Segmentation of Objects in Underwater Video

Jerome Mariette, University of Poitiers

Mentors: Duane Edgington & Danelle Cline
Summer 2005

Keywords: AVED, Segmentation, Adaptive thresholding, Canny edge detection, Homomorphic filtering.

ABSTRACT

Using Remotely Operated Vehicles (ROVs) to study marine fauna allows analyzing animal diversity, distribution and abundance in a non-invasive way and offers significant advantage over traditional techniques. Thus, the development of an Automated Visual Events Detection (AVED) system could be a powerful tool for professional annotators, because it would offer a significant decrease in the time required to analyze the numerous hours of videos recorded during ROVs dives.

The AVED system is based on a neuromorphic-selective attention algorithm, designed on the human vision system. This algorithm pre-selects objects which will be identified and flagged as interesting events if they are successfully tracked over several frames.
The system gives excellent results for the detection of objects (92% correlation with professional annotations); however, detected objects are not always detected in their entirety. This is a great obstacle for the ultimate aim of the project: the automated detection, recognition and classification of underwater animals.

Results reveal the necessity to implement the most adapted segmentation algorithm to the AVED system.
The system detects more objects and tracks them longer in benthic habitat video when it uses a background canny algorithm than the current one.
The homomorphic canny algorithm gives good results with the midwater habitat because it improves the tracking and the detection quality. The homomorphic filter is probably the reason of these results because of the non-uniform lightning correction, problem which appears in midwater videos.

INTRODUCTION

For oceanographic research, the development of ROVs was a real breakthrough because it allowed underwater fauna analysis without using conventional methodologies that may be destructive to fragile organisms (Robison, 2000). However, manual processing of the abundant Quantitative Video Transects (QVTs) obtained by ROVs dives is an extremely labor-intensive method limiting marine ecological research.

An AVED system is currently in development at Monterey Bay Aquarium Research Institute (MBARI). The AVED system is basically a system used to detect events in the video transects. The pre-selection of events by the system could improve productivity and efficiency of professional annotators by reducing the time required to analyze videos.

Once completed, AVED would be able to automatically detect, recognize and classify organisms during ROVs dives in real time (Edgington et al., 2004).

The AVED system is based on a neuromorphic-selective attention algorithm, modeled on the human vision system (Itti and Koch, 1998), proved to be robust for target detection in a variety of natural scenes (Itti and Koch, 1999; Itti and Koch, 2000). The saliency map (locations of potential events) obtained by the attention selection module, coupled to the segmentation allows the detection of salients objects. After what the tracking algorithm, by comparing frame by frame a couple of these objects parameters, is able to identify and mark interesting events in the video (Edgington et al., 2003).

Because the AVED tracking relies on a well segmented scene, that removes the background from objects of interest, the segmentation needs to be efficient. Currently, the AVED system yields a 92% correlation with professional annotations (Kerkez, 2004), however, detected objects are not always detected in their entirety, and some times two objects are grouped into one. This is a great obstacle for the automated recognition and classification of underwater fauna, because the classification is trained on single instances of animals, not groups of animals or partial animals. In this way, the segmentation improvement could be a key for the project.

MATERIALS AND METHODS

1. MATERIALS

MBARI has a team of ROVs, the *Ventana* and the *Tiburon*. The *Ventana*, deployed from the *Point Lobos*, is equipped with a Sony HDC-750 HDTV (1035i30, 1920x1035 pixels) camera and uses a DVW-A500 Digital BetaCam video tape recorder. On the ROV *Tiburon*, launched from the *Western Flyer*, the video data is acquired by a Panasonic WVE550 3-chip CCD (625i50, 752x582 pixels) camera, and records video with a DVW-A500 Digital BetaCam video tape recorder.

To digitalize Video acquired by the ROV, a Targa 3000 video capture card in a Pentium Xeon 2.0 GHz was used. Adobe Premier 6.5, running on Windows XP platform, captured the video in AVI movie files at a resolution of 720x480 pixels and 30 frames per second.
Then, videos were uncompressed and converted to Quicktime format for a perfect compatibility with the AVED system.

2. ALGORITHMS

Applying the AVED system in underwater habitats has demonstrated difficulties due to non-uniform lighting and the presence of noise from high-contrast organic debris in the water column (‘marine snow’) (Edgington et al., 2004). The followings algorithms are methods used to improve underwater images and to remove background from objects of interest.

2.1 Adaptive Thresholding

Thresholding is used to segment an image by setting all pixels whose intensity values are above a threshold to a foreground value and all the remaining pixels to a background value.

Background Substraction

In order to create a more uniform background than what is observable in ROV frames, a background subtraction is performed. For this step, a morphological opening method is used to estimate the background intensity. This method relies on a non-linear filter obtained by erosion followed by dilation with the same structured element. This method allows the elimination of small objects (here the foreground) to keep big ones (here the background).

Contrast Limited Adaptive Histogram Equalization

Moreover, to improve the contrast and objects visibility a histogram adjustment is performed. The method used is based on a Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm. CLAHE maximizes the contrast throughout an image by adaptively enhancing the contrast of each pixel relative to its local neighborhood. This process calculates histograms for small regions areas of pixels, leading to local histograms which will be equalized (Etta D. Pisano, 2000).

Adaptive Thresholding

The Thresholding algorithm used is based on Otsu's method.
In this way, pixels of a given image are represented in L gray Levels [1,2,...,L]. The number of pixels at level i is donated by ni, and the total number of pixels by $N=n1+n2+...+nL$. Then dichotomization of pixels into two classes C_0 and C_1 denotes respectively pixels with [1...k] and [k+1...L]. The method determines the threshold by determining the grey level that maximizes the between-class variance of the gray level histogram.

Then, in order to connect objects, a dilatation operator is used. The structured element used depends on the kind of video processed. For midwater video a line shaped structured element is used, and for benthic video, a circle shaped structured element is preferred.

Figure 1: Pre-processing and processing steps to process Adaptive Thresholding algorithm.

2.2 Canny Edge Detection

Edges in images are areas with strong intensity contrast: a jump in intensity from one pixel to the next. Edge detecting an image significantly reduces the amount of data and filters out useless information, while preserving the important structural properties in an image. In this particular method of segmentation, areas delimited by boundaries represent the segmented object.
In order to improve the Canny detector sensitivity to edges and in that way to improve the segmentation, a pre-treatment on images is required. Two methods will be tested: Background Subtraction and a Homomorphic filtering.

Background Subtraction

This method is based on a morphological opening in the same way as described in Adaptive Thresholding *(section 2.1.)*.

Homomorphic Filtering

Scenes illuminated by ROVs artificial lightning, during their dives, often suffer from non-uniform lightning. One method that attempts to reduce this effect relies on Homomorphic filtering (J.R. Parker, 1997). This process attempts to separately analyze illumination and reflectance information. Enhancing the reflectance properties of an image can increase contrast in areas of low and high illumination levels. This, in particular prevents the appearance of saturation areas that occur with the application of gain.

Crucial for the success of the homomorphic approach is the selection of an appropriate frequency-domain filter function in order to modify the illumination and reflectance components of an image differently. For this approach, the Butterworth type high pass equations is known to be far superior to other frequency domain filter functions (Adelmann, 1998), including Gaussian equations, making the Butterworth high pass suitable for use with the homomorphic filtering approach.

An image is given by :

$f(x,y) = i(x,y)r(x,y)$, with $i(x,y)$ the illumination and $r(x,y)$ the reflectance information.

Homomorphic method uses a logarithm function to separate reflectance and illumination information in an image.

$$g(x,y) = ln(i(x,y).r(x,y))$$

Using g(x,y), the Fourier transform can be applied to an additive signal rather than a multiplicative signal in the following way :

$$g(x,y) = ln(i(x,y)) + ln(r(x,y))$$
$$G(u,v) = I(u,v) + R(u,v)$$

Applications of the Butterworth filter to enhance changes in reflectance and suppress changes in illumination.

The High-pass Butterworth filter is giving by:

$$H(u,v) = \frac{1}{1 + [D0/D(u,v)]^2}, \text{ with } D(u,v) = \sqrt{u^2 + v^2}$$

Then,

$$\overline{G}(u,v) = H(u,v)I(u,v) + H(u,v)R(u,v)$$

Compute the inverse Fourier transform gives:

$$\overline{g}(x,y) = \overline{i}(x,y) + \overline{r}(x,y)$$

To reverse the logarithm used at the beginning of the process, compute an exponential function:

$$\overline{k}(x,y) = exp(\overline{i}(x,y)) + exp(\overline{r}(x,y)).$$

Where k is the image after enhancement.

Figure 2: Homomorphic filtering algorithm.

Canny Edge Detection

The Canny operator works in a multi-stage process. First, the image is smoothed by Gaussian convolution, then a simple 2-D first derivative operator is applied to the smoothed image to highlight areas of the image with high first spatial derivatives. Edges lead to the creation of ridges in the gradient magnitude image. The algorithm then tracks along the top of the ridges and sets to zero all pixels that are not actually on the ridge top. The tracking process uses two thresholds in order to detect the both directions of the edge.

Finally, in order to obtain segmented objects, a flood-fill algorithm is performed.

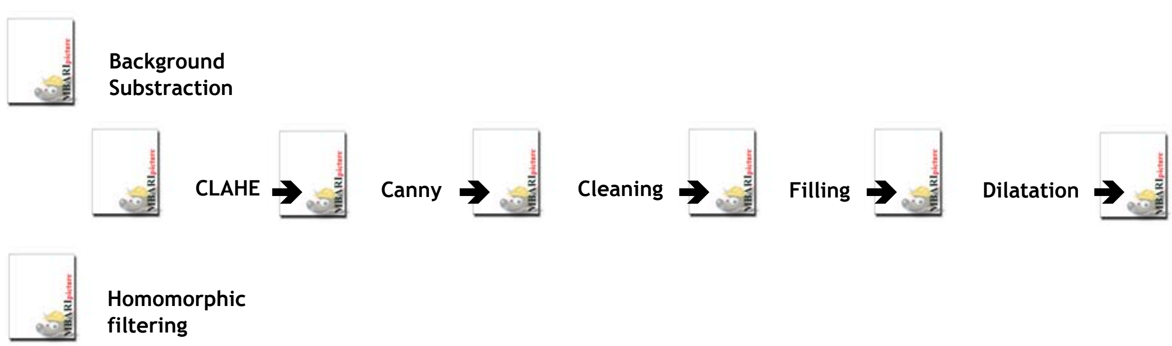

Figure 3: Pre-processing and processing steps to process Adaptive Thresholding algorithm.

RESULTS

The visual characteristics of the ocean benthic habitat differ greatly from the midwater environment, thus, results obtained for this two habitats will be analyzed separately.
In order to know if the segmentation algorithms improve the current one, we will focus on one hand on the number of objects detected, and on the other hand on the system ability to track objects. This one could be studied counting:

_the number of tracking frames, higher is the number, better is the tracking.
_the number of events needed for the detection of one object, a good tracking will use one event for one object.

1. BENTHIC results

In a processed 874 frames segment of video, the total number of objects detected is calculated for each algorithm.
When the AVED system using the current segmentation detects 19 of the 21 objects present in the video, it detects 20 objects with an adaptive thresholding and a background canny, and only 17 with a homomorphic canny segmentation (figure 1).

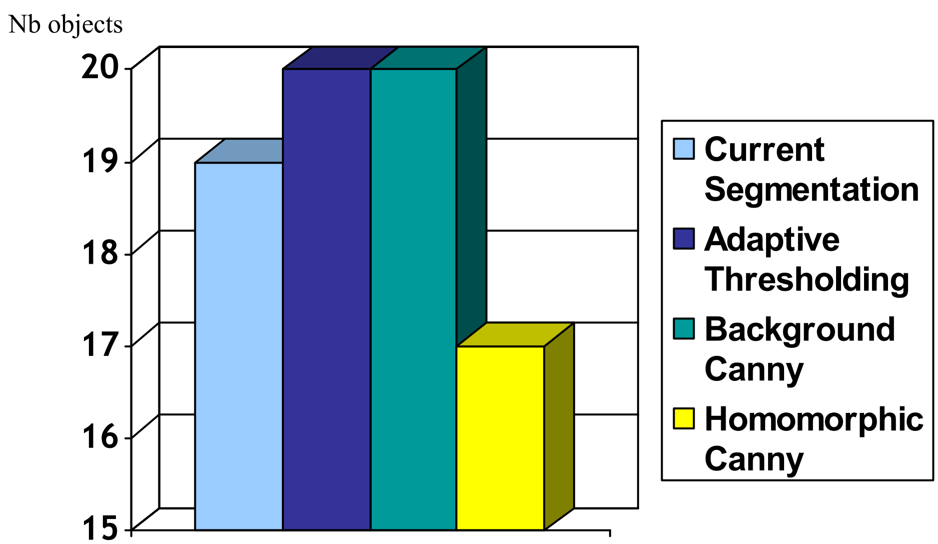

Figure 1: Number of objects detected by the system with each algorithm tested on a processed 874 frames segment of video.

Moreover, in order to study the tracking efficiency of the system, the number of frames of objects tracking is determined.
The AVED system tracks objects in 1036 frames for the 19 objects detected with the current segmentation. The adaptive thresholding leads to a 996 frames tracking for the 20 objects detected, where a background canny allows a track for 1153 frames for the same number of objects. For the homomorphic canny, the number of frames doesn't go upper than 929 frames for the only 17 objects detected (figure 2).

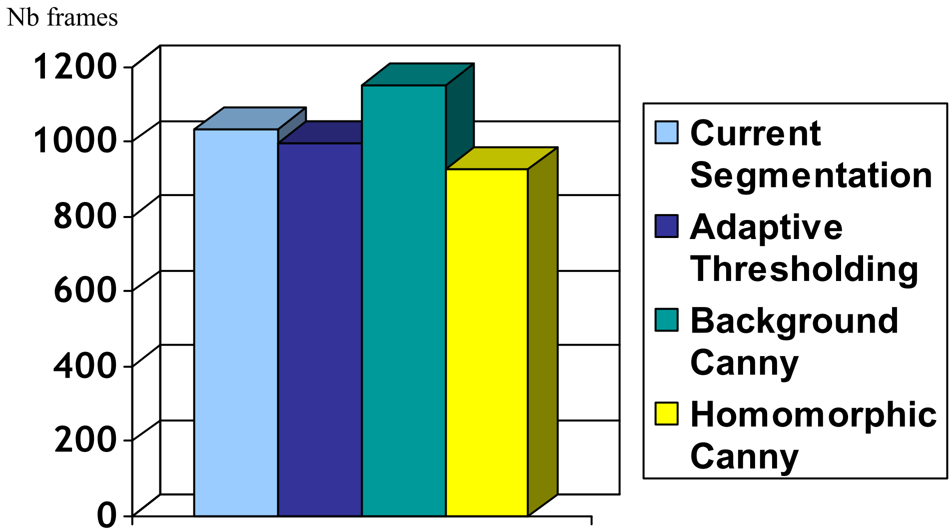

Figure 2: Number of frames of objects tracking for each algorithm tested on a 874 frames segment of video.

Finally, the video shows a grouping of two objects in one event (figure 3) or a non-detection of this objects when objects are superposed; problem appeared with all algorithms tested.

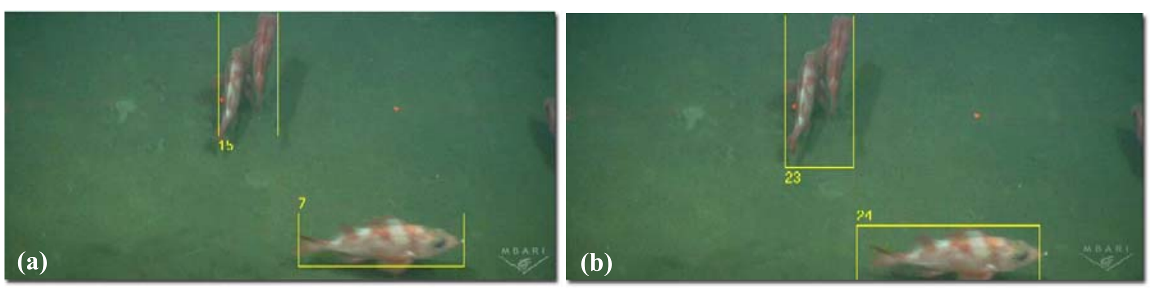

Figure 3: Frame 110 sample of objects superposition. (a) AVED with current segmentation, (b) AVED with background canny algorithm.

2. MIDWATER results

For midwater results, two kinds of video are tested: video recorded in a transect way (the camera is in motion) and in a non-transect way (the camera is focused on an object).

2.1. NON-TRANSECT results

For this results, just one object appear on the 1001 frames segment of video, thus, the results shown are results for the object considered.
The AVED system tracks the object during all the 1001 frames of the video with each algorithm tested (figure 4).

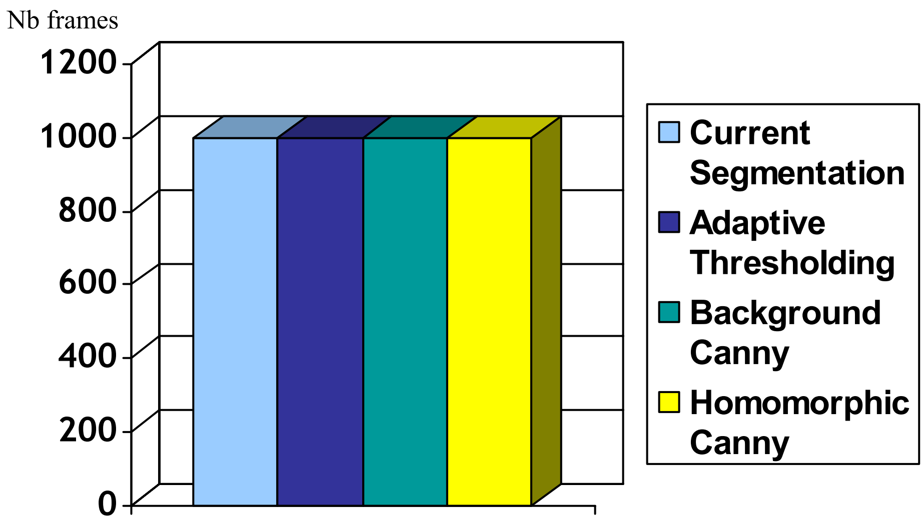

Figure 4: Number of frames of objects tracking for each algorithm tested on a 1001 frames segment of video.

However, in this case, the number of frames is not enough; indeed, the number of events needed for the detection of one object is a good clue of the efficiency of the tracking system.
Where the AVED system needs three events to detect the considered object with the current segmentation, it needs only one with a canny segmentation (for the both pre-treatment), and two with an adaptive thresholding (figure 5).

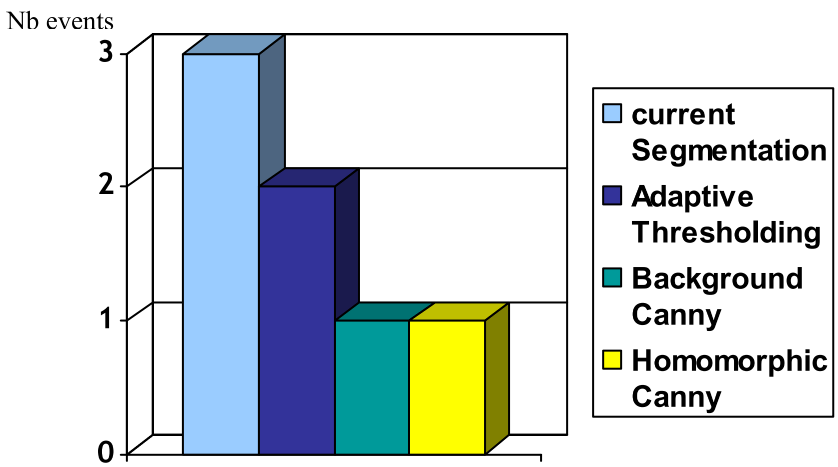

Figure 5: Number of events needed for the object detection.

Finally, the visualization of the video shows a more stable detection (figure 6) with the homomorphic canny algorithm during all the video, phenomena more observable with a lightning change. Indeed, the boundaries box jumps less than with the others algorithms tested.

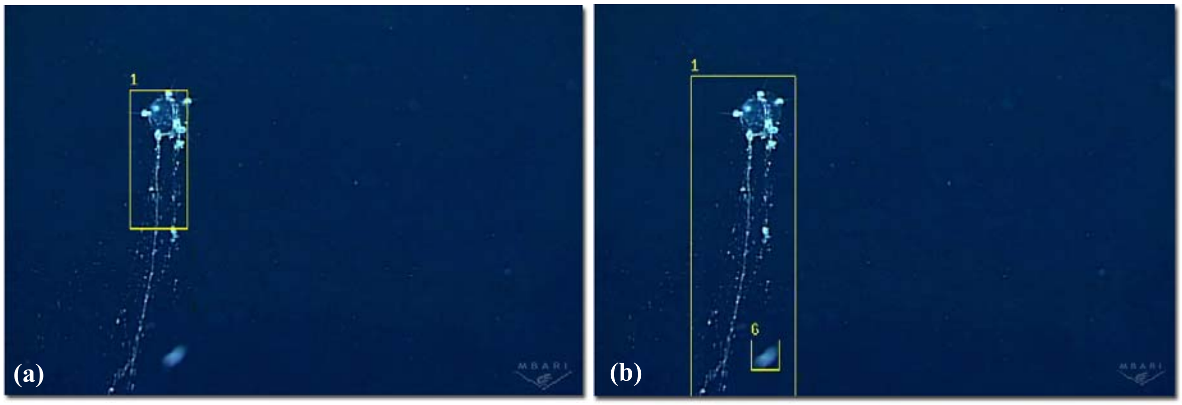

Figure 6: Frame 200 sample of the difference of detection stability. (a) AVED with current segmentation, (b) AVED with homomorphic canny algorithm.

2.1. TRANSECT results

For this result, the adaptive thresholding algorithm was not tested, this one gave bad segmented images.

In a processed 310 frames segment of video, the AVED system detects all objects present in the video. But where the system using the current segmentation tracks objects in 160 frames, it tracks objects for 210 frames with a background canny segmentation and 215 frames with a homomorphic canny.

In this video, few objects appear, thus, results shown focus on one particular object representative of almost objects in the video.

The AVED system tracks this object for 74 frames with the current segmentation, 94 with the background canny, and 110 frames with the homomorphic canny (figure 7).

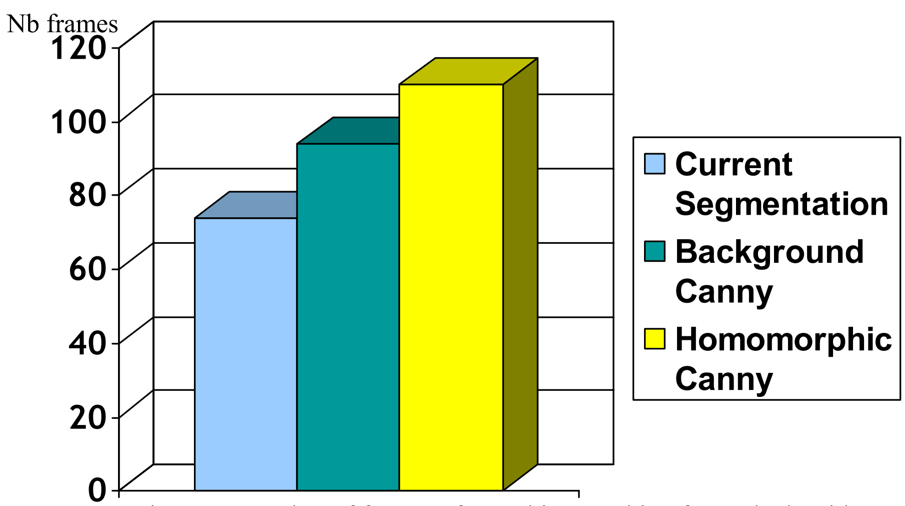

Figure 7: Number of frames of one object tracking for each algorithm tested on a 310 frames segment of video.

This object is detected by 6 different events when the system uses a current segmentation, where only 2 events are used with the both canny algorithms (figure 8).

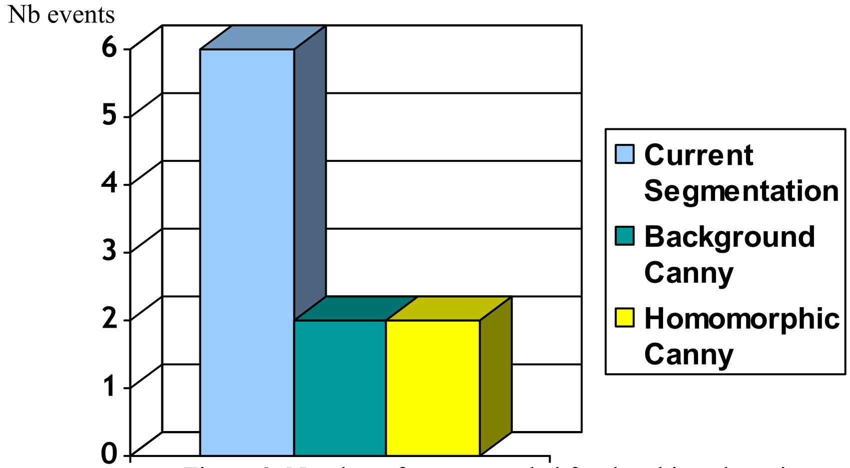

Figure 8: Number of events needed for the object detection.

The video visualization shows a better stability of the detection with the homomorphic canny algorithm than with the current segmentation (figure 9).

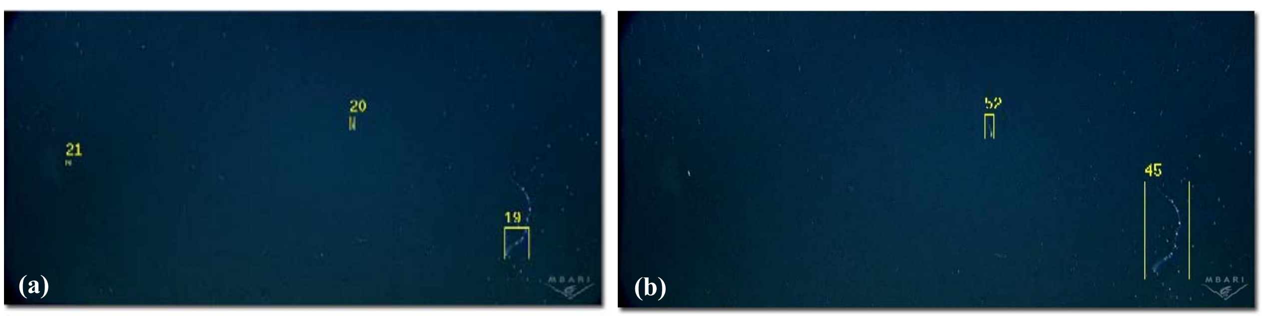

Figure 9: Frame 50 sample of the difference of detection stability. (a) AVED with current segmentation, (b) AVED with homomorphic canny algorithm.

DISCUSSION

The comparison of results issued from a system like the AVED system is a pretty hard job, because the appreciation of detection and tracking quality is subjective. In this way, parameters used to analyse results are probably not enough to conclude on the quality improvement of the system. Anyway, this work shows a real difference either for the detection or the tracking of the AVED system when this one uses different segmentation algorithms. This first result reveals the interest of this work, and the necessity to implement the most adapted segmentation algorithm into the AVED system.

Results obtained for benthic habitat show a little improvement of the detection when this one uses a canny segmentation (20 events against 19 with the current segmentation). This same algorithm allows a longer tracking when a background subtraction is preformed, indeed, the AVED system tracks objects in 1036 frames with the current segmentation, where a background canny allows a track for 1153 frames. However, no algorithms tested solved the problem of objects superposition.
The system detects more objects and tracks them longer in benthic habitat video when it uses a background canny algorithm than with others algorithms tried, even the current one.

For the midwater habitat, few objects appear on videos tested, thus, the number of objects detected is not a pertinent parameter. However, the tracking quality could be analyzed in two ways: length of tracking and number of events needed to track one object.
Thus, the homomorphic canny algorithm gives the best results (for both transect and non-transect recording). This one allows a better tracking of objects by the system increasing the number of frames (110 frames against 74 with the current segmentation) and decreasing the number of events needed to detect one object (2 events against 6 needed with the current segmentation).
Moreover, the visualisation of videos reveals a more stable detection with the homomorphic canny algorithm than with the current one, because objects are detected in their entirety. This phenomena is more observable with the lightning changes appeared in the non-transect video.
However, algorithms tried did not solve objects superposition in videos tested.
The homomorphic canny algorithm gives good results with the midwater habitat because it improves the tracking and the detection quality. The homomorphic filter is probably the reason of these results because of the non-uniform lightning correction, problem which appears in midwater videos.

The homomorphic filtering gives good results for non-uniform lightning and a good response to lightning changes appearing in midwater videos. A segmentation based on a Canny Edge Detection allows an improvement of the performances of the tracking system, and the quality of the detection.

Algorithms tried improve the current one; in this way objectives are reached, but results were tried only on few videos. To ensure adequate testing of the processing system longer videos with a greater number of species and application testing under a variety of conditions, are required.

CONCLUSIONS/RECOMMENDATIONS

Nowadays, there are several segmentation algorithms, and they become uncountable if alternatives ones are considered. Finding the most adapted algorithm is a pretty hard problem because there are no ways to evaluate segmentation results. Thus, for the AVED project, the only one way was to integrate the algorithm in the project.

However, no algorithms implemented succeed to solve the superposition problem, for what binaries segmentations can't alone be the solution.
Moreover, before to process a Canny Edge Detection, a CLAHE is performed. But this algorithm is probably not the best one; indeed, a Partial Differential Equations algorithm (PDE) allows enhancing edges, what could be a powerful tool before Edge Detection operation.
Colour segmentation were not tried, this kind of segmentation could be helpful for objects with strong colour information. Finally, active contour algorithm was tried but was slow to compute; this algorithm has to be more considered because give excellent results in several overviews dealing about medical image processing [C. Xu and J. L. Prince (2000), "Global Optimality of Gradient Vector Flow", *Proc. of 34th Annual Conference on Information Sciences and Systems (CISS'00)*, Princeton University.]

The AVED system needs a good segmented scene to detect objects, but this is not the only aim of the improvement work. Indeed, the recognition module relies on the segmentation because it needs a single and complete instance of objects.
The AVED system yields a 92% correlation with professional annotations (Kerkez, 2004) what is a strong foundation for the project. Once operational, AVED will be considered as a breakthrough by the world of marine ecology. It will allow to study in real time the underwater fauna and will give to oceanographic research several tools for organisms' assessments.

ACKNOWLEDGEMENTS

I wish to thank Duane Edgington for allowing me to work on such an exciting project; Danelle Cline for her guidance, her encouragement and her patience during the entire summer. I'd also like to thank Dorothy Oliver for all her help this summer.

Thanks also to George Matsumoto for making this internship a unique and unforgettable experiment (I hope the survival course is a requirement for 2006). Thanks to the fifteen others intern for giving me the possibility to improve my mediocre English. Good luck to all 2005 interns and don't forget to visit me when you come to France.

References:

Adelmann HG. (1998). Butterworth equations for homomorphic filtering of images.

Edgington, D.R., D. Walther, K.A. Salamy, M. Risi, R.E. Sherlock, and C.Koch (2003). Automated event detection in underwater video. *Proc. MTS/IEEE Oceans Conference*, San Diego, CA.

Edgington, D.R. et al (2004). MBARI internal project proposal, *Moss Landing, California*.

Etta D. Pisano, et al (2000). Image Processing Algorithms for Digital Mammography : A Pictorial Essay. *Imaging & Therapeutic Technology*.

Itti, L., and C. Koch (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, **40(10-12)**: 1489-1506.

Itti, L., C. Koch, and E. Nieber (1998). A model for saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20(11):** 1254-1259.

Itti, L., and C. Koch (1999). Target detection using saliency-based attention. *In Proceedings RTO/SCI-12 Workshop on Search and Target Acquisition (NATO Unclassified)*, Utrecht, The Netherlands.

Kerkez. (2004). Automated event detection in video of Benthic Habitat, *MBARI internal project proposal*, Moss Landing, California.

Parker. (1997). The homomorphic filter – Illumination. *Algorithms for Image Processing and Computer Vision*, **6:** 241-246.

Robison, B.H. (2000). The co evolution of undersea vehicles and deep-sea research. *Marine Technology Society Journal*, **33**: 69-73.