Using the Motion Perceptibility Measure to Classify Points of Interest for Visual-based AUV Guidance in a Reef Ecosystem

Lourdes Labastida-Valdés, L. Abril Torres-Méndez
Robotics and Advanced Manufacturing Group
CINVESTAV Campus Saltillo
Ramos Arizpe, Coahuila, MEXICO
Emails: lourdes.labastida; abril.torres@cinvestav.edu.mx

Seth A. Hutchinson
Dept. of Electrical and Computer Engineering
University of Illinois
Urbana Champaigne, USA
Email: seth@uiuc.edu

***Abstract*—In this paper we present a novel method for classifying relevant points in a sequence of images of a distant target to autonomously guide an underwater vehicle towards it. Feature points are classified by using a measure called motion perceptibility, which relates the magnitudes of the rate of change between matched feature points at different image frames (in distance), thus inherently considering the change in feature's position. This measure helps to detect which feature points are most likely to leave the field of view of the camera, thus indicating that they do not belong to the target region. By using a visual attention model adapted to underwater images, relevant points are detected and tracked using a visual servoing approach. Preliminary results on sea trials demonstrate the feasibility of our methodology.**

I. INTRODUCTION

Marine ecosystem is one of the ecosystems with the greatest number of living organisms. Particularly, coral reefs serve as a refuge and breeding site of a huge number of species. The correct gathering of information from this type of environment is crucial in order to avoid damaging it. Since prolonged stays underwater for humans is not feasible, the use of Autonomous Underwater Vehicles (AUVs) is an alternative for underwater exploration since allow, by means of performing tasks such as image capturing, the evaluation of reef structures and wildlife that inhabits it. Finding objects, monitoring, and recognition of the ocean floor are among the applications of autonomous underwater vehicles.

One of the main challenges in underwater robotics is location of the robots as it is very difficult to achieve under water. A simple GPS sensor cannot be used under water and some other technologies such as the SONAR can be invasive to the ecosystem. Furthermore, in aquatic environments there are not structures with defined geometric shape. This is why the use of visual information is useful for guiding a robot on exploration tasks.

We can get many information from an image captured by a vision system, such as coordinates of points of interest in the image plane, perimeters, areas and centroids of figures, among others. However, it is necessary to ensure that the obtained information is appropriate for the desired application, especially when working with images in aquatic environments. The main challenge that the algorithms for detection and tracking of points in underwater environments face is that often the composition of the water causes a limited visibility. Therefore, methods to obtain information from images in underwater scenes must be robust. With the right information, an aquatic vehicle, equipped with a vision system, can make connections between the real world and the image, and can be guided in exploration tasks.

In this paper we present a method to classify relevant points in aquatic images to guide a robot, driven by a visual servo control system, towards a distant static target. The classification is accomplished by implementing the motion perceptibility measure [1] in the detection of the points that are most likely to leave the field of view of the camera, positioning them in the target region. To detect and track the relevant points we used a robust method based on a visual attention model.

II. RELATED WORK

There have been some studies regarding the detection and tracking of points in underwater images such as the one conducted by Maldonado *et al.* [2]. A model of visual attention aims to imitate the behavior of the human eye when it detects an object in its field of view that catches its attention. Usually the most interesting elements to the human eye are defined based on color and intensity. Similarly, the method proposed by Maldonado *et al.* processes an image, obtaining different color maps and scales, and giving a preference to a certain color feature by means of assigning a weight to the map of each color. The final map is a gray-scale image in which the brightest points in the image are the most relevant regions of the original image. In this way the detection of a set of relevant points is accomplished and the most relevant region is selected and is designated as a focus of attention (FOA). In the next frame, new points are detected and the difference in position and color between each new point and the FOA of the previous frame. The closest point is chosen as a new FOA. In order to storage the color and position of pixels the image is segmented into superpixels [3]. Each superpixel contains a descriptor with the information of all the pixels that it groups.

The visual attention method of Maldonado *et al.* [2] proved to be robust enough to detect and track relevant points in aquatic scenes, even when visibility is poor. We extended the method in order to select more than one relevant region as

978-0-933957-43-5 ©2015 MTS

FOA and track each of these FOAs in a sequence of images. The tracking is accomplished by computing the distance in color and position between each relevant point in a new frame and each FOA in the previous frame. Thus, each new FOA corresponds to a point with the shortest distance relative to a FOA of the previous frame.

The results obtained with the second video presented an issue for tracking the FOAs from a long distance, an aspect that is essential in order to meet the objectives of this thesis. In Section 5, a method used to deal with this problem is explained.

III. Motion perceptibility measure.

As we stated previously, in computer vision it is possible to obtain information from images such as color of pixels, perimeter and centroid of Figures, etc.This information is known as image feature parameters. One of the most used feature parameter is the projection onto the image plane of a point in three-dimensional space [4]. The correspondence is given by:

$$\begin{aligned} u &= f\frac{x}{z}, \qquad (1) \\ v &= f\frac{y}{z}. \end{aligned}$$

The Image Interaction matrix, also known as Image Jacobian matrix, relates the changes with respect to time in the trajectory of a camera and the changes with respect to time in the image feature parameters. Hutchinson *et al.* [4] present a model of an image Jacobian matrix based on the coordinates of points in the image plane. Such matrix is defined as follows:

$$\mathbf{J}_v = \begin{bmatrix} \frac{f}{z} & 0 & -\frac{u}{z} & -\frac{uv}{f} & -\frac{f^2+u^2}{f} & -v \\ 0 & \frac{f}{z} & -\frac{v}{z} & \frac{-f^2-v^2}{f} & \frac{uv}{f} & u \end{bmatrix}. \qquad (2)$$

When we use a set of points as feature parameters it is necessary to stack the interaction matrix of each point to have the interaction matrix of the whole set.

Sharma and Hutchinson exploit the properties of the Image Jacobian and introduce the motion perceptibility measure [1]. Motion perceptibility is an scalar quantity that provides a sense of how well a camera can detect the movement of objects in its field of view. In other words, motion perceptibility relates the velocity of objects in 3D space with the velocity of the features in the image that captures such objects.

The motion perceptibility measure is computed depending on the number of image feature parameters available. Let k be the number of image feature parameters and m be the task space dimension, the motion perceptibility is given as follows:

- With $k > m$, $\omega = \sqrt{\det(\mathbf{J}_v^T \mathbf{J}_v)}.$
- With $k < m$, $\omega = \sqrt{\det(\mathbf{J}_v \mathbf{J}_v^T)}.$
- With $k = m$, $\omega = |\det(\mathbf{J}_v)|$.

In that work, the motion perceptibility measure was analyzed in two cases involving a redundant manipulator robot and non-redundant one. In both cases, the robot motion is detected by an active camera in a hand/eye configuration. This analysis shows that there exists an important relation between the motion perceptibility measure and the relative position of the camera and the manipulator. It is also demonstrated the effectiveness of the motion perceptibility measure applied in different visual servo control tasks such as optimal positioning of a camera, trajectory planning of a robot and an active camera, identify critical directions of movement, among others.

In order to understand the behavior of the motion perceptibility measure, we computed the measure in different configurations using an active camera with six degrees of freedom which was focused on static visual features. One of the configuration consisted of a paper sheet with colored circles as a static target. The camera moved towards the target perpendicularly to the projection axis and also moved in semicircle in order to capture the paper sheet with a variety of orientations. Five circles were selected as image features. The perceptibility of each feature and of the entire set was obtained at different known distances between the camera and the target and with the projection axis in different known angles with respect to an axis perpendicular to the image plane. The coordinates of the image features were obtained by inspection. The experimental setup can be seen in Fig. 1.

Figure 2 shows the changes in the overall perceptibility with respect to the distance camera-target and the camera orientation. It was observed that ω increases when the camera is centering the target. This shows that there exists an important dependence of the perceptibility on the orientation of the camera.

Fig. 1. Experimental Setup 2 used to compute the perceptibility.

Although the relationship between the perceptibility and the distance camera-features is clear, there is an important restriction: it is necessary to know (or at least have an estimate) such distance to use the interaction matrix of Eq. 2. However, to know how far objects are using only visual information in an aquatic environments is a complex task. The variation of the perceptibility with respect to the orientation is useful to deal with this problem. It is possible to relate the velocity of points on the image with the angular velocity of the camera considering only the part of the interaction matrix that performs the mapping between these two variables. For example, the following interaction matrix relates the velocity

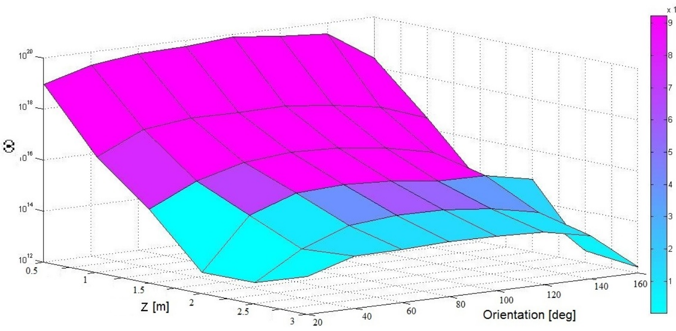

Fig. 2. Changes in the overall perceptibility with respect to the distance camera-target and the orientation of the camera.

of the image features with the angular velocity of the camera along the x and y axes:

$$\mathbf{J}_v = \begin{bmatrix} -\frac{uv}{f} & -\frac{f^2+u^2}{f} \\ \frac{-f^2-v^2}{f} & \frac{uv}{f}r \end{bmatrix}. \tag{3}$$

IV. Implementation of the motion perceptibility measure to classify points of interest.

Visual servo control is characterized by using visual information from images to control the movement of a robot. In most cases, the information obtained from an image are features of interest in a 3D scene which are projected in the image. However, it is essential that these features remain in the field of view of the camera during the control task. Some studies such as the one is presented in [5] have analyzed the use of the motion perceptibility measure, introduced by Hutchinson and Sharma, to keep a set of mobile features within the field of view. In this work, authors show that maximizing the motion perceptibility of a set of features, those features will be more likely to remain in the field of view

A potential application of the motion perceptibility measure is the feature classification. From a set of image features it is possible to select those whose velocity in the image plane may be related to the camera velocity throughout a wider range of directions. This implies that the motion perceptibility may be useful to select the features that are most likely to remain in the fieleld of view of the camera while it can move in different directions. A classification of this type is helpful to guide the movement of a camera towards a static target which is at long distance, as explained in the following paragraphs.

Let ω_{k+1} be the perceptibility of a set of k features and ω_{k+1} be the perceptibility of $k+1$ features. Consider now a set of n features, each one of the features is a candidate to be added to the set of k features. In order to maximize the perceptibility, the new feature should be the one that when added provides the highest perceptibility to the whole set. In other words the new feature must be one that helps to satisfy the following condition:

$$\omega_{k+1} = \max(\omega^j_{k+1}), \tag{4}$$

where ω^j_{k+1} is given by:

- For $k+1 > m$,

$$\omega^j_{k+1} = \sqrt{\det\left(\begin{bmatrix} \mathbf{J}_{v_k} \\ \mathbf{J}^j_v \end{bmatrix}^T \begin{bmatrix} \mathbf{J}_{v_k} \\ \mathbf{J}^j_v \end{bmatrix}\right)}, \ j = 1, \ldots, n.$$

- For $k+1 < m$,

$$\omega^j_{k+1} = \sqrt{\det\left(\begin{bmatrix} \mathbf{J}_{v_k} \\ \mathbf{J}^j_v \end{bmatrix} \begin{bmatrix} \mathbf{J}_{v_k} \\ \mathbf{J}^j_v \end{bmatrix}^T\right)}, \ j = 1, \ldots, n.$$

- For $k+1 = m$,

$$\omega^j_{k+1} = \left|\det\left(\begin{bmatrix} \mathbf{J}_{v_k} \\ \mathbf{J}^j_v \end{bmatrix}\right)\right|, \ j = 1, \ldots, n.$$

In the previous expression, ω^j_{k+1} is the perceptibility computed by stacking the overall interaction matrix of the set of k features and the interaction matrix of each feature of the set of n features. The parameter m refers to the dmension of the task space.

A. Use of the convex hull to optimize the classification.

The method to classify features presented above requires a high computational cost since it is necessary to compute the perceptibility with all the features belonging to the set n at each time that we want to add a new feature to the set of k features. For example, suppose a control task in which we want to track a set of $k = 5$ features, which can be chosen from a total of $n = 10$. This implies that the perceptibility should be computed at first for $k = 1$. In order to do that we take as a possible new feature each one of the $n = 10$ available, and determine which one provides the highest perceptibility. The feature that provides highest perceptibility is selected to be part of the desired set. After that, it is necessary to compute the perceptibility with $k+1$ features, stacking the interaction matrix of the feature already selected and the interaction matrix of each one of the candidate features. The previous procedure should be repeated until the set of $k = 5$ features is completed. Then, the approximate number of floating point operations is:

$$\text{FPO} \approx n[\ p(4k+3) + p(p!)], \tag{5}$$

with p as the number of columns of J_{v_k}.

From Eq. 5, we can infer that when decreasing the number of candidate features we can decrease the computational cost of the method. We added the estimation of the convex hull of the set of candidate features to address this problem.

Consider a set of points in a plane (for example the image plane), denoted as S, then, the convex hull of S is a convex polygon that contains all the points of S. The points on the boundary of the convex hull are called vertices [6]. Figure 3 shows a graphic representation of the convex hull.

By finding the vertices of the convex hull of the set of candidate features, we can apply the classification method only to those vertices, since the vertices are the points that most likely will leave the field of view of the camera. Thus, the decreasing of the number of candidate features is achieved. It is noteworthy that introduction of the convex hull in the method involves a restriction: the features must be coplanar.

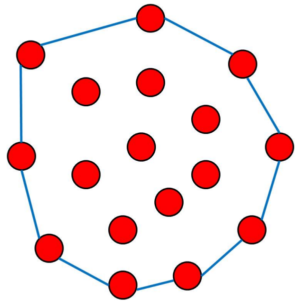

Fig. 3. Graphic representation of the convex polygon.

We used an additional strategy for choosing the features that maximize the perceptibility. The following describes the method step by step.

1) It was computed the perceptibility with all the features of the set of n features except one of the vertices of the convex hull.
2) Process of Step 1 was repeated with every of the remaining vertices. In this way, we can create a list of values of motion perceptibility measure. Each perceptibility value has a corresponding removed vertex.
3) An average of the perceptibility values in the list was obtained. Such average was used as a threshold.
4) We observed which perceptibility values were above the threshold.
5) The vertices that were removed to obtain perceptibility values above the threshold were selected as the features that most likely will leave the field of view of the camera.

We used the above method for classifying features according to their likelihood of leaving the field of view of a camera to guide an aquatic robot towards a static target in a coral reef environment. The center of the convex hull was useful to establish a visual error, and thus, implement a visual servoing system to produce the movement of the robot. Experiments and results are described in the next section.

V. FIELD EXPERIMENTAL RESULTS

We performed several experiments in a coral reef environment using an aquatic robot called Mexibot, which belongs to the series of robots Aqua [7]. We previously chose a particularly prominent area of reef as a target region. The visual attention algorithm was used to make one of the robot cameras to detect a set of relevant features in the target region. We then implemented the classification method to detect and remove those features in which the motion perceptibility was higher a given threshold. Those features are removed as they have a high probability of leaving the field of view of the camera while the robot is moving to the target. The starting point of the trajectory was set at a distance of 8 meters from the target region. At this starting point, the robot camera focused directly on the desired objective, *i.e.*, the initial image was well conditioned. We used the image interaction matrix of Eq. 3 since the distance between the robot and the target is unknown during most of the trajectory.

With the purpose of computing an accurate convex hull, we always keep the same number of FOAs, to do this, the removed FOAs are replaced by those FOAs that are close to the most relevant point. Robot motion in yaw is achieved by using a simple visual servoing system. The distance between the center's coordinates of the image and the center's coordinates of the convex hull was used as a visual error to compute the control commands. We describe the visual servoing system with more detail in Section 6.

As we discussed in Section 5.1, the features used to compute the convex hull must be coplanar. To deal with that restriction, we defined a squared mask inside the initial image, with the same coordinate center. The pixels inside this mask contain the target region (detected at the beginning), and as the robot moves towards the target, its size is increased gradually, in such a way that the final size of the mask matched the size of the original image. In Fig. 4, we show an illustrative representation of the mask sampling process. By doing this, we assure that the mask always contains the target region and thus the features are more likely to be found.

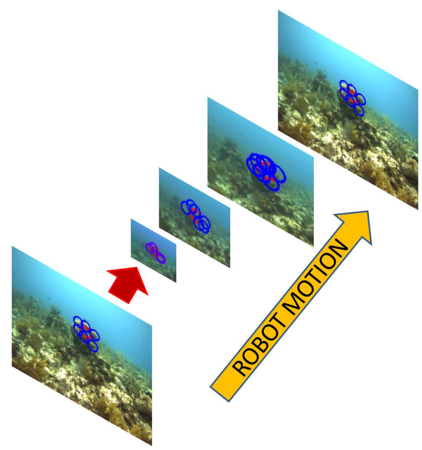

Fig. 4. Increase in the size of the subimage with respect to the original image.

As it was expected, the robot took the straight path until it reached the target region. This target region was automatically centered by the robot's camera image during almost the entire trajectory. The results obtained were divided into three segments, depending on the distance between the robot and the target region and the size of the mask. In the first segment, the robot is far from the target and the mask is $1/3$ of the input image size. In the second segment, the robot is at medium distance to the target region and the window is half of the input image. At the final segment the robot is very close to the target and the window has the same size of the original image. Figure 5 shows the mask of one frame of each segment with an additional image taken from the outside of robot. Also, it is shown in the figure the resulting vertices and the center of the convex hull in one frame of each segment. The vertices are marked with blue circles, while the center of the convex is marked with a red circle.

The segments were chosen such that at least 3 FOAs remained in the target region for a period of time.

Since in the segment 1 the subimage is small, there are no significant changes in the FOAs coordinates, so there is no significant increase in the overall perceptibility that can be

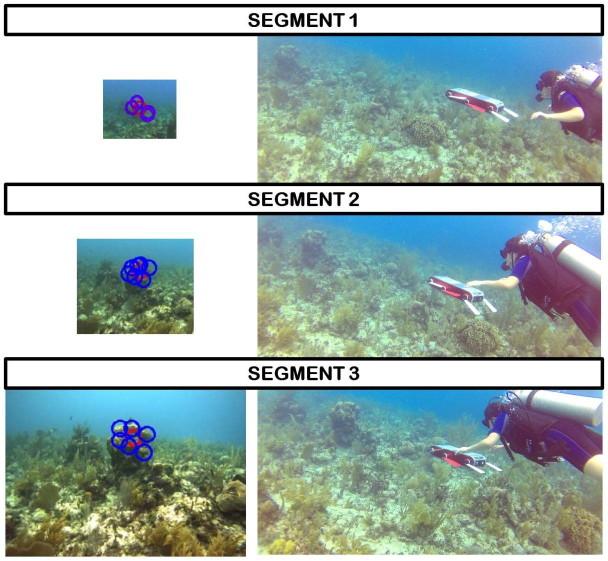

Fig. 5. Resultant vertices and center of convex hull in one frame of each segment.

related to the decrease in the distance from the robot to the target.

Figures 6-9 show the results obtained from Segment 2. Again, the mask is still small, thus, the changes in the coordinates of the FOAs are small and there is no significant increase in the overall perceptibility that can be related to the decrease in the distance from the robot to the target. The overall visibility and the removing of the vertices have similar results to those obtained in Segment 1. In Segment 3, the robot is very close to the target region, the mask and the original input image have the same size, thus, we can see that the perceptibility increases as the distance robot-target decreases.

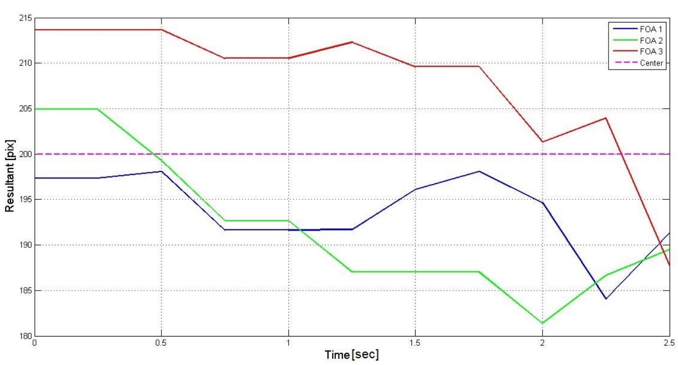

Fig. 6. Resultant of the FOAs coordinate vector in Segment 2.

VI. Conclusion and Future Work

A method to classify relevant points in aquatic images to guide the navigation of an underwater robot was presented in this document. The classification was accomplished by the use of the motion perceptibility measure. The detection and tracking of relevant points was done by using a visual attention algorithm. To drive the motion of the robot we used a PD visual servo control. The method resulted successful to

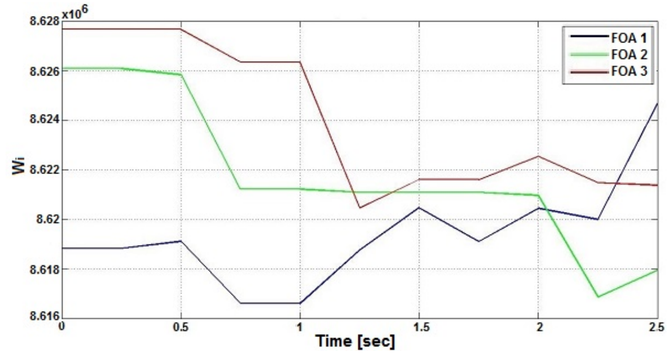

Fig. 7. Changes in the individual perceptibility of FOAs with respect to the time in Segment 2.

classify features in a navigation task with a distant target region from the starting point, while the visual servo control was sufficient to produce convenient motion of the robot. However, the method can be improved since the detected features can be positioned in the target region by means of having a notion of the distance between the robot and the features and selecting those features that are at the desired distance.

Acknowledgment

We would like to thank CONACyT for their support and project funding (CB-220540). We also thank Mar Adentro Diving, Mahahual, for their support during our sea trials.

References

[1] R. Sharma and S.Hutchinson, “Motion perceptibility and its application to active vision-based servo control,” *IEEE Transactions on Robotics and Automation*, vol. 13, pp. 607–617, 1997.

[2] A. A. Maldonado-Ramrez, L. A. Torres-Mndez, and E. A. Martnez-Garca, “Robust detection and tracking of regions of interest for autonomous underwater robotic exploration,” *COGNITIVE: The Sixth International Conference on Advanced Cognitive Technologies and Applications*, vol. IARIA, 2014.

[3] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, , and S. Susstrunk, “Slic superpixels,” EPFL, Tech. Rep., 2010.

[4] S. A. Hutchinson, H. D. G, and P. I. Corke, “A tutorial on visual servo control,” *IEEE Transactions on Robotics and Automation*, vol. 12, pp. 482–495, 1996.

[5] N. R. Gans, G. Hu, K. Nagarajan, , and W. E. Dixon, “Keeping multiple moving targets in the field of view of a mobile camera,” *Transactions on Robotics and Automation*, vol. 16, pp. 482–495, 2011.

[6] M. T. Goodrich and R. Tamassia, *Algorithm Design*. John Wiley & Sons Ltd., 2002.

[7] G. Dudek, M. Jenkin, C. Prahacs, A. Hogue, J. Sattar, P. Giguere, A. German, H. Liu, S. Saunderson, A. Ripsman, S. Simhon, L.-A. Torres, E. Milios, P. Zhang, and I. Rekletis, “A visually guided swimming robot,” *Intelligent Robots and Systems*, 2005.

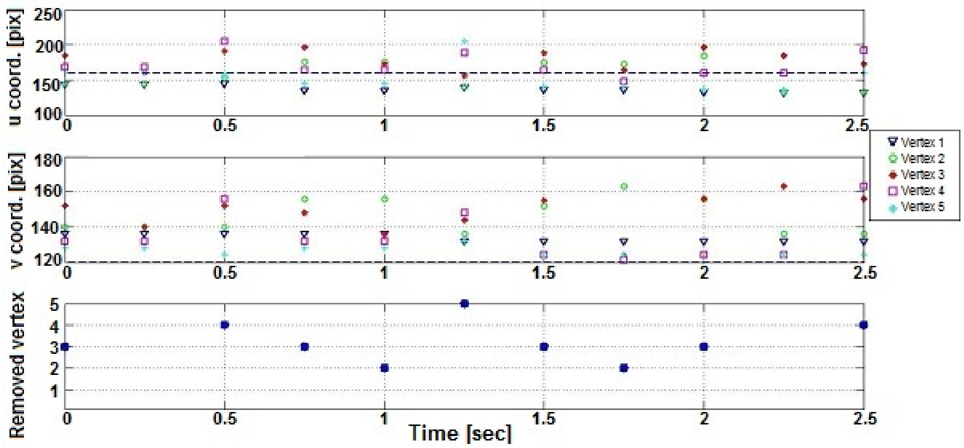

Fig. 8. Removed vertices with respect to their coordinates in Segment 2.

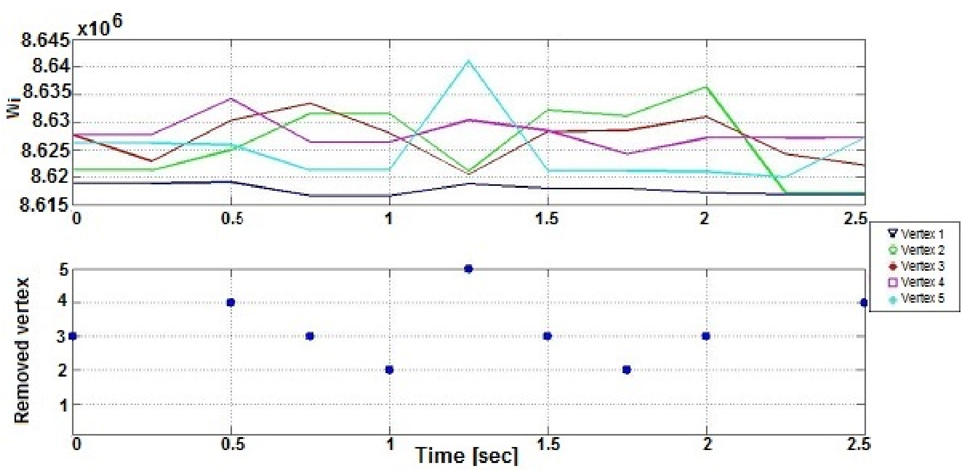

Fig. 9. Removed vertices with respect to their perceptibility in Segment 2.