Proceedings of the 5th WSEAS Int. Conf. on SIGNAL, SPEECH and IMAGE PROCESSING, Corfu, Greece, August 17-19, 2005 (pp303-308)

Fast Denoising for Moving Object Detection by An Extended Structural Fitness Algorithm

ALBERTO FARO, DANIELA GIORDANO, CONCETTO SPAMPINATO
Dipartimento di Ingegneria Informatica e Telecomunicazioni
Facoltà di Ingegneria, Università di Catania
Viale Andrea Doria, 6 - 95125
ITALY

Abstract: - Aim of the paper is to propose an optimized method to remove from the images taken by a camera the most common noises, i.e., the ones due to "waving trees" or to bad digitalization. This makes a real time identification of the objects traversing windy scenes or scenes taken with cameras affected by salt and pepper noise possible. The method is based on an extension of the Structural Fitness (SF) algorithm to avoid the analysis of those pixels that can be safely considered to belong to the background according to a simple statistical formula proposed in the paper. The experiments discussed in the paper demonstrate how the denoising method works and its excellent time performance and accuracy in typical engineering applications.

Key-Words: - Structural Fitness, Object detection, Background, Foreground

1 Introduction

When cameras are used for detecting moving objects or the appearance of an event, ideally one would have a noiseless background so that the objects or events to be detected could be easily obtained by subtracting the current image from the background. However, various sources of noise usually corrupt the images taken by the cameras for object detection, so that even in the absence of a new object traversing the scene the intensity of the pixels of an image fluctuates around a mean value.

In indoor applications, such as the surveillance of an apartment or the robot networking for domestic automation, this is essentially caused by electromagnetic fields, whereas in outdoor applications, such as vehicle detection or anti-firing surveillance, this is mainly due to the natural movement of some objects of the background (e.g., trees, electrical cables, etc.). Thus denoising images is a significant problem in object detection systems, especially when it has to be performed within real-time constraints.

One of the most wide used method for object detection in noisy environment is based on the statistical generation of the *average background* to be subtracted to the current image in order to point out the foreground object [1]. Such background is obtained at end of a training phase concerning the processing of N consecutive frames, without objects on the foreground, as the image whose pixel p_{ij} are characterized by the following average luminosity value:

$$\mu_{ij} = \frac{1}{N}\sum_{t=1}^{N} v_{ij}(t) \quad \forall\, p_{ij} \qquad (1)$$

where $v_{ij}(t)$ is the value of the luminosity of the pixel $\mathbf{p_{ij}}$ belonging to the frame taken at time t. From the knowledge of μ_{ij} it is possible to evaluate if a pixel p_{ij} of the frame taken at time t_0 belongs or not to the background as follows:

$$p_{ij} \in Foreground\,, if \left| v_{ij}(t_0) - \mu_{ij} \right| \geq T \qquad (2)$$

$$p_{ij} \in Background\,, if \left| v_{ij}(t_0) - \mu_{ij} \right| < T \qquad (3)$$

where $v_{ij}(t_0)$ is the current value of the luminosity of the pixel p_{ij} and T is a suitable threshold to be fixed experimentally. The main problem of this method of object detection is that the average background becomes meaningless when we have to detect foreground objects traversing windy scenes especially if such scenes contain trees or suspended objects.

As an example fig.1b shows the difference between the current scene (fig.1a) and the average background for a scene affected by waving trees. This difference should result in a quasi zero luminosity frame, whereas fig.1b shows a great number of luminous pixels that will be confused with foreground objects.

Another typical noisy situation that cannot be managed by the average background method is the one shown in fig.1c dealing with a frame affected by salt and pepper noise. Also in this case the difference between the current frame (fig.1c) and the average background does not result in a quasi zero luminosity frame, as pointed out in fig.1d, thus making difficult the detection of possible objects traversing this scene.

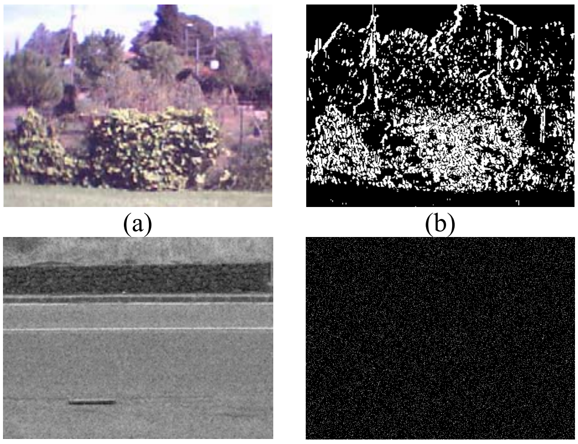

Fig. 1 – (a) Test Image with "waving trees", (b) Output of the model, (c) Test Image with "Salt and Pepper Noise", (d) Output of the model

Aim of the paper is to propose an optimized method to remove from the images taken by a camera the most common noises, i.e., the ones due to "waving trees" or to bad digitalization. This makes a real time identification of the objects traversing windy scenes or scenes taken with cameras affected by salt and pepper noise possible. For this reason, the paper proposes an extended form of the Structural Fitness (SF) algorithm [2].

In short, in the SF algorithm, outlined in sect.2, we have to analyze all the pixels according to a structural fitness formula in order to decide if the pixel belongs or not to the background. Our extension, presented in sect.3, optimizes the SF algorithm by a) a straightforward better organization of the computation steps proposed in the SF algorithm to make it more useful for engineering applications, and b) limiting the number of pixels to be handled with the SF algorithm in order to avoid the analysis of those pixels that can be safely considered to belong to the background, This latter extension will be done by means of two statistical formulas proposed in the paper depending on the noise type.

The experimental results, shown in sect.4, demonstrate that the proposed extension allows us to detect very accurately the foreground objects belonging to windy scenes or taken by cameras affected salt and pepper noise. Sect.4 will demonstrate also that the proposed denoising method is not time consuming thus allowing us to interpret video-sequences concerning objects traversing a given scene as requested by the indoor and outdoor surveillance applications or by the traffic flows monitoring.

2 Structural Fitness (SF) Algorithm

The aim of the structural fitness algorithm, proposed in [2], is to eliminate the noise, like waving trees, by exploiting the tendency of any pixel to take a value similar to the ones of its neighbors (fig.2a).

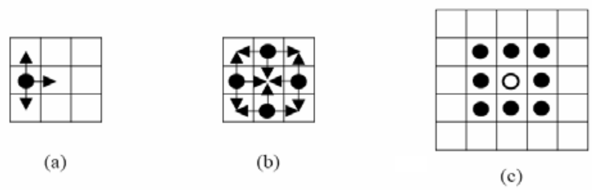

Fig. 2 - Structuring elements: pixel neighbors (a), compound structure, (b), cell-based structure (c).

According to this method a time consuming fitness analysis has to be performed for each pixel in order to evaluate if it belongs to the background or to the foreground (i.e., to an object traversing the scene or to a new event of the scene).

In fact the analysis of any pixel p_{ij} implies to compute the value of its structural fitness sf_{ij} by the expression $sf_{ij} = \Sigma_{Cij}\ f_{rs}$ where:

- the value of the function f_{rs} dealing with pixel p_{rs} depends on the four pixels $p_{r-1,s}$ $p_{r+1,s}$ $p_{r,s-1}$ $p_{r,s+1}$ that constitute a structure known as compound structure (fig.2b);
- the summation Σ_{Cij} usually extends overthe 24 pixels p_{rs} belonging to the 25x25 pixel cell Cij around pixel p_{ij} (fig.2c).

A pixel belongs to the foreground if $\Sigma_{Cij}\ f_{rs} > T_{ij}$, where T_{ij} is a suitable threshold. Formally the SF algorithm can be expressed as follows:

$\forall$ *pixel* $\boldsymbol{p_{ij}} \in$ *Image*

center the cell based structure $\boldsymbol{C_{ij}}$ *in* $\boldsymbol{p_{ij}}$

$\boldsymbol{sf_{ij}} = 0$

$\forall\ \boldsymbol{p_{rs}} \in \boldsymbol{C_{ij}}$

center the compound structure in $\boldsymbol{p_{rs}}$

compute $\boldsymbol{f_{rs}}$

$\boldsymbol{sf_{ij}} = \boldsymbol{sf_{ij}} + \boldsymbol{f_{rs}}$

if $\mathbf{sf_{ij}} < \mathbf{T_{ij}}$ => *pixel* $\boldsymbol{p_{ij}}$ *belongs to the background*

3 Extended SF (ESF) Algorithm

It is easy to understand that the outlined structural fitness algorithm is not compatible with real time applications, so the aim of the paper is to optimize the SF algorithm by reducing the processing time and increasing precision. This will be obtained:

- by a better organization of the processing steps to compute the mentioned functions f_{rs} with the aim of avoiding redundant computations, and
- by means of some statistical rules that avoid the application of the SF algorithm to the pixels that have low probability of being affected by noise.

The proposed algorithm is called Extended Structural Fitness (ESF) algorithm.

3.1 Preprocessing

The first optimization of the 'structural fitness' algorithm derives from observing that the computation of the function f_{rs} does not depend on the cell **Cij** but on the values of the four pixels located at north, south, east and west of pixel p_{rs}, thus it is more convenient to perform a pre-processing of the image by computing a matrix F whose general element is f_{rs} and then obtaining sf_{ij} by using the values of the matrix F without repeating their computation for every cell **Cij**.

The SF algorithm modified according to this rule is as follows:

$\forall$*pixel* $p_{ij} \in$ *Image*

center the compound structure in p_{rs}

compute f_{rs}

$F(r,s) = f_{rs}$

$\forall$*pixel* $p_{ij} \in$ *Image*

$sf_{ij} = \Sigma_{Cij} F(r,s)$ *where r and s are the labels of the pixels* p_{rs} *belonging to* C_{ij}

if $\mathbf{sf_{ij}} < \mathbf{T_{ij}}$ => *pixel* p_{ij} *belongs to the background*

This allows us to save about fifteen computation for every pixel of an image assuming that the structural algorithm uses a 25x25 pixel cell.

3.2 Temporal analysis

Another possibility to further increase accuracy and to significantly decrease the execution time of the mentioned SF algorithm is to limit the number of pixels to be evaluated by avoiding to process the set Ω of pixels that with high probability have not been modified because of the passage of some object in the scene.

Such pixels p_{ij} are the ones that may have a luminosity value $v_{ij}(t)$ that differs from the value at previous instants, i.e., $v_{ij}(t) \neq v_{ij}(t-1)$, but whose variation in time shows a statistical behavior that is compatible with the one of the pixels of the background corrupted by a noise due to the movement of some background objects of the scene.

To estimate the upper and lower bounds for each pixel that allows us to decide if the pixel 'surely' belongs to the background or if it may belongs to the foreground we have to advance some assumption about the statistical model of each pixel of the image in absence of objects on the foreground. In the paper we assume that the 'natural' variation of the pixels, i.e., its variation in absence of relevant objects but in presence of noise, follows a Gaussian distribution as shown in Fig.3.

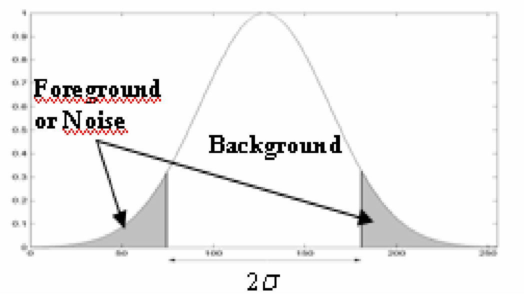

Fig.3 – Statistical model of a general pixel p_{ij}

Thus the problem of finding the mentioned set Ω can be easily solved after having found the average value μ_{ij} and the variance σ_{ij}^2 of the assumed statistical model as follows:

$$if\ v_{ij} \in \left[\mu_{ij} - \sigma_{ij}\ , \mu_{ij} + \sigma_{ij}\ \right] \rightarrow p_{ij} \in \Omega \quad (4)$$

If the pixel p_{ij} belongs to Ω we don't apply the structural fitness algorithm, otherwise the SF algorithm provided with the proposed preprocessing will be applied. The experimental results will show that the temporal analysis is particularly useful to manage noise such as the 'weaving tree' noise.

3.3 Spatial analysis

An additional way to estimate the pixels that belongs to the background is the one of performing a spatial analysis of the image. This analysis is based on the assumption that if the system is noiseless, the value v_{ij} of the general pixel p_{ij} would differ from the one of its neighbors by a fixed quantity q_{ij}, whereas in presence of a noise such quantity will vary according to a Gaussian statistics.

Thus saying μ'_{ij} and $\sigma^{2}{}'_{ij}$ the average value and the variance of the quantity q_{ij} in a kernel of dimensions N , the pixels that are within the interval $[\mu'_{ij} - \sigma'_{ij}, \mu'_{ij} + \sigma'_{ij}]$ may be considered to belong to the background, i.e.:

$$v_{ij} \in \left[\mu'_{ij} - \sigma'_{ij}, \mu'_{ij} + \sigma'_{ij}\right] \rightarrow p_{ij} \in \Omega \qquad (5)$$

where μ'_{ij} and $\sigma^{2}{}'_{ij}$ are defined in the following formulas (6) and (7):

$$\mu_{ij}' = \frac{1}{N} \sum_{x=\left(i-\frac{N-1}{2}\right)}^{x=\left(i+\frac{N-1}{2}\right)} \sum_{y=\left(j-\frac{N-1}{2}\right)}^{y=\left(j+\frac{N-1}{2}\right)} v_{xy} \qquad (6)$$

$$\sigma^{2}{}'_{ij} = \frac{1}{N} \sum_{x=\left(i-\frac{N-1}{2}\right)}^{x=\left(i+\frac{N-1}{2}\right)} \sum_{y=\left(j-\frac{N-1}{2}\right)}^{y=\left(j+\frac{N-1}{2}\right)} \left(v_{xy} - \mu_{xy}\right)^{2} \qquad (7)$$

Also in this case, if the pixel p_{ij} belongs to Ω we don't apply the structural fitness algorithm, otherwise the SF algorithm will be applied. The experimental results will show that the spatial analysis is particularly useful to manage the salt and pepper noise due to a bad digitalization of the camera.

4 Experimental results

In order to point out the practical advantages of the proposed ESF algorithm, in this section we show a significant decrease of the processing time and some increase of accuracy of the ESF algorithm with respect to the SF algorithm.

For this reason the section will discuss some relevant cases dealing with images of 320x240 pixels affected by 'weaving tree' noise and bad digitalization. The image processing according to the mentioned SF and ESF algorithms has been performed by a PC provided with an AMD Athlon 3GHz processor and a DDR memory of 1 Gbyte.

The processing time ***t*** has been measured by the time interval needed for processing a single image, whereas accuracy ***a*** has been computed by the following formula:

$$a = \left(1 - \frac{n}{n_{tot}}\right) * 100 \qquad (8)$$

where n_{tot} is the total number of the pixels of the image, and n is the number of noisy pixel that have been not discovered by the algorithm. With reference to the images analyzed in this section, we have obtained that the processing time of the SF algorithm is about 10 seconds whereas its accuracy is very close to 100%.

Since the SF algorithm accuracy is very high but its speed is very low, in the following we have to demonstrate that by the ESF algorithm we may achieve better time performances at the same high accuracy. This demonstration will be done by taking into account three relevant situations, i.e., a) windy scenes with many trees, b) windy scenes with few trees, and c) scenes affected by salt and pepper noise. Finally this section will demonstrate that the proposed extension of the SF algorithm does not interfere with the detection of an object traversing the scene, i.e., no part of the shape of the foreground object is considered belonging to the background.

4.1 Windy scenes with many trees

Fig.4 show the results of the application of the three methods proposed in the previous section (i.e., SF powered by preprocessing, SF with preprocessing and temporal analysis, and SF with preprocessing and spatial analysis) to analyze the windy scene shown in fig.1a. The ESF algorithm provided with only the pre-processing phase has the same accuracy of the SF algorithm but a processing time of 1797 msec that is an order of magnitude lower than the previous one. The ESF algorithm powered by the proposed preprocessing and temporal analysis has shown a further increase of the accuracy and a significant decrease of the processing time. If we use the spatial analysis, we obtain the same time performance of the temporal one but a lower accuracy.

The better time performance of the temporal and spatial analysis with respect to the SF algorithm depends on the limited number of pixels to process with the SF formula. In fact the number of pixels to process is 76800 (320x240) pixels for the SF algorithm, 1500 for the ESF algorithm powered by the temporal analysis, and 500 for the ESF algorithm powered by the spatial analysis.

Let us note that the spatial analysis is less accurate of the temporal one because it is not able to detect all the pixels that are affected by the waving noise. In fact, the test used in the spatial analysis to verify if the pixel belongs or not to the background is three times more time consuming than the one adopted by the temporal analysis. Consequently, although the number of pixels to process in the spatial analysis is less that the one of the temporal analysis, the total time performance time of the spatial analysis is more or less the same of the temporal one. Let us note also that the 100% accuracy of the temporal analysis simply means that the Gaussian statistical formula is enough to model the movements of the leafs under the experimented wind conditions

Algorithm	Output Image	Proc. Time	Accuracy
SF with pre-processing		**1767 msec**	**98,64 %**
SF with temporal analysis		**234 msec**	**100%**
SF with spatial analysis		**277 msec**	**82.5%**

Fig.4 – Experimental results dealing with windy scenes characterized by many trees

4.2 Windy scenes with few trees

Fig.6 shows the results of the application of the three methods proposed in the previous section to analyze the windy scene with few tress shown in fig.5. The results confirm the analysis performed in the previous paragraph

(a) **(b)**

Fig.5 - (a) Test Image with low waving trees, (b) Image difference between mean background and current frame

Algorithm	Output Image	Proc. Time	Accuracy
SF with pre-processing		**1767 msec**	**99,72 %**
SF with temporal analysis		**99 msec**	**100%**
SF with spatial analysis		**187 msec**	**97.5%**

Fig.6 – Experimental results dealing with windy scenes characterized by few trees

4.4 Scenes with moving foreground objects

To illustrate that the proposed ESF algorithm does interfere with the detection of the foreground objects [3] and [4], let us consider the situation shown in fig.9b dealing with the monitoring of the cars traversing a windy scene in which some electrical cables oscillate. The output of the ESF algorithm powered by temporal analysis shows the entire shape of the car traversing the scene without detecting the moving cables (fig.9d). On the contrary, by subtracting the current image (fig.9b) to the average background (fig.9a) it is not possible to eliminate completely the waving noise.

Of course in this case any algorithm for motion detection will be able to count the cars traversing the scene without taking into account the waving noise. However, if the noise increases or if the shape of the cars is smaller than the one shown in fig.8, some significant error could arise and the proposed ESF algorithm should be used if one is interested in a real time monitoring of the traffic flow. In fact, although a typical video sequence consists of 30 frames per second, both the traffic flow monitoring and the surveillance system can be safely implemented by using a lower frame rate, such as for example ten frames per second, and our method.

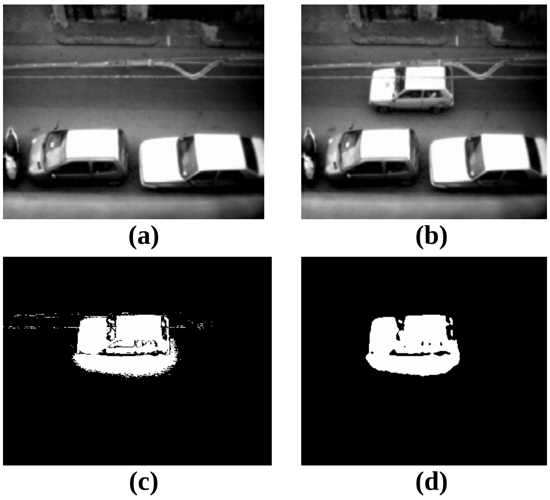

Fig. 9 - (a) Average Background, (b) Test Image, (c) Image difference between Test Image and Average Background, (d) Output of SF with temporal analysis.

5 Concluding remarks

The paper has shown that a better organization of the computation steps is important to increase the time performance of the SF algorithm for detecting objects in noisy scenes. However, this is not enough. In fact the temporal and spatial analysis proposed in the paper have to be performed to achieve a time performance compatible with real time application such as surveillance and traffic monitoring. In particular, the temporal analysis is more suitable for managing windy scenes, whereas the spatial analysis is better for eliminating in real time the salt and pepper noise.

The processing time could be more conveniently decreased by using advanced hardware platforms such as DWA-Destination Word Accumulation or Grid architectures depending on the image resolution and on the time reaction requested by the object detection applications.

Some further analysis should be performed to deeply study if the turbulent variations of the wind influence the Gaussian formula assumed in the paper to model the pixel variations. This analysis, particularly needed for object detection in extreme environmental conditions, is outside the scope of the paper.

References:

[1] Gonzales R.C. and Woods R.E., Digital Image Processing- Prentice Hall 2002

[2] Bevilacqua A., Roffilli M., Robust Denoising and Moving Shadow Detection in Traffic Scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) -Technical Sketches* , Kauai Marriot, Hawai, USA, December 2001, pp.1-4.

[3] Di Stefano L., Neri G., Varani E., Analysis of Pixel-Level Algorithms for Video Surveillance Applications, In *Proceedings of 11thInternational Conference on Image Analysis and Processing, ICIA 2001*, Palermo, Italy, September 2001, pp. 542 – 546.

[4] Faro A., Giordano D. and Spampinato C., Soft-computing agents processing web cam images to optimize metropolitan traffic systems. In *Proceedings of Int. Conf. on Computer Vision and Graphics (ICCVG)* Warsaw, 2004.