Appearance-based odometry and mapping with feature descriptors for underwater robots

Silvia Silva da Costa Botelho[1]*, Paulo Lilles Jorge Drews Junior[2], Mônica da Silva Figueiredo[1], Celina Haffele Da Rocha[1], Gabriel Leivas Oliveira[1]

[1]Universidade Federal do Rio Grande – FURG, Rio Grande, RS, Brazil

[2]Universidade Federal de Minas Gerais – UFMG, Belo Horizonte, MG, Brazil

Abstract: The use of Autonomous Underwater Vehicles (AUVs) for underwater tasks is a promising robotic field. These robots can carry visual inspection cameras. Besides serving the activities of inspection and mapping, the captured images can also be used to aid navigation and localization of the robots. Visual odometry is the process of determining the position and orientation of a robot by analyzing the associated camera images. It has been used in a wide variety of non-standard locomotion robotic methods. In this context, this paper proposes an approach to visual odometry and mapping of underwater vehicles. Supposing the use of inspection cameras, this proposal is composed of two stages: i) the use of computer vision for visual odometry, extracting landmarks in underwater image sequences and ii) the development of topological maps for localization and navigation. The integration of such systems will allow visual odometry, localization and mapping of the environment. A set of tests with real robots was accomplished, regarding online and performance issues. The results reveals an accuracy and robust approach to several underwater conditions, as illumination and noise, leading to a promissory and original visual odometry and mapping technique.

Keywords: *Robotics, Computer Vision, Underwater Vehicles,Topological Maps, Self-localization and mapping.*

1. Introduction

In mobile robot navigation, classical odometry is the process of determining the position and orientation of a vehicle by measuring the wheel rotations through devices such as rotary encoders. While useful for many wheeled or tracked vehicles, traditional odometry techniques cannot be applied to robots with non-standard locomotion methods. In addition, odometry universally suffers from precision problems, since wheels tend to slip and slide on the floor, and the error increases even more when the vehicle runs on non-smooth surfaces. As the errors accumulate over time, the odometry readings become increasingly unreliable.

Visual odometry is the process of determining equivalent odometry information using only camera images. Compared to traditional odometry techniques, visual odometry is not restricted to a particular locomotion method, and can be utilized on any robot with a sufficiently high quality camera.

Autonomous Underwater Vehicles (AUVs) are mobile robots that can be applied to many tasks of difficult human exploration[8]. In underwater visual inspection, the vehicles can be equipped with down-looking cameras, usually attached to the robot structure[11]. These cameras capture images from the deep of the ocean. In these images, natural landmarks, also called keypoints in this work, can be detected allowing the AUV visual odometry.

In this paper we propose a new approach to AUV localization and mapping. Our approach extract and map keypoints between consecutive images in underwater environment, building online keypoints maps. This maps can be used to robot localization and navigation.

We use Scale Invariant Feature Transform (SIFT), which is a robust invariant method to keypoints detection[16]. Furthermore, these keypoints are used as landmarks in an online topological mapping. We propose the use of self-organizing maps (SOM) based on Kohonen maps[15] and Growing Cell Strutures (GCS)[9] that allow a consistent map construction even in presence of noisy information.

First the paper presents related works on visual odometry and mapping. Section 3 presents a detailed view of the our approach with SIFT algorithm and Self-Organizing Maps, followed by the implementation, test analysis and results with different undersea features. Finally, the conclusion of the study and future perspectives are presented.

*e-mail: silviacb@ee.furg.br

2. Related Works

Localization, navigation and mapping using vision-based algoritms use visual landmarks to create visual maps of the environment. In the other hand the identification of landmarks underwater is a complex task due to the highly dynamic light conditions, decreasing visibility with depth and turbity, and image artifacts like aquatic snow. The extent to which the robot navigates, the map grows in size and complexity, increasing the computational cost and difficult to process in real time. Moreover, the efficiency of the data association, an important stage of the system, decreases as the complexity of the map augment. It is therefore important for these systems, extract a few, but representative, features/keypoints (points of interest) of the environment.

The development of a variety of keypoint detectors was a result of trying to solve the problem of extracting points of interest in image sequences, Shi and Tomasi[23], SIFT[16], Speeded up robust features Descriptor (SURF)[2], affine covariant etc. These proposals have mainly the same approach: extraction of points which represents regions with high intensity gradient and texture. This region represented by them are highly discriminatory and robust to noise and changes in illumination, point of view of the camera, etc.

Some approaches using SIFT for visual indoor Simultaneous Localization and Mapping (SLAM) were made by Se and Lowe[21, 22]. They use SIFT in a stereo visual system to detect the visual landmarks, together with odometry, using ego-motion estimation and the Kalman filter. The tests were made in structured environments with knew maps.

Several AUVs localization and mapping methods are based on mosaics.[10, 13] Mahon and Willians[17] propose a visual system for SLAM in underwater environment, using the Lucas-Kanade optical filter and extended Kalman filter (EKF), with aid of a sonar. Nicosevici et al.[18] propose an identification of suitable interest points using geometric and photometric cues in motion video for 3D environmental modeling.

Booij[3] has the most similar approach to the presented in this work. They do visual odometry with classical topological maps based on appearance. In this case, the SIFT method is used in omnidimentional images. However, this approach is validated only with mobile robots in terrestrial environment. The use of both i) SIFT to extract visual underwater features; and ii) SOM topological maps with Growing Cell Strutures for mapping and localization on underwater environment was not found in the literature.

3. A System for Visual Odometry

Figure 1 shows an overview of the approach proposed here. First, the underwater image is captured and pre-processed to removal of the radial distortion and others distortions caused by water diffraction. With the corrected image, keypoints are detected and local descriptors for each one of these points are computed by SIFT. Each keypoint has a n dimensional local descriptors and global pose informations. A matching stage provides a set of correlated keypoints between consecutive images. Considering all correlated points found, outliers are removed, using RANSAC[7] and LMedS[20] algorithms.

The relative motion between frames is estimated, using the correlated points and the homography matrix.

In addition, the keypoints are used to create and train the topological maps. A growing cell strutures algorithm is used to create the nodes and edges of the SOM. Each node has a n-dimensional weight. After a trainning stage, the system provides a topological map, where its nodes represent the main keypoints of the environment.

During the navigation, when a new image is captured, the system calcules its local descriptors, correlating them with the nodes of the current trainned SOM. To estimate the pose of the robot (center of the image), we use use the correlated points/nodes and the homography matrix concept. Thus, it is obtained the global position and orientation of the center of the image, provinding the localization of the robot.

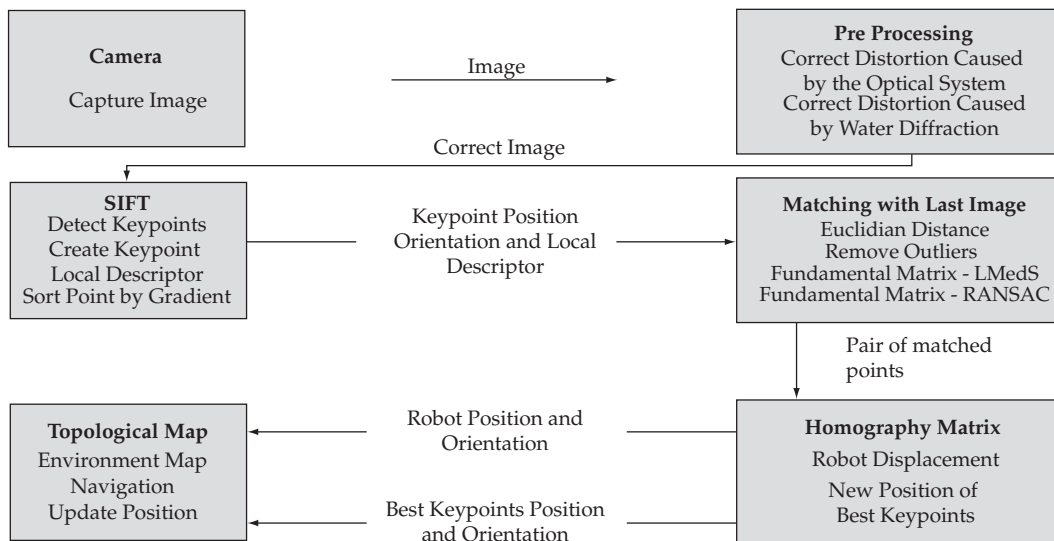Next, it is detailed each module of the proposed approach.

Figure 1. Overview of the system proposed.

3.1. Pre-processing

The distortion caused by the camera lenses can be represented by a radial and tangential approximation. As the radial component causes higher distortion, most of the works developed so far corrects only this component[12].

In underwater environment, there is an additional distortion caused by water diffraction. Equation 1 shows one method to solve this problem[27], where m is the point without radial distortion with coordinates (m_x, m_y), and m_0 the new point without additional distortion; u_0 and v_0 are the central point coordinates. Also, $R = \sqrt{m_x^2 + m_y^2}$ and R_0 are defined by 2 with focal distance f.

$$m_0 x = m_x + \frac{R_0}{R}(m_x - u_0)$$
$$m_0 y = m_y + \frac{R_0}{R}(m_y - v_0) \tag{1}$$

$$R_0 = f \tan(\sin^{-1}(1.33 * \sin(\tan^{-1}\frac{R}{f}))) \tag{2}$$

3.2. Scale Invariant Feature Transform – SIFT

The Scale Invariant Feature Transform (SIFT) is an efficient filter to extract and describe keypoints of images[16]. It generates dense sets of image features, allowing matching under a wide range of image transformations (i.e. rotation, scale, perspective) an important aspect when imaging complex scenes at close range as in the case of underwater vision. The image descriptors are highly discriminative providing bases for data association in several tasks like visual odometry, loop closing, SLAM, etc.

First, the SIFT algorithm uses the Difference-of-Gaussian filter to detect potential interesting points in a space invariant to scale and rotation. The SIFT algorithm generates a scale space $L(x, y, k\sigma)$ by convolving repeatedly an input image $I(x, y)$ using a variable-scale Gaussian, $G(x, y, \sigma)$, Equation 3:

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y) \tag{3}$$

SIFT analyzes the images at different scales and extracts the keypoints, detecting scale-invariable image locations. The keypoints represent scale-space extrema in the difference-of-Gaussian function $D(x, y, \sigma)$ convolved with the image, see 4:

$$D(x,y,\sigma) = (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y) \tag{4}$$

where k is a constant multiplicative factor.

After the keypoints extraction, each feature is associated with a scale and an orientation vector. This vector represents the major direction of the local image gradient at the scale where the keypoint was extracted. The keypoint descriptor is obtained after rotating the nearby area of the feature according to the assigned orientation, thus achieving invariance of the descriptor to rotation. The algorithm analyses images gradients in 4 × 4 windows around each keypoint, providing a 128 elements vector. This vector represents each set of feature descriptors. For each window a local orientation histogram with 8 bins is constructed. Thus, SIFT maps every feature as a point in a 128-dimension descriptor space.

A point to point distances computation between keypoints in the descriptors space provides the matching. To eliminate false matches, it is used an effective method to compare the smallest match distance to the second-best distance[16], where through a threshold it is selected only close matches.

Furthermore, outliers are removed through RANSAC and LMedS, fitting an homography matrix H^1. In this paper, this matrix can be fitted by both RANSAC and LMedS methods[26]. Both methods are considered only if the number of matching points is bigger than a predefined threshold t_m.

3.3. Estimating the homography matrix and computing the camera pose

We use the homography concept to provide the camera pose. A homography matrix H is obtained from a set of correct matches, transforming homogeneous coordinates into non-homogeneous. The terms are operated in order to obtain a linear system[14], considering the keypoints $(x_1, y_1)…(x_n, y_n)$ in the image I and $(x_1', y_1')…(x_n', y_n')$ in the image I' obtained by SIFT.

The current global pose of the robot can be estimated using Equation 5, where $^1H_{k+1}$ is the homography matrix between image I_1 in the initial time and image $I_k + 1$ in the time k+1. The matrix $^1H_1$ is defined by the identity matrix 3x3 that consider the robot in the beginning position (0,0).

$$^1H_{k+1} = \prod_{i=1}^{k} {}^i H_{i+1} \tag{5}$$

Thus, the SIFT provides a set of scale invariant keypoints, described by a feature vector. A frame has a m keypoints, and each keypoint, X_i, has 128 features, $f_1,…,f_{128}$ and the pose and scale (x, y, s):

$$X_i = f_1, f_2, …, f_{128}, x, y, s, i = 1, …, m \tag{6}$$

These m vectors are used to obtain a topological map, detailed in the next section.

3.4. Topological maps

In this work, the vectors extracted from SIFT are used to compose a topological map. This map is obtained using a self-organizing mapping (SOM) based on Kohonen Neural Networks[15] and the Growing Cell Structures (GCS) method[9]. Like most artificial neural networks, SOMs operate in two modes: training and mapping. Training builds the map using input examples. It is a competitive process, also called vector quantization. A low-dimensional (typically two dimensional) map discretizes the input space of the training samples. The map seeks to preserve the topological properties of the input space. A structure of this map consists of components called nodes or neurons. Associated with each node is a weight vector of the same dimension as the input data vectors and a position in the map space. Nodes are connected by edges, resulting in a (2D) grid.

3.4.1 Building the map

Our proposal operates in Scale Invariant Feature vectors Space, SIFT space, instead of image space, in other words, our space has n = 131 values (128 by the SIFT's descriptor vector and 3 by the feature's pose). A Kohonen map must be created and trained to represent the space of descriptors. To build the map, feature vectors are presented to the SOM. The learning algorithm is based on the concept of nearest-neighbor learning using KD-Tree algorithm[16]. When a new input arrives, the topological map determines the feature vector of the reference node that best matches the input vector. As our system uses several feature vectors associated with each captured image, the nearest-neighbor algorithm is applied to each feature vector separately. The results of the nearest-neighbor algorithms are combined with a simple scheme based on unanimous voting.

The Growing Cell Structures method allows the creation and removal of the nodes during the learning process. The algoritm constrains the network topology to k-dimensional simplices whereby k is some positive integer chosen in advance. In this work, the basic building block and also the initial configuration of each network is a k = 2-dimensional simplex. For a given network configuration a number of adaptation steps are used to update the reference vectors of the nodes and to gather local error information at each node. This error information is used to decide where to insert new nodes. A new node is always inserted by splitting the longest edge emanating from the node q with maximum accumulated error. In doing this, additional edges are inserted such that the resulting structure consists exclusively of k-dimensional simplices again.

After a set of tranning steps, the kohonen map represents the descriptors space. This SOM can be used to locate the robot during the navigation.

3.4.2. Location the robot on the map

New frames are captured during the navigation. For each new frame F, SIFT calculates a set of m keypoints X_i, see equation 6. A n = 131 dimensional descriptor vector is associated to each keypoint. We use the trainned SOM to map/locate the robot in the environment. A mapping stage is runned m times. For each step i there will be one single winning neuron, N_i: the neuron whose weight vector lies closest to the input descriptor vector, X_i. This can be simply determined by calculating the Euclidean distance between input vector and weight vectors. After the m steps we have a set of m winner nodes, N_i, associated with each one feature descriptor, X_i. With the pose information of m pairs (X_i, N_i), we can use the homography concept to obtain a linear matrix transformation, H_{SOM}. Equation 7 gives the map localization of the center of the frame, $X_c' = (x_c', y_c')$:

$$X_C' = H_{SOM} * X_C, \tag{7}$$

where X_c is the position of the center of the frame.

Moreover the final topological map allows the navigation in two ways: through target positions or visual goals. From the current position, graph search algorithms like Dijkstra[6] or A* algorithm[5] can be used to search a path to the goal.

4. System Implementation, Tests and Results

In this work, it was developed the robot presented in Figure 2. This robot is equipped with a Tritech Typhoon Colour Underwater Video Camera with zoom, a miniking sonar and a set of sensors (altimeters and accelerometers)[4]. Due this robot is experimental phase, it is impossible to put it to work in the sea. The acquition of some reference to experiements is very hard in this kind of environment,too. Considering this situation, this work use a simulated underwater conditions proposed by Arredondo and Katia[1]. Using it, different undersea features were applied in the images, like turbidity, sea snow, non-linear illumination, and others, simulating different underwater conditions. Table 1 shows the applied features (filters).

The visual system was tested in a desktop Intel Core 2 Quad Q6600 computer with 2Gb of DDR2-667 RAM. The camera is NTSC standard using 320x240 pixels at a maximum rate of 29.97 frames per second.

4.1. The method in different underwater features

The visual system was tested in five different underwater environments, corresponding the image without distortion and first four filters presented in Table 1 (the effects were artificially added to the images). Figure 3 enumerates the detected and matching keypoints obtained in a sequence of visual navigation. Even though the number of points and correlations has diminished with the quality loss because of underwater conditions, it is still possible to localize the robot, according Figure 4. In this figure, the motion referential is represented in blue (legended as "Odometry"), executed
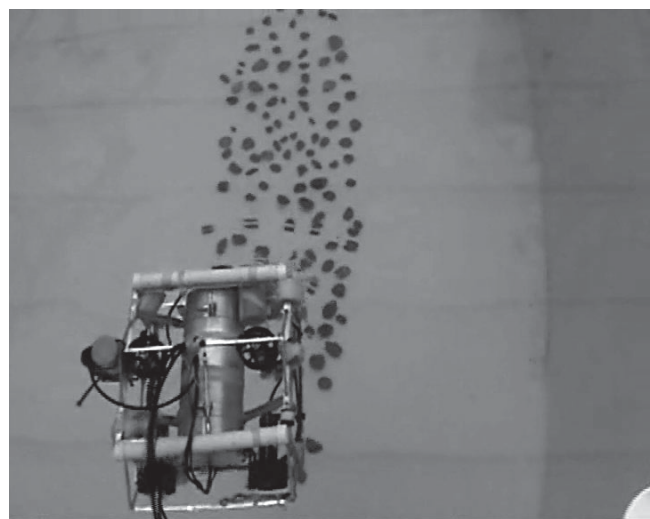

Figure 2. ROVFURGII in test field.

Table 1. Undersea Features For Each Distortion using in the tests.

Distortion	1	2	3	4	5
Light Source Distance (m)	0.2	0.22	0.25	0.25	0.3
Attenuation Value (%)	0.05	0.05	0.06	0.05	0.05
Gaussian Noise (σ)	2	2	2	4	4
Gray Level Minimum	20	30	20	20	20
Number of Flakes of Sea Snow	30	30	30	30	30

Figure 3

Number of keypoints in image vs. Time (s)

(a)

Number of correct Matches vs. Time (s)

(b)

— Vision Normal   — Vision Distortion 3

— Vision Distortion 1   — Vision Distortion 4
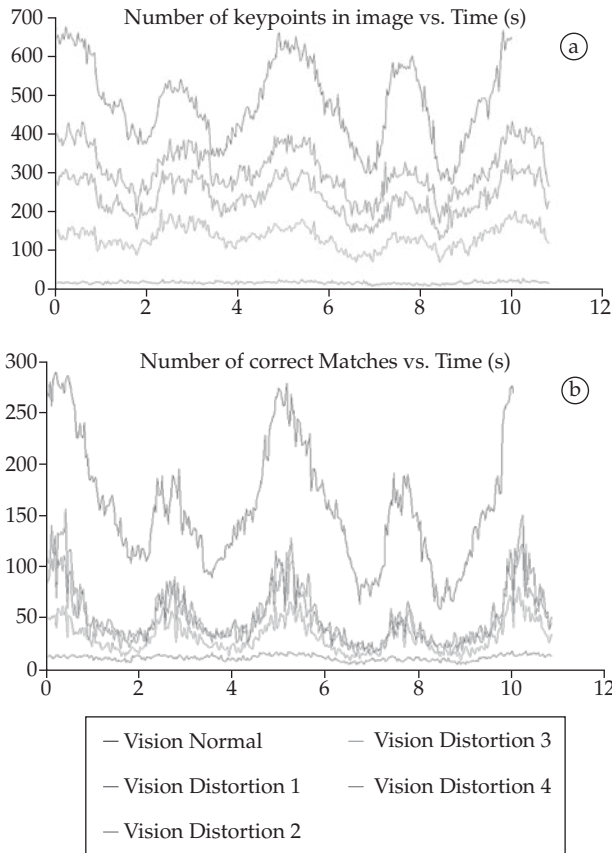
— Vision Distortion 2

Figure 3. Number of keypoints detected and true correlation during the robotic arm movement.

by a robotic arm composed by an Harmonic Drive PSA-80 actuator with a coupled encoder supplying angular readings in each 0.651 ms, with a camera coupled to this. It allows the reference system a good precision, 50 0pulses per revolution. Therewith, it is possible to see that our approach is robust to underwater environment changes. All graphics showed in this paper use centimeter as metric unit, including Figure 4.

4.2. Online robotic localization

Tests was performed to evaluate the SIFT algorithm performance considering a comparison with another algorithm for robotic localization in underwater environment: KLT[19, 25, 24, 23].

Figure 5 shows the performance results using SIFT and KLT methods. SIFT has obtained an average rate of 4.4 fps over original images, without distortion, and a rate of 10.5 fps with the use of filter 5, the worst distortion applied. KLT presented higher averages, 13.2 fps and 13.08 fps, respectively. Note that SIFT has worst performance in high quality images because the large amount of detected points and, consequently, because the higher number of descriptors to be processed. The KLT, instead, keeps an almost constant performance. However, due to the slow dynamic associated with undersea vehicle motion, both methods can be applied to online AUV SLAM. The green cross represent the real final position and the metric unit is centimeter.
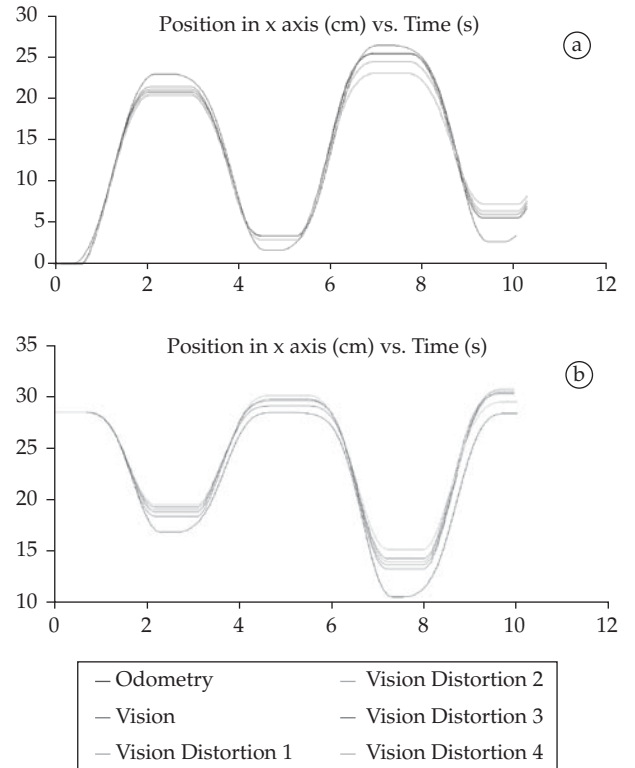
Position in x axis (cm) vs. Time (s)

(a)

Position in x axis (cm) vs. Time (s)

(b)

— Odometry   — Vision Distortion 2

— Vision   — Vision Distortion 3

— Vision Distortion 1   — Vision Distortion 4

Figure 4. Position determined by the robotic arm odometry and a visual system, without and with distortion.

Position of Vehicle Without Distortion

(a)

····· KLT
— SIFT

Position of Vehicle With Distortion 5

(b)

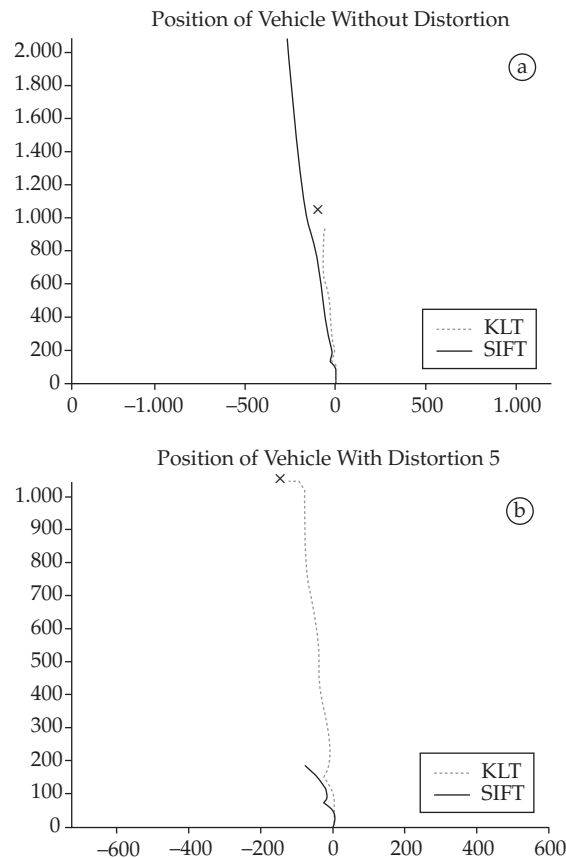····· KLT
— SIFT

Figure 5. Real Robot Localization in online system, without and with artificial distortion.

Botelho S et al.

The SIFT results related to the robot localization were considered satisfactory, even with extreme environment distortions (filter 5). In the other hand, KLT gives unsatisfying results for both cases, once it is too much susceptible to the robot's depth variation, or image scale, that occurs constantly in the AUV motion, despite the depth control.

4.3. Robustness to scale

Tests were performed to estimate the robustness of the proposed system to the sudden scale variation. In this case, a translation motion with height variation was performed with the camera to simulate a deeper movement of the robot in critical conditions.

The Figure 6 shows the SIFT results, considered satisfactory, even in critical water conditions. Considering the use of
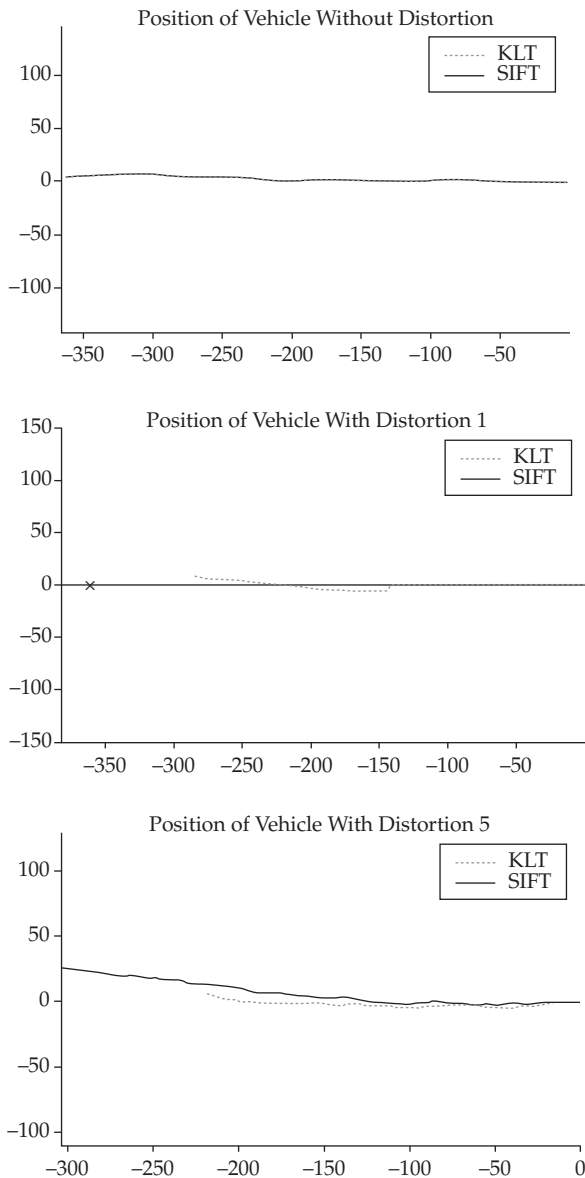
some filters in extreme conditions, SIFT is superior to KLT although it shows an inexistent movement in Y axis. Over the tests, SIFT has shown an average rate of 6,22 fps over original images captured by the camera and a rate of 7.31 fps using filter 1 and 10.24 fps using filter 5. The KLT have shown 12.5, 10.2 and 11.84 fps, respectively. The green cross represent the real final position, is the same for all graphics in the Figure 6, the metric unit is centimeter.

4.4. Topological maps

Tests to validate the mapping system proposed were performed. For example, during a navigation task a set of 1.026 frames were captured. From these frames, a total of 40.903 vectors are extracted from SIFT feature algorithm.

To build the map, 1026 frames and 40.903 keypoints are presented to the SOM. Figure 7 shows the final 2D map, discretizing the input space of the training samples.

4.4.1. Building the map

When a new keypoint arrives, the topological map determines the feature vector of the reference node that best matches the input vector. The Growing Cell Structures (GCS) method allows the creation and removal of the nodes during the learning process. Table 2 shows intermediate GCS adaptation steps with number of frames, keypoints and SOM
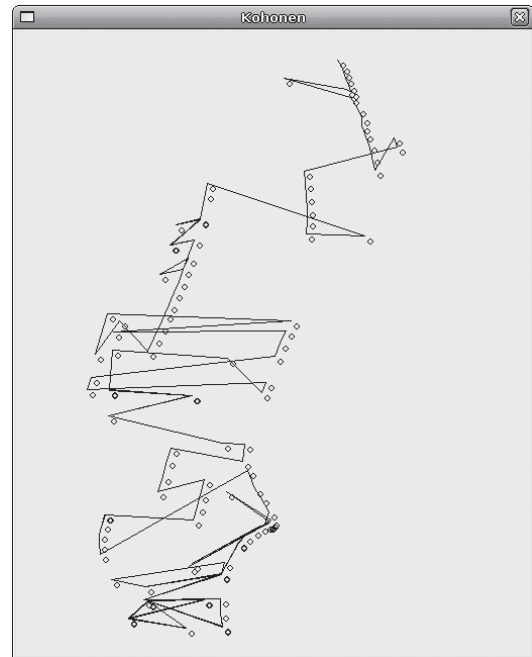

Figure 6. Localization with translation and scale movement without and with distortion.

Figure 7. Topological Map generated by ROVFURGII in movement.

Table 2. Building the map with GCS algorithm.

Frames	Keypoints	Nodes
324	20.353	280
684	35813	345
1026	44903	443

nodes. After the tranning stage (1026 frames), the kohonen map represents the relevant and noise tolerant descriptors space using a reduced number of nodes. This SOM can be used to locate the robot during the navigation.

4.1.2. Location of robot on the map

New frames are captured during the navigation. We use the trainned SOM to map/locate the robot in the environment. Figure 8 shows the estimated position of a navigation task. In this task the robot crosses three times the position 0.0. In this figure we can see the position estimated by both the SOM map (blue) and only by visual odometry (red). In the crossings, Table 3 shows the normalized errors of positioning in each of the methods. The reduced error associated with the SOM localization validates the robustness of topologycal approach.
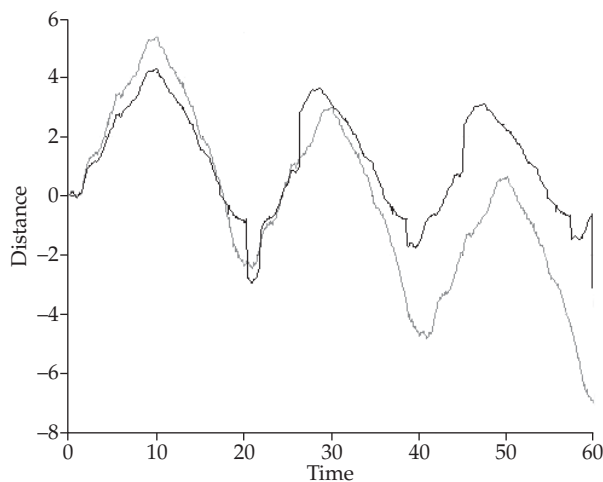

Figure 8. Distance Y generated by ROVFURGII in movement.

Table 3. Normalized Localization errors of only Visual Odometry and SOM.

Visual Odometry	SOM
0.33	0.09
0.68	0.35
1.00	0.17

5. Conclusion

This paper proposed a new approach to visual odometry and mapping of a underwater robot using only online visual information. This system can be used either in autonomous inspection tasks or in control assistance of robot closed-loop, in case of a human remote operator.

A set of tests were performed under different underwater conditions. The effectiveness of our proposal was evaluated inside a set of real scenarios, with different levels of turbidity, snow marine, non-uniform illumination and noise, among others conditions. The results have shown the SIFT advan-

tages in relation to others methods, as KLT, in reason of its invariance to illumination conditions and perspective transformations. The estimated localization is robust, comparing with the vehicle real pose.

Considering time performance, our proposal can be used to online AUV SLAM, even in very extreme sea conditions.

The correlations of interest points provided by SIFT were satisfying, even though with the presence of many outliers, i.e., false correlations. The proposal of use of fundamental matrix estimated in robust ways in order to remove outliers through RANSAC and LMedS algorithms shows good results.

The original integration of SIFT and topological maps with GCS for AUV navigation is a promising field. The topological mapping based on Kohonen Nets and GCS showed potential to underwater SLAM applications using visual information due to its robustness to sensory impreciseness and low computational cost. The GCS stabilizes in a limited number of nodes sufficient to represent a large number of descriptors in a long sequence of frames. The SOM localization shows good results, validating its use with visual odometry.

As future work, we propose to detail the analysis of our topological mapping system, executing a set of tests with different scenarios and parameters. We intend to fusion different sensor information. The utilization of stereoscopic vision is also a possibility in order to provide more accuracy to the system. Finally, nowadays, tests with SURF algoritm are being done with similar results that SIFT.

References

1. Arredondo M and Lebart K. A methodology for the systematic assessment of underwater video processing algorithms. *Oceans* 2005; 1:362-367.

2. Bay H, Tuytelaars T, Booktitle L and Gool L Van. Surf: speeded up robust features. In: *Proceedings of 9 European Conference on Computer Vision*; 2006; Graz, Austria. Springer: Lecture Notes in Computer Science; 2006. P. 404-417.

3. Booij O, Terwijn B, Zivkovic Z and Krose B. Navigation using an appearance based topological map. In: *Proceedings of IEEE International Conference on Robotics and Automation*; 2007; Roma, Italy. Amsterdam: Publications of the Universiteit van Amsterdam; 2007. p. 3927-3932.

4. Centeno M. *Rovfurg-II*: projeto e construção de um veículo subaquático não tripulado de baixo custo. [Master thesis]. Rio Grande: Universidade Federal do Rio Grande; 2007.

5. Dechter R and Pearl J. Generalized best-first search strategies and the optimality af a*. *Journal of the Association for Computing Machinery* 1985; 32(3):505-536.

6. Dijkstra EW. A note on two problems in connexion with graphs. *Numerische Mathematik* 1959; 1:269-271.

7. Fischler M and Bolles R. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* 1981; 24(6):381-395.

8. Fleischer SD. *Bounded-error vision-based navigation of autonomous underwater vehicles*. [PhD thesis]. Stanford: Stanford University; 2000.

9. Fritzke B. *Growing cell structures*: a self organizing network for unsupervised and supervised learning. Berkeley: University of California; 1993. (Technical report).

10. Garcia R, Cufi and Carreras M. Estimating the motion of an underwater robot from a monocular image sequence. In: *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*; 2001; Maui, Hawaii. Girona, Spain: Institute of Informatics and Applications,University of Girona; 2001. p. 1682-1687. (v. 3).

11. Garcia R, Lla V and Charot F. VLSI architecture for an underwater robot vision system. In: *Proceedings of IEEE Oceans Conference*; 2005; Brest, France. Girona, Spain: Institute of Informatics and Applications,University of Girona; 2005. p. 674-679. (v. 1)

12. Gracias N, van der Zwaan S, Bernardino A and Santos-Vitor J. Results on underwater mosaic-based navigation. In: *Proceedings of IEEE Oceans Conference*; 2002. Biloxi, Mississippi. Lisboa, Portugal: Instituto Superior Técnico & Instituto de Sistemas e Robótica; 2002. p. 1588-1594. (v. 3).

13. Gracias N and Santos-Victor J. Underwater video mosaics as visual navigation maps. *Computer Vision and Image Understanding*. 2000; 79(1):66-91.

14. Hartley R and Zisserman A. *Multiple View Geometry in Computer Vision*. Cambridge: Cambridge University Press; 2004.

15. Kohonen T. *Self-organizing maps*. Secaucus: Springer-Verlag; 2001.

16. Lowe D. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*. 2004; 60(2):91-110.

17. Mahon I and Williams S. Slam using natural features in an underwater environment. In: *Proceedings of International Conference on Control*, *Automation*, *Robotics and Vision*; 2004, Kunming, China. NSW Austrália: University of Sydney; p. 2076-2081. (v. 3).

18. Nicosevici T, García R, Negahdaripour S, Kudzinava M and Ferrer J. Identification of suitable interest points using geometric and photometric cues in motion video for efficient 3-d environmental modeling. In: *Proceedings of International Conference in Robotic and Automation*; 2007; Roma, Italy. p. 4969-4974.

19. Plakas K and Trucco E. Developing a real-time, robust, video tracker. In: *Proceedings of MTS/IEEE Oceans Conference and Exhibition*; 2000; Providence, RI, USA. Edinburgh, UK: Heriot-Watt University; 2000. p. 1345-1352. (v. 2).

20. Rousseeuw P. Least median of squares regression. *Journal of the American Statistics Association*. 1984; 79(388):871-880.

21. Se S, Lowe D and Little J. Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks. *The International Journal of Robotics Research*. 2002; 21(8):735-758.

22. Se S, Lowe D and Little J. Vision-based global localization and mapping for mobile robots. *IEEE Transactions on Robotics*. 2005; 21(3):364-375.

23. Shi J and Tomasi C. Good features to track. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*; 1994; Seattle, WA, USA. NY, USA: Cornell University Ithaca; 1994. p. 593-600.

24. Tomasi C. and Kanade T. *Detection and tracking of point features.* Pittsburgh: Carnegie Mellon University; 1991. (Technical report).

25. Tommasini T, Fusiello A, Roberto V and Trucco E. Robust feature tracking in underwater video sequences. In: *Proceedings of MTS/IEEE Oceans Conference and Exhibition*; 1998; Nice, France. IT: Università di Udine; 1998. p. 46-50. (v. 1).

26. Torr PHS and Murray DW. The development and comparison of robust methodsfor estimating the fundamental matrix. *International Journal of Computer Vision*. 1997; 24(3):271-300.

27. Xu X and Negahdaripour S. Vision-based motion sensing for underwater navigation and mosaicing of ocean floor images. In: *Proceedings of MTS/IEEE Oceans Conference and Exhibition*; 1997; Halifax, NS, Canada. Coral Gables, FL: University of Miami; 1997. p. 1412-1417. (v. 2).