

Figure 1. Example of FAST corner detection. Squares represent pixels in the image, respectively the gray numbers inside are the pixel values. Red square is the central pixel, and the purple numbers are the indices of a pixels. The green squares show the pixels where criteria of pixel value in consecutive pixels to be higher/lower than the central pixel. The values in empty squares are not taken into account in the shown iteration.

In order to achieve invariance to scale which is crucial for high-quality key points, we are searching for key points not only in the image plane, but also in scale-space using the FAST score as a measure for saliency. For this purpose the image is sampled up in multiple steps, where at each step an interpolation is applied, as shown on figure 2[7].

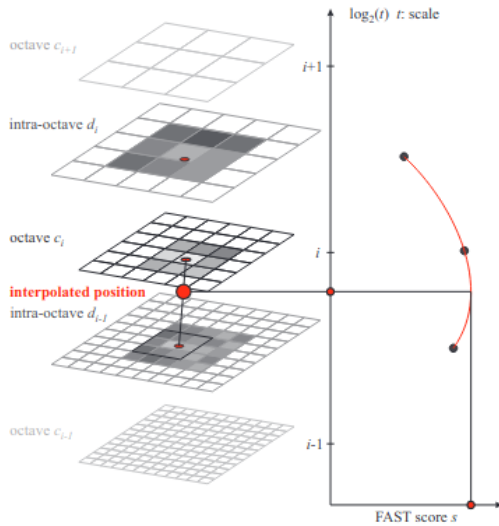


Figure 2. Finding key points in different scale

Achieving rotation invariance is done using a pattern for sampling the neighbor pixels around the key-point. The pattern is illustrated in figure 3.

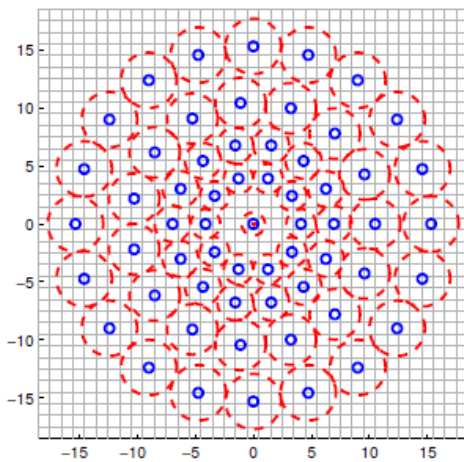


Figure 3. BRISK sampling pattern with 60 points, where the blue circles are the sampling locations and the red ones are the standard Gauss deviation

Let's for a key point k , consider N . $(N - 1)/2$ sampling points (p_i, p_j) with smoothed intensity values in the image at these points $I(p_i, \sigma_i)$ and $I(p_j, \sigma_j)$. Then local gradients can be written as[8]:

$$g(p_i, p_j) = (p_j - p_i) \cdot \frac{I(p_j, \sigma_j) - I(p_i, \sigma_i)}{\|p_j - p_i\|^2} \tag{4}$$

There will be S short pairs and L long pairs. Iterating over every short pair we can estimate the overall pattern direction of the gradient to be:

$$g = \begin{pmatrix} g_x \\ g_y \end{pmatrix} = \frac{1}{L} \cdot \sum g(p_i, p_j) \tag{5}$$

the orientation around each key point is calculated by the formula

$$\alpha = \text{artanh}(g_y, g_x) \tag{6}$$

The bit vector descriptor d_k is constructed based on the short distance intensity between points $(p_i^\alpha, p_j^\alpha) \in S$ in a way the each bit takes value:

$$b = \begin{cases} 1, & I(p_j^\alpha, \sigma_j) > I(p_i^\alpha, \sigma_i) \\ 0, & I(p_j^\alpha, \sigma_j) \leq I(p_i^\alpha, \sigma_i) \end{cases} \tag{7}$$

Finally the descriptor points are associated by searching minimal Hamming distance.

2.3. ORB

ORB image descriptor is very similar to BRISK as it uses FAST algorithm to find key points, but instead of up sampling and interpolating the key features it down sample the original image to three layer pyramid, and at each layer, a Gaussian filter is applied[9].

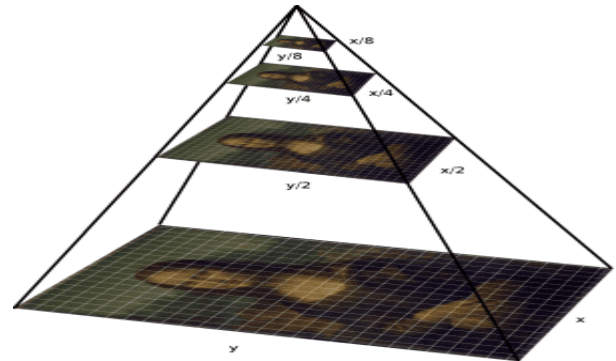
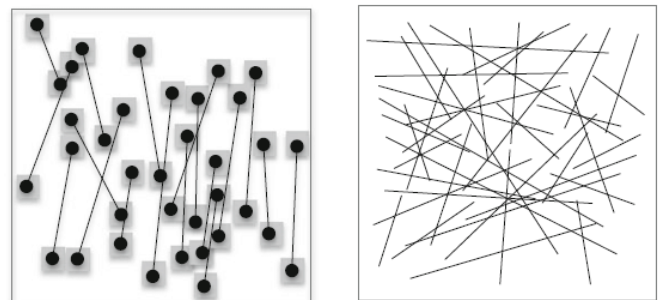


Figure 4. Down sampled image

The down sampling is done in order to increase the scale invariance and centroid method is used to increase rotation invariance. For every feature point the given point is considered as an origin, of the centroid and a circular area is constructed around said origin. Then weighted average is calculated for the area defined by the circle and the given point is considered center of mass of the centroid. Usually the origin and the center of mass do not match and are shifted from one another and the centroid position C is calculated in the neighborhood S . Based on that shift the angle between the origin O and the center of mass C is calculated by formula 6. This way to every key point is added direction. Based on that direction for every two pairs of point a correction is added. The selected pairs are further reduced, by keeping only the pairs with the lowest



correlation, as shown in figure 5[10].

Figure.5 On the left side is shown reduced set of pair from the whole set on the right.

Further the descriptor matching is done as in BRISK.

3. Experimental results

The article compare the performance of the depicted visual descriptors on underwater images. For the purpose a set of underwater photos are chosen from free data set U45. Each photo is paired with the same one, but rotated on 15 degrees counterclockwise. The purpose is to test rotation invariance of the said descriptors. Future work will include translation and scale in addition to the rotation.

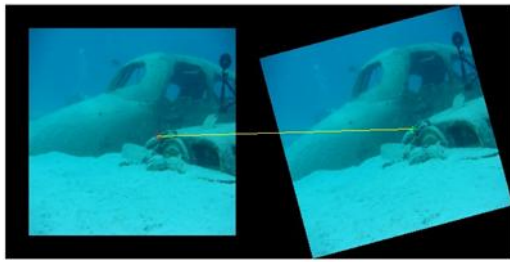


Figure. 6a – Descriptor matching of plane foliage using Harris detector

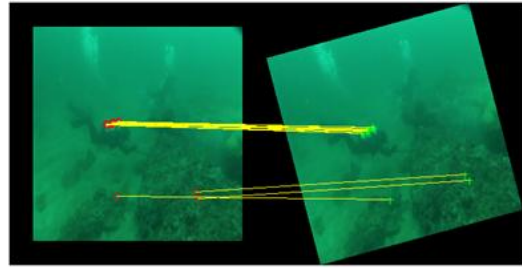


Figure. 7c – Descriptor matching of diver using ORB detector

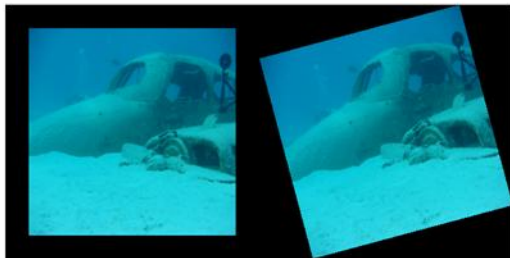


Figure. 6b – Descriptor matching of plane foliage using BRISK detector

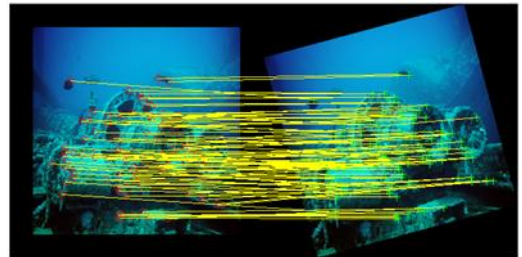


Figure. 8a – Descriptor matching of using Harris detector

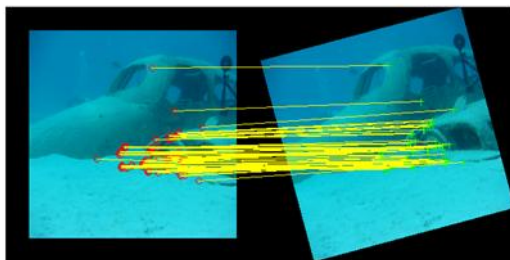


Figure. 6c – Descriptor matching of plane foliage using ORB detector

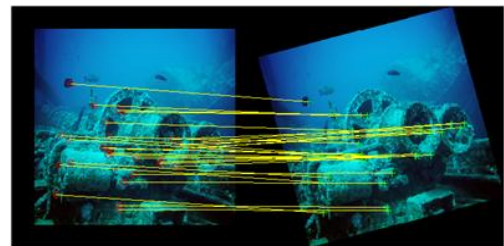


Figure. 8b – Descriptor matching of using BRISK detector

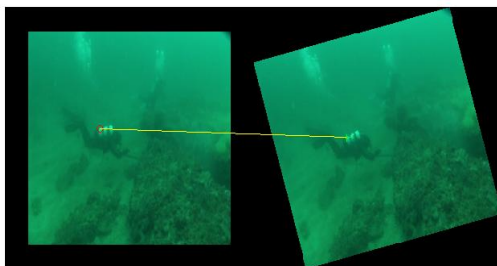


Figure. 7a – Descriptor matching of diver using Harris detector

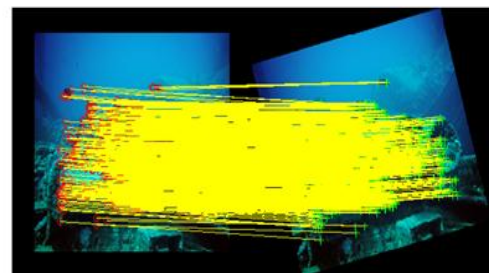


Figure. 8c – Descriptor matching of ship deck using ORB detector

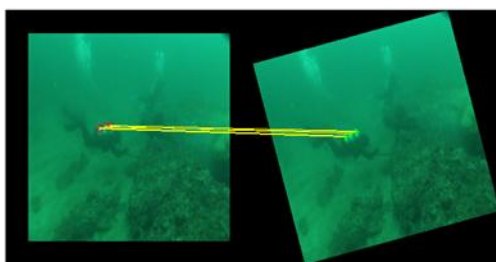


Figure. 7b – Descriptor matching of diver using BRISK detector

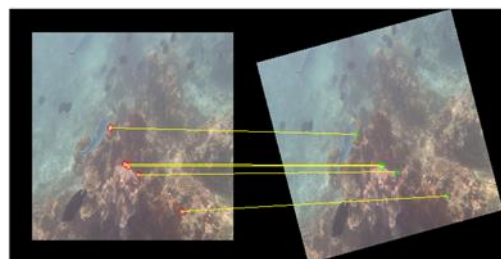


Figure 9a – – Descriptor matching of underwater sea bed using Harris

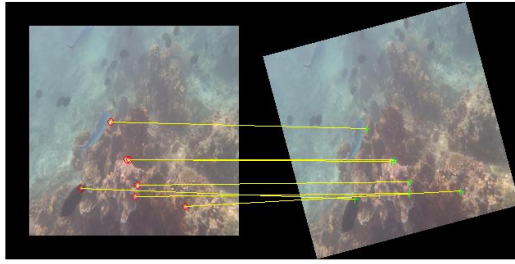


Figure 9b -- Descriptor matching of underwater sea bed using BRISK

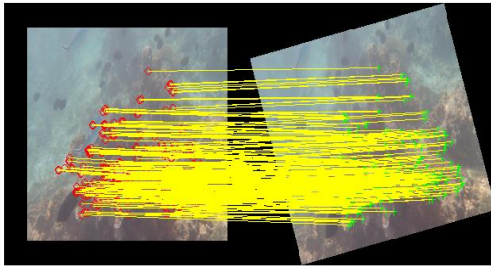


Figure 9c -- Descriptor matching of underwater sea bed using ORB

4. Conclusion

From key points association on figures 6 to 9 it becomes evident that ORB image detector is perform best on underwater images, by large margin. Future work will include addition to other descriptors as well as comparison in respect to errors and time. The pictures were hand-picked to test situation where the object that has been viewed lacks texture like the plane from figure 7 as well as where pictures are subjected to "haze", like in pictures 8 and 9, which result in blurry image and again lack of texture. However ORB showed satisfying result in all scenarios.

5 References %ORB%

1. Wen Liu, Shuo Wang, Zhongliang Deng, (Senior Member, IEEE) Tingting Fan, Mingjie Jia, Hong Chen, A Review of Image Feature Descriptors in Visual Positioning, Proceedings International Conference on Indoor Positioning and Indoor Navigation, 29 Nov. – 2 Dec. 2021, Lloret de Mar, Spain
2. Demirhan M , Premachandra C . Development of an Automated Camera-Based Drone Landing
3. System. IEEE Access 2020, 8:202111-202121. Rublee E , Rabaud V , Konolige K , et al. ORB: An efficient alternative to SIFT or SURF[C]// 2011 International Conference on Computer Vision. IEEE, 2012.. %Harris%
4. A. Bhatia, W. E Snyder, and G. Bilbro, Stacked integral image, in IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2010, pp. 1530–1535. <https://doi.org/10.1109/ROBOT.2010.5509400>.
5. Javier S´anchez¹, Nelson Monz´on², Agust´in Salgado¹ An Analysis and Implementation of the Harris Corner Detector Submitted on 2018–06–04, accepted on 2018–09–18. ISSN 2105–1232 c 2018 IPOL & the authors CC–BY–NC–SA
6. C. Harris and M. Stephens, A combined corner and edge detector, in Alvey Vision Conference, vol. 15, Manchester, UK, 1988, pp. 10–5244.
7. Leutenegger, Stefan, Siegwart, Roland, BRISK: Binary Robust Invariant Scalable Keypoints, IEEE ISBN:978-1-4577-1102-2 26 January 2012
8. M. Calonder, V. Lepetit, C. Strecha, and P. Fua. BRIEF: Binary Robust Independent Elementary Features. In Proceedings of the European Conference on Computer Vision , (ECCV), 2010. 1, 2, 3, 4
9. Chuan Luo et al 2019 Overview of Image Matching Based on ORB Algorithm. et al 2019 J. Phys.: Conf. Ser. 1237 032020
10. Li X H, Xie C M, Jia Y Z, Zhang G F. Fast target detection algorithm based on ORB feature[J], Journal of Electronic Measurement and Instrument, 2013, 27(5):455-460.