BACKGROUND DIFFERENCING TECHNIQUE FOR IMAGE SEGMENTATION BASED ON THE STATUS OF REFERENCE PIXELS

Toshiyuki YOSHIDA

Dept. Information Science, Fukui Univ., Fukui 910-8507, Japan

ABSTRACT

Background frame differencing is one of the most efficient image segmentation techniques in case that a static background frame is available. The technique, however, involves two major problems: one is that decision with a single global threshold often produces poor segmentation results and the other is that an estimation for the optimal threshold is actually difficult. To overcome these problems, this paper proposes a very simple yet effective algorithm that refers to the status of previously judged neighboring pixels: more specifically, the proposed algorithm varies the threshold value in accordance with the status of the reference pixels. The experimental results illustrate that the proposed technique gives better segmentation results than the simple background differencing one with the optimal global threshold.

1. INTRODUCTION

Object extraction from an image is required in such areas as object-based coding, video surveillance[1], collaboration applications such as a video conferencing in an immersive environment[2], and has been recognized as an important technique. The final goal of this technique is to discriminate target objects from the background, which is equivalent to assigning the label"F" (representing "foreground") to object pixels, and the label B (representing "background") to background pixels. Although much efforts have been made for this problem, it is still left unsolved.

The object extraction techniques proposed thus far can be classified into two categories: automatic and semiautomatic. The former works without any human support, while the latter requires user's interaction, for instance, in setting an initial region around the target object. In a real-time application, however, the automatic technique is exclusively employed. In this case, if the static background frame is available from a fixed video camera, the background frame differencing technique[2]–[6] can be utilized.

The fundamental idea of the background frame differencing is very simple. Let $b(i,j)$ and $t(i,j)$ be the background frame and the target frame including the object to be extracted, respectively, and let $L(i,j)$ represent a 2-D array for storing the judged labels (note that $b(i,j)$ and $t(i,j)$ represent the luminance of frames. An extension to color images will be discussed in Sect. 2). The decision rule for the labeling is described as

$$\begin{cases} |\,e(i,j)\,| \le T & \text{then} \quad L(i,j) = \mathrm{B} \\ |\,e(i,j)\,| > T & \text{then} \quad L(i,j) = \mathrm{F} \end{cases}, \tag{1}$$

where $e(i,j)$ is the frame difference given by

$$e(i,j) = b(i,j) - t(i,j)\,, \tag{2}$$

and T is a prescribed threshold.

A direct application of this algorithm, however, involves some difficulty caused mainly by camera noise. First, the discrimination of pixels with a single global threshold T does not yield satisfactory results because the frame difference e is corrupted by camera noise. Second, the presence of noise makes the optimal decision of the threshold T very difficult.

To alleviate these problems, Refs. [3]–[5] propose improved techniques based on Markov random field. Ref. [6] gives an automatic threshold decision algorithm. This paper attempts to overcome the problems by introducing reference pixels into the background differencing algorithm. The proposed technique actually varies the threshold T in Eq. (1) in accordance with the status of the reference pixels.

The rest of this paper is organized as follows. Sect. 2 explains the details of the proposed technique. Sect. 3 discusses a derivation of the optimal threshold values for the proposed algorithm. Sect. 4 gives some experimental results and makes a comparison with a fixed threshold approach. Finally, Sect. 5 gives conclusions.

2. PROPOSED ALGORITHM

2.1. Outline of the Proposed Algorithm

The decision rule (1) is often applied to the target frame by the raster scan order as shown in Fig. 1 (a). Note that in this case the label $L(i,j)$ in the shaded region can be used for the decision on the current pixel "A". Furthermore, since neighboring pixels in a natural image are highly correlated, it is considered that neighboring labels in $L(i,j)$ are also highly correlated if they are ideally assigned.

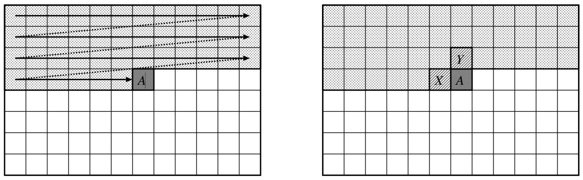
These facts imply that decision error for each pixel can be reduced by referring to the labels of the neighboring pixels. In this paper, the pixels marked as "X" and "Y" shown in Fig. 1 (b) are used as the reference pixels for the current pixel "A", and the threshold value T in Eq. (1) is varied in accordance with the combination of the labels for X and Y.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

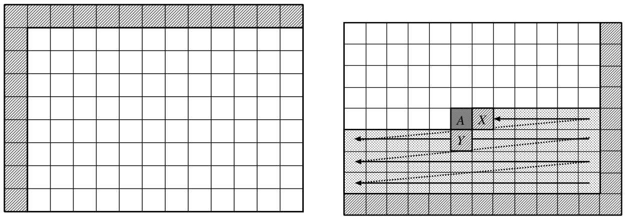
Let T_{FF}, T_{FB}, T_{BF}, and T_{BB} represent the threshold values for $(X,Y) = (F,F)$, $(X,Y) = (F,B)$, $(X,Y) = (B,F)$ and $(X,Y) = (B,B)$, respectively. Then, the decision rule (1) can be modified as

$$\begin{cases} |\,e(i,j)\,| \leq T_{XY} & \text{then} \quad L(i,j) = \mathrm{B} \\ |\,e(i,j)\,| > T_{XY} & \text{then} \quad L(i,j) = \mathrm{F} \end{cases} , \tag{3}$$

where $X = L(i-1,j)$ and $Y = L(i,j-1)$. Before starting the decision on the target frame, the label B is given for each pixel just outside the frame as shown in Fig. 1 (c).



(a) Scanning order (b) Reference pixels



(c) Initialization (d) Reversed scanning order

Fig. 1 : Proposed algorithm.

Assume that the reference pixels have been judged as $X = F$ and $Y = F$, and that the the current pixel A belongs to the object region. In this case, setting a relatively small value for T_{FF} impedes the decision of $A = B$, which contributes to a reduction of decision error. In case of $(X,Y) = (B,B)$, T_{BB} should be relatively large so that the decision $A = B$ is encouraged. A derivation for the optimal threshold values will be discussed in Sect. 3.

2.2. Detailed algorithm

We first modify the decision rule (3) so that it can be applied to RGB color images. Let $b_C(i,j)$ and $t_C(i,j)$ ($C = R$, G, B) represent one of the RGB color components of the background and the target frames, respectively. Although several extensions for color images are possible, here we use the following decision rule:

$$\begin{cases} |\,e_R(i,j)\,| \leq T_{XY} \text{ and } |\,e_G(i,j)\,| \leq T_{XY} \text{ and} & \\ \qquad |\,e_B(i,j)\,| \leq T_{XY} \quad \text{then} & L(i,j) = \mathrm{B} \\ \text{Otherwise} & L(i,j) = \mathrm{F} \end{cases} , \tag{4}$$

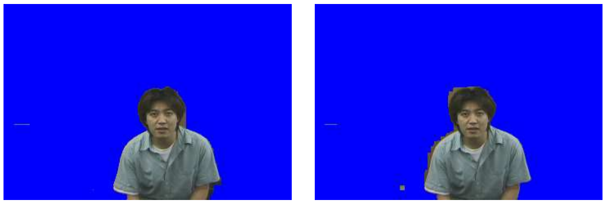
where

$$e_C(i,j) = f_C(i,j) - t_C(i,j) \qquad (C = R,\, G,\, B)\,.$$

In Eq. (4) the reference pixels X and Y are the same as in Eq. (3): $X = L(i-1,j)$ and $Y = L(i,j-1)$ (see Fig. 1 (b)).

Fig. 2 (a) shows the result of the proposed algorithm applied to the background and target frames given in Fig. 5 (a) and (b), respectively. Compared with Fig. 5 (d) which is the optimal result by the fixed threshold approach (see Sect. 4 for details), the accuracy of extraction is actually improved in Fig. 2 (a). The introduction of the reference pixels, however, causes error propagations as can be seen in Fig. 2 (a), especially in the right hand side of the target object. This is because the proposed algorithm tends to select T_{FF} at the right side boundary of the object, which works to prevent the decision from passing from F into B. This error further propagates with the forward scanning order making some groups of F pixels at the right side.

To overcome the problem, we apply the proposed algorithm again to the target frame with the reversed scanning order shown in Fig. 1 (d), and then take the intersection of the F regions in the both results. Fig. 2 (b) gives the result by the backward scanning order. By taking the intersection of Fig. 2 (a) and (b), the final result of extraction is obtained, which is shown in Fig. 5 (c). Although this algorithm requires twice the computational cost of the simple background differencing algorithm, the total computational cost is still low enough to be employed in a real-time application.



(a) Result with the forward scanning order (b) Result with the backward scanning order

Fig. 2 : Examples of Error Propagation.

3. A DERIVATION OF THE OPTIMAL THRESHOLD VALUES

It is widely accepted that camera noise can be well modeled by the Gaussian distribution. Here we assume that a Gaussian noise n with zero mean and the variance σ^2 is added independently to each pixel through the camera. Then, the observed frame difference can be written as

$$e(i,j) = (b(i,j) + n_b) - (t(i,j) + n_t)\ . \tag{5}$$

Since the noise n_b and n_t are assumed to be mutually independent, noise in the frame difference $e(i,j)$ is also Gaussian having the variance $2\sigma^2$. The conditional probability density function for e (for simplicity the index (i,j) is dropped later on) can be written as

$$f(e|B) \;=\; \frac{1}{\sqrt{2\pi \cdot 2\sigma^2}}\, e^{-\dfrac{e^2}{2 \cdot 2\sigma^2}} \tag{6}$$

$$f(e|F) \;=\; \frac{1}{\sqrt{2\pi \cdot 2\sigma^2}}\, e^{-\dfrac{(e-\Delta)^2}{2 \cdot 2\sigma^2}}\ , \tag{7}$$

where $f(\,\cdot\,|B)$ and $f(\,\cdot\,|F)$ represent the assumptions that the pixel under consideration belongs to B and F, respectively, and Δ denotes the difference between the foreground and background pixel values. Fig. 3 illustrates Eqs. (6) and

(7) with $\Delta > 0$ and the threshold T. p_1 in Fig. 3 corresponds to the probability that a B pixel is judges as F by mistake, while on the contrary $p_2(\Delta)$ is the probability that an F pixel is judged as B by mistake. Here the notation $p_2(\Delta)$ is used to indicate that it depends on the difference between the background and foreground Δ. To get the average value of $p_2(\Delta)$, we assume that Δ distributes uniformly in the range $[-255, 255]$, and take the expectation value $\overline{p_2}$.

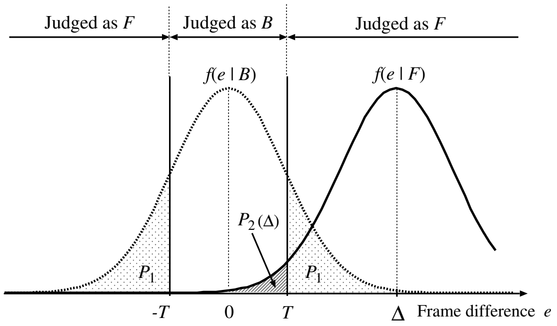


Fig. 3 : Distributions of the frame difference e.

Our goal here is to determine the optimal threshold values in a certain sense. To this end, we attempt to calculate the threshold values that minimizes the total error probability. Let $P(F|XY)$ and $P(B|XY)$ denote the probabilities that the pixel under consideration belongs to the foreground F and background B, respectively, under the observation of the reference pixels X and Y.

The total error probability P_{XY} under the observation of the reference pixels X and Y is then calculated as

$$P_{XY} = P(B|XY) \cdot p_1 + P(F|XY) \cdot \overline{p_2}\,. \tag{8}$$

Since P_{XY} depends on the threshold value T in Fig. 3 , we define the optimal threshold T_{XY} as

$$T_{XY} = \arg\min_{T} \left(P(B|XY) \cdot p_1 + P(F|XY) \cdot \overline{p_2}\right)\,. \tag{9}$$

To evaluate Eq. (9) numerically, we first estimated the probabilities $P(F|XY)$ and $P(B|XY)$ for several test images, and obtained

$$\begin{aligned} &P(F|FF) = 0.97, \qquad P(B|FF) = 0.03 \\ &P(F|FB) = P(F|BF) = P(B|FB) = P(B|BF) = 0.5 \\ &P(B|BB) = 0.97, \qquad P(F|BB) = 0.03 \end{aligned}$$

By substituting these values into Eq. (9), the optimal threshold values T_{FF}, T_{FB}, T_{BF}, and T_{BB} are calculated for the variance of camera noise σ^2. Fig. 4 shows the result of the calculation, which can be utilized in determining the threshold values in an actual application.

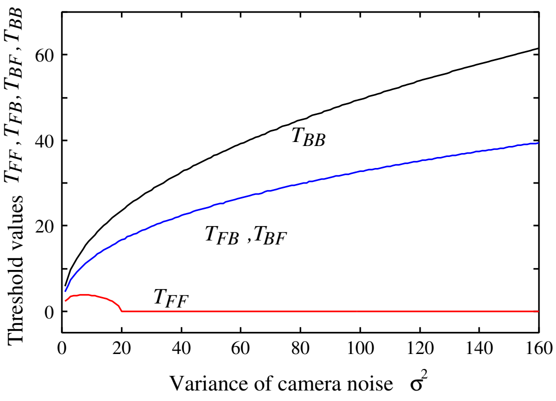


Fig. 4 : Calculated optimal threshold values.

4. EXPERIMENTAL RESULTS

This section gives several experimental results by the proposed technique and compares them with those obtained by the fixed threshold approach.

Three sets of test sequences were prepared, each of which consists of 50 static background frames and a single target frame including foreground object. Figs. 5 and 6 show the test sequences I and II, respectively. For the test sequence III, we used the noise-added version of the test sequence II, where Gaussian noise with zero mean and $\sigma^2 = 60$ was added independently to each frame of the test sequence II.

For each test sequence, the average background frame and the optimal thresholds were obtained in the following way. First, the 50 static backgrounds were averaged to obtain the average background frame, which is shown in Fig. 5 (a) and Fig. 6 (a). By using the same set of the static backgrounds, the variances of R, G, and B components were independently calculated for each pixel. The maximum value of all the calculated variances was then extracted as the variance of camera noise σ^2. Finally, the optimal thresholds T_{FF}, T_{FB}, T_{BF}, and T_{BB} were obtained from Fig. 4. Table. 1 lists the variance of camera noise and the optimal threshold values obtained for each of the test sequences I, II, and III.

The results by the proposed technique are shown in Fig. 5, 6, and 7 (c). To make a comparison, the results by the fixed threshold approach are also given in Figs. 5, 6, and 7(d), where the threshold T was determined optimally by minimizing the error

$$E = \frac{\#(\text{error pixels})}{\#(\text{object pixels in the ideal mask})}\,. \tag{10}$$

As the ideal mask in Eq. (10), a manually segmented object mask was utilized.

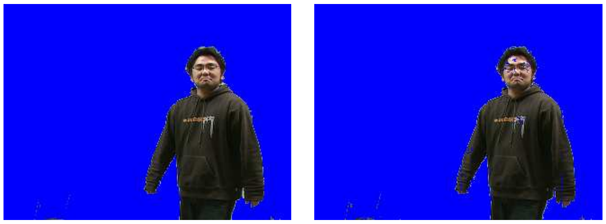
Table 2 summarizes the errors E evaluated for the proposed technique and the fixed threshold approach. Figs. 5–7 as well as Table 2 indicate how the proposed technique improves the accuracy of extraction. This clearly shows effectiveness of the proposed technique.

(a) Background frame (b) Target frame

(c) Proposed (d) Fixed threshold

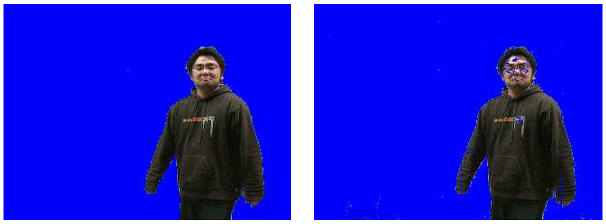
Fig. 5 : Results for the test sequence I.

(a) Background frame (b) Target frame



(c) Proposed (d) Fixed threshold

Fig. 6 : Results for the test sequence II.



(c) Proposed (d) Fixed threshold

Fig. 7 : Results for the test sequence III (Noise added).

Table 1 : Variance of camera noise σ^2 and the optimal threshold values.

Test seq.	σ^2	T_{FF}	$T_{FB} = T_{BF}$	T_{BB}
I	44	1	23	34
II	41	1	23	33
III	152	1	39	60

Table 2 : Error E in [%] for the proposed technique and the fixed threshold approach.

Test seq.	Proposed	Fixed threshold (optimal T)
I	1.81	8.50 (18)
II	2.02	2.95 (28)
III	1.92	3.35 (36)

5. CONCLUSIONS

This paper proposed a novel background frame differencing technique that referes to decisions previously made for neighboring pixels In the proposed algorithm, the threshold value for discrimination is varied in accordance with the status of the reference pixels. Although the introduction of the reference pixels actually reduces the isolated decision errors, it tends to produce error propagation both in the object and background regions. By applying the proposed algorithm twice in mutually opposite directions and taking the intersection of these results, the problem was alleviated. Furthermore, the optimal threshold values for the proposed technique was also derived.

In the proposed algorithm, two pixels are currently used as the reference pixels. However, by optimizing the positions of the reference pixels, the accuracy of extraction can be further improved. This is left for future work.

6. REFERENCES

[1] G. L. Foresti, Petri Mä"hönen, and Carlo S. Regazzoni, "Multimedia video-based surveillance systems, " Kluwer Academic Publishers, Norwell, Massachusetts, 2001.

[2] J. Pan, C. Gu and M.-T. Sun, "An MEPG-4 virtual video conferencing system with robust video object segmentation," 2nd MPEG-4 Workshop and Exhibition, 2001.

[3] A. Aach and A. Kaup, "Statistical model-based change detection in moving video, " Signal Processing, vol. 31, pp. 165–180, 1993.

[4] A. Aach and A. Kaup, "Bayesian algorithms for adaptive change detection in image sequences using Markov random fields, " Signal Processing: Image Communication, vol. 7, pp. 147–160, 1995.

[5] R. Mech and M. Wollborn, "A noise robust method for 2D shape estimation of moving objects in video sequences considering a moving camera, " Signal Processing, vol. 66, pp. 203–217, 1998.

[6] Y.-W. Huang, S.-Y. Chien, B.-Y. Hsieh, and L.-G. Chen, "Automatic threshold decision of background registration technique for video segmentation, " Visual Communications and Image Processing 2002, pp. 552–563, 2002.