ROBUST BACKGROUND MAINTENANCE FOR DYNAMIC SCENES WITH GLOBAL INTENSITY LEVEL CHANGES

Kapje Sung Youngbae Hwang In-So Kweon
Department of Electrical Engineering, KAIST, Daejeon, Korea
kjsung@rcv.kaist.ac.kr unicorn@rcv.kaist.ac.kr iskweon@kaist.ac.kr

Abstract – In background subtraction, the maintenance of significant backgrounds is very critical under various scene changes. In this paper, we propose a background maintenance method for dynamic scenes including global intensity level changes caused by the changes of illumination conditions and camera settings. If global intensity level changes abruptly, the conventional background model cannot discriminate true foreground pixels from the background. The proposed method immediately modifies the background model based on the estimation of level changes by mean-shift. Saturated pixels are handled by an additional scheme. In real-world experiments for dynamic scenes, our proposed method outperforms previous methods by adaptive background maintenance and handling of saturated pixels.

Keywords – Background maintenance, background subtraction, non-parametric estimation

1. Introduction

Background subtraction is widely used for various applications such as tracking moving objects, video surveillance systems and traffic monitoring. It maintains background images or background models to classify new observations as background parts or foreground parts. Since the classification of foregrounds totally depends on the correctness of the background model, the maintenance of the background model is the very crucial issue.

There have been many works that tried to detect true foregrounds with suppression of false alarms. Wren et. al. [1] used a single Gaussian model to represent the distribution of the background. However, it is insufficient to represent the background intensity by using only one Gaussian distribution when the background is complex. To deal with complex backgrounds, Stauffer and Grimson [2] proposed frameworks using mixture of Gaussians. This approach can have multiple hypotheses for the background so that it can be adapted for complex scenes such as waving trees, streaming waters and refreshing monitors. In [3], Elgammal et. al. proposed a non-parametric approach to represent backgrounds by recent background samples. To maintain the recent background samples, they updated the background by combining selective and blind methods. Mittal and Paragios [4] combined the non-parametric approach with motion information as optical flow to deal with persistent dynamic behavior in time.

The problem of previous works is that they cannot deal with abrupt global intensity level changes because their background update schemes generally assume that the background is gradually changed. However, there are various global level changes varied by illumination conditions and camera settings (camera gain, aperture and shutter speed, etc.). In that situation, conventional methods have many false alarms because previous updated backgrounds become insignificant. Although the further update process recovers the background model, it needs sufficient frames.

In this paper, we propose a background maintenance method for dynamic scenes including global intensity level changes. The proposed method immediately modifies the background model based on the estimation of intensity level changes. Although foreground pixels cannot support to estimate level changes, we find significant ratio between previous and current frames by mean-shift iteration. Since there are saturated pixels when the pixel intensity exceeds the dynamic range, we handle those pixels by an additional scheme. We test our proposed method when camera settings and illumination conditions often change.

This paper is organized as follows. In Section 2, we explain our proposed background maintenance method which consists of intensity level estimation and saturation pixel handling. Section 3 shows some experimental results to validate the proposed approach applied to image sequences in the indoor environment. Finally, we present our conclusions in Section 4.

2. PROPOSED MODEL

2.1 Background modeling using kernel density estimation

The background model in kernel density estimation [3] is non-parametric. It does not need parameter estimation. It does not have a parametric form contrary to mixture of Gaussians [2]. It constructs a probability density function with recent background samples for every pixel as:

$$\Pr(x_t) = \frac{1}{N}\sum_{i=1}^{N}\prod_{j=1}^{d}\frac{1}{\sqrt{2\pi\sigma_j^2}}e^{-\frac{1}{2}\frac{(x_{t_j}-x_{i_j})^2}{\sigma_j^2}} \tag{1}$$

where N is the number of samples, d is the number of color channels, x_{i_j} is background sample, σ is kernel bandwidth and x_{t_j} is value of j-th dimension in color space for x_t.

Using this kernel density estimator, a new pixel is considered a foreground pixel if $\Pr(x_t) < threshold$ where the threshold is a global threshold over entire image.

2.2 Handling of global intensity level changes

We considered the global intensity level changes as a ratio between previous frame and current frame, g_j. If we know the ratio, g_j, between two successive frames, we can update the background samples by multiplying g_j. This process adapts the previous samples to current intensity level. Because sample intensities are multiplied by g_j, standard deviation of noise will be changed with same rate. Thus, standard deviation should be also multiplied by g_j. Then modified kernel density estimation function is as follows,

$$\Pr(x_t) = \frac{1}{N}\sum_{i=1}^{N}\prod_{j=1}^{d}\frac{1}{\sqrt{2\pi(g_j \cdot \sigma_j)^2}} e^{-\frac{1}{2}\frac{(x_{t_j} - g_j \cdot x_{i_j})^2}{(g_j \cdot \sigma_j)^2}} \quad (2)$$

The g_j can be estimated by averaging all ratios of each pixel if all pixels belong to the background. But if there are foreground regions, we have to ignore the foreground region pixels when we calculate average ratios. So if we know the background region exactly, g_j can be estimated as,

$$g_j = \frac{1}{M}\sum_{n \in Background}\frac{x_{t_j}^n}{x_{t-1_j}^n} \quad (3)$$

where M is the number of background pixels and n is pixel index.

But it is impossible to find the foreground region when abrupt global intensity level change is occurred. Many pixels which do not belong to foreground region will be considered as foregrounds in this situation. To solve this problem, we used mean-shift algorithm [5]. First, we get the intensity ratios r_j^i between previous frame and current frame for every pixel. And we compute the mean of ratios iteratively,

$$m(r) = \frac{\sum_{i=1}^{M} r_j^i K\left(\frac{|r - r_j^i|}{h}\right)}{\sum_{i=1}^{M} K\left(\frac{|r - r_j^i|}{h}\right)} \quad (4)$$

where $K(\cdot)$ is kernel function, h is size of the kernel and r is current mean. At the each iteration, we get a mean of samples in the current kernel. Next iteration, we move the kernel to the mean we get at the previous iteration. We repeat this process until the kernel doesn't move any more. Then final center of the kernel is result. Because kernel size is limited, samples outside the kernel are ignored. So we can eliminate the effect of foreground region.

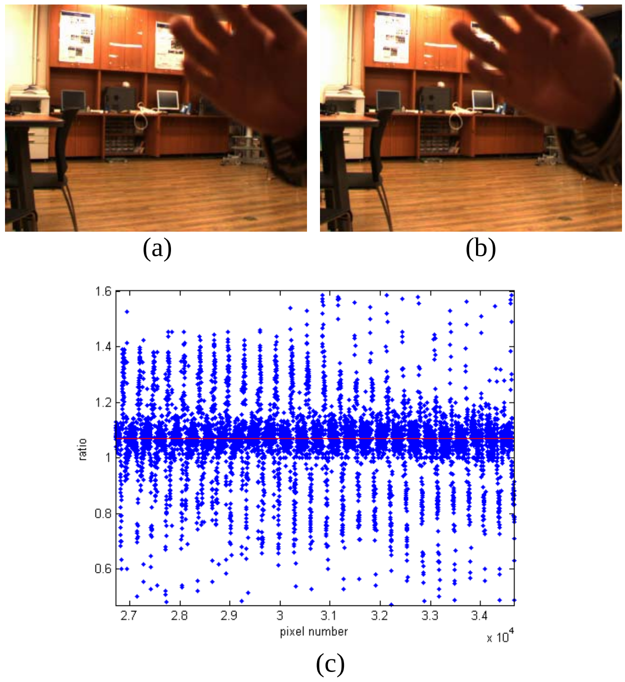


Fig. 1. R value ratio estimation result. (a),(b) successive two frame including ratio change; (c) blue points are ratios of R values, red line is estimated ratio.

Fig.1 shows the R value ratio estimation result. There are many incorrect ratios because of large foreground regions. But we can estimate the reasonable ratio with the proposed method. The estimation of ratios for R, G and B values between the previous frame and the current frame is shown in fig.2. One person comes and goes 4 times in the sequence and there are 4 big jumps of g_j's at the each time.

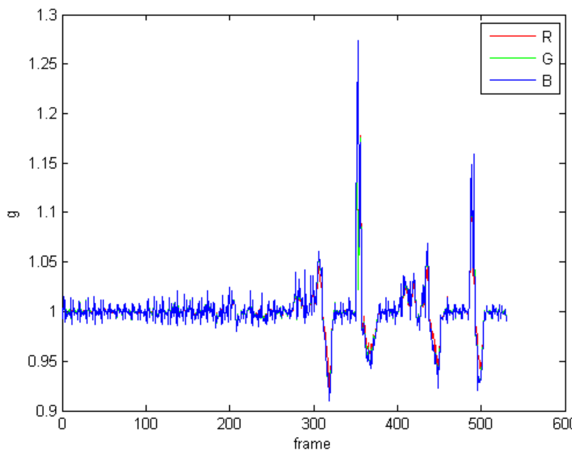


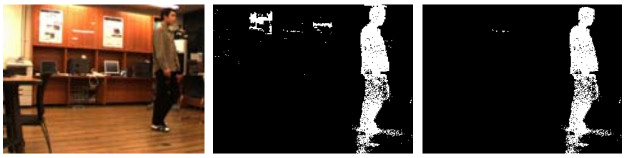
Fig. 2. Ratio between previous frame and current frame.

2.3 Saturation Problem

When g_j is lower than 1, we cannot estimate exact values of saturated samples. After ratio normalization, a saturated sample can have range of $[255g_j\ 255]$. So we modified the kernel function for saturated samples when g_j is lower than 1.

$$K(x_{t_j} - x_{i_j}) = \begin{cases} \dfrac{1}{\sqrt{2\pi\sigma_j^2}} e^{\frac{1}{2}\frac{(x_{t_j} - g_j \cdot x_{i_j})^2}{\sigma_j^2}} & , x_{t_j} < g_j \cdot x_{i_j} \\ \dfrac{1}{\sqrt{2\pi\sigma_j^2}} & , x_{t_j} \geq g_j \cdot x_{i_j} \end{cases} \quad (5)$$

By using this modified kernel function, we can improve result of bright regions. In fig.3-(b), the result of the poster on a bookcase is bad because it has high intensity. But these false alarms were reduced in fig.3-(c) with proposed method.

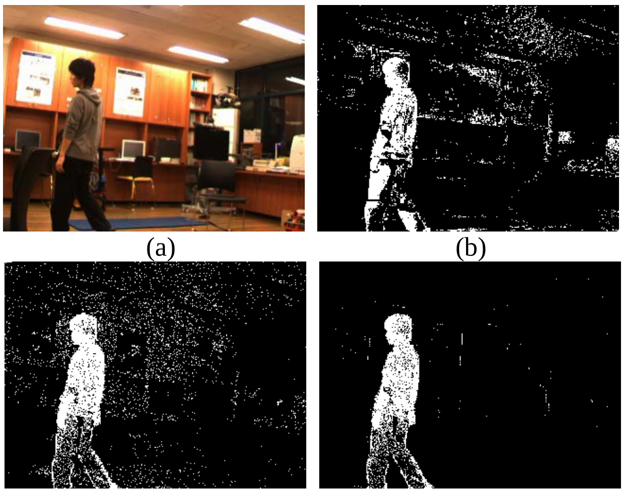


(a) (b) (c)

Fig. 3. (a) Original Image; (b) result with saturation problem; (c) result after saturation problem solved.

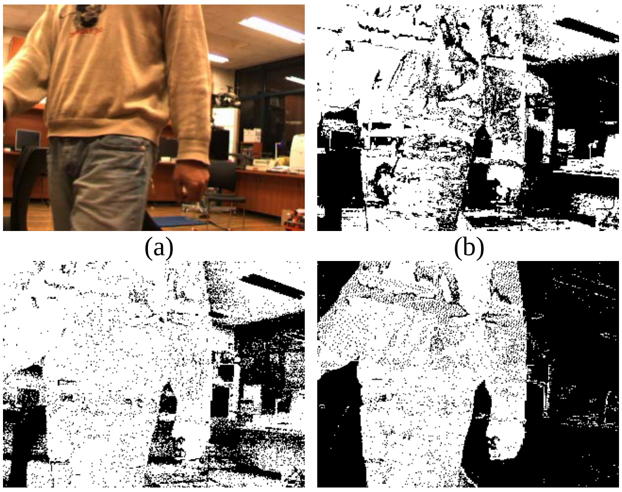
3. EXPERIMENTAL RESULTS

All experiments were conducted under automatic camera setting changes. When a person appears in a scene or disappears from a scene or moves toward the camera, global intensity level is changed rapidly. We used



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

Fig.4. (a) A frame with abrupt intensity level decrease. detection results using (b) Mixture of Gaussians model, (c) Kernel Density Estimation, (d) proposed method.



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

Fig.5. (a) A frame with abrupt intensity level increase and large foreground. detection results using (b) Mixture of Gaussians model, (c) Kernel Density Estimation, (d) proposed method.

sequences taken from a Flea camera and image size is 320×240. A threshold of kernel density estimator is set to 10-8 for all experiments. The uniform kernel is used for mean-shift and kernel radius is 0.1.

We employed two widely-used methods, Mixture of Gaussians [2] and Kernel Density Estimation [3] to compare with the proposed method. For fare comparison, we did not perform post processing for all experiments.

Fig.4 shows detection results in case of abrupt intensity level decrease. Because one person disappeared from the scene, intensity average was increased. So global intensity level was decreased automatically. In this situation, other two methods cannot recover the sudden intensity change immediately because the background model is not changed. But our method changes the previous samples to adapt current intensity condition. So it can catch the change right after intensity level change. But there are missing pixels in the leg region of a person. Because his pants are black and homogeneous, the system considers the black color as a background.

Fig.5 shows results when global intensity level was increased. In this sequence, a person moved toward camera. Although the foreground region is larger than background region, we can find the proper ratio with proposed method. The number of false positive pixels of our method is smallest. And the number of missing pixels is a little poor than Kernel Density Estimation model.

Fig.6 and fig.7 also show another results. The distribution of missing pixels for kernel density estimation model and proposed method are very similar. It is because proposed method also uses kernel density estimation. But proposed method is better at false positive ratio.

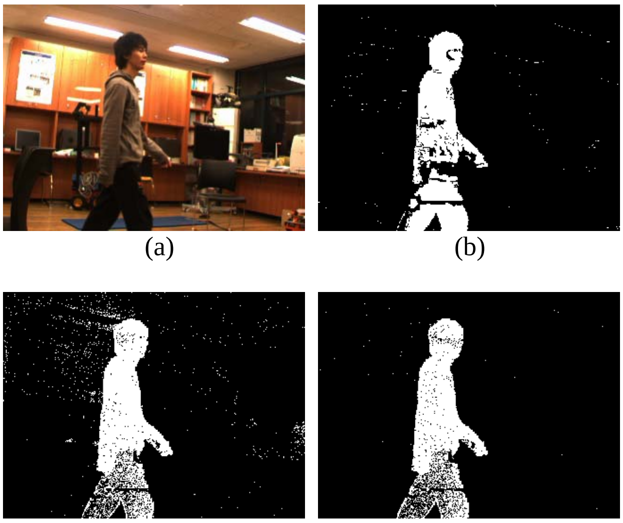


Fig.6. (a) A frame with abrupt intensity level change. detection results using (b) Mixture of Gaussians model, (c) Kernel Density Estimation, (d) proposed method.

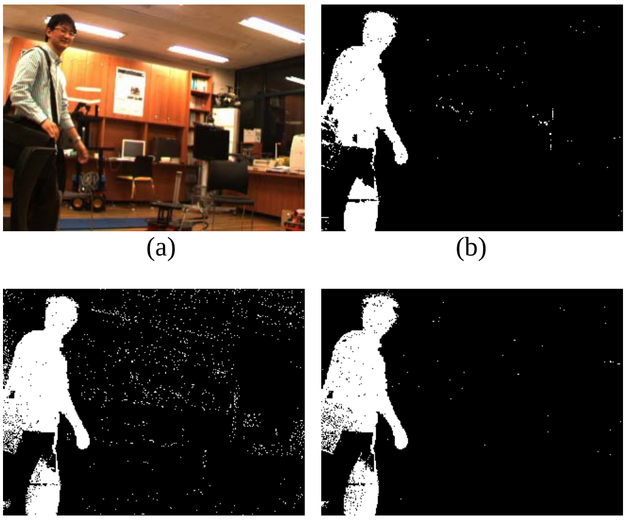


Fig.7. (a) A frame with abrupt intensity level change. detection results using (b) Mixture of Gaussians model, (c) Kernel Density Estimation, (d) proposed method.

4. CONCLUSION

In this paper, we propose an effective background maintenance method when global intensity level changes exist. In general, conventional methods update background statistics gradually. Thus, they cannot deal with abrupt intensity level changes. By using mean-shift algorithm, we find significant changes of intensity level although foreground changes exist. Additionally, we modified the kernel function to solve saturation problem when the intensity level changes. We show that intensity level estimation adapts to real sequence well. Consequently, foreground extraction results for video sequences illustrate the good performance compared with other methods.

References

[1] C. Wren, A. Azarbaygaui, T. Darrel, and A. Pentland, “Pfinder: Real-time tracking of the human body,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 19, pp. 780–785, July 1997.

[2] C. Stauffer and W. Grimson, “Learning patterns of activity using real-time tracking,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 22, pp. 747–757, 2000.

[3] A. Elgammal, D. Harwood, and L. Davis, “Nonparameteric model for background subtraction,” in European Conf. on Computer Vision, 2000, pp. 751–767.

[4] A. Mittal and N. Paragios, “Motion-based background subtraction using adaptive kernel density estimation,” in IEEE conf. on Computer Vision and Pattern Recognition, 2004, pp. 302–309.

[5] K. Fukunaga and L. Hostetler, “The estimation of the gradient of a density function, with applications in pattern recognition,” in IEEE IT, vol. 21, pp. 32–40, 1975.