A Sensor for Vision-based Navigation in Underwater Path Tracking with Color and Edge Segmentation

Mario A. Jordán [1,2,⋆], Emanuel Trabes[1], Carlos E. Berger[1,2]. Jorge L. Bustamante[1,2]

[1] Argentinean Institute of Oceanography (IADO-CONICET). Florida 8000, Complejo CCT, Edificio E1, B8000FWB, Bahía Blanca, ARGENTINA.

[2] Dto. Ingeniería Eléctrica y de Computadoras- Univ. Nac. del Sur (DIEC-UNS).
No Institute Given

Abstract. This paper aims the design and implementation of a vision-based sensor for navigation of underwater vehicles with adaptive attributes. The objective pointed out is a sensor for tracking of underwater lines. The sensor employs a basic structure with a pixel-wise AND operation of binarized frames of separated channels HSV and an edge-segmented frame. The basic sensor performs well by good illuminated scenes. By significant drops of luminance, the efficiency falls. So an adaptive sensor is proposed over the basic structure. It operates on the brightness channel carrying out a maximization of contains in the accumulator bins of a Hough transformation. It has proven to enhanced the identification of the tracked line increasing the success rate.
Key words: Vision-based sensor, underwater path tracking, Hough Transform, HSV model, pipeline identification.

1 Introduction

Vision-based systems for underwater navigation currently employ optic or acoustic devices which provides a video or an intermittent image for own localization and rate estimation with respect to the surrounding environment (Jordán *et al.*, 2010). They basically extract visual features of the seafloor by image segmentation, searching for chromatic and topological characteristics like colors, texture, shapes and contours.

Image segmentation is a widely studied area with hundreds of different algorithms (Law *et al.*, 2004). The existing techniques to perform image segmentation are quite numerous, containing basic procedures like thresholding, clustering, data compression, histogram, edge detection, region-growing, spliting-and-merging, level set, fast marching, graph partitioning, watershed transformation, primal sketch, semi-automatic and trainable segmentation among the most referenced (Shapiro and Stockman, 2001; Szeliski, R., 2010).

⋆ Corresponding Author: Mario A. Jordán: E-mail: mjordan@criba.edu.ar. Address: CCT-CONICET. Florida 8000, B8000FWB, Bahía Blanca, ARGENTINA

When processing videos in real time for visual servoing, the main requirement is the reliability of the vision-based feedback to the controller which must produce ultimately the course corrections in a stable way upon this information. An important application in the area of underwater navigation is the pipeline tracking. Here, the changes of scene illumination may alter drastically the performance of the vision-based sensor. Many of the problems associated with the performance drop are the consequence of improper selections of threshold levels for image segmentation. Even when there exists useful techniques for adaptive or automatic settings of such levels (Cheng *et al.*, 2001; Harrabi and Ben, 2011), the sensor functioning fails if there does not exist a sufficient number of additional (sometimes redundant) attributes involved in the segmentation.

While most edge detection techniques have been developed for gray-scale images, color images can provide additional information, for instance, noticeable edges between iso-luminant colors are useful cues but fail to be detected by gray-scale edge operators (Szeliski, R., 2010).

The central contribution of this papers aims the image segmentation from a moving camera in order to detect and track in real time objects with certain specified attributes of color and shape. Particularly, the object we are focusing in this work is a line stretch (actually a duct) framed in a vision field on the seafloor. One assumes the stretch has certain morphological attributes like a dominant rectilinear profile, at least piecewise linear. However, other topological characteristics like connectedness are not necessary ensured, for instance by temporal occlusion of the line.

We will emphasize the experimental observation that the success in the line identification is very sensible on how the image codification is structured and that the techniques has to restrict the information starting from inside to outside in this structure. The goal of the paper is to describe a new methodology for color/shape segmentation that provides a continuous identification of a line on a luminously disturbed and regular-quality video aimed to the goal of path tracking.

2 Pipeline position

We start the description of an underwater scene and the object type to be identify and track over it. In Fig. 1 the typical appearance of a duct (pipe or cable) on the seafloor is illustrated. We can say that the physical pipeline is identified at a time point when we can establish a tangent line with slant α and distance x_L from a reference point (pivot point) to the vertical axis of the frame.

From a chromatic viewpoint and from the psychophysical perception of the image too, it is clear that almost monochromatic aspect of the image masks the detection of the true color composition of the scene. It is well-known that under low-level-illuminated scenes, colors in red, orange, yellow (and so on in decreasing wavelength) will go losing brightness, and the more accentuated the deeper is the reflecting object. So the dynamic range of luminosity remains an

important issue in the use of underwater highly sensitive cameras and artificial illuminators.

Generally, not only color attributes for the trackeable object are taken into account, but also for the scene. In case of the line, these will be somehow appropriately detected together with attributes of the profile.

In formal mathematical terms, let F a set of connected pixels and a segmentation of F be defined as a partition into connected subsets S_i with $F = \cup S_i$ with $S_i \cap S_j = \Phi$. Then let P be a predicate of homogeneity (or uniformity) defined on groups of connected pixels such that it is valid

$$P(S_i) = \textbf{true} \text{ for all } S_i \tag{1}$$
$$P(S_i \cup S_j) = \textbf{false}, \text{ with } S_j \text{ a neighbor set}$$

The predicate may be a hue value interval in the color model HSV, a histogram, a close curve, and so on. The predicates we are searching for are related to color and profile (edge above all) of the (pipe)line.

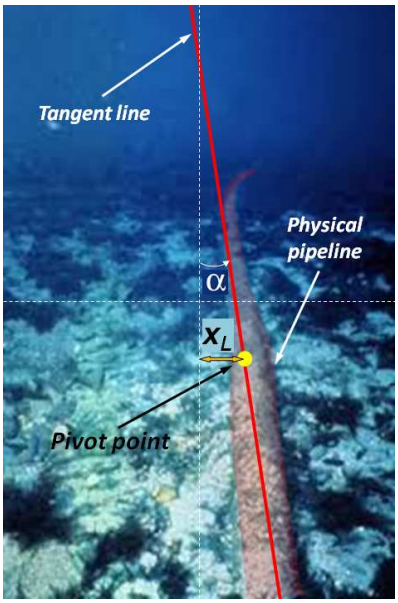

Fig. 1 - Subaquatic pipeline and identification of the AUV realtive position for line tracking

2.1 Color predicate

One way to isolate color attributes consists in dealing with some of the cylindrical-coordinate representations of pixels, namely HSL, HSV and HSI from a RGB color model. Let us consider a hue channel, the saturation channel and brightness channel separately for values of hue, saturation and luminance (or value or intensity), respectively.

We define the color predicate $P_c(.)$ as the color attributes of a initial patch picked up from the frame in which object to be tracked is depicted. In the Computing Vision jargon, the patch is usually defined as a ROI (Region Of Interest). This commissioning process for patch hues is done in a calibration phase. So

$$P_c(S_{patch}) \triangleq \textbf{true} \text{ if for all } p_{ij} \in S_{patch} \text{ we have} \\ H(p_{ij}) \, \epsilon \, [H_0 - \delta_H, H_0 + \delta_H] \tag{2}$$

where p_{ij} is a pixel of the patch S_{patch} and $H(.)$ is the H-value (between 0^o and 360^o) of the HSV channel for some pixel. The parameter H_0 is the mean H-value of the patch given by the hue channel and the δ_H is the dispersion of that hue value. This last value is a design parameter that can be appropriately selected in the commissioning phase. For instance, δ_H may be the maximal dispersion of the H-value on the patch of some integer value larger.

2.2 Edge predicate

One of the usually employed attributes in subaquatic navigation is related to the shape. Here we define a predicate of homogeneity for the 2D-pofile of the trackeable object (which is a detectable stretch of line) as

$$P_e(S_{line\_stretch}) \triangleq \textbf{true} \text{ if for all } p_{ij} \in S_{line\_stretch} \\ \text{with } S_{line\_stretch} \equiv S_{Hough} \tag{3}$$

where $S_{line\_stretch}$ is a pixel set coincident with S_{Hough}, which is a set containing the pixels of the Hough Transform on the frame that recognizes a straight edge of the line stretch present in the frame.

2.3 Hough Transform for a straight edge

Employing a Hough parametric space, any shape can be written in parametric equation form (Antolovic, 2008). For instance, a straight line detection is based on the point-line duality relation, which is a singularity-free alternative with respect to other parametrization forms. Computationally, it is advantageous for a line in the Hough transform to employ

$$y = -\left(\frac{1}{\tan\theta}\right) x + \frac{r}{\sin\theta} \tag{4}$$

with parameters r and θ, where r is the distance from the frame origin to the line and θ is the angle of the vector from the origin to this closest point. So the pair (r, θ) is unique to the line that crosses that closest point. All points (x_i, y_i) in the space x-y that are on the line, are mapped in a unique point (r_0, θ_0) in the parameter space r-θ. This point represents the crosspoint of the trigonometric curves $r(\theta) = x_i \sin\theta + y_i \cos\theta$ in the parameter space. The purpose of the Hough

transform is to address this problem by making it possible to perform groupings of edge points into object candidates by performing an explicit voting procedure over a set of parameterized image objects.

For each pixel and its neighborhood, the Hough Transform algorithm determines if there is enough evidence of a straight edge at that pixel. The way this evidence raises is the following. The pixel is mapped into the parameters space (r, θ), providing a new element of a cluster of small dimensions that is progressively growing and is related to the straight edge on which this pixel would be. Then an associated accumulator bin is incremented by one. By finding the bins with the highest number of values, typically by looking for local maxima in the accumulator space, the most likely lines can be extracted, and their (approximate) geometric definitions read off.

Formally, the accumulator bin named b_i produces $b_i(p_j) = n_i$, it is n_i votes from n_i pixels p_j and it is valid

$$\begin{aligned} b_i(p_j) = n_i \rightarrow S_{Hough} = \{p_j\} \\ \max_{i=1...N} b_i(p_j) = \max_{i=1...N} n_i \rightarrow S_{Hough} \equiv S_{line\_stretch}, \end{aligned} \tag{5}$$

where N is the number of bins. The approach was significantly improved, above all in the performance of the voting scheme, producing a transform more robust to the detection of spurious lines (Fernandes and Oliveira, 2008).

3 Vision-based Sensor

Now, we can argue that color and shape attributes may provide together a solid basis of information to detect the (pipe)line under variable environmental conditions of light. In Fig. 2 we present the basic sensor algorithm for position estimation. As indicted in Fig. 1, the estimation parameters are α and x_L and correspond to a vision space (not to the physical space with Euclidean metrics).

It is noticing that the algorithm employs a gray scale for detecting the edges of objects by means of a Canny edge detector. Parallel, the RGB information is separated into channels, namely: the hue H, the saturation S, and the value V, which are binarized in isolation. Color attributes are searched for in the hue channel only where a threshold is used in the binarization equal to the mean value in the hue patch of the commissioning phase.

One of the innovations of this algorithm is the implementation of the pixel-wise AND operation between all binarized frames. After the stage, the recognition of the line stretch has been proven to be much more successful than the shape segmentation alone.

4 Adaptive Mechanisms for the Sensor

Good luminance conditions of the scene favors the right functioning of the sensor. However, by drops in the external luminance, the brightness of colors decreases

significantly (actually, the histogram of the value shifts to the left). Even when the hue histogram is less insensible to luminance changes in comparison with saturation and brightness, the AND operation may fail due to the darkening of the frame composition that masks any vestige of edge. In such cases we notice that color and shape predicates are false.

In order to slump the level of sensibility of the pattern recognition to changing scene lighting, some warnings of sensor malfunctioning has to be diagnosed. Additionally, we expect that some added adaptive properties of the sensor will allow an enhancement of the precision in the line estimation too. We described the elements of an adaptive sensor separately before summarizing the complete design of the algorithm in the closure of this section.

4.1 Generation of performance errors

For a stable navigation the control system employs a memory of the line parameters (x_L, α) based on the past history of the trajectory. Under the hypothesis that the (pipe)line may have a smooth curvature, then the actual position of the line can not be too dissimilar from the previous position. So, one can propose a trust zone in the vision space about the previous tangent line and argue that the actual estimated tangent line has to be inside (see Fig. 3).

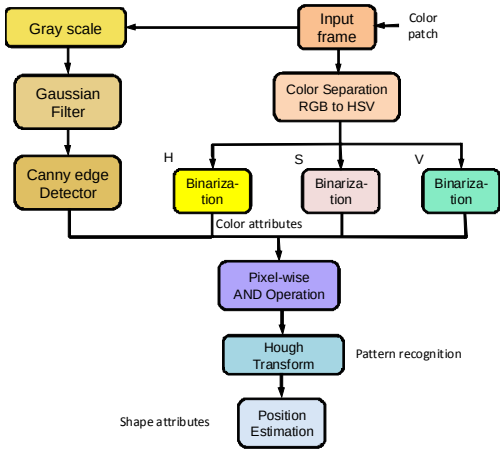

Fig. 2 Vision-based Sensor

In this sense we postulate a flag of error ε_p for the success of the position estimation according to

$$\begin{gathered} \text{if } \|\alpha_0 - \alpha_1\| + \|x_{L_0} - x_{L_1}\| \leq \delta_p \text{ then } \varepsilon_p = 1 \\ \text{if the actual line is undetected then } \varepsilon_p = 0, \\ \text{otherwise } \varepsilon_p = 0, \end{gathered} \tag{6}$$

where δ_p is a real-valued number, for instance $\delta_p = 25$, which results for the Euclidean norm with position differences of $|\alpha_0 - \alpha_1| \leq 15^o$ and $|x_{L_0} - x_{L_1}| \leq 20$ pixels.

Another postulation of the performance error can be done according to the results of the Hough Transform. One can observe the bins b_i of the accumulators and take the parameters (r_0, θ_0) and (r_1, θ_1) for the actual and previous estimated tangent lines to state

$$\begin{gathered} \text{if } \|r_0 - r_1\| + \|\theta_0 - \theta_1\| \leq \delta_p \text{ then } \varepsilon_p = 1 \\ \text{if there is almost empty bins then } \varepsilon_p = 0, \\ \text{otherwise } \varepsilon_p = 0, \end{gathered} \tag{7}$$

similarly δ_p is a real number, for instance $\delta_p = 25$, which results for the Euclidean norm with position differences of $|r_0 - r_1| \leq 20$ pixels and $|\theta_0 - \theta_1| \leq 15^o$. We consider that the Hough Transform gives a positive result (line detected) if at least there is one accumulator bin which has more than a critical number of pixels, for instance 5.

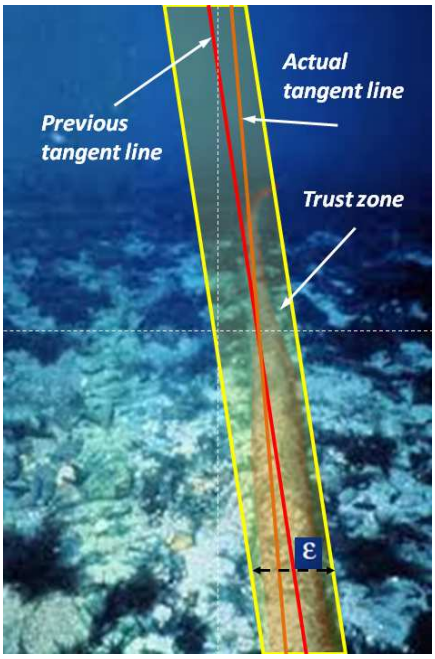

Fig. 3 - Trust zone of the sensor of width ε

Both postulations (eqs. (6) and (7)) has similar efficiency, nevertheless the second postulation is more simple for programming the algorithm code.

4.2 Indicator of optimal detection

The certainty of right line detection is given by the indicator flag state $\varepsilon_p = 1$. Additionally, we can ensure certain optimality for both the right detection and a more precise identification. With this end in mind, we will design a particular indicator based on the Hough Transform.

As said in our observation from many videos, large variation of the scene lighting causes significative variation of the value in the channel V (of the HSV model). Also the pixel-wise AND operation is decisive in its contribution to the success of the line detection. Accordingly we can take the threshold of this value as a sensitive variable to maximize the probability of the success and enhance the precision of the estimates α and x_L.

Thus, we maintain the hypothesis of the existence of an optimal cut point in the binarization of V for which

$$\begin{aligned} P_{\max} &= \max_{\text{threshold of } V} P(\varepsilon_p = 1) \text{ occurs if} \\ n_{\max} &= \max_{\text{threshold of } V} b_i(p_j) \text{ with} \\ S_{Hough} &= \{p_j\} \text{ and } S_{Hough} \equiv S_{line\_stretch}, \end{aligned} \tag{8}$$

where $P(.)$ is the probability function of an event, $P_{\max}$ the maximal probability of $\varepsilon_p = 1$ and $n_{\max}$ the maximal contain of pixels that can only correspond to the bin associated to the line. We are then claiming the hypothesis that both maximizations are directly related and that there exists and optimal threshold for V that maximizes the bin contain corresponding to the right line detection. The confirmation of that hypothesis is carried out through the analysis of numerous case studies.

4.3 Iterative adjustment of brightness threshold

Now, the adaptive adjustment of the sensor brightness threshold can be best illustrated in Fig. 4. After color and edge segmentations, the algorithm generates the error ε_p. By $\varepsilon_p = 1$, the algorithm employs a variation of the threshold of V in certain stipulated initial quantity δ_V, for instance 30% of the initial cut point and repeats the procedure of segmentations. In the new steps the contains of the accumulator bins are checked, particularly such one n_i that gives $\varepsilon_p = 1$ according to (7). By a gradient-based procedure new quantities δ_V are proposed in order to maximized n_i in new iterations. After a stipulated maximal number of iterations or when the quantity δ_V has bee reduced under a minimal tolerance, the algorithm is aborted. In this step, the tangent line which marks the pipeline position is estimated.

The experimental analysis with the proposed sensor determines the superiority of it in the success rate with respect to other procedures that do not employ illuminance adaptation.

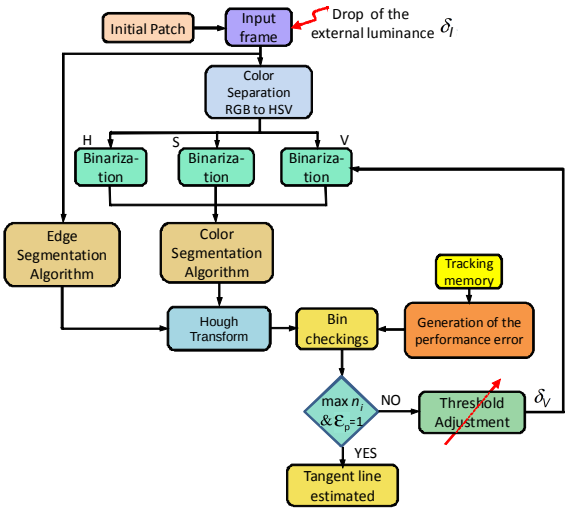

Fig. 4 - Adaptation mechanism to maintain the sensor performance

5 Experimental case study

For the validation of the sensor performance, experimental setups has specifically been mounted. A scene with a color line (to be tracked) was deposit on the floor in a water tank. The scene was arranged in such a form that other objects like colored stones and additional colored lines with dissimilar hues be present looking approximately like natural scenarios. The scene was external illuminated with large variations of the luminance. A subaquatic camera was conducted at certain altitude along the line and the resulting videos were processed in the framework of the proposed sensor. The commissioning phase consists above all in picking up a patch of the line to extract color attributes.

We have selected one case study which is deemed to be representative of the whole experimental tests carried out in this work. We chose three particular frames of the video containing the line tracking test. In order to compare the adaptive sensor performance, a sensor with fixed (but good tuned) parameters is developed with such purpose.

We begin the illustration with Fig. 5 which depicts the evolution of the performance errors ε_p for the two sensors. Clearly the adaptive sensor has shown a better performance in comparison.

In Fig. 6, we see, on the left, three selected frames and, on the right, the three homologous frames with augmented reality. These illustrate three instances of the navigation environment, namely a poor, a normal and a good illuminated scene, respectively. In each frame on the right, there are 3 tangent lines described together, namely a yellow line, a blue line and a green line characterizing the estimated previous line, the estimated line with adaptive sensor and the estimated line with fixed-parameter sensor, respectively.

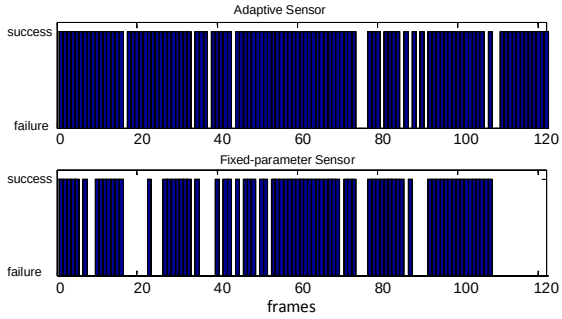

Fig. 5 - Evolution of the performance errors in the adaptive sensor (top) and the fixed-parameter sensor (bottom)

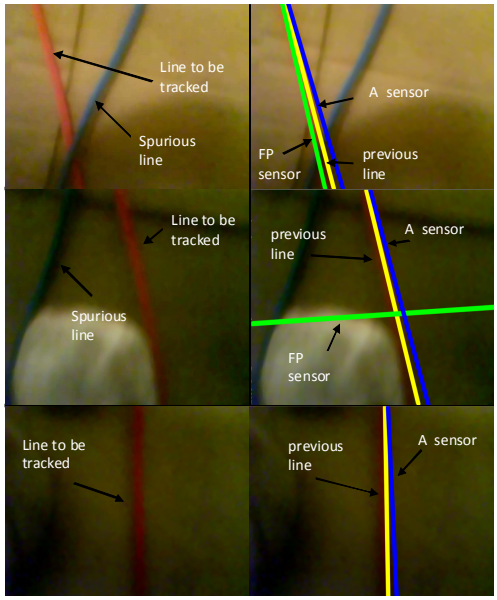

Fig. 6 - Video frames of the line tracking. Left: raw frames. Right: frames with identified lines (in augmented reality)

The first instance (it is,the first row) agrees with a right identification of the line by the two sensors. This means that in both estimations, the approximative coincidence caused $\varepsilon_p = 1$. In the second row, the adaptive sensor is able to estimate the line correctly (yellow and blue lines concur nearly) while the fixed-parameter sensor (green line) does not. The third row depicts a case with low level of illumination. Here, the fixed parameter sensor fails and the green line does not appear in the frame with augmented reality.

Fig. 7 shows the results of the pixel-wise AND operation (*cf.* Fig. 2). Again there are three rows that cut across the three frames of the Fig. 6.

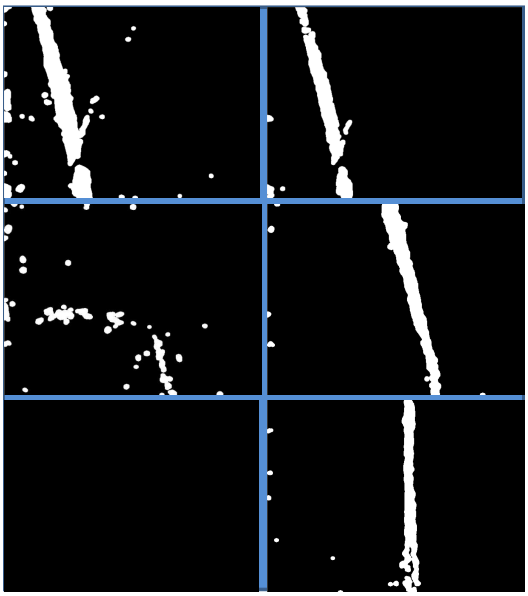

Fig. 7 - Detection of straight objects. Left: for the fixed-parameter sensor (FP). Right: for the adaptive sensor (A)

As we see, the first row illustrates the rectilinear objects with high degree of resemblance with the line, conducing in this way to the right line identification. On the other side, we notice in the second row, that the rectilinear object of the fixed-parameter sensor possesses a quite different slant in comparison with that of the adaptive sensor. The third row shows clearly the absence of a straight object in the frame of the fixed-parameter sensor, pointing out consequently a failure. However, the high drop of luminance existing in this case, does not affect the performance of the adaptive sensor which shows success.

Finally, the histograms for the analyzed three frames are given in Fig. 8 in separated channels. Also the histogram of the initial patch is reproduced. Comparing with the initial histogram of the patch, those of the saturation and value channels in the frame 1 (it is, by normal luminance) suffer from a shift to the right, but they retain their standard deviations approximately. Moreover, the brightness in the V-channel is more affected because it moves over much more that the histogram of saturation.

In frame 2, when the scene luminance decays, the brightness (value) shifts completely to the right and the saturation moves over to the left but only a little. In frame 3, by the same conditions of darkness, the value histogram is quite on the right. It is worth noticing that the hue histogram in all frames practically shows no significant change. This is an important fact that was exploited in the design of the adaptive algorithm.

6 Conclusion

In this work a vision-based sensor for navigation of underwater vehicles is designed and implemented. The objective pointed out is an adaptive sensor for

tracking of underwater lines. Several experimental validation tests has shown that the sensor superior in performance that other types of algorithm that do not employ luminance adaptation. The sensor is composed of a basic structure that performs a pixel-wise AND operation of binarized frames of separated channels HSV and an edge-segmented frame. This basic sensor performs well by good illuminated scenes. By significant drops of luminance, the efficiency falls. So an adaptive sensor is proposed over the basic structure. It operates on the brightness channel carrying out a maximization on contains in an accumulator bins of the Hough transformation. It has proven to enhanced the identification of the tracked line increasing the success rate.

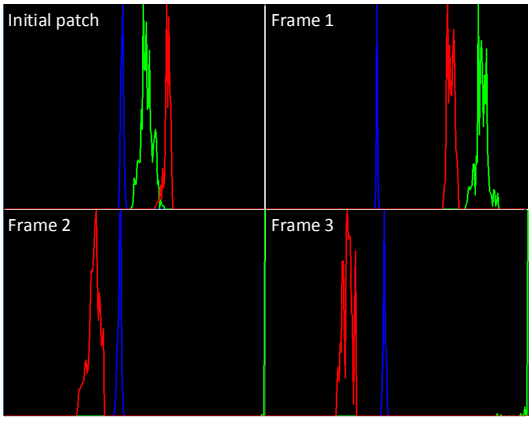

Fig. 8 - Histograms of the initial patch and of the separated channels in HSV model. In blue: H-channel, in green: S-channel and in red: V-channel

References

1. Antolovic, D. (2008). "Review of the Hough transform method, with an implementation of the fast Hough variant for line detection". Technical Report TR663, April 2008, Indiana University.
2. Cheng, H.D.; Jiang, X., Sun, A. and Wang, J. (2001). "Color image segmentation: Advances and prospects". Pattern Recognition 34 (12).
3. Fernandes, L.A.F. and Oliveira, M.M. (2008). "Real-time line detection through an improved Hough transform voting scheme". Pattern Recognition, Volume 41, Issue 1, January 2008, Pages 299–314.
4. Harrabi, R. and Ben, B.E. (2011). Color image segmentation using automatic thresholding techniques. SSD'2011, Tunisia 2011, 1-6.
5. Jordán, M.A., Berger, C.E., Bustamante, J.L. and Hansen, S. (2010). Path Tracking in Underwater Vehicle Navigation - On-Line Evaluation of Guidance Path Errors via Vision Sensor. *49th IEEE Conf. on Decision and Control*. Atlanta, USA.
6. Law, M.H.C.; Figueiredo, M.A.T.; and Jain, A.K. (2004). Simultaneous Feature Selection and Clustering Using Mixture Models, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 26, NO. 9.
7. Shapiro, L.G. and Stockman G.C. (2001). "Computer Vision", pp 279-325, New Jersey, Prentice-Hall, ISBN 0-13-030796-3.
8. Szeliski, R. (2010), Computer Vision: Algorithms and Applications, Sep. 3, 2010 draft, Springer.