Comparison of Background Subtraction Methods for a Multimedia Application

F. El Baf, T. Bouwmans, B. Vachon
LMA Laboratory
University of La Rochelle
Avenue M. Crepeau, La Rochelle, 17000, France
Phone: (0033) 0546458322 - E-mail: felbaf@univ-lr.fr

Keywords: Multimedia Video Processing, Background Subtraction

Abstract - This article presents the Aqu@theque application and a comparison of background subtraction methods to improve the performance of this application. The Aqu@theque application consists in elaborating an information system dedicated to aquariums in an interactive learning area. In particular, this article presents our comparison of different background subtraction methods to detect fish in video sequences and the improvement for the Aqu@theque application.

1. INTRODUCTION

This system allows a visitor of an aquarium to select on an interactive interface fishes that are filmed on line by a remote video camera. This interface is a touch screen divided into two parts. The first one presents the list of fishes that are in the tank and is useful all the time, independently of the video produced by the camera. The filmed scene is visualized in the remaining part of the screen. The computer can supply information about fishes selected by the user with his finger. A fish is then automatically identified and some educational information about the fish is put on the screen. The user can also collect each identified fish whose virtual representation is shown on another screen. This second screen is a virtual tank reproducing the natural environment where the fish lives in presence of fishes from different species that live close to it. The behavior of every fish in the virtual tank is modeled. The project is based on two original elements: the automatic detection and recognition of fish species in a remote tank of an aquarium and the behavioral modeling of virtual fishes by multi-agents methods. First, we present the principle and the different steps met in this multimedia application called Aqu@theque. In a second step, we will present our study of statistical background subtraction algorithms in the context of this multimedia application.

2. AN INTERACTIVE LEARNING SPACE: THE AQU@THEQUE APPLICATION

We present the principle of this interactive learning space. More information can be found in [1] [2] [3].

The system integrates three different functional blocks

1) The interactive part, with an interface drawn with Macromedia Director, and using a touch screen.
2) The recognition system.
3) The 3D real-time engine, coupled with a mechanism of behavior modeling.

Video stream, issued from the camera which films in live the fish tank, is integrated into the interface (figure 2) leaving the possibility to the user to act on the video: this functionality was written in Lingo language (the language script of Director).

So the visitor can:

- Act (move, zoom ...) in live, on the video stream.
- Select a fish on the touch screen
- Appoint it by its name in the menu
- Create his virtual tank.
2.1 Fish Identification

When a fish is selected on the touch screen, the recognition system is launched. To allow a real-time and automatic fish recognition, our system consists of:
- A segmentation step allowing to extract the main regions corresponding to fishes in video sequences. This step uses a background subtraction method.
- A features extraction step based on segmentation’s results.
- A classification step of fishes with respect to the different species those are present in the tank.
More information can be found in [1] [2].

2.2 Information about identified fishes

If the chosen fish is well identified, the user can choose by mean of the interface to reach the following information:
1) Educational information: multimedia information about the selected fish is proposed to the user. This information is given in the form of indexed pedagogic cards (Fig 2), pictures, real videos on the way of life, the origin and the environment of animals as well as their protection. The educational information is accessible by menu, conceived with the same hierarchy for all the species (hierarchical menu from subjects as the biology, the species, its protection, the behavior, the environment ...).

2) 3D representation: the chosen fish can also be represented in 3D under all the angles (with zoom, rotations ...), thanks to computer generated images and the technology cult3D. The user can then manipulate it in order to observe all its details (Fig 3).

2.3 Virtual Tank

The user can build up his virtual fish tank. Fig 4 shows porkfish and shark in the virtual tank. The virtual tank manages the behavior of each fish. More details can be found in [3].

3. COMPARISON OF BACKGROUND SUBTRACTION METHODS

The automatic detection of fish species is made using background subtraction. The principle consists in the difference between the current frame and a background computed with a chosen model. The method proposed in the previous work [1] [2] represents the background by the median model which fails in some situation met in aquatic scene. So, we have compare and evaluate three algorithms which are more sophisticated using statistical representation. These algorithms are the following: Simple Gaussian [4], Mixture of Gaussians [5] and Kernel Density Estimation [6].

3.1 Single Gaussian (SG)

Wren [4] developed an algorithm to model each background pixel according to normal distribution characterized by its mean value $\mu$ and its standard deviation $\sigma$ in the YUV color space. This model requires a number $t$ of frames to compute $\mu$ and $\sigma$ in each color component:
\[ \mu(x, y, t) = \sum_{i=1}^{t} \frac{p(x, y, i)}{t} \]  
\[ \sigma(x, y, t) = \sqrt{\sum_{i=1}^{t} \frac{p^2(x, y, i)}{t} - \mu^2(x, y, t)} \]  

where \( p(x, y, i) \) is the current intensity value of the pixel at the position \((x, y)\) at time \(i\). After that, a pixel is considered as belonging to a foreground object according to the rule:

\[ |\mu(x, y, t) - p(x, y, t)| > c \sigma(x, y, t) \]  

where \(c\) is a certain constant.

This method adapts to indoor scene with little gradual illumination changes, but it fails in several cases: sudden illumination changes and moving background objects like trees, flags or algae.

### 3.2 Mixture of Gaussians (MOG)

It was proposed in [5] that the colors of each background pixel are modeled by a mixture of \(K\) Gaussians, which is given by the following formula in the multidimensional case:

\[ \Pr(x_i) = \sum_{j=1}^{K} \omega_j \frac{1}{(2\pi)^{d/2}|\Sigma_j|^{1/2}} \exp\left(-\frac{1}{2}(x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j)\right) \]  

where \(\omega_j\) is a weight associated to the \(j\)th Gaussian, with mean \(\mu_j\) and standard deviation \(\Sigma_j\), according to the time proportion of colors appearance.

A pixel matches a Gaussian if:

\[ \sqrt{(x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j)} < \delta \]  

To handle multimodality in the background, Stauffer used as criterion the ratio \(r_j = \omega_j / \sqrt{|\Sigma_j|}\), which supposes that a background pixel corresponds to a high weight with a weak variance due to the fact that the background is more present than moving objects and that its value is practically constant. The foreground detection is made by ordering the \(K\) distributions by their \(r_j\) and the first \(B\) Gaussians which exceed certain threshold \(T\) are retained for the background:

\[ \arg \min_{\omega_j} \left\{ \frac{\sum_{j=1}^{b} \omega_j}{\sum_{j=1}^{k} \omega_j} > T \right\} \]  

### 3.3 Kernel Density Estimation (KDE)

Elgammal [6] estimated the probability density function for each pixel colour using the kernel estimator \(K\) for \(N\) recent sample of intensity values as:

\[ \Pr(x_i) = \frac{1}{n} \sum_{i=1}^{N} K(x_i - x_i) \]  

The foreground detection is done according the following rule:

If \(\Pr(x_i) < Th\), the pixel is belonging to the foreground else is belonging to the background. Like the Mixture of Gaussians, the Kernel Density Estimation is also adapted to handle the multimodality of the background but, unlike the Mixture of Gaussians, it doesn't need to estimate the parameters of the Gaussians.

### 4. QUALITATIVE EVALUATION

Performance evaluation contains several senses. It can be required in term of time consuming and memory consuming or in terms of how well the algorithm detects the targets with less false alarms. To evaluate performance in the first sense, the time and the memory used can be measured easily by instruction in line code. A first qualitative comparison is showed in Table.1

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>MOG</th>
<th>KDE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed</td>
<td>Fast</td>
<td>Interm.</td>
<td>Slow</td>
</tr>
<tr>
<td>Memory</td>
<td>Interm.</td>
<td>Interm.</td>
<td>High</td>
</tr>
</tbody>
</table>

Table 1: Qualitative Comparison

### 5. QUANTITATIVE EVALUATION

Using Aqu@theque sequence, the performance evaluation can be done quantitatively by the sense of ROC (Receiver Operating Characteristic) Analysis, or by a measurement presented by Li [7] from the comparison of the segmentation results with the “ground truths”. Roc evaluation is centralized around the tradeoff of miss detection rate (FNR) and false alarm rate (FPR), where the similarity measure of Li integrates the false positive and negative errors in one measure. Let \(A\) be a detected region
and $B$ be the corresponding ground truth, the similarity between $A$ and $B$ is defined by Li as:

$$S(A,B) = \frac{A \cap B}{A \cup B}$$

$S(A,B)$ lies between 0 and 1. If $A$ and $B$ are the same, $S(A,B)$ approaches 1, otherwise 0 if $A$ and $B$ have the least similarity.

The SG, the MOG and the KDE were implemented using OpenCV. We present the comparison of the three algorithms in Table 2. The MOG algorithm used is the one proposed by KaewTraKulPong [8] which is an improved version of Stauffer's algorithm [5].

<table>
<thead>
<tr>
<th></th>
<th>SG (5 gaussians)</th>
<th>MOG (5 gaussians)</th>
<th>MOG (3 gaussians)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FPR</td>
<td>0.0005</td>
<td>0.0068</td>
<td>0.007</td>
</tr>
<tr>
<td>FNR</td>
<td>0.6419</td>
<td>0.4001</td>
<td>0.3683</td>
</tr>
<tr>
<td>Detection Rate</td>
<td>0.3580</td>
<td>0.5998</td>
<td>0.6317</td>
</tr>
<tr>
<td>$S(A,B)$</td>
<td>0.3564</td>
<td>0.5710</td>
<td>0.6004</td>
</tr>
</tbody>
</table>

Table 2: Quantitative evaluations by using the ROC analysis and $S(A,B)$ measure of a sequence Aquarium.

The sequence image is a sequence containing 2600 images of size 640*480 in RGB. The results obtained are shown in the Fig 5.

Fig 5: a) Image 201, b) Groundthruth, c) SG Foreground Mask, d) MOG Foreground Mask and e) KDE Foreground Mask

6. CONCLUSION

The Aqu@theque project enhances the visit of an aquarium by providing an interactive learning space where educational information is available. A user creates dynamically a virtual aquarium according to the selected fishes on the interface. Aqu@theque brings to the visitor an additional dimension by allowing him to become an actor instead of staying a passive spectator.

In this paper, we addressed particularly the problem of background subtraction which is one of the key steps in the system. A first qualitative evaluation showed that the MOG is more efficient. In the future, we will make a quantitative evaluation using ROC Curves and a sequence test of the VSSN 2005 [7]. Qualitative tests showed that the use of the MOG in the background subtraction step enhance the percentage of recognition improving the performance of our interactive learning space.

Furthermore, the principle of Aqu@theque can be used in E-Learning environments. The background subtraction method must be choice following the critical situations (illumination changes) met in the sequence used in the application.

REFERENCES


