레터논문-07-12-3-09

Background Initialization by Spatiotemporal Similarity

Gooman Park[a)‡]

Abstract

A background initialization algorithm based on the spatiotemporal similarity measure in a motion tracking system is proposed. 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 co-located regions share the similarity.

The outputs from each subinterval are filtered by second stage median filter. The proposed method showed good results even in the busy and crowded sequences where the real background does not exit.

Keyword: moving object tracking, background, motion detection

1. Introduction

Many motion detection systems rely on the technique of background subtraction. By comparing incoming frames to a reference background updated at every frame time, some local regions which have changed are efficiently located. Furthermore, the background update and the moving object detection strongly depend on how correct the background initialization is. Our goal is to create a robust and adaptive background initialization method flexible enough to handle variations in lighting, background jitter, and static objects. We have proposed a new background initialization method based on spatiotemporal similarity and two stage median filtering.

There are previous works regarding the background initialization that use the subinterval similarity[1,2,3]. In a training sequence, some stable intervals are located, and then the most likely one is selected for initialization. These methods only use pixel information, but we used the regional information which helps to find the more accurate stable interval. Stauffer and Grimson[4] 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 deal with lighting changes, slow-moving objects, and introducing or removing objects from the scene. But this system takes longer time to achieve the true background image.

Firstly, we set up several subintervals and measure the regional motion quantity between a base frame and the other frames in the subinterval. The most likely stable subsequence is selected in the subinterval in which co-located regions share the similarity. The median is applied to individual pixels in this subsequence. Secondly the outputs of each subinterval are filtered by second stage median.

a) 서울산업대학교 공과대학 매체공학과
Department of Media Engineering, Seoul National University of Technology
‡ 교신저자 : 박구만(gmpark@snut.ac.kr)

2. Background initialization by regional similarity measure

If we apply the median filter over the long period of sequence, we can obtain the initialized background, but it contains pepper and salt noises in the vicinity of moving objects. To avoid it, we divide a training sequence into several subintervals at first. Regional motion quantity is measured between base frame and the rest frames in a subinterval. The weighted motion quantity is accumulated in the motion template as shown in Fig.1.

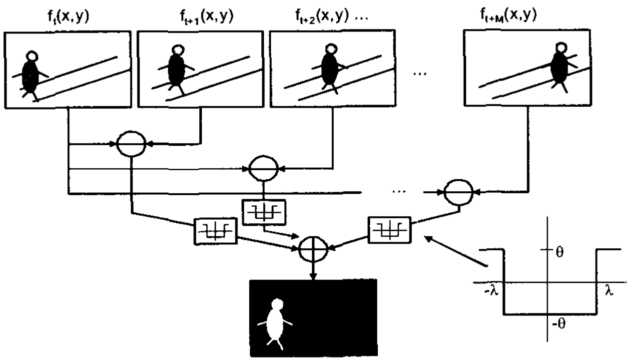

Fig.1. Weighted sum of differences in each subinterval

This template indicates the regions to be filtered in a selective way. Based on the motion template, the most likely stable subsequence is selected in which the co-located regions share the similar means and variances. The median filter is applied for each subinterval independently. We use secondary filter for the temporary output from each subinterval.

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

$$d_i(x,y) = f_0(x,y) - f_{0+i}(x,y) \tag{1}$$

where 0 is the first frame, and i =1,...,M $f_{0+1}(x,y)$ is the i th frame in a subinterval. M is the number of frames in a subinterval.

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. j = 1,⋯,N and N is the number of subintervals.

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

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

After removing the small fractional regions or particles caused by random noise, and after the region growing and segmentation there result 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 k_{th} region in the template $D_j(x,y)$. k = 1,⋯,P.

For all the co-located kth regions in temporal direction, we calculate the mean $m_{k,i}$and variance $\sigma^2_{k,i}$ of ith frame in jth subinterval. The subsequence is determined when all the frames starting from frame a to frame b in a consecutive way satisfy the following condition expressed by (4):

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

where, a<b, and a and b are in the range of 1,...,M. γ_1

and γ_2 are the tolerance values. We find the longest time range lab as follows:

$$l_{ab} = \max\{b-a\} \qquad (5)$$

We applied the median filter for this range and obtained initialized background $B_j(x,y)$. We repeated it for all k.

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

We used as short subintervals as possible. If there are static objects, then the background does not produce satisfactory results. To overcome static object problem, we applied (7) for all the temporary background Bj(x,y) where j = 1,⋯,N.

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

Then the desired initialized background $B_{init}(x,y)$ is finally obtained as in (7).

3.Experimental Result and Analysis

We applied the scheme in various surveillance video sequences that represent those environments such as offices, busy roads and crowded places. The video sequence has the resolution of 352x288 and rate of 15 frames per second. In the experiment, we used the 7 subintervals and 50 frames per subinterval. We set λ to 15, θ to 5, γ_1 to 2 and γ_2 to 2. Fig.2 shows the result of background initialization for the road sequence. In the sequence, there are many vehicles passing, but our method showed good results. Because they move fast regardless of their dense occupation in the scene, they were removed easily. Fig.2(g) shows the example of motion template for the first subinterval. Fig.3 shows the result for the subway station sequence.

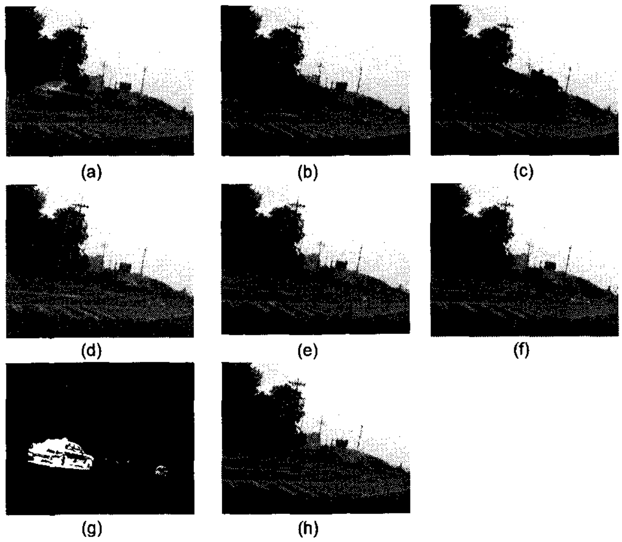

Fig.2 The result of background initialization(road sequence) (a) frame 0, (b) frame 150, (c) frame 200, (d) initialized background in subinterval 0, (e) initialized background in subinterval 3, (f) initialized background in subinterval 4, (g) motion template for initialization in subinterval 0, (h) The result of background initialization after second stage median.

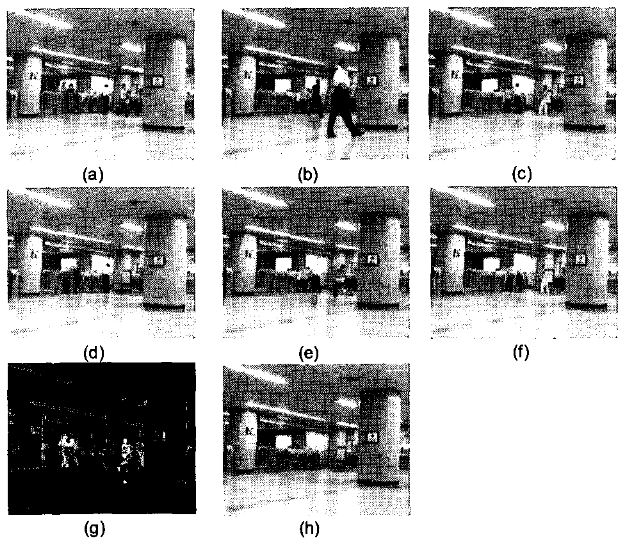

Fig.3 The result of background initialization(subway station) (a) frame 0, (b) frame 150, (c) frame 200, (d) initialized background in subinterval 0, (e) initialized background in subinterval 3, (f) initialized background in subinterval 4, (g) motion template for initialization in subinterval 0, (h)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

This video contains a large number of people crossing and some unmoving people at the entrance looking for their tickets for several seconds. The output of first stage shows initialization errors in Fig.3(d), Fig3.(e) and Fig.3(f). The static objects stay for longer time than the duration of a subinterval. By help of the second stage median, those errors are removed as shown in Fig.3(h).

Fig.4 shows the results for background initialization between the two methods. The simple median filtering contains some salt and pepper noise.

(a) Proposed initialization method (b) Simple median filtering over whole interval

Fig.4 Comparison of initialized backgrounds

4. Conclusion

We suggested two stage median filtering algorithm based on spatiotemporal similarity. Much of the computation is focused on the motion area. The two stage median operation proved to be effective in case that some subinterval outputs include the static objects in it because of the short length of training sequence. Our method showed excellent results even in the crowded sequence which has no true background all over the training period. For the future work, we would increase the successful selection rate of subsequence that represents true backgrounds.

References

[1] 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.

[2] 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.

[3] W. Long and Y.H. Yang, "Stationary Background Generation: An Alternative to the Difference of Two Images," Pattern Recognition, vol.23, no.12, pp.1351-1359, 1990.

[4] 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.