A Proposed Method for Background Subtraction and adding Scribbles in Video Matting

Mervat El-Seddek[1], Hossam El-Din Moustafa[2]
[1]*Electronics and Communications Department - Misr Engineering and Technology Institute*
[2]*Electronics and Communications Department – Faculty of Engineering – Mansoura University*
(E-mail: Hossam_moustafa@hotmail.com)

Abstract: Identifying moving objects from a video sequence is a fundamental task in most computer-vision applications. It is essential for visual tracking systems. In this paper, a new algorithm for background subtraction has been proposed and evaluated. Decision-level fusion of three background subtraction algorithms; frame difference, approximated median, and Mixture of Gaussians; has been utilized. It has been found that the percentage error for each frame has reduced remarkably when applying the proposed technique.Moreover, an entirely-automatic algorithm for adding scribbles in video matting is proposed to improve video matting results depending on the suggested background subtraction algorithm. The overall performance is evaluated. The average error has been reduced effectively and the overall performance has been enhanced.

Keywords-Background subtraction; frame difference; mixture of Gaussians; approximated median scribbles; video matting

I. INTRODUCTION

Foreground detection in a video sequence is one of the most important tasks in video surveillanceand tracking systems [1]. Background subtraction is utilized for discriminating moving objects from the background scene. In background subtraction, the current image is subtracted from a reference one. This leaves new objects in the scene [2]. Fig. (1) shows the four maincomponents of a background subtraction system. These are preprocessing, background modeling, foreground detection, and data validation [3].

The simplest background modeling algorithm is frame differencing [4]. It uses the video frame at time (t –1) as the background model for the frame at time t. Frame differencing performs poorly if the background is not truly static (e.g., fluttering leaves, water, waves). The generalized Mixture of Gaussians (MoG) has been utilized to model complex, non-static backgrounds [2]. However, backgrounds with fast variations cannot be accurately modeled with just a few Gaussians, causing problems for sensitive detection.

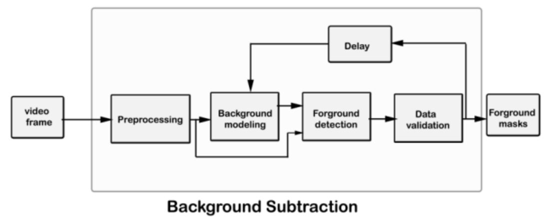


Fig. 1. Basic Steps for Background Subtraction Algorithms

Median filtering is one of the most commonly-used backgrounds modeling techniques [5], [6], [7]. The background estimate is defined to be the median at each pixel location of all frames in the buffer.The assumption is that the pixel stays in the background for more than half of the frames in the buffer. The approximated median filter was proposed such that the running estimate of the median is incremented by one if the input pixel is larger than the estimate, and decreased by one if small [8].

The present paper proposes a new combinational method for background modeling. It relies on introducing some modification in each of the three previously explained techniques and then merging the outputs of the three techniques to obtain the output of the proposed fusion technique.

The organization of the paper is as follows: Section 2 presents the proposed algorithm and a comparative study of its performance with the three traditional methods. Section 3 introduces the matting problem. In section 4, the proposed method for adding scribbles is presented. Section 5 is the conclusion.

II. THE PROPOSED ALGORITHM

The proposedtechnique relies on two consecutive steps. The former utilizes image processing techniques to enhance results of each single background subtraction method.The most important step is the second one where data fusion techniques are utilized to combine the outputs of each technique into a single output as shown in Fig. 2.

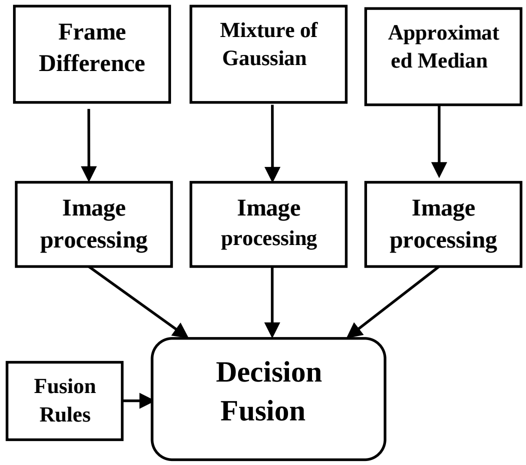


Fig. 2. Main Steps of the Proposed Algorithm

A. The Image Processing Step

In order to circumvent the problem of holes and to minimize the percentage error; a number of image processing steps was carried out to the output of each of the three basic algorithms as follows:

1. Edge Detector

An edge detection algorithm is applied to the resultant of each background subtraction algorithm separately. Sobel function [9] was chosen to detect frame transitions.

2. Dilate The Image

Some gaps are observed in the lines surrounding the object. These linear gaps can be eliminated if the Sobel image is dilated using linear structuring elements, which can be created using the Strel function.

3. Fill Interior Gaps

The dilated image shows the outline of the object quite clearly, but some holes still exist in the interior of the object. To fill these holes, the 'imfill' function in MATLAB has been utilized.

4. Smooth The Object

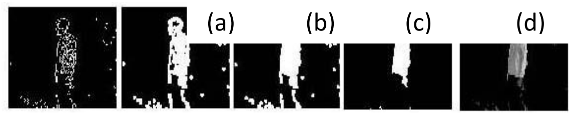
In order to make the segmented object look natural, it must be smoothed. This can be achieved by eroding the image twice with a diamond structuring element using the 'Strel' function [10]. The appearance of the object is now clear.

Fig.s 3, 4, 5show the results obtained from the three basic background subtraction algorithms after applying image processing procedures. It can be noticed that the suggested image enhancement methodhas given a better and clearer appearance for the object.

(a) (b) (c) (d) (e)

Fig. 3.Frame Difference Results

(a) Edge Detector, (b) Dilate the Image,(c) Fill Interior Gapes, (d) Smooth the Object, and (e) The Extracted Foreground



(a) (b) (c) (d) (e)

ults

or Gapes, (d) Smooth the Object, and (e) The Extracted Foreground

(a) (b) (c) (d) (e)

Fig. 5. Approximated Median Results

(a) Edge Detector, (b) Dilate the Image, (c) Fill Interior Gapes, (d) Smooth the Object, and (e) The Extracted Foreground

B. The Decision Level Fusion Algorithm

The processed frames obtained from the three background subtraction algorithmsare then merged using data fusion criteria. This step improves the resultant frame and reduces the percentage error as well. Therefore, a decision level fusion algorithm was utilized using the following fusion rules:

$$If\ ((V1\ \&\ V2)\ OR\ (V1\ \&\ V3)\ OR\ (V2\ \&\ V3))\ \text{is FG then } O(i,j) = FG \quad (1)$$

$$If\ ((V1\ \&\ V2)\ OR\ (V1\ \&\ V3)\ OR\ (V2\ \&\ V3))\ \text{is BG then } O(i,j) = BG \quad (2)$$

Where V_1, V_2, V_3 are the outputs of frame difference, Mixture of Gaussian, and approximated median, while O(i,j) represents the output pixel.

The percentage error was calculated with the aid of two videos using the following equation [11].

$$PEP(q) = \sum_{i=1}^{M} \sum_{j=1}^{N} |\, o(i,j) - f(i,j)\, | \quad (3)$$

Where:

o(i,j) is the output pixelin black and white form

f(i,j) is the original pixel of input video

M& N represent number of rows and columns in each frame

q is the total number of frames in input video

PEP is the average percentage error

Fig. (6) and Fig. (7) show the calculated error for the three algorithms of background subtraction as well as the proposed algorithm. It is obvious that the percentage error of the proposed algorithm has the least magnitude compared withthose obtained from the three methods.

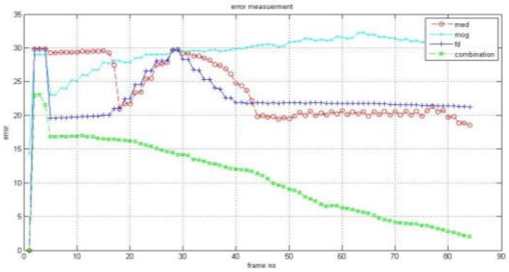


Fig. 6. Percentage Error as a Function of Frame Number (Video 1)

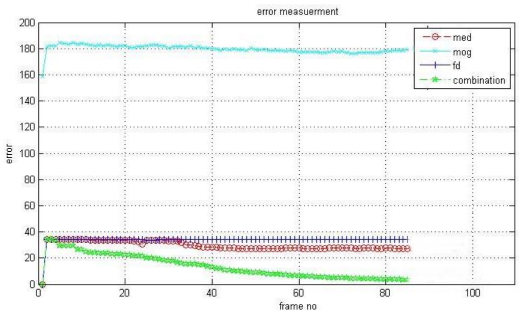


Fig. 7. Percentage Error as a Function of Frame Number (Video 2)

III. THE MATTING PROBLEM

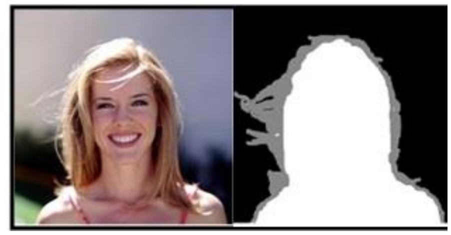
The matting problem was previously introduced by Porter and Duff in 1984 [12]. The image I_Z is seen as a mix of foreground F_Z and background B_Z using a parameter called "alpha matte", α_Z ,as follows [13]:

$$I_Z = \alpha_Z F_Z + (1-\alpha_Z) B_Z \qquad (4)$$

Where the range of α_Z lies between zero and unity. The unity value of α_Z means that the pixel is pure foreground, whereas the zero value indicates that it is a pure background. If $0 < \alpha_Z < 1$,the pixel will be mixed. Evaluation of alpha for mixed pixels is essential to separate the foreground.

Matting is considered as a difficult mathematical problem as it involves seven unknown values while there are only three given parameters. The three known values are the elements of the vector I_Z while the seven unknowns are α_Z and the three components of foreground F_Z as well as the three components of the background B_Z.

To overcome this mathematical difficulty, a trimap can be utilized. This task is simple a segmentation of input image into three areas: a foreground area by the white colour, a background area by the black colour, and unknown area represented by grey colour [13]. Fig. (8) shows an example of a trimap.



(a) (b)

Fig. (8) (a): Original Frame, and (b) The Trimap

Scribbles can be considered as an alternative choice to trimap. The user determines a small number of scribbles to extract a matte [13]. All unmarked pixels are then considered as unknowns. The procedure of adding scribbles at each frame will be investigated.

IV. THE PROPOSED ALGORITHM

The result of the decision fusion technique controls the scribble addition procedure. The automatic addition method is summarized in the following steps.

1. Evaluate the output of the decision fusion technique as shown in Fig. (9).

Fig.(9) The output of decision fusion algorithm

2. Get a user defined picture that has only two colors; blue background with small white rectangles on it, as shown in Fig. (10).

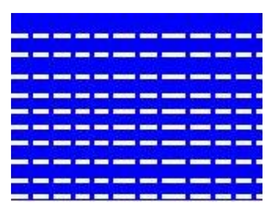


Fig. (10) The Fabricated picture

3. Use rgb2gray function and a suitable threshold to get the black and whiteversion of the pictureas shown in Fig. (11).

Fig.(11) Fabricated picture(black and white)

4. The output of the decision fusion technique is then multiplied by the recently obtained black and white picture, as shown in Fig. (12).

Fig.(12) The output of the Multiplication step

5. Use the following decision rule to get the final frame by comparing the output of step 4 with picture in step 2:

Let I_2: the output of step 2
I_4: the output of step 4
I_5: the output of step 5, then
If (I_4=white &I_2=white) then I_5=white
If (I_4=black &I_2=white) then I_5=black
If (I_4=black &I_2=blue) then I_5=blue

Fig. (13) shows the output of step 5.

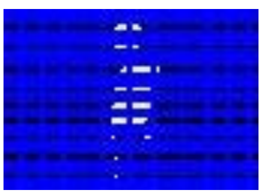


Fig.(13) The output of the comparison step

6. If you have a blue pixel, replace it with the original frame pixelis to get theoriginal image with the added scribbles, as shown in Fig. (14).

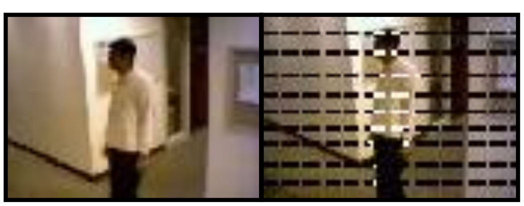


Fig. (14) The resultant image with added Scribbles

V. PERFORMANCE EVALUATION

Performance have been evaluated on the basis of the average error per pixel in each frame using a set of video sequences and Eq. (5). Results of two video examples are shown in Fig. (15).

$$\mathrm{AEP}(q) = \frac{1}{mn}\sum_{i=1}^{n}\sum_{j=1}^{m}|\,o(i,j) - f(i,j)\,| \qquad (5)$$

VI. CONCLUSION

In this paper, an algorithm was proposed to detect foreground objects from their background. It consists of two main steps. Experimental results on two fabricated videos had shown that that the proposed combinational algorithm had eliminated the holes significantly.It has been shown that the application of a decision level fusion criterion to the three processed foregrounds had improved greatly the resultant foreground. Moreover, the calculated percentage error had been greatly reduced.

In addition, a proposedtechnique of scribbles addition was introduced.It utilizes the results of the investigated data fusion algorithm and a user-defined picture. It is argued that the proposed algorithm of adding scribbles had shownbetter performance with minimized average error.

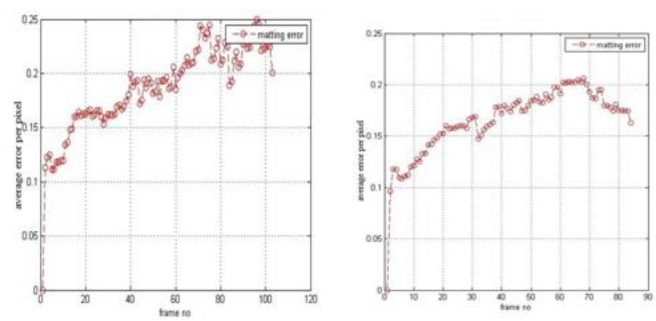


Fig. (15) Results for 2 video sequences

REFERENCES

[1] Mingjun Wu., XianrongPeng”Spatio-temporal context for codebook-based dynamic background subtraction” Science Direct, Elsevier GmbH 2009.

[2] S. Cheung and C. Kamath."Robust techniques for backgroundsubtraction in urban traffic video".in Proc. of the VCIP, 2004.

[3] J. Mike McHugh, Janusz Konrad, Venkatesh Saligrama, and Pierre-Marc Jodoin ”Foreground- Adaptive Background Subtraction”, IEEE Signal Proc.2009.

[4] Wan Zaki "Moving object detection using key point’s reference model", EURASIP Journal on Image and Video Processing 2011.

[5] R. Cutler and L. Davis, “View-based detection," in Proceedings Fourteenth International Conference on Pattern Recognition, 1, pp. 495-500, (Brisbane, Australia), Aug 1998.

[6] R. Cucchiara, M. Piccardi, and A. Prati, “Detecting moving objects, ghosts, and shadows in video streams," IEEE Transactions on Pattern Analysis and Machine Intelligence 25, pp. 1337-1342, Oct 2003.

[7] B. Lo and, S. Velastin, “Automatic congestion detection system for underground platforms," in Proceedings of 2001 International symposium on intelligent multimedia, video, and speech processing, pp. 158-161, (Hong Kong), May 2001.

[8] N.J.B. McFarlane and C.P. Schofield, ”Segmentation and tracking of piglets in images” Machine Vision and Applications 8-187-193,1995.

[9] Tapas Kumar and G. Sahoo,” A Novel Method of Edge Detection using Cellular Automata”, International Journal of Computer Applications (0975 – 8887) Volume 9– No.4, November 2010.

[10] “Detecting A Cell Using Image Segmentation” ,5/5/2011, http://www.docstoc.com/docs/78776244/Image_Segmentation

[11] Vassilios Chatzis, Adrian G. Bor¸s, and Ioannis Pitas,” multimodal decision-level fusion for person authentication”, IEEE transaction on systems, man , and cybernetics, VOL. 29, NO. 6, NOV 1999.

[12] T. Porter and T. Duff, “Compositing digital images,” Proc. of A special interest group of the Association for Computing Machinery (ACM SIGGRAPH), vol.18, no.3, pp. 253–259, July 1984.

[13] Jue Wang and Michael F. Cohen, “Image and Video Matting: A Survey,”

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.190.1825&rep=rep1&type=pdff