COMPARISON OF BACKGROUND SUBTRACTION METHODS FOR A MULTIMEDIA LEARNING SPACE

F. El Baf, T. Bouwmans and B. Vachon
Laboratory LMA, University of La Rochelle, Avenue M. Crépeau, La Rochelle, France

Keywords: Multimedia Application, Video Processing, Background Subtraction.

Abstract: This article presents, at a first time, a multimedia application called Aqu@theque. This project consists in elaborating a multimedia system dedicated to aquariums which gives ludo-pedagogical information in an interactive learning area. The reliability of this application depends of the segmentation and recognition steps. Then, we focus on the segmentation step using the background subtraction principle. Our motivation is to compare different background subtraction methods used to detect fishes in video sequences and to improve the performance of this application. In this context, we present a new classification of the critical situations which occurred in videos and disturbed the assumptions made in background subtraction methods. This classification can be used in any application using background subtraction like video surveillance, motion capture or video games.

1 INTRODUCTION

The proposed system allows a visitor of an aquarium to select on an interactive interface fishes that are filmed on line by a remote video camera (Figure 1). This interface is a touch screen divided into two parts. The first one shows the list of fishes present 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 it is put on the screen. The user can also select 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 it preys and predators. 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 a new classification of the critical situations met in videos and our study of statistical background subtraction algorithms in the context of this multimedia application.

Figure 1: Aqu@theque interactive learning space.

2 AN INTERACTIVE LEARNING SPACE: THE AQU@THEQUE APPLICATION

In this section, we present the principle of this interactive learning space. More information can be found in our previous papers (Semani, SPR 2002), (Semani, IVRCIA 2002), (Desfieux, IVRCIA 2002). The system integrates three different functional blocks:

- The interactive part, with an interface drawn with *Macromedia Director*, and using a touch screen;
- The recognition system;

El Baf F., Bouwmans T. and Vachon B. (2007).
COMPARISON OF BACKGROUND SUBTRACTION METHODS FOR A MULTIMEDIA LEARNING SPACE.
In *Proceedings of the Second International Conference on Signal Processing and Multimedia Applications*, pages 149-154
DOI: 10.5220/0002134201490154

- The 3D real-time engine, coupled with a mechanism of behavior modeling;

2.1 Fish Identification

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:

- Make action (move, zoom ...) in real time, on the video stream;
- Select a fish on the touch screen;
- Select it by its name in the menu;
- Create his virtual tank;

This functionality was written in *Lingo* language (the language script of *Director*).

2.2 Fish Recognition

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 the extraction of 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 which are present in the tank;

More information can be found in our previous papers (Semani, SPR 2002), (Semani, IVRCIA 2002).

2.3 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:

- Educational information: multimedia information about the selected fish is proposed to the user. This information is given in the form of indexed pedagogic cards (Figure 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...);
- 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 (Figure 3).

Figure 2: Educational information.

Figure 3: The virtual Porkfish.

2.4 Virtual Tank

The user can build up his virtual fish tank which manages the behavior of each fish. More details can be found in (Desfieux, IVRCIA 2002).

2.5 Discussion

The fish identification is the main step of the Aqu@theque application because it determines its reliability. Furthermore, this step must be in real time as much as possible because the user must have to wait as less as possible to know the specie's name of the selected fish. In the following, we present our study to perform the segmentation step which uses the background subtraction principle.

3 COMPARISON OF BACKGROUND SUBTRACTION METHOD

The automatic detection of fish species is made using background subtraction technique based on the difference between the current frame and a background reference frame. In the previous work (Semani, SPR 2002) (Semani, IVRCIA 2002), the background was modeled by a median but this model fails in some critical situations met in aquatic scene. So it should be necessary to study these situations to handle their effects. But before, we

propose a general classification of the critical situations which may appear in any video and then we focus on the critical situations occurred in the case of aquatic scenes. After that, we compare and evaluate three background subtraction algorithms which are more sophisticated than the median. This algorithm are the following: Single Gaussian (Wren, 1997), Mixture of Gaussians (Stauffer, 1999) and Kernel Density Estimation (Elgammal, 2000).

3.1 Assumptions and Critical Situations

There are three main assumptions which assure a good functioning of the background subtraction methods: the camera is fixed, the illumination is constant and the background is static, i.e pixels have a unimodal distribution and no background objects are moved or inserted in the scene. In these ideal conditions, background subtraction gives good results. In practice, some critical situations may appear and perturb this process. In (Toyama, 1999), the author identifies 10 critical situations in the field of the video surveillance. To be more general, we explain and improve this classification in a most possible abstract way and gives examples naturally in the fields of video surveillance but in more in the fields of motion capture and multimedia applications. In final, we identify 13 critical situations (CS) which are the following:

- CS1 - Noise image due to a poor quality image source;
- CS2-1 - Camera jitter;
  CS2-2 - Camera automatic adjustments: Many modern cameras have auto focus, automatic gain control, automatic white balance and auto brightness control;
- CS3 - Gradual illumination changes: It can be illustrated by illumination change during a day in an outdoor scene;
- CS4 - Sudden illumination changes: It can be illustrated by a light switch in an indoor scene;
- CS5 - Bootstrapping: During the training period, the background is not available in some environments;
- CS6 - Camouflage: A foreground object's pixel characteristics may be subsumed by the modeled background;
- CS7 - Foreground aperture: When a moved object has uniform colored regions, changes inside these regions may not be detected. Thus, the entire object may not appear as foreground;
- CS8 - Moved background objects: A background object can be moved. These objects should not be considered part of the foreground;
- CS9 - Inserted background objects: A new background object can be inserted. These objects should not be considered part of the foreground;
- CS10 - Movement in the background: Backgrounds can vacillate and this requires models which can represent disjoint sets of pixel values. For example:
  - Waving trees, moving bushes in outdoor scene for video surveillance application (Toyama, 1999).
  - Ocean waves in outdoor scene for video marine surveillance application (Culibrk, 2007).
  - Moving algae in aquatic scene for Aqu@theque (Semani, 2002).

  In this case, the background is multimodal and dynamic (Stauffer, 1999) (Elgammal, 2000).
- CS11 - Beginning moving foreground object: When an object initially in the background moves, both it and the newly revealed parts of the background called ghost are detected. For example:
  - A person or a car begins to move in a video surveillance application (Toyama, 1999);
  - A fish begins to move in the Aqu@theque application (Semani, 2002);
- CS12 - Sleeping foreground object: A foreground object that becomes motionless cannot be distinguished from a background object and then it will be incorporated in the background;
  - A person or a car stops to move for video surveillance application (Toyama, 1999);
  - A fish stops to move for Aqu@theque (Semani, 2002).

  This critical situation depends of the context. Indeed, in some applications, motionless foreground object must be incorporated (Huerta, 2006) and in others no (Semani, 2002).
- CS 13 – Shadows: Shadows can be detected as foreground. A complete study on shadow detection can be found in (Cucchiara, 2003);

In aquatic scenes, all these situations occurred. For example, illumination changes (CS3, CS4) in aquatic scenes are owed to the ambient light, the spotlights which light the tank from the inside and from the outside, the movement of the water due to fish and the continuous renewal of the water.

These illumination changes can be local or global following their origin. Furthermore, the constitution of the aquarium (rocks, algae) and the texture of fishes amplify the consequences of the brilliant variations.

So, the statistical background subtraction methods tested in the following section must deal specifically with these critical situations.

3.2 Statistical Background Subtraction Algorithms

3.2.1 Single Gaussian (SG)

Wren (Wren, 1997) developed an algorithm to model each background pixel according to normal distribution characterized by its mean value μ and its standard deviation σ in the YUV color space.

This model requires a number t of frames to compute μ and σ in each color component:

$$\mu(x,y,t) = \sum_{i=1}^{t} \frac{p(x,y,i)}{t}$$

$$\sigma(x,y,t) = sqrt\left(\sum_{i=1}^{t} \frac{p^2(x,y,i)}{t} - \mu^2(x,y,t) \right)$$

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 (CS3) but it fails in several cases: sudden illumination changes (CS4) and moving background objects like trees, flags or algae (CS10).

3.2.2 Mixture of Gaussians (MOG)

It was proposed in (Stauffer, 1999) 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:

$$\Pr(x_t) = \sum_{j=1}^{K} \omega_j . f_j(x_t; \mu_j, \Sigma_j)$$

$$\Pr(x_t) = \sum_{j=1}^{K} \pi_j \left(\frac{1}{(2\pi)^{p/2} |\Sigma_j|^{1/2}} \exp\left(-\frac{1}{2}(x_t - \mu_j)^T . \Sigma_j^{-1} . (x_t - \mu_j) \right) \right)$$

where ω_j is a weight associated to the jth Gaussian, with mean μ_j and covariance $\sum_j$, according to the time proportion of colors appearance

A pixel matches a Gaussian if: $sqrt\left((x_t - \mu_j)^T . \Sigma_j^{-1} . (x_t - \mu_j)\right) < \delta$

To handle multimodality in the background (CS10), Stauffer used as criterion the ratio $r_j = \omega_j / |\Sigma_j|^{1/2}$, 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:

$$\arg\min_b \left(\frac{\sum_{i=1}^{b} \omega_{i,t}}{\sum_{i=1}^{K} \omega_{i,t}} > T \right)$$

The MOG deals with lighting changes (CS3), multimodal background with low frequency variations (CS10), moved background objects (CS8) and sleeping foreground objects (CS12) but still sensitive to the initialization of the parameters (μ_0, $\sum_0$) and to their estimation at each step.

3.2.3 Kernel Density Estimation (KDE)

Elgammal (Elgammal, 2000) estimated the probability density function for each pixel color using the kernel estimator *K* (Gaussian kernel) for *N* recent sample of intensity values as:

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

The foreground detection is done according the following rule: If $\Pr(x_t) < Th$, 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 (CS10) but, unlike the Mixture of Gaussians, it adapts very quickly to high frequency variations in the background, and it doesn't need to estimate the parameters of the gaussians because it makes any assumption about the form of the underlying distributions. In the following section, we compare these three methods.

4 PERFORMANCE EVALUATION

The three algorithms were implemented using OpenCV. The MOG's algorithm used is the one proposed by KaewTraKulPong (KaewTraKulPong, 2001) which is an improved algorithm of (Stauffer, 1999). For the three algorithms, post processing is

used to eliminate isolated pixels corresponding to false positive detection. To evaluate the performance of the statistical background subtraction algorithms, we have made first a qualitative performance evaluation to evaluate the impact of the choice on the user's waiting. Indeed, this criterion is important because the system must give the identification in quasi real time. After, we made a quantitative evaluation to evaluate the performance in term of quality of the segmentation. In the section 4.3, we discuss the results of the qualitative and quantitative evaluation.

4.1 Qualitative Evaluation

Performance evaluation contains several senses. Performance evaluation 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 the Table 1.

Table 1: Qualitative comparison.

	SG	MOG	KDE
Speed	Fast	Intermediate	Slow
Memory	Intermediate	Intermediate	High

Because of the KDE is too slow and the need of the treatment speed is essential, we will thus neglect the KDE in the quantitative evaluation.

4.2 Quantitative Evaluation

After experimenting in the previous section, the KDE seemed to be too slow for the application, so we decide to evaluate quantitatively the SG and the MOG. The sequence image used is Aqu@theque sequence and contains 2600 images of size 640*480 in RGB. The results obtained are shown in the Figure 4.

The evaluation of these two methods has been done quantitatively by the sense of ROC (Receiver Operating Characteristic) Analysis, and by a measurement presented by Li (Li, 2004) 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 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 ground truths are marked manually.

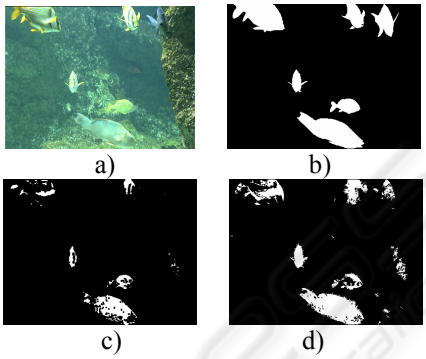

Figure 4: a) Image 201, b) Image Ground Truth, c) SG - Foreground Mask and d) MOG - Foreground Mask.

The evaluation results are shown in Table 2, which reveals that the best results are obtained for the MOG when K=3 Gaussians.

Table 2: Quantitative evaluations using the $S(A,B)$ measure and Roc analysis.

	SG	MOG (k=5)	MOG (k=3)
FPR	0.0005	0.0068	0.007
FNR	0.6419	0.4001	0.3683
Similarity	0.3564	0.5710	0.6004

4.3 Discussion

The qualitative evaluation shows that the SG and the MOG offer good performance in time consuming and memory requirement. The KDE algorithm is too slow for the application and requires too much memory. The quantitative evaluation shows that the MOG gives better results than the SG. Then, the MOG is revealed to be the method which is the most adapted between the three compared methods to the Aqu@theque application.

5 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 the one of the key steps in the system. First, we identified the critical situations met in video and improved the classification made by (Toyama, 1999).

This classification can be used in any application which uses background subtraction like video surveillance, motion capture or video games. In a second step, we made a comparison between three statistical background subtraction methods in the context of video sequence acquired from aquatic scenes.

A first qualitative evaluation showed that the MOG is more efficient, without adding time to the user's request. Quantitative tests confirm that the MOG enhance the percentage of detection. So, the recognition was improved and the performance of our interactive learning space too. In the future, we can test more background subtraction methods and make sophisticated evaluation using ROC Curves and the PDR method developed by Kim (Kim, 2006) on sequences test of the VSSN 2006 (VSSN, 2006).

Furthermore, the principle of Aqu@theque can be used in any multimedia environments that need this type of interaction. The background subtraction method must be chosen according to the critical situations met in the sequence used in the application.

REFERENCES

Semani D., Saint-Jean C., Frélicot C., Bouwmans T., Courtellemont P., January 2002, Alive Fish species characterization for on line video-based recognition, *Proceeding of the SPR 2002*, Windsor, Canada, pages 689-698.

Semani D., Bouwmans T., Frélicot C., Courtellemont P., July 2002, Automatic Fish Recognition in Interactive Live Videos, *Proceeding of the IVRCIA 2002*, volume XIV, Orlando, Florida, pages 94-99.

Desfieux J., Mascarilla L., Courtellemont P., 2002, Interactivity and Educational Information in Virtual Real-Time 3D Videos, *Proceeding of the IVRCIA 2002.*

Wren C., Azarbayejani A., Darrell T., Pentland A., , July 1997, Pfinder : Real-Time Tracking of the Human Body, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 19, No. 7, pages 780 – 785.

Stauffer C., 1999, Adaptive background mixture models for real-time tracking, *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, pages 246-252.

Elgammal A., Harwood D., Davis L., June 2000, Non-parametric Model for Background Subtraction, *6th European Conference on Computer Vision 2000*, Dublin, Ireland.

Toyama K., Krumm J., Brumitt B., 1999, Wallflower: Principles and Practice of Background Maintenance, *Proceedings of the International Conference on Computer Vision 1999*, pages 255-261.

Cucchiara R., Grana C., Piccardi M., Prati A. , 2003, Detecting Moving Objects, Ghosts and Shadows in Video Streams", *IEEE Transactions on Pattern Analysis and Machine Intelligence 2003*, volume 25, no 10, pages 1337-1342, 2003.

Culibrk D., Socek D. , Marques O., Furht B., March 2007, Automatic kernel width selection for neural network based video object segmentation, *VISAPP 2007*, Barcelona, Spain.

KaewTraKulPong P., Bowden R., September 2001, An Improved Adaptive Background Mixture Model for Real-time Tracking with Shadow Detection, *Proceedings 2nd European Workshop on Advanced Video Based Surveillance Systems, AVBS 2001*, Kingston, UK.

Kim K., Chalidabhongse T., Harwood D., Davis L. , 2006, PDR: Performance Evaluation Method for Foreground-Background Segmentation Algorithms, *EURASIP Journal on Applied Signal Processing*.

Huerta I., Rowe D., Gonzàlez J., Villanueva J., October 2006, Improving Foreground Detection for Adaptive Background Segmentation, *First CVC Internal Workshop on the Progress of Research and Development (CVCRD 2006)*.

VSSN 2006 - Algorithm Competition in Foreground/Background Segmentation: http://www.imagelab.ing.unimo.it/vssn06/.