Journal of the Korean Institute of Illuminating and Electrical Installation Engineers Vol. 21, No.4, pp. 11~18 May 2007

Tracking Object Movement via Two Stage Median Operation and State Transition Diagram under Various Light Conditions

Goo-Man Park*

Abstract

A moving object detection algorithm for surveillance video is here proposed which employs background initialization based on two-stage median filtering and a background updating method based on state transition diagram.

In the background initialization, the spatiotemporal similarity is measured in the subinterval. From the accumulated difference between the base frame and the other frames in a subinterval, the regions affected by moving objects are located. The median is applied over the subsequence in the subinterval in which regions share similarity. The outputs from each subinterval are filtered by a two-stage median filter.

The background of every frame is updated by the suggested state transition diagram. The object is detected by the difference between the current frame and the updated background.

The proposed method showed good results even for busy, crowded sequences which included moving objects from the first frame.

Key Words : Background initialization, Background update, Moving object tracking

1. Introduction

The detection of motion in many tracking system, including the method described here, relies on the technique of background subtraction. By comparing incoming image frames to a reference background updated at every frame time, local regions which have changed can be efficiently located. The use of background subtraction is popular in motion detection and tracking. To identify the correct moving objects, background subtraction must be accurate. Therefore, the correct background initialization and background update are essential.

Problems that frequently occur in surveillance systems include camera perturbation on the road, sudden or gradual illumination changes, shadows from clouds, and moving objects at rest. Due to camera noise or gradual illumination changes, these changes occur in static video differently from normal motion. This means that the

* Main author : Dept. of Media Engineering, Seoul National Univ. of Technology
Tel : +82-2-970-6430, Fax : +82-2-974-6123
E-mail : gmpark@snut.ac.kr
Date of submit : March, 15, 2007
First assessment : March, 29, 2007
Completion of assessment : April. 12, 2007

background changes continuously and requires updating.

There are several types of backgrounds with respect to moving objects: a bare background without any moving objects, a background with normal moving objects, a background with initially motionless objects, and a background with idle objects in the middle of the sequence after normal motion. A motion tracking system must describe these different situations. Unmoving object causes difficulty in background updating. In an extreme case the inactive objects do not move during most of the video recording time, it can be erroneously regarded as background.

1.1 Previous work

Many different systems have been proposed for background initialization and updating in motion tracking problems.

Multiple hypotheses of the background value at each pixel are generated by locating periods of stable intensity in the sequence[1]. The likelihood of each hypothesis is then evaluated using optical flow information from the neighborhood around the pixel, and the most likely hypothesis is chosen to represent the background. A Local Image Flow algorithm is similar to adaptive smoothness in that it generates hypotheses by locating intervals of relatively constant intensity. When static objects exist in the scene, the initialization procedure may produce incorrect output. When objects are stationary for a long period of time, the algorithm uses information about motion in the vicinity of a pixel as an additional heuristic. The optical flow is considered for the neighborhood surrounding each pixel. If the direction of optical flow in the neighborhood is toward the pixel, then it is likely that a moving object is approaching the pixel. However, if the majority of optical flow is directed away from the pixel, it is likely that a moving object is leaving the area. A Local Image Flow algorithm is a kind of adaptive smoothness in that it generates hypotheses by locating intervals of relatively constant intensity. When stationary objects stay in place long enough to have maximum likelihood, the measured background model will likely be incorrect.

Stauffer and Grimson[2] proposed an adaptive on-line parametric color model in which the background color of each pixel is modeled as a multiple Gaussian mixture(MGM). This is one of the most commonly used approaches. They modeled the history of each pixel in the image by K Gaussian density distributions. Their system can accommodate lighting changes, slow-moving objects, and the introduction or removal of objects from the scene.

Heikkila and Pietilainen[3] proposed a modified threshold scheme based on the LBP(local binary patterns) model. Textural information is utilized when modeling a background and the LBP is selected as the measure of this texture. Due to the addition of offset to the threshold, the discriminative power of the LBP operator has increased.

The Median operation process has been employed for the problem of background initialization. However, the Median has a breakdown point of only 50[%]. Hanzi and David Suter[4] proposed a new, robust method which can tolerate more than 50[%] of noise and foreground pixels in the background initialization process.

C.R. Wren, et al[5] have assumed that the pixels, over a window of time at a particular image location, are Gaussian distributed. After the background value of the pixel is obtained, exponential smoothing is employed to update for slow or gradual change in the background scene. Such approaches do not address scenes with dynamic backgrounds or where foreground objects are present in the training stage.

1.2 Proposed Methods

The goal of this study is to create a robust, adaptive moving object detection system that is flexible enough to handle variations in lighting, background jitter, static objects, illumination changes and other factors.

This study proposes a new background initialization method based on the spatiotemporal similarity measures and a background updating method based on a state transition diagram.

The background initialization method must understand situations in which moving objects are already in the scene or in which there is only background. In background initialization, firstly, several subintervals which were independently inputted to a median filter were set. Before filtering, the quantity of accumulated motion was measured between the base frame and the rest frames in a subinterval, which indicated the selective filtering regions. Secondly, the output of each subinterval was filtered by a two-stage median filter.

The background of every frame was updated by a suggested state transition diagram. An object was detected by the difference between the current frame and the updated background.

The proposed method showed good results even for busy, crowded sequences which included moving objects from the first frame.

This paper is organized as follows. Section 2 explains the proposed background initialization method and its characteristics. A two stage median filtering that depends on motion quantity is introduced. In section 3, a background update method is suggested. Specifically, the state transition diagram is explained. Section 4 describes object detection by the subtraction method. The experimental results and analysis is given in Section 5. Finally, the system is summarized in section 6.

2. Background Initialization

Background initialization is an important fundamental procedure for tracking moving objects in video sequences. Uncovered background helps to find moving objects accurately. In many practical tasks, such as a busy road or in a public area, a real background image seldom exits. By background initialization, a real background not containing any moving objects was extracted.

2.1 Two stage median filtering

By applying the median operation over a long period of sequence, the initialized background can be obtained. This, however, contains salt-and-pepper noises along the motion trajectory. In order to avoid this, a training sequence was first divided into several subintervals. Regional motion quantity was measured between the base frame and the rest frames in a subinterval. The weighted motion quantity was accumulated in a motion template as shown in Fig. 1.

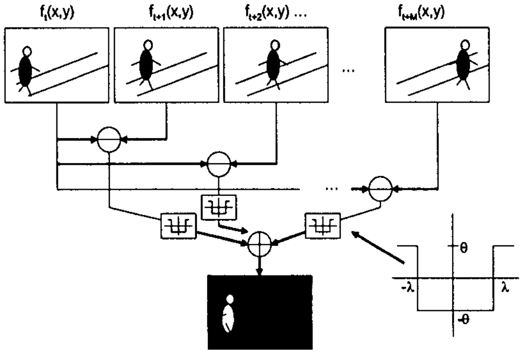


Fig. 1. Weighted sum of differences in each subinterval

This template indicates the regions that are selectively filtered. Based on the motion template, the most likely stable subsequence was selected in which the co-located regions share similar means and variances. The median filter was applied to each subinterval independently. A secondary filter was used for the temporary output from each subinterval.

2.2 Regional motion quantity measurement

The regional motion quantity between the base frame and the rest frames in a subinterval helped to selectively filter the regions which had high motion. Unnecessary median operations could be reduced by selecting the region. From the motion template only, consecutive frames in which co-located regions share similar mean and variance were selected. The illustration is given as follows.

Let $f_t(x,y)$ be the pixel value at the spatial coordinate (x,y) in time. $d_i(x,y)$ is the difference signal obtained by subtracting the first frame and the other frame in the interval as in (1).

$$d_i(x,y) = f_t(x,y) - f_{t+i}(x,y) \tag{1}$$

where t is the first frame time. i =1,...,M and j = 1,⋯,N. $f_{t+i}(x,y)$ is the ith frame in a subinterval. M is the number of frames in a subinterval and N is the number of subintervals.

By the threshold, a motion weighted value $y_i(x,y)$ in a frame is obtained as follows:

$$y_i(x,y) = \begin{cases} \theta, & d_i \geq \lambda \ or \ d_i \leq -\lambda \\ -\theta, & elsewhere \end{cases} \tag{2}$$

Let $D_j(x,y)$ denote the motion template in jth subinterval.

$$D_j(x,y) = \sum_{i=1}^{M} y_i(x,y) \tag{3}$$

$D_j(x,y)$ is obtained by adding or subtracting the weighted motion strength θ. If $D_j(x,y)$ is less than zero, let $D_j(x,y)$=0, otherwise, $D_j(x,y)$=255. This variable shows the position of moving objects in the background.

Removal of the small fractional regions or particles caused by random noise, and the region growth and segmentation therein resulted in P regions in which the background was covered by moving objects. P is the number of regions in the template. $R_{k,i}$ is the kth region in the template $D_j(x,y)$. k=1,⋯,P.

For all the co-located kth regions in a temporal direction, the mean $m_{k,i}$ and variance $\sigma^2{}_{k,i}$ of ith frame in jth subinterval was calculated. The subsequence was found in which the means and variances shared similar values within a tolerance bound. This subsequence $C(a,b)$ is obtained by (4):

$$\begin{aligned} C(a,b) = \{ & |m_{k,a} - m_{k,a+1}| \leq \gamma_1 \wedge |m_{k,a+1} - m_{k,a+2}| \leq \gamma_1 \\ & \wedge |m_{k,a+2} - m_{k,a+3}| \leq \gamma_1 \ \ldots \wedge |m_{k,b-1} - m_{k,b}| \leq \gamma_1 \\ & \wedge |\sigma^2{}_{k,a} - \sigma^2{}_{k,a+1}| \leq \gamma_2 \wedge |\sigma^2{}_{k,a+1} - \sigma^2{}_{k,a+2}| \leq \gamma_2 \\ & \wedge |\sigma^2{}_{k,a+2} - \sigma^2{}_{k,a+3}| \leq \gamma_2 \ \ldots \wedge |\sigma^2{}_{k,b-1} - \sigma^2{}_{k,b}| \leq \gamma_2 \} \end{aligned} \tag{4}$$

where, a<b, and a and b are in the range of 1,...,M. γ_1 and γ_2 are tolerance values. The longest time range l_{ab} was found as follows:

$$l_{ab} = \max\{b-a\} \tag{5}$$

Apply the median filter for this range and obtain the initialized background Bj(x,y). This was repeated for all k.

$$B_j(x,y) = \arg\min \sum_{\substack{v=a \\ u=a\ to\ b}}^{b} distance\{f_u(x,y), f_v(x,y)\} \tag{6}$$

As short subintervals as possible were used. If static objects are present, then the background does not produce satisfactory results. To overcome the problem of static objects, (7) was applied for all the temporary background $B_j(x,y)$ where $j=1,\cdots,N$.

$$B_{init}(x,y) = \arg\min \sum_{\substack{v=1 \\ u=1\ to\ N}}^{N} distance\{B_u(x,y), B_v(x,y)\} \tag{7}$$

Then the desired initialized background $B_{init}(x,y)$ is obtained.

3. Background Update

A background update in each frame time is essential to obtain the moving objects of a sequence. As the background is updated more accurately, the object can also be more accurately determined. Even though the neighbor frames in a temporal direction have a high correlation, there are small changes which should be taken into consideration. The changes in the scene are caused by the illumination change, sudden motion intrusion, shadows, swaying branches and leaves blown by the wind, among other factors.

A pixel based motion detection and background update method was proposed as most background updating techniques operate at the pixel-level. Once the change in grey value in each pixel was measured between the initialized background and input frame, the meaning of the changes could be classified into three states. These states describe one of the following events whether the transition was from background to background, from foreground to background, or background to foreground, or a transition between two different foreground objects. The block level motion information is utilized in pixel level background updating.

3.1 Changes caused by random noise

Meaningful changes must be distinguished from those which are affected by noise sources. Of the two types of changes, the first one is due to the moving objects in the scene; the other is from random noise sources such as irregular light sources or the thermal noise of the camera. The luminance changes for a 16x16 block were calculated and a true motion block was distinguished.

If the number of pixels whose difference is larger than a threshold exceeds the predefined percentage in each block, it was regarded as a motion block. A block with dense random noise can be mistakenly regarded as a motion block, as well.

After selecting the blocks having large differences between consecutive frames, the difference signal from these were scanned in rectangular order from left to right in a row and from upper to lower lines. The difference signal indicates the difference between two blocks. The distance between the two large difference values is accumulated within the block. The sum of all the distances is divided by the number of pixels which are larger than the threshold. If the result is larger than another threshold, then the block is of true motion.

3.2 State transition diagram

Every pixel is labeled according to its own changed pattern. There are three states. The state transition diagram is shown in Fig. 2.

There are eight different situations concerning the changes in background, as indicated by the arrows. Depending on these conditions, pixels in the background hold current information or copy a new value from the input frame. The three states are as follows:

S_0 : at time t-1, no change;
at time t, no change.
True background continues and no motion occurred.

S_1 : at time t-1, no change;
at time t, change.
A moving object is newly detected.

S_2 : at time t-1, change;
at time t change.
Objects are moving continuously.

In Fig. 2, '0' and '1' represent 'no change' and 'change', respectively. The 'h' represents 'holding the current information' and, conversely, 'c' means 'copying the pixel value of the current frame'.

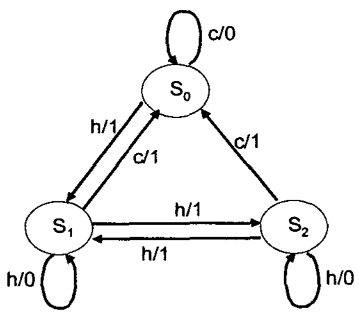


Fig. 2. The state transition diagram that describes the change patterns in a scene

These states have the following relationship and continue the transition of states within the diagram.

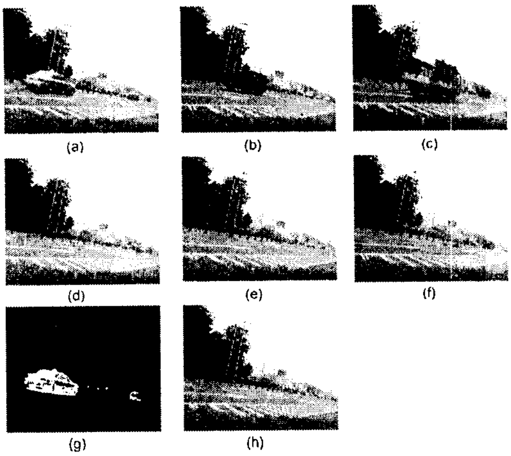
4. Moving Object Separation

After the background update, the foreground was obtained by separating the background from each input. The regions of the input frame that do not match to the background are considered foreground. By simply subtracting the background image from the input frame, the foreground can be obtained. However, when a low quality camera was used, scattered noises was visible in the foreground; this made object description inaccurate. In this plan, the shape feature was used to eliminate noise regions.

5. Experimental Results and Analysis

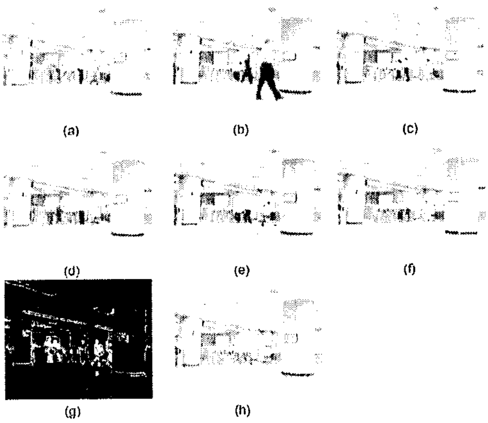
The plan was applied in various surveillance video sequences that represented environments such as offices, busy roads and crowded places. The video sequence had a resolution of 352×288 and a rate of 15 frames per second obtained by a PC camera. In the experiment, 7 subintervals and 50 frames per subinterval were used. λ was set to 15, θ to 5, γ_1 to 2 and γ_2 to 2. Fig. 3 shows the result of background initialization for the road sequence. In this sequence, there were many passing vehicles, but this method displayed good results. Because these objects moved rapidly regardless of their dense occupation of the scene, they were easily removed. Fig. 3 (g) shows the example of motion template for the first subinterval.

Fig. 4 shows the results of a sequence in a subway station. This video contains a large number of people crossing as well as some people at the entrance looking for their tickets who remain stationary for several seconds. The output of the first stage shows the initialization errors in Fig. 4 (d), Fig. 4 (e) and Fig. 3 (f). The static objects remain for a longer time than the duration of a subinterval. By the help of the second stage median, those errors are removed as shown in Fig. 4 (h).



(a) frame 0, (b) frame 150, (c) frame 200,
(d) initialized background in subinterval 0,
(e) initialized background in subinterval 3,
(f) initialized background in subinterval 4,
(g) motion template for initialization in subinterval 0,
(h) The result of background initialization after the second stage median

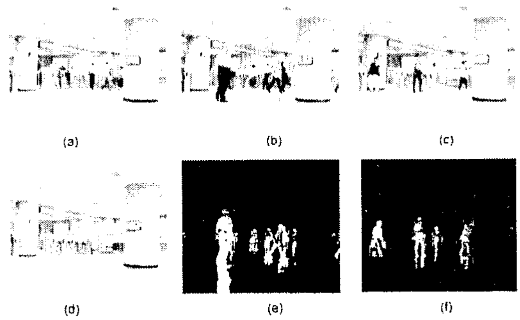
Fig. 3. The result of background initialization (road sequence)



(a) frame 0, (b) frame 150, (c) frame 200,
(d) initialized background in subinterval 0,
(e) initialized background in subinterval 3,
(f) initialized background in subinterval 4,
(g) motion template for initialization in subinterval 0,
(h) The result of background initialization after the second stage median

Fig. 4. The result of background initialization (subway station)

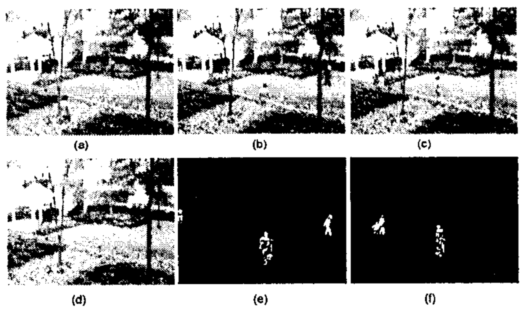
In the background update simulation, this plan retained the true background images over the whole sequence. In Fig. 5, the updated background were used to detect moving objects in the subway sequence as shown in Fig. 5 (e) and Fig. 5 (f).



(a) frame 0, (b) frame 94, (c) frame 429,
(d) initialized background,
(e) detected object of (b),
(f) detected object of (c)

Fig. 5. The subway station sequence

In Fig. 6, the updated background was used to detect moving objects for the garden sequence as shown in Fig. 6(e) and Fig. 6(f).



(a) frame 0, (b) frame 85, (c) frame 371,
(d) initialized background,
(e) detected object of (b),
(f) detected object of (c)

Fig. 6. The garden sequence

Through the results of the above simulation, the satisfactory performances of the proposed background initialization, background update, and the moving object detection were proven.

6. Conclusion

In this paper, a new method for background initialization and background update as well as an object detection method was presented.

For the background initialization method, a two stage median filtering algorithm based on spatiotemporal similarity was suggested. Much of the computation focused on the area of motion. The two stage median operation proved to be effective in cases in which some subinterval outputs include static objects because of the short length of the training sequence. This method displayed excellent results even in crowded sequences which had no true background for the duration of the training period.

A background update method using a state transition diagram, which showed the true background over the whole sequence, was proposed.

The proposed background initialization and update method produced accurate object detection outputs as a result.

The sequences used in this experiment show that this algorithm performs excellently in practical and real situation even in video sequence with a variety of noise. The experimental result showed its robustness despite noise. This algorithm can be used in many real situations where a true background does not exist at any given moment.

For future work, a successful selection rate of subsequence that represents a true background in the background initialization step should be increased.

References

[1] D. Gutchess, et al., "A Background Model Initialization Algorithm for Video Surveillance," Proceedings of Eighth IEEE International Conference on Computer Vision, 2001, ICCV 2001, Vol.1, pp.733-740, July, 2001.

[2] C. Stauffer, W.E.L. Grimson, "Adaptive Background Mixture Models for Real-time Tracking," Proc.IEEE Conference on Computer Vision & Pattern Recognition, vol.2, pp.246-252, June, 1999.

[3] Marko Heikkila, and Matti Pietikainen, "A Texture-Based Method for Modeling the Background and Detecting Moving Objects," IEEE Trans. Pattern Analysis and Machine Intelligence, vol.28, no.4, pp.657-662, April, 2006.

[4] Hanzi Wang and David Suter, "Background Initialization with A New Robust Statistical Approach," Proceedings of 2nd Joint IEEE International Workshop on VS-PETS, pp.153-159,October,2005.

[5] C.R. Wren, et al., "Pfinder: Real-time Tracking of the Human body," PAMI 19(7): pp.780-785, 1997.

Biography

Goo-Man Park

BS. Electronics Engineering, Hankuk Aviation University, Kyunggi, Korea(Mar. 1980~Feb. 1984).
MS. Electronics Engineering, Yonsei University, Seoul, Korea(Mar. 1984~Feb. 1986).
Ph.D Electronics Engineering, Yonsei University, Seoul, Korea(Mar. 1986~Feb. 1991).
Associate Professor(Aug. 1, 1999~Present): Dept. of Media Engineering, Seoul National University of Technology, Seoul, Korea.
Research Scholar(Jan. 31, 2006~July. 31, 2007): Dept.of Electrical and Computer Engineering, Georgia Institute of Technology.
Editor-in-chief(Jan. 1, 2003~Dec., 2004): Korea Society Broadcast Engineers Magazine.
Assistant Professor(Sep. 1, 1996~July. 31, 1999): Dept. of Electronics Engineering, Honam University, Kwangju, Korea.
Senior Researcher(Mar. 1, 1991~Aug. 31, 1996): Signal Processing Lab., Samsung Electronics Co. Suwon, Korea.