A FAST AND ACCURATE SCOREBOARD ALGORITHM FOR ESTIMATING STATIONARY BACKGROUNDS IN AN IMAGE SEQUENCE

Andrew H. S. Lai and Nelson H. C. Yung

Department of Electrical & Electronic Engineering

The University of Hong Kong

Chow Yei Ching Building, Pokfulam Road, Hong Kong

Tel : (852) 2857-6414 Fax : (852) 2559-8738 Email : hslai@eee.hku.hk

ABSTRACT

This paper presents a stationary background estimation algorithm for color image sequence. The algorithm employs the running mode and running average algorithms, which are two commonly used algorithms, as the estimation core. A scoreboard is used to kept the pixel variations in the image sequence and is used to select between the running mode or the running average algorithm in each estimation. Our evaluation results show that by selecting, intelligently, the estimation core between the two algorithms according to the scoreboard values, the proposed background estimation algorithm has excellent performance in terms of estimation accuracy and speed.

1. INTRODUCTION

Background estimation methods are not uncommon in areas such as video compression, visual traffic surveillance and other applications involving the extraction of a moving object from an almost stationary background[1-6]. Although many have attempted to characterize and extract moving objects directly[7-9], to be able to handle this complexity of moving objects can sometimes be overwhelming, requiring rigorous mathematical and analytical techniques which demand enormous computing resources, and yet, not always successful. For this reason, background estimation methods are being considered by a number of groups[2,3] because of the inherent simplicity in estimating the almost stationary background, and the relatively small computing resource involved. Most of these background estimation algorithms consider the statistical behavior such as the running mode or running averages of the image sequence, as the basis for estimation. Algorithms based on the running mode are reputed to be quite accurate but their computing speed is so slow that none of them are considered for real-time applications. On the other hand, running average algorithms are inherently fast but their estimation accuracy is poor. Therefore, the challenge is to research and develop an algorithm that can accurately estimate the stationary background in an image sequence, without requiring too much computing resources.

Using this as our goal, a new background estimation algorithm is proposed where a scoreboard is introduced to keep a record of the intensity variations of pixels between the current image and the previously estimated background. By assigning a positive score to a small variation and a negative score to a large variation, the score for each pixel accumulates and the score of a pixel is used for estimating the current background at the same location. When doing so, the weighted running average is calculated for positive score, and the running mode is calculated for negative score. The argument is that positive score indicates small variation, and therefore, weighted running average should be good enough to estimate the background, with reasonable accuracy. However, for large variation or negative score, only the running mode method would give acceptable accuracy. Assuming that there are more background pixels in an image, this proposed method would have similar accuracy to the running mode method, but at the computing speed close to the running average method. In fact, our evaluation results show that the proposed scoreboard algorithm has comparative estimation accuracy to the running mode algorithm, while it's computing speed is 15 times faster.

This paper is organized as follows: Section 2 discusses the running average and running mode methods; Section 3 describes the details of the proposed algorithm; Section 4 evaluates the proposed algorithm and the running mode and running average algorithm using a normal traffic image sequence, by considering their visual quality of the estimation, mean-square errors (MSE) and computing speed; and the paper is concluded in Section 5.

2. BACKGROUND ESTIMATION USING RUNNING MODE AND AVERAGE

In theory, a stationary background may be defined by the pixels that do not change much in intensity or color throughout the image sequence. To capture this *stationary* characteristic in pixels, numerous methods can be employed. One of them is to consider the probability density function of a pixel location (x, y) over the entire image sequence, as defined as

$$B_{N-1}(x, y) = \text{mode}\{I_k(x, y)|k = 0,...,N-1\}, \qquad ...(1)$$

where $B_{N-1}(x, y)$ is the estimated background pixel at (x, y); $I_k(x, y)$ is the kth image in a sequence of N images and

$$\text{mode}\{I_k(x, y)\} = \max\{\rho_{x,y}(r)|0 \leq r \leq (L-1)\}, \qquad ...(2)$$

where $\rho_{x,y}(r)$ is the number of pixels in (x, y) with intensity r and L is the number of intensity levels. In practice, the nth estimated background at (x, y) is defined as

$$B_{mode,n}(x, y) = \text{mode}\left\{I_k(x, y)\middle| k = \max\left[(n - M + 1), 0\right], \ldots, n\right\}$$

$$\ldots(3)$$

where M (≥ 1) is the number of previous images considered when calculating the mode. When $n < (M - 1)$, all the previous images are used for the calculation, i.e. $k = 0, \ldots, n$. Clearly, if more previous images are used to calculate the mode, the more accurate the estimation is. For the same reason, longer computing delay would result.

For the running average algorithm, instead of the mode, the average of the pixel values at (x, y) from the previous images is determined as the estimated background, as given below:

$$B_{avg,n}(x, y) = \frac{1}{n}\left[(n - 1) \cdot B_{avg,n-1}(x, y) + I_n(x, y)\right] \qquad \ldots(4)$$

where $B_{avg,0}(x, y) = I_0(x, y)$. In general, the running average algorithm is fast as only the averages are calculated. The problem lies in that the estimation error is large when there is rapid movement in the image. Only if we can harness the estimation accuracy offered by the running mode method and the speed of the running average method, then we would be able to challenge the problem as stated.

3.   SCOREBOARD ALGORITHM

The idea of the scoreboard algorithm is that if we can detect the pixels that are suffering from large intensity variations, then the running mode algorithm can be used to estimate the background, for better accuracy. Conversely, for pixels that do not have large variations, the running average algorithm is used for better speed. As majority of pixels would have small variations in most image sequences, both accuracy and speed are expected to be improved.

To implement the above, a scoreboard is introduced to keep a record of the intensity variations of pixels between the current image and the previously estimated background. By assigning a positive score to a small variation and a negative score to a large variation, the score for each pixel accumulates and the score of a pixel is used for estimating the current background at the same location. When doing so, the weighted running average is calculated for positive score, and the running mode is calculated for negative score. The argument is that positive score indicates small variation, and therefore, weighted running average should be good enough to estimate the background, with reasonable accuracy. However, for large variation or negative score, only the running mode method would give acceptable accuracy. A weighted running average (Eqt. 7) is employed here because it improves from the running average method. Figure 1 depicts the block diagram of the scoreboard algorithm.

Initial conditions of the scoreboard algorithm are that, $B_{scoare,0}(x, y) = I_0(x, y)$ and the score for each (x, y) is zero. To update the score for (x, y) in the scoreboard, $I_n(x, y)$ is compared with the previous estimated background, $B_{score,n-1}(x, y)$, according to the following equation:

$$d(x, y) = 1 - \frac{\left|I_n(x, y) - B_{score,n}(x, y)\right|}{T} \qquad \ldots(5)$$

where T is a threshold value chosen to accommodate the small intensity variations (10-20 in the intensity scale of 256) introduced to the image sequence as a result of the camera, rather than the motion due to the scene. Therefore, T is typically chosen to lie between 10 and 20. For convenience, the value of $d(x, y)$ is further bounded to $(-1.0 \ldots 1.0)$ by limiting $d(x, y)$ in the negative axis, i.e. if $d(x, y) < -1$ then $d(x, y) = -1$. Thus, $d(x, y)$ can be interpreted as:
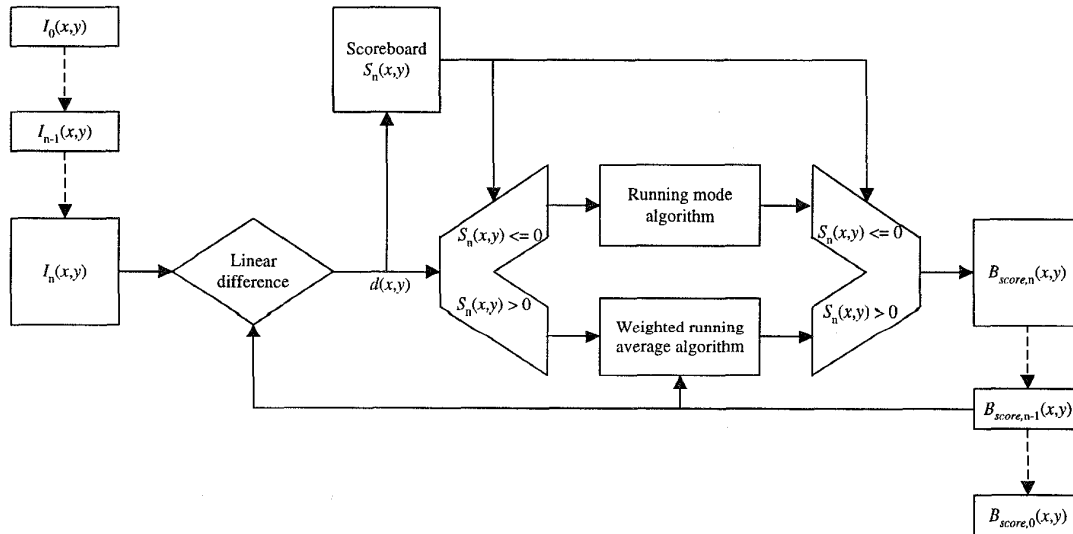

Figure 1 - Block diagram of the scoreboard algorithm.

Condition	Variation
$d(x,y)=1 \rightarrow I_n(x,y)=B_{score,n-1}(x,y)$	no
$1 < d(x,y) < 0 \rightarrow \left\| I_n(x,y) - B_{score,n-1}(x,y) \right\| < T$	small
$d(x,y)=0 \rightarrow \left\| I_n(x,y) - B_{score,n-1}(x,y) \right\| = T$	critical
$0 < d(x,y) < -1 \rightarrow \left\| I_n(x,y) - B_{score,n-1}(x,y) \right\| > T$	large

From Eqt. 5, the score for the current image $I_n(x,y)$ at (x,y) is defined as

$$S_n(x,y) = S_{n-1}(x,y) + d(x,y) \qquad ...(6)$$

Therefore, positive $S_n(x,y)$ indicates the intensity variation at (x,y) is small, where the weighted running average given in Eqt. 7 can be used for the estimation. In Eqt. 7, if the current score is negative (large variation), then the last estimated background is used instead of estimating a new one (Eqt 7(b)). If this repeats until $S_n(x,y)$ becomes negative, then the running mode method is used instead.

For $d(x,y) > 0$:

$$B_{score,n}(x,y) = \frac{1}{c_n(x,y)} \cdot \left[c_{n-1}(x,y) \cdot B_{score,n-1}(x,y) + d(x,y) \cdot I_n(x,y) \right]$$

where $c_n(x,y) = c_{n-1}(x,y) + d(x,y)$. $\qquad ...(7a)$

For $d(x,y) \le 0$:

$$B_{score,n}(x,y) = B_{score,n-1}(x,y)$$

where $c_n(x,y) = c_{n-1}(x,y)$. $\qquad ...(7b)$

On the other hand, negative $S_n(x,y)$ indicates a large variation at (x,y), where Eqt 3 is used for the estimation instead. Moreover, for pixels with large intensity variation, their cumulative scores can be reset to $S_n(x,y) = 1$ and $c_n(x,y) = 1$ so as to minimize the cumulative estimation errors.

4. EVALUATION

For evaluation purpose, the background of a traffic scene image sequence was being estimated by the proposed scoreboard algorithm and compared with the running mode and running average algorithms. There are altogether 70 frames in this short sequence of a busy road taken from a camera mounted on an overhead foot bridge. A typical frame of the image sequence is depicted in Figure 2, where five moving vehicles can be seen as well as the road work on the left. Only the vehicles are seen to be moving in this image, and the rest is consider to be the background. During the evaluation, the running mode algorithm has M set to 30, the running average algorithm used all the previous frames for estimation, while the scoreboard algorithm has T set to 15 and M set to 10.

Figure 3 depicts the 30th frame of the estimated backgrounds by the three algorithms. As seen in this Figure, all three
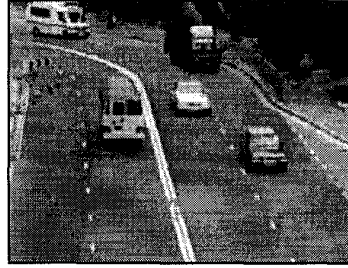

Figure 2 - Typical frame of the test sequence.

algorithms are able to estimate the background, although their differences in accuracy can be detected visually. For instance, the running averaged background is noticeably smoothed, particularly objects at a distance. The other two algorithms do not suffer from this, but rather, small amount of impulses can be identified alone the center line of the right hand lane. It appears that Figure 3(a) has more of these impulses, and not as smooth as Figure 3(c). Overall, the scoreboard background is slightly better than the running mode background, and the running average background is the worst.

Figure 4 depicts the 50th frame of the estimated backgrounds. The contrast between the results from the three algorithm becomes more obvious, with the running average algorithm having a rather smoothed background, while the running mode background has a lot of noise-like impulses on the center line of right lane. The scoreboard background appears to be slightly smoothed, but the overall quality is good.

Considering the mean-square errors (MSE) of the estimated backgrounds by the three algorithms using a reference background, Figure 5 shows their error functions over 70 frames. It can be observed that the running mode algorithm gives the smallest MSE (most accurate) almost for all cases. The scoreboard algorithm gives MSE close to the smallest, except for around the 30th frame. In fact, it's MSE is the smallest at around the 50th frame. For the running average algorithm, it's MSE is consistently poor, although the gap tends to narrow at higher frame number in this case.

The three algorithms were evaluated on a Pentium 100MHz machine under Linux and the dimension of the images are 320×240 pixels. The instantaneous computing delay for each frame was measured and the total delay determined. From the total delay, the average delay per frame for each algorithm was calculated and tabulated in the following:

	Running mode	Running average	Scoreboard
Avg. delay per frame	10.19 sec.	0.32 sec.	0.69 sec.

Table 1 - Average delay per frame

From Table 1, the running mode algorithm required over 10 seconds per estimation, while both the running average and scoreboard algorithms required less than a second. In fact, the running average is twice as fast as the scoreboard, even though it's results are not comparable. On the other hand, the scoreboard algorithm seems to provide the best solution to the background estimation problem as it is accurate and fast.
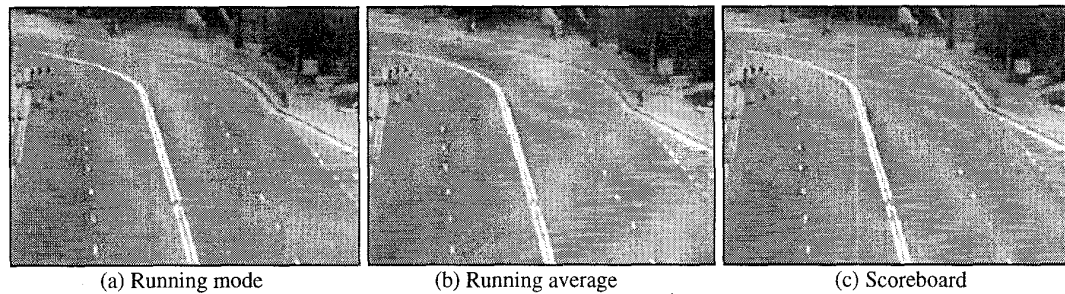
| (a) Running mode | (b) Running average | (c) Scoreboard |

Figure 3: Estimated background - 30th frame

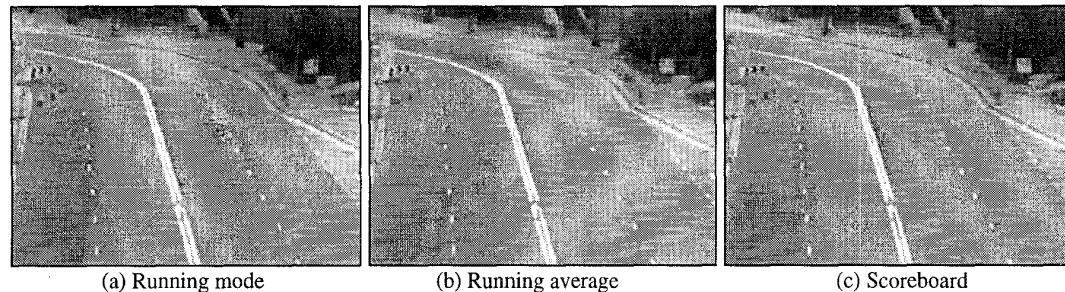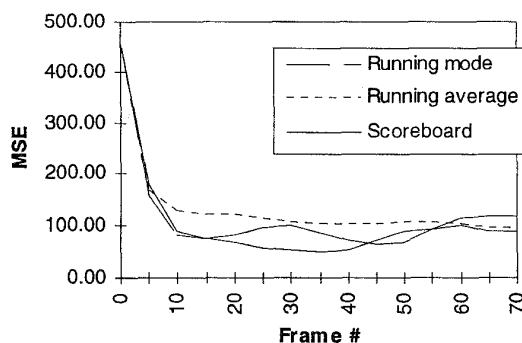| (a) Running mode | (b) Running average | (c) Scoreboard |

Figure 4: Estimated background - 50th frame

Figure 5 - MSE of the three background estimation algorithms

5. CONCLUSION

In conclusion, the scoreboard algorithm proposed for estimating the stationary background in an image sequence exhibits some positive and attractive properties that compares favorably with existing techniques using running mode or running average. The visual appearance of the background estimated by the scoreboard algorithm is on a par with the running mode estimated background, if not better, and it's MSE is almost as good as the running mode MSE. When computing delay is concerned, the delay required by the scoreboard algorithm is only twice that of the running average algorithm, but close to 15 times faster than the running mode algorithm. From these, we can conclude that the scoreboard algorithm satisfies the goal we were set out to achieve in the first place.

6. REFERENCES

[1] Neil Hoose, "Computer Image Processing in Traffic Engineering", *Research Studies Press Ltd.*, London, 1991.

[2] Brian Gloyer, Hamid K. Aghajan, Kai-Yeung Siu, and Thomas Kailath, "Video-based Freeway Monitoring System using Recursive Vehicle Tracking", *SPIE Proceeding*, Vol. 2421, pp. 173-180, 1995.

[3] J. Malik, S. Russell, J. Weber, T. Huang and D. Koller, "A Machine Vision Based Surveillance System for California Roads", *PATH Project MOU-83 Final Report*, University of California, November, 1994.

[4] A Neri, S. Colonnese, G. Russo, "Video Sequence Segmentation for Object-based Coders using Higher Order Statistics", *Proc. ISCAS'97*, vol. II, pp. 1245-1248, 1997.

[5] Aleksej Makarov, "Comparison of Background Extraction Based Intrusion Detection Algorithm", *Proc. ICIP'96*, vol. I, pp. 521-524, 1996.

[6] J. A. Nicholls, Don Monro, "Scaleable Video with Background Segmentation", *Proc. ICIP'96*, vol. II, pp. 529-532, 1996.

[7] Marie-Pierre Dubuisson Jolly and Anil K. Jain, "Contour Extraction of Moving Objects in Complex Outdoor Scenes", *International Journal of Computer Vision*, 14, pp.83-105, 1995.

[8] W. F. Gardner & D. T. Lawton, "Interactive Model-based Vehicle Tracking", *IEEE Trans. on PAMI*, Vol.18, No.11, Nov. 1996, pp.1115-1120.

[9] D. Koller, et al, "Model-Based Object Tracking in Traffic Scenes", *Proc. 2nd European Conference on Computer Vision*, Italy, May 1992, pp.437-452.