Robust Detection and Tracking of Regions of Interest for Autonomous Underwater Robotic Exploration

A. Alejandro Maldonado-Ramírez, L. Abril Torres-Méndez
Robotics and Advanced Manufacturing Group
CINVESTAV Campus Saltillo
Ramos Arizpe, Coahuila, MEXICO
Emails: alejandro.maldonado.ramirez@gmail.com, abril.torres@cinvestav.edu.mx

Edgar A. Martínez-García
Lab. de Robótica, Inst. of Eng. and Tech.
Universidad Autónoma de Cd. Juárez
Juarez, Chihuahua, Mexico
Email: edmartin@uacj.mx

***Abstract*—Autonomous robotic exploration of unstructured and highly dynamic environments is a complex task, particularly, in underwater environments. An underwater robot needs to quickly detect a region of interest and then track it for a certain period of time in order to plan for the next trajectory; all of these while keeping its motion control stable. In this paper, we present a novel approach that robustly detects and tracks regions of interest in underwater video streams at frame rate. First, to detect relevant regions in an image, our approach combines two existing visual attention schemes with some improvements to adjust it to underwater scenes. Second, a scaled version of the resulting image is segmented by using a superpixel segmentation algorithm, and each relevant point is associated to a superpixel descriptor. The descriptor helps to track the same region as long as it results interesting for the visual attention algorithm. The experimental results demonstrate that our approach is robust when tested on different videos of underwater explorations.**

Keywords-visual attention models; regions of interest; superpixel segmentation; feature tracking; underwater vision

I. INTRODUCTION

The development of simple sensory motor skills for tracking an object of interest starts at early stages of life, this involves motion of eyes, head and even tongue and/or hands (in newborns), which together with cognitive skills, direct the way to explore the surrounding environment [3], [15]. As we grow and get more mobility, we develop more sophisticated exploratory skills, which can be transferred and adapted to new objects or scenes. It is at this point that the exploration fully involves active perception and navigation skills [6]. However, the goal of exploration is not just to navigate and look around in the environment but to build hypotheses about the data, in other words, to build knowledge about what it is sensed. This knowledge depends on the type of environment and the application for which the exploration is required [5]. For a robotic system, for example, the goal of exploring a natural habitat may be to prevent natural disasters. In any case, a key aspect in the exploration task is to know what features are relevant in an environment in order to learn about it and take important decisions while interacting with it.

The detection and tracking of relevant regions in an scene is a fundamental part of any autonomous robotic exploration task [16]. Particularly, in underwater environments, it may result complex. On one hand, the inherent physical properties of marine environments cause geometrical distortions, such as color distortions, dynamic lighting conditions and suspended particles (known as "marine snow"), resulting in poor visibility thus hindering computer vision tasks. On the other hand, this type of environments are unstructured and highly dynamic. Since exploration is implicitly linked to motion, the tracking of relevant features must be stable enough to allow for smooth movements for the suitable control of the robotic system.

In this research work, we present a real-time visual attention model to robustly detect and track relevant underwater features with the aim of exploring coral reefs. The real-time characteristic in robotics applications is fundamental since the tracked relevant features will help to direct the exploration trajectories in subsequent captured images while estimating the relative robot pose.

The outline of the paper is as follows. Section II presents the related work. Section III describes our model and its implementation. The experimental results and discussion are presented in Section IV. Finally, in Section V the conclusions and future work are given.

II. RELATED WORK

The use of visual attention models to find regions of interest in images is a common preprocessing tool for a variety of applications. However, for practical applications, the main challenge for designing these systems lies in their real-time performance requirements. Particularly, when applied to video streams at frame rate. Two of the more popular ones, due to their easy implementation, flexibility and fast computation are the Neuromorphic Vision Toolkit (NVT) proposed by Itti *et al.* [12] and the attention system called Visual Object detection with a CompUtational attention System (VOCUS) by Frintrop *et al.* [10]. The Focus Of Attention (FOA) is the place in the image that draws the attention of the system. Itti *et al.* [12] obtained the FOA by using a Winner-Take-All neural network. Frintrop *et al.* [10] simply find the point with the highest saliency value by scanning every point, and the most salient region is determined by seed region growing.

Recently, visual attention models have been used in robotic applications [5]. There has been likewise underwater applications of these models. For example, Walther *et al.* [17] and Edgington *et al.* [7] detect objects and potentially interesting visual events for humans in order to label the frames of a video

 ISBN: 978-1-61208-340-7

stream as interesting or boring. In both research work, the NVT [12] model is used and the videos were recorded by a Remotely Operated Vehicle (ROV). Barat and Rendas [4] present a visual attention system for detection of manufactured objects. Their model is based on the minimum description length test for detecting the motion of contrasting neighboring regions. After that a statistical snake is adapted to determine the boundary of the object. Lobato *et al.* [13] use intensity, motion and edge maps as features for their visual attention model to detect the Norway lobsters and help scientist to quantify them.

In all these works, the visual attention models are used for aiding humans in the task of analyzing video streams. In our case, we want that the visual attention model leads the robot motion by automatically detecting and tracking features that are considered of interest for exploration. Particularly, we are interested in transferring abilities to an Autonomous Underwater Vehicle (AUV) in order to detect regions of interest without human supervision while successfully navigating the environment. Visual attention models for autonomous underwater exploration require an strict real time performance.

As hardware limitations in underwater robots are still an issue, we need to rely on fast computational algorithms.

III. METHODOLOGY

In this section, we describe our visual attention model for underwater scenes.

The general structure of our methodology (see Fig. 1) combines two methods with some improvements to adjust it to be used in underwater scenes – the NVT [12] and the VOCUS method [10].

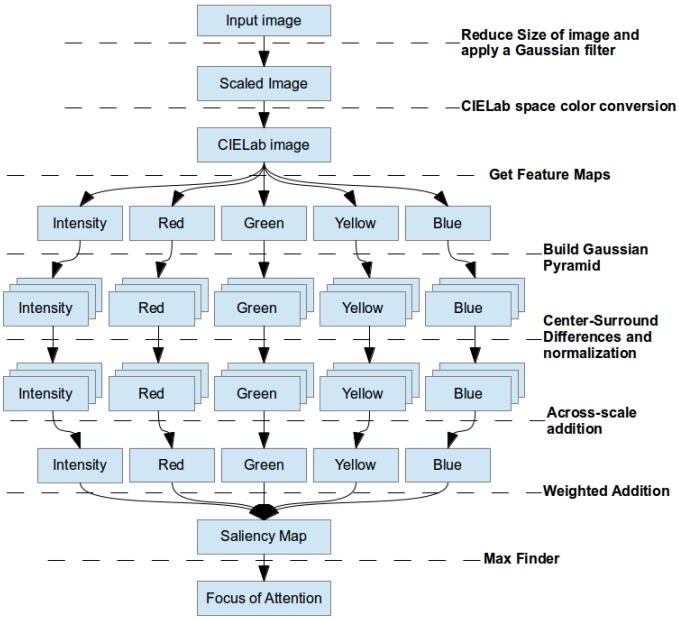


Figure 1. General diagram of our visual attention algorithm.

Given that underwater environments are unstructured, *i.e.*, the existing objects lack of specific orientation and shape, our attention model relies strongly on the intensity and color information. Therefore, the use of a color space capable of highlighting colors that are different from the color of seawater is crucial. As pointed out, this represents a challenge given that visibility conditions in open water are not always ideal and color tonality tends to diminish significantly. The space color we choose is the CIELab. This color model has the characteristic of being perceptually uniform and also that its a and b channels naturally encompass the green-red and blue-yellow contrast colors, which turns to be perfect for underwater image analysis and processing. By assigning weights to each of the five color channels (intensity, green, red, blue and yellow) the focus of attention can be directed to a certain object or characteristic. This is specially important since water absorbs color abruptly with depth thus limiting its detection.

Furthermore, our model needs to be robust to dramatic changes in illumination, which are very common for underwater images or videos. By considering only the chromatic channels a and b of the CIELab color space, our model does not consider abrupt changes in brightness in images as these tend to be smooth in the chromatic channels. This turns out to be fine for our purposes.

An important aspect to consider in any computational visual attention system is how to highlight the relevant parts of each feature. This is usually done by using a center-surround mechanism (also called *center-surround difference*), which is inspired in cells of the human visual receptive field [14].

For exploration tasks, keeping track of the same focus of attention, or one near a previous one at different image frames, is of particular importance. First, it allows the robot to lead its motion in a smooth manner avoiding sudden maneuvers, which may cause the system to become unstable. Secondly, for the navigation part, it is important to have an estimate of the current pose of the robot, thus, by tracking the same feature (natural landmark in our case), it would allow the robot to estimate its relative pose by means of triangulation.

Our strategy to achieve this is based on the fact that once a region of interest is identified as FOA, this region should be kept as FOA as long as it results interesting for the attention system. In other words, the robot needs to robustly keep track of the same or very similar FOA for a certain period of time in order to make inferences about it, to estimate its relative pose, and finally, to plan the motion to the next relevant region to be explored. To achieve this behavior we apply a superpixel segmentation technique based on the Simple Linear Iterative Clustering (SLIC) algorithm [2]. The information captured at each superpixel forms a *descriptor*, which helps to discriminate the FOA at the current frame by considering the FOAs in the previous frame. By doing this, the algorithm tries to keep the same region as FOA in consecutive frames.

In the following sections, we describe in more detail each of the steps involved in our visual attention model.

A. Preprocessing of the image

The input RGB image is scaled to a size of 320×240 pixels and then blurred with a Gaussian filter. After that, the image is converted to the CIELab color space to extract a particular color from an image (Section III-B).

B. Getting the features maps

We use intensity and color (red, yellow, green and blue) as features. The intensity map corresponds to the $L-$channel of the CIELab image. The colors are extracted from the a and b

channels, as described in [8], as follows,:

$$F_i(x,y) = V_{max} - \|ab(x,y) - p\|, \quad (1)$$

where F_i is the i^{th} feature map, $V_{max} = 255$ in 8-bit depth images, $p = (a_d, b_d)$ is the desired color to extract (only the a and b component are used) and $ab(x,y)$ is the ab-channel of the image. The color feature maps are gray-scale images in which the intensity indicates how near is the desired color to the original color of the pixel. As it was previously mentioned, there is no need of using the orientation feature, since underwater environments are unstructured.

C. Getting the conspicuity maps

The conspicuity maps tell us where the most relevant regions are in an specific feature map. We are going to have five conspicuity maps at the end of this process. One for intensity and four for each of the colors.

The first step to calculate the conspicuity maps is to build a Gaussian pyramid. The number of levels used in the pyramid depends on the size of the image and the size of the relevant regions to be found [9]. In the algorithm, we use a $3-$level pyramid, *i.e.*, three scales $s_n = \frac{1}{2^n}$ with $n = \{0,1,2\}$. We obtain three maps F_{in} for each feature map.

Once the pyramids are built, the center-surround differences are applied. The center-surround differences are implemented as in [9], but using filter operations.

$$D_{in\sigma}(x,y) = center - surround \quad (2)$$

$$center = F_{in}(x,y) \quad (3)$$

$$surround = K(\sigma) * F_{in}(x,y) \quad (4)$$

$$K(\sigma) = \frac{1}{(2\sigma+1)^2 - 1} \begin{bmatrix} 1 & \dots & 1 \\ \vdots & 0 & \vdots \\ 1 & \dots & 1 \end{bmatrix}_{(2\sigma+1)\times(2\sigma+1)} \quad (5)$$

where $*$ is the convolution operator. We use $\sigma = \{3,4\}$.

After the center-surround differences are applied, each of the resultant maps $D_{in\sigma}(x,y)$ is normalized in a range of $[0,M]$ (in our case, we set $M = 1$). Then, all the obtained maps from the same feature are added across-scale in s_2. This way, we obtain a conspicuity map C_i for each feature.

D. Getting the saliency map

To calculate the saliency S map we normalize each of the conspicuity maps, then we weigh each with a value w_i and add them into a single saliency map S. By changing the values of w_i, we can give a preference to certain color feature.

$$S(x,y) = \sum_i w_i C_i(x,y). \quad (6)$$

It is worth noting that the saliency map is a gray-scale image in the scale s_2. The most relevant parts of the image appear brighter in the saliency map.

The way to fuse the maps into a single one (the scaled feature maps into a conspicuity map and conspicuity maps into a saliency map) is called a *naive* strategy [11]. We have also implemented the normalization operator $N(\cdot)$ [12],[11] to fuse the maps.

Finally, in order to give priority to relevant points that are close to the most relevant point in a saliency map, each value of saliency in the map is weighted as follows:

$$w = e^{-a\sqrt{(x_c-x)^2+(y_c-y)^2}}, \quad (7)$$

where (x_c, y_c) are the coordinates of the most relevant point **c**, (x,y) are the coordinates of the other points of the image and a is a positive value. This way the points nearer to **c** are more likely to be chosen as the next relevant points by the algorithm.

We compared the relevant regions obtained with our method to those obtained with the model of the non-iterative $N(\cdot)$ normalization and the model of the iterative normalization using a dataset of non-underwater images, containing natural and man-made objects. Two examples of natural outdoor scenes are shown in Fig. 2 (top and middle rows). As our interest is in underwater scenes, we carry on this comparison using a dataset of underwater images containing only natural structures (rocks, coral reefs, fishes). The last row of Fig. 2 shows an example with an underwater scene. Each of the relevant regions detected is surrounded by a circle. It can be observed that the relevant regions detected by our model are all on the rock formation whereas the regions detected with the other models are mostly on areas like sand or sea water, which are not of relevance for exploration tasks.

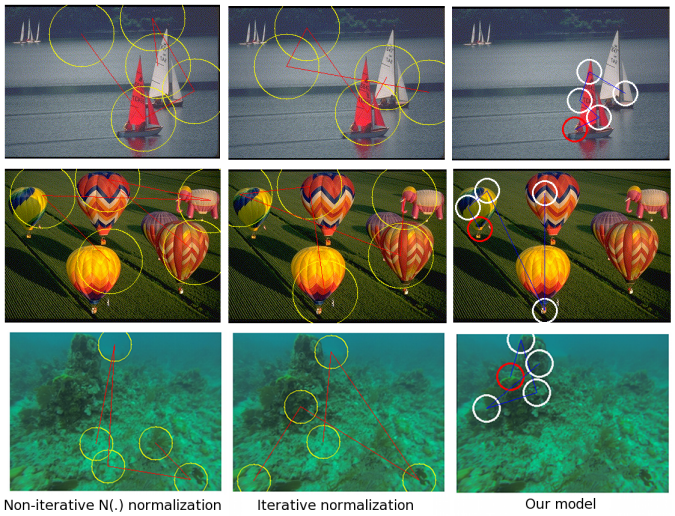


Figure 2. Comparison results on the detection of relevant regions (indicated by circles) in non-underwater an underwater scenes by using the $N(\cdot)$ normalization (left), the iterative normalization (middle) and our visual attention model (right).

In general, good results are obtained for still images, however for a video sequence of a coral reef, the results were not satisfying for our purpose (underwater robotic exploration) because the focus of attention changed its position arbitrarily from one frame to the next one. Therefore, a robust tracking of FOA is fundamental (more details in Section III-E).

E. Robust tracking of focus of attention

We find the most relevant point (FOA) by scanning all the values of saliency in the map and choosing the one with the highest value, then we set the surrounding points to zero in a given radius. We repeat this process until we find the n most saliency points. It is important to recall that we want

 ISBN: 978-1-61208-340-7

the algorithm to find a relevant region and be able to find the same relevant region in the next image frame, that is, we want to keep track of the relevant region for few more subsequent images if and only if this region is still among the n most saliency ones. We are interested in this behavior because it will lead the movement of a robot during an exploration. Having abrupt changes of the FOA from one frame to next one may cause an erratic motion.

To tackle this problem, we segment the smallest image in the 3-level pyramid (*i.e.*, an image of 80×60 for an input image of 320×240) in m superpixeles using the SLIC algorithm [2]. Each superpixel is a set of pixels with similar features and it is characterized by a $5-$dimensional vector of the form $\mathbf{c}$=$[L_c, a_c, b_c, x_c, y_c]$. The n most relevant points are described with a descriptor $\mathbf{c}$ of the superpixel they belong. Once we have the descriptor of each saliency point, we choose the closest (the most similar) to the descriptor of the FOA from the previous frame. The chosen region become the FOA of the current frame. The distance (similarity) measure is based on the SSD metric as in [2], but without using the L component:

$$dist(p_1, p_2) = \sqrt{\left(\frac{d_s}{n_s}\right)^2 + \left(\frac{d_c}{n_c}\right)^2}, \quad (8)$$

$$d_s = \sqrt{(a_{p_1} - a_{p_2})^2 + (b_{p_1} - b_{p_2})^2},$$

$$d_c = \sqrt{(x_{p_1} - x_{p_2})^2 + (y_{p_1} - y_{p_2})^2},$$

where n_c and n_s are the normalization factor for the distance in the color space and the image space, respectively. These values were set as described in [1]. Fig. 3 illustrates the use

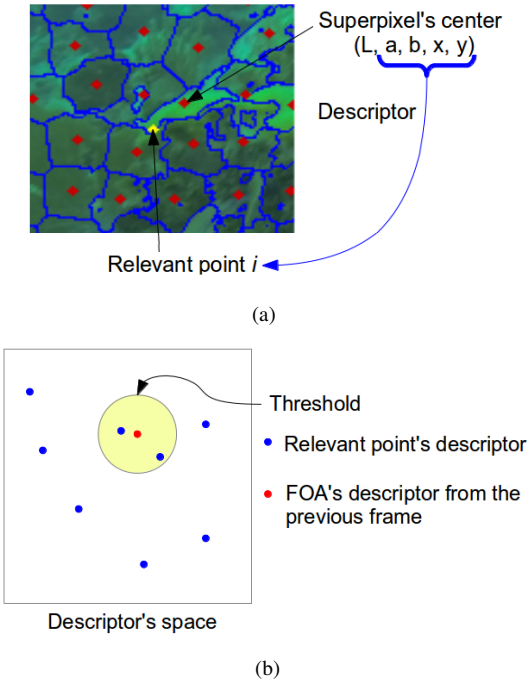


Figure 3. Finding the next focus of attention. (a) Superpixels are used to associate a descriptor to each relevant point. (b) The distance defined in (8) is used to find the next focus of attention. The relevant point descriptors inside the yellow circle represent the FOA candidates.

of superpixels to achieve a stable tracking of similar FOAs in a region of interest. If the distance from the closest saliency descriptor to the previous FOA descriptor is greater than a defined threshold (yellow circle in Fig. 3b), the distances are ignored and the point with the highest saliency value is chosen as the new FOA.

IV. Experimental Results

Before conducting the experiments, we tested our algorithm with different image resolutions to analyze their performance. The complexity of the algorithm is $O(N)$, where N is the total number of pixels in the image. The average processing time, in a 2.1GHz dual-core processor, for an image of 640×480 is 184 ms. We select to use the 320×240 resolution (49 ms) as the behavior of the detected FOA was better (with less abrupt changes). Also, some parameters, related to regions considered as relevant in underwater scenes, needed to be tuned before running our algorithm. We conducted visual tests with 32 people (16 women and 16 men) in an age range of $20-30$ years old. In the experiment, each person was asked to select the region(s) in an underwater image that attracted more their attention. A set of eight images were shown to each person.

Fig. 4a depicts some examples of the regions of interest chosen by people. Each column shows the regions selected in an image. With this information, we train our algorithm by assigning to each color feature a weight. In Fig. 4b, some regions of interest chosen by our visual attention algorithm are shown. It can be observed that the regions detected by the algorithm resembles the ones detected by people. We carried

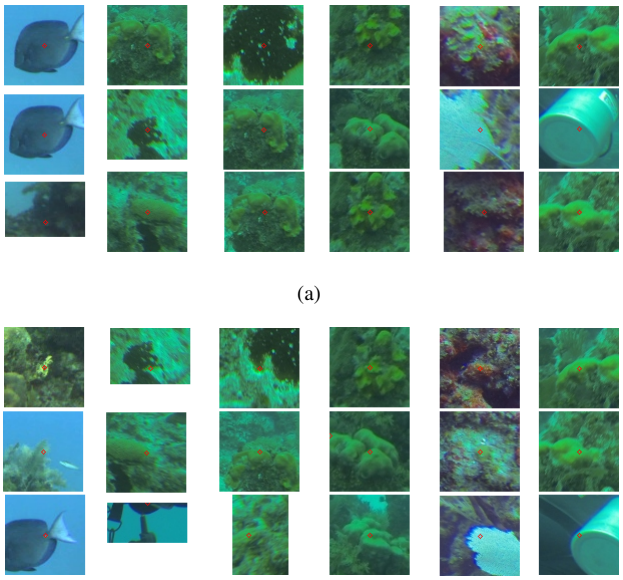
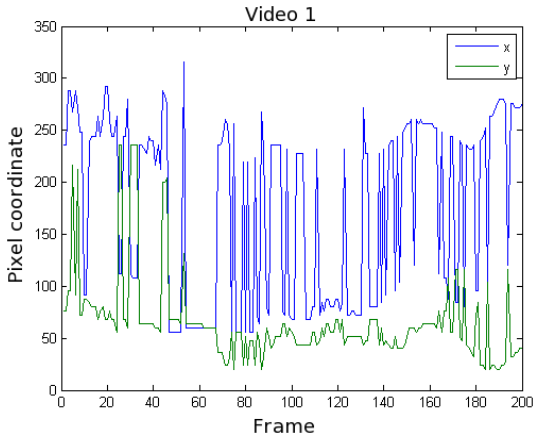


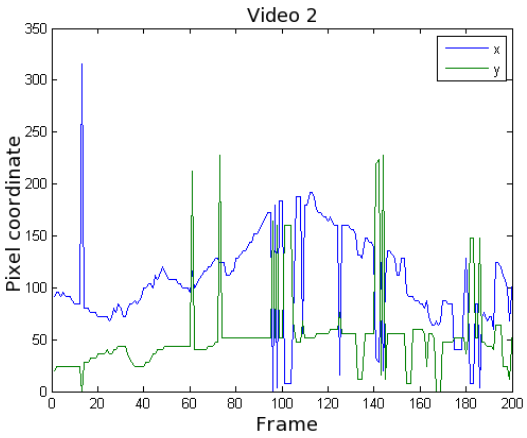
Figure 4. Some of the region of interests selected by (a) people and by (b) our visual attention algorithm. A column shows the snapshots considered as FOAs in each of the images.

out a set of experiments using two different underwater videos taken during a dive exploration of the coral reef of Mahahual, Costa Maya. The first video (Video 1) is a 30 fps video in which the diver's camera motion is mainly forward. The second video has also a frame rate of 30 fps but the camera's motion is mainly a rotation around its vertical axis and it is

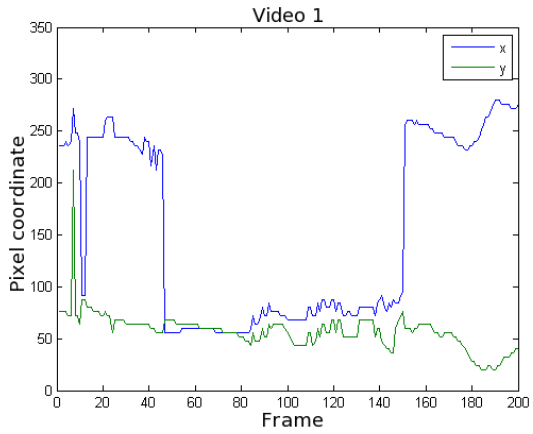
 ISBN: 978-1-61208-340-7



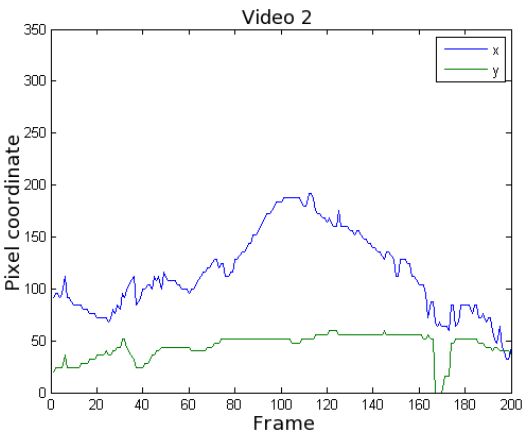
(a) Position of FOA during forward motion.



(b) Position of FOA during rotational motion.



(c) Position of FOA during forward motion (using superpixel descriptors for each saliency point).



(d) Position of FOA during rotational motion (using superpixel descriptors for each saliency point).

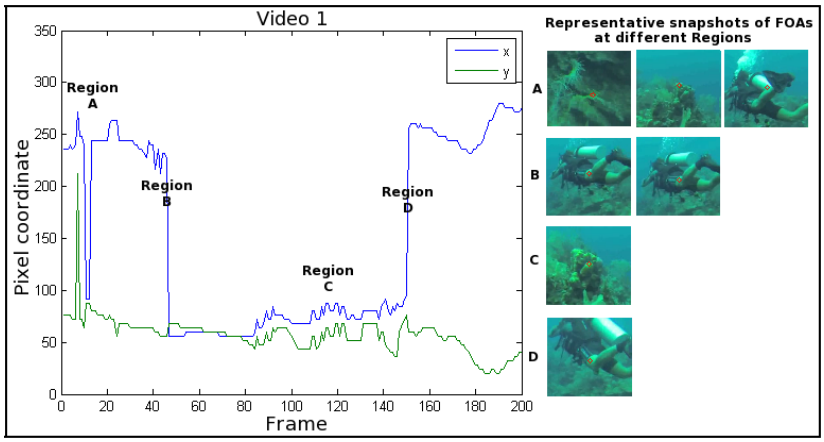
Figure 5. Position of the FOA on the image plane. (a) The FOA is chosen by using directly the point with the highest saliency map. (b) The FOA is chosen using the superpixel descriptors for each saliency point in order to keep a FOA in the same region on a given number of consecutive frames, avoiding with this, an erratic motion in the robot.

slower than Video 1. We use Video 1 to test the performance of the algorithm in forward motion and the Video 2 for rotational motion. The experiments were done in 200 consecutive frames on each video.

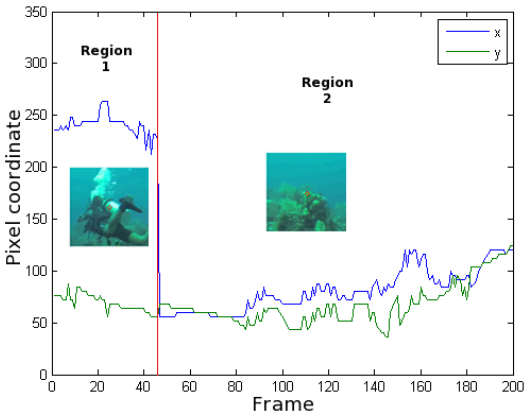
In the first experiment, the FOA was chosen by using only the information provided by the saliency map obtained from the visual attention algorithm, *i.e.*, the point with the highest saliency value on the map. Figures 5a-b show the graphs of the (x, y) images coordinates of the FOA obtained at each frame for Videos 1 and 2, respectively. As we can see, there are some abrupt changes in the position of the FOA from one frame to the next one, especially on forward motion. In general, this can be considered as a good behavior because an exploration task implies that the focus of attention changes over time. However, when the FOA only stays for a very short period of time (*i.e.*, in very few consecutive frames), then this may become a problem as the abrupt uncontrolled changes of position of the FOA may cause an erratic movement in the robot. To solve this, we segment the image in superpixels. From the saliency map we take the n most relevant points and describe them with the a, b, x and y component of the superpixel they belong to. Figures 5c-d show the results of the FOA obtained at each frame for Videos 1 and 2 but now using the superpixel descriptors.

Once we have the descriptor of each saliency point, we choose the closest (the most similar) to the FOA of the previous frame. The similarity measure is calculated using (8). It is important to mention that if the distance from the closest saliency point to the previous FOA is greater than a defined threshold then we ignore the distances and the point with the highest saliency value is chosen as the new FOA. It can be observed that the FOA still takes arbitrary regions in the image, but once a new FOA is obtained, it stays almost in the same region on several consecutive frames. The effect of the improvement to the visual attention algorithm can be seen more clearly on the plot of Video 1. We can obtain less abrupt changes in the FOA by adjusting the threshold for the similarity measure between a previous FOA and the new saliency points. In the previous experiment, for Video 1, this threshold was 2 (Figures 5c and 6a). We observed that the greater the value of the threshold the less abrupt changes in the FOA. Fig. 6b shows results when applying a threshold value of 2.5 and using the naive normalization. Each time the distance overpasses the threshold, we extract a snapshot of the FOA and its surrounding region (pixels) to see how different these

 ISBN: 978-1-61208-340-7

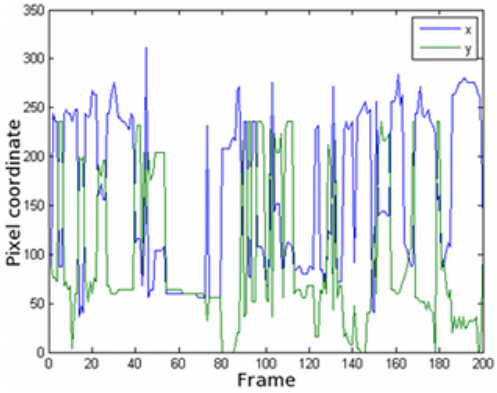


(a) Position of FOAs with their representative snapshots at each change (indicated by regions A, B, C and D) when using a threshold of 2.

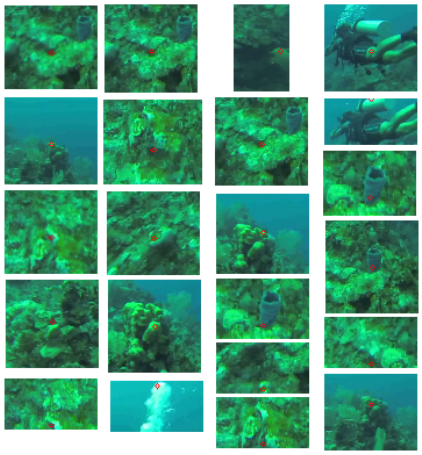


(b) Position of FOAs with two relevant regions when using a threshold of 2.5.

Figure 6. Relevant regions captured at each change of FOA.



(a) Position of FOA points.



(b) Relevant regions detected.

Figure 7. Position of FOA points with some of the associated relevant regions when using the $N(\cdot)$ operator.

regions are. As it can be seen in Fig. 6a, when the threshold was set to 2 the algorithm found 10 relevant regions, although many of them represent almost the same scene. Instead, when the threshold was set to 2.5 (see Fig. 6b) the algorithm found only 2 relevant regions.

We carried out an additional experiment, in which our algorithm was tested using the $N(\cdot)$ operator in order to compare the results with those obtained using the naive normalization (Fig. 6b). We can observe in the plot of Fig. 7a that the position of the FOA changes abruptly when the $N(\cdot)$ operator is used; about 57 relevant regions were found using this operator (Fig. 7b). Even so, this may be useful in an offline program to find all the possible relevant regions or in a training phase.

Finally, as poor visibility is a common problem in underwater environments, we want to see how our algorithm performs in this type of conditions. Fig. 8 shows the position of the FOA and a snapshot of the relevant regions found when running our algorithm. Despite of the poor visibility, parts like the hand of the diver, the blue triangle in the red ball or the yellow tube are detected.

V. Conclusion and Future Work

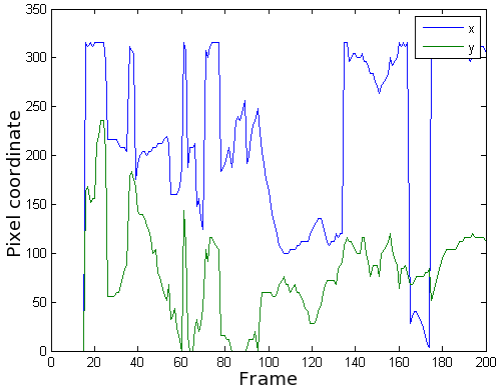
We have presented a novel approach which robustly detects regions of interest in underwater video streams and tracks them under forward and rotational movements. As it is shown in the experiments, the robustness of this approach is mainly due to two parts. The first part is the visual attention model that can determine relevant regions of an underwater image even if the geometry or shape of the environment to explore is unknown, making it ideal when dealing with unstructured environments.

The other important part is the use of the superpixels as a descriptor, because it summarizes the information of color and position of the similar set of pixels, thus reducing the computational time significantly.

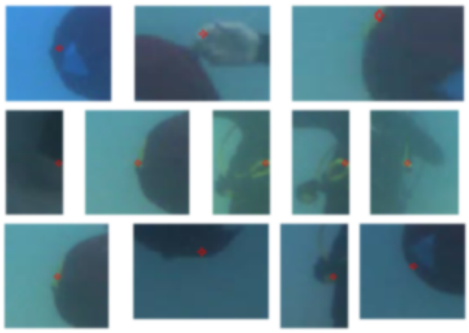
Our approach turns out to be also robust for tracking saliency zones even in scenes with poor visibility conditions.

As future work, we want to test our approach in real underwater explorations performed by our robotic system. Also, further analysis on the training process is needed in order to determine the concurrence of FOAs annotated by the users with those detected by our algorithm. Finally, the approach could be extended to track more that one region of attention.

 ISBN: 978-1-61208-340-7



(a) Position of FOA points.



(b) Relevant regions detected.

Figure 8. Results of applying our visual attention algorithm in underwater environments with poor visibility.

This will help to plan ahead the robot's trajectories, which leads to a better exploration.

ACKNOWLEDGMENT

The authors would like to thank CINVESTAV and CONACyT for funding this project, and to the Robotics Lab of Universidad Autónoma de Ciudad Juárez for its collaboration.

REFERENCES

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," IEEE Transactions on PAMI **34** (2012), no. 11, 2274–2282.

[2] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels", Tech. report, EPFL, 2010.

[3] James Mark Baldwin, *Mental development in the child and the race: Methods and processes*, Macmillan, 1906.

[4] C. Barat and M.-J. Rendas, "A robust visual attention system for detecting manufactured objects in underwater video," OCEANS, 2006, pp. 1–6.

[5] M. Begum and F. Karray, Visual attention for robotic cognition: a survey," IEEE Transactions on Autonomous Mental Development **3** (2011), no. 1, 92–105.

[6] Bennett I. Bertenthal, *Origins and early development of perception, action, and representation*, Annual review of psychology **47** (1996), no. 1, 431–459.

[7] D.R. Edgington, K.A. Salamy, M. Risi, R. E. Sherlock, D. Walther, and C. Koch, "Automated event detection in underwater video," OCEANS, vol. 5, 2003, pp. P2749–P2753 Vol.5.

[8] S. Frintrop, "VOCUS: a visual attention system for object detection and goal-directed search", Ph.D. thesis, Rheinische Friedrich-Wilhelms-Universität Bonn, 2006.

[9] S. Frintrop, "Computational visual attention," Computer Analysis of Human Behavior (A. A. Salah and T. Gevers, eds.), Springer London, 2011, pp. 69–101.

[10] S. Frintrop, G. Backer, and E. Rome, "Goal-directed search with a top-down modulated computational attention system," Pattern Recognition (W. G. Kropatsch, R. Sablatnig, and A. Hanbury, eds.), Lecture Notes in Computer Science, vol. 3663, Springer Berlin Heidelberg, 2005, pp. 117–124.

[11] L. Itti and C. Koch, "A Comparison of Feature Combination Strategies for Saliency-Based Visual Attention Systems," Journal of Electronic Imaging **10** (1999), 161–169.

[12] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence **20** (1998), no. 11, 1254–1259.

[13] P. Lobato-Correia, P. Y. Lau, P. Fonseca, and A. Campos, "Underwater video analysis for Norway lobster stock quantification using multiple visual attention features," 15th European Signal Processing Conference, 2007.

[14] S. Palmer, "Vision Science, Photons to Phenomenology," The MIT Press, 1999.

[15] J. Piaget, *The origins of intelligence in children*, New York: Int. Univ. Press, 1952.

[16] S. Thrun, S. Thayer, W. Whittaker, C. Baker, W. Burgard, D. Ferguson, D. Hannel, M. Montemerlo, A. Morris, Z. Omohundro, and C. Reverte, *Autonomous exploration and mapping of abandoned mines*, IEEE Robotics Automation Magazine **11** (2004), no. 4, 791.

[17] D. Walther, D. R. Edgington, and C. Koch, "Detection and tracking of objects in underwater video," Proc. of the IEEE Computer Society Conference onComputer Vision and Pattern Recognition (CVPR), vol. 1, 2004, pp. I–544–I–549.

 ISBN: 978-1-61208-340-7