Statistical background subtraction based on the exact per-pixel distributions

Youngbae Hwang[†], Hanbyul Joo[†], Jun-Sik Kim[‡], In-So Kweon[†]

[†] Dept. of Electrical Engineering, KAIST, Daejeon, Korea

[‡] Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA

E-mail: unicorn@rcv.kaist.ac.kr

Abstract

Most of background subtraction methods represent background statistics using probabilistic unified frameworks such as the Gaussian mixture model or kernel density estimation. But these models cannot define the exact difference between two pixels. It causes misclassification such as false alarms and misses. We presented a new sensor noise model appropriate for general CCD cameras. Based on this, we propose a novel background subtraction method. Our noise modeling needs a line estimation step to relate image intensities with parameters of the noise distribution. This paper describes a new line estimation algorithm given two consecutive static images, and from which can have a well-fitted distribution for each pixel according to intensity of the pixel. In addition, we present a background update method to deal with the continuous variation of the background. We can estimate accurate foregrounds by adapting the estimated per-pixel distributions and background updates.

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. The difference between a current frame and the background model should be well-measured and the background model should be covered the characteristic of the background well.

There have been many works that tried to obtain accurate foregrounds with suppression of false alarms. Wren et. al. [2] used a single Gaussian model to model 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, Friedman and Russel [3] and Stauffer and Grimson [4] proposed frameworks using mixture of Gaussians. These approaches 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. However, they can not deal with the uncertainties in the correct manner. In [5], Elgammal et. al. proposed a non-parametric approach to deal with the uncertainties in an accurate manner. Mittal and Paragios [6] 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 can not define the exact difference for each pixel individually. To deal with this, they used the approximated distributions to represent background statistics. But we can define the exact difference based on our previous sensor modeling. It provides us a correct criterion to classify the foreground.

In our previous work [1], we proposed a new noise model for the difference between two neighboring pixels. We apply the noise model to measure the difference of pixels in two consecutive frames. We showed that intensity is linearly related with parameters of noise. It means that if we estimate the line between intensity and parameters, the noise distributions are determined for each individual pixel. To estimate this line, we propose a new estimation algorithm for two frames with no scene change. This estimation method can reduce required images for the line estimation. For background subtraction, we maintain a background image to adapt the scene changes by illumination and conversion from dynamic objects to static objects (e.g. a car stops for a while). We classify foreground pixels statistically based on noise distributions.

This paper is organized as follows. In Section 2, we briefly review the previous noise modeling method [1]. Section 3 presents a new line estimation method in the temporal domain to reduce required static images. In Section 4, we explain the strategy for background subtraction. Section 5 shows some experimental results to validate the proposed approach. Finally, we present our conclusions in Section 6.

2. Noise modeling using the Skellam distribution

In [7], several sources of image noise are listed. But most of the sources can usually be much reduced by appropriate design of manufacturers and also can be ignored because the total SNR of a complete system is typically dominated by the smallest SNR. The most dominant noise component is photon noise in CCD or CMOS cameras. In [1], we assume that the image noise is caused from dominant photon noise.

The photon noise is usually modeled by a Poisson distribution [7]. When we assume that image intensity of each pixel follows the Poisson distribution, the probability distribution for p photons in an observation time interval T seconds is known to be a Poisson as

$$P(p \mid \rho, T) = \frac{(\rho T)^p e^{-\rho T}}{p!} \qquad (1)$$

where ρ is a rate parameter measured in photons per second. The mean and standard deviation are given by

$$\mu = \rho T \qquad (2)$$

$$\sigma = \sqrt{\rho T} \qquad (3)$$

Since the number of photons determines the intensity of a pixel, photon noise is not independent of the signal. In addition, photon noise is neither Gaussian nor additive. As shown in (2), μ means the number of photons over interval T. It is natural that the number of photons in a brighter pixel is larger than that in a darker pixel.

We use a Poisson distribution directly to represent the distribution of intensity difference. The difference of two Poisson random variables is defined a Skellam distribution [8]. The probability mass function (pmf) of a Skellam distribution is a function of **k** which means difference between two Poisson random variables, and expressed as

$$P(k; \mu_1, \mu_2) = e^{-(\mu_1 + \mu_2)} \left(\frac{\mu_1}{\mu_2} \right)^{k/2} I_k(2\sqrt{\mu_1 \mu_2}) \quad (4)$$

where μ_1 and μ_2 are means or expected values of two Poisson distributions and $I_k(z)$ is the modified Bessel function of the first kind.

We can estimate the Skellam parameters of intensity difference easily by using statistics of Skellam distribution. The mean μ_S and the variance σ_S^2 of a Skellam distribution are given as

$$\mu_S = \mu_1 - \mu_2 \qquad (5)$$

$$\sigma_S^2 = \mu_1 + \mu_2 \qquad (6)$$

From (5) and (6), we can calculate the parameters directly as

$$\mu_1 = (\mu_S + \sigma_S^2)/2 \qquad (7)$$

$$\mu_2 = (-\mu_S + \sigma_S^2)/2 \qquad (8)$$

μ_S and σ_S^2 are obtained from images of a static scene like

$$\mu_S = \frac{\sum_t (x_t(i,j) - x_{t+1}(i,j))}{n} \qquad (9)$$

$$\sigma_S^2 = \frac{\sum_t (\mu_S - (x_t(i,j) - x_{t+1}(i,j)))^2}{n} \qquad (10)$$

where $x_t(i,j)$ means the intensity of (i,j) position at frame t and n is the number of total images.

As noted previously, the parameters, μ_1 and μ_2 are the number of photons so that there might be a certain relationship between Skellam parameters and intensity. In [1], we showed that Skellam parameters and intensity have linearity. We called this line as the Intensity-Skellam line. If we estimate Intensity-Skellam line, we can estimate Skellam parameters for each pixel according to intensity of R, G, B channels.

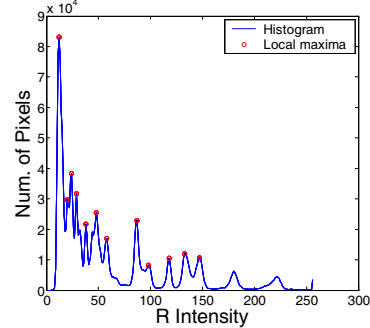Figure 1 Test image for Skellam-Intensity line estimation

Figure 2 Intensity histogram and local maxima

3. Intensity-Skellam line estimation in the temporal domain

When we obtain some pairs of a Skellam parameter and intensity, we can fit a line to set of pairs. If we estimate Skellam parameters at a certain pixel in the temporal domain, we need more 10,000 images to ensure that the calculated statistics is sufficiently stable. But it is impractical to capture so many static images of indoor or outdoor images. We propose a new Intensity-Skellam line estimation method to reduce the required static images to two images. It is more practical compared with capturing over 10,000 images. When we assume that each pixel in an image is mutually independent, we can regard the intensity difference between two corresponding pixels in two frames as set of intensity difference in the temporal domain. Therefore, we have sufficient pixels at certain intensity for estimating Skellam parameters.

Our calibration algorithm for a channel is as follow.

1. Generate a histogram of intensity in first frame and find local maxima x_m^1, $m = 0, 1, \ldots, M$

2. Find set of corresponding pixels in two frames around the local maxima as
$$X_m = \{x^1(i,j), x^2(i,j) \mid x_m - \varepsilon < x^1(i,j) < x_m + \varepsilon\}.$$

3. Calculate the mean and the variance of a Skellam distribution for each set X_m, as
$$\mu_S \mid X_m = \frac{\sum_{X_m} (x_k^1(i,j) - x_k^2(i,j))}{n},$$
$$\sigma_S^2 \mid X_m = \frac{\sum_{X_m} (\mu_S \mid X_m - (x_k^1(i,j) - x_k^2(i,j)))^2}{n}.$$

4. Calculate Skellam parameters, μ_1 and μ_2 from (7) and (8).

5. Fit two lines to pairs of (x_m^1, μ_1) and (x_m^1, μ_2) using conventional RANSAC methods.

(a) R channel   (b) G channel
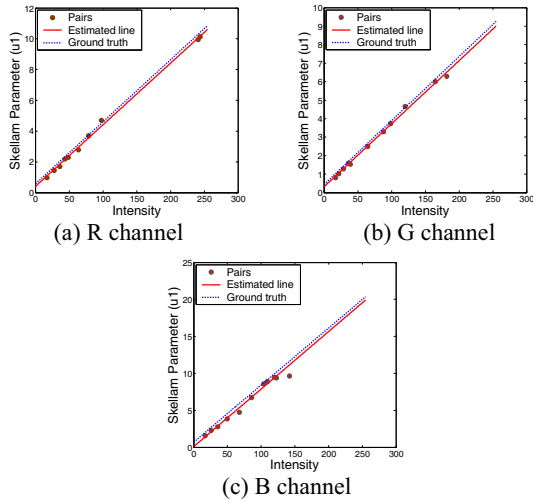

(c) B channel

Figure 3 Intensity-Skellam line estimation results

We carry out an experiment on the algorithm for two real images as shown in Figure 1. We capture color patterns because we can obtain accurate Intensity-Skellam line from a large number of static images as ground truth. Figure 2 shows the histogram of intensity in the R channel and detected local maxima. From the local maxima, we have the pairs of intensity and Skellam parameters. We set the value of ε to one. The Intensity-Skellam line is determined by RANSAC. Figure 3 shows an estimation result for each channel. For comparison, we plot ground truth lines obtained by 10,000 static images. Our Intensity-Skellam line is very accurate compared with the ground truth. Most of the pairs used for fitting lines are laid around the fitted line. It means that we can estimate the line by using small number of pairs.

4. Background subtraction

4.1. Maintenance of the background image

When we measure the difference between only consecutive frames, we can not deal with various situations that can be occurred in general indoor and outdoor environment such as abrupt or gradual illumination changes, position changes of furniture and moving vehicles are stopped or parked. Therefore, we should maintain a background image to deal with such environmental changes.

There are two conventional mechanisms to update background. First one is a selective update which adds the new pixel to the model only if it is classified as a background pixel. From this approach, we can correct the part of backgrounds occluded by moving objects at previous frames. But the problem is that erroneous detection results or temporary static objects may make permanent incorrect background model. It causes the incorrect foreground detection at incorrect background parts. Second one is a blind update which just adds the new sample to the model. This update can reduce the previous problem of the selective update since it restores the incorrect background model to correct one. But it causes another problem of updating incorrect background model by moving objects and false alarms.

In order to deal with these problems of conventional update methods, we propose another background model-

ing strategy. The conventional update methods [5] used only background images and current upcoming images. But we utilize difference between two consecutive images to update the background model. Our algorithm of maintaining the background model is as follows:

1. Initialize a background image, $I_B^0(p)$, as an image in the first frame.

2. For each frame t, store history images of length T and update background mask as

$$M^t(p_B) = 1,\ M^t(P_F) = 0.$$

(P_B, P_F: Classified as a background and a foreground pixel in consecutive images, respectively)

3. If $t \geq T$, update the background model as

$$I_B^t(p) = \sum_{t'=t-T}^{t} I_H^{t'}(p), \qquad \text{if } \prod_{t'=t-T}^{t} M_{t'}(p) = 1$$

$$I_B^t(p) = I_B^{t-1}(p), \qquad\qquad \text{otherwise}$$

(I_H^t is a stored history image at frame t).

Our algorithm can deal with the problems of selective and blind updates. We can correct the part of backgrounds occluded by moving objects the same as the selective update. The problem of a blind update is solved because we update the incorrect background parts when there is no foreground during T frames. Therefore, we don't need the blind update to restore the incorrect parts. The only defect is that it takes T frames to update the background correctly. When we set T=1, our algorithm became totally same as a selective update. Otherwise, our algorithm corrects the erroneous parts within T frames.

4.2. Background subtraction strategy

From Section 3, we have exact distributions for each pixel in the image. We can classify foregrounds to measure the difference between a current image and the maintained background image from Section 4.1. The update parameter T is set to five. The classification is performed by a statistical hypothesis test [9]. Our noise modeling is so precise that subtraction results have few false alarms as shown in Figure 4. Therefore, we do not need to any further processing as shown in other methods [5]. Our background update algorithm is applicable since the result is almost the true foregrounds.

5. Experimental results

We apply our algorithm to an outdoor sequence. We use PointGrey Flea camera with a 640x480 resolution. Our algorithm runs by 4Hz in AMD Athlon64 X2 2.41GHz and 2GB memory. In Figure 4(b), we show updated background images given current images as shown in Figure 4(a). Although there are moving objects, our background model can adapt the background well. Based on updated backgrounds we can detect foreground parts by measuring the difference between current upcoming images and background images. In Figure 4(c), detection results are so accurate that our method can de-

tect almost true boundaries for moving humans and cars except for a few shadows. Compared with the result by mixture of Gaussian [4], our results have smaller missing pixels as well as smaller false alarms because mixture of Gaussian can not guarantee exact distribution of pixel color differences but have rough Gaussian distributions for background estimation. The third column result shows a scene change by an abrupt illumination change. Since our background model is promptly adapted, false alarms are much reduced than the result from mixture of Gaussian. Furthermore, our all of pixels have exact distribution according to intensity, we can detect moving objects in the dark condition as shown in Figure 5.

It is worthy noted that all of our detection results have pretty small false alarms as shown in Figures 4 and 5. It is very advantageous to apply surveillance system. Most of subtraction algorithm suffered from false alarms as shown in third column of Figure 4(d). It is difficult to discriminate the real foreground caused by moving human or cars from the false alarms caused by misclassification. But, our results have small false alarms although we do not use any further post processing. Therefore, we can determine that there is some foreground pixels in our detection result, most of them are caused by real moving objects.
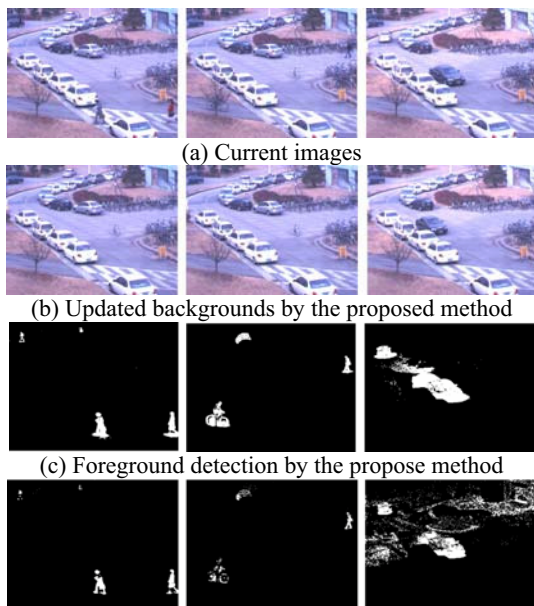

(a) Current images

(b) Updated backgrounds by the proposed method

(c) Foreground detection by the propose method

(d) Foreground detection by Mixture of Gaussian [4]
Figure 4 Background subtraction results (daylight)

6. Conclusions

We propose a new background subtraction method by using previously proposed noise modeling and a new background update strategy. Other previous methods deal with background subtraction with approximated distributions because they can not define an exact difference for classification. But we can define an exact difference based on sensor noise modeling. To estimate Intensity-Skellam line, we propose an estimation method for two static images. The line estimation result is so accurate that there is little difference with the available ground truth. We propose a background update approach. Conventional methods used only the difference between an input image and a background image to update the

background model. In the contrary, we utilize the difference between consecutive frames for updating backgrounds. Our background update method adapts the change of backgrounds very well. Based on estimated noise statistics and updated background models, we can detect foreground parts precisely in sequences. The advantage of our method is that there are few false alarms without post processing and we can apply our algorithm for dark conditions.
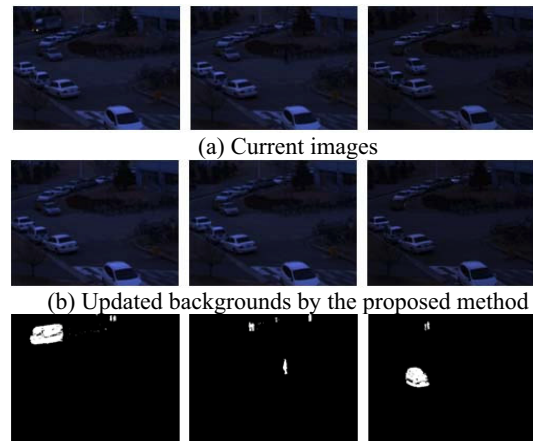

(a) Current images

(b) Updated backgrounds by the proposed method

(c) Foreground detection by the propose method
Figure 5 Background subtraction results (dark)

Acknowledgement

References

[1] Youngbae Hwang, Jun-Sik Kim and In-So Kweon, "Sensor noise modeling using the Skellam distribution: Application to the color edge detection", In Int. Conf. on CVPR, 2007, to appear

[2] C. Wren, A. Azarbayejani, T. Darrel, and A. Pentland, "Pfinder: Real-time tracking of the human body", IEEE Trans. on PAMI, vol.19, no.7, pp.780-785, 1997

[3] N. Friedman and S. Russel, "Image segmentation in video sequences: A probabilistic approach", In Conf. on Uncertainty in Artificial Intelligence (UAI), pp.175-181, 1997

[4] C. Stauffer and W. Grimson, "Learning patterns of activity using real-time tracking", IEEE Trans. on PAMI, vol. 22, pp.747-757, 2000

[5] A. Elgammal, D. Harwood and L. Davis, "Non-parametric model for background subtraction", pp.751-767, In European Conf. on Computer Vision, 2000

[6] A. Mittal and N. Paragios, "Motion-based background subtraction using adaptive kernel density estimation", In Int. Conf. on CVPR, pp.302-309, 2004

[7] I. Young, J. Gerbrands, and L. van Vliet, "Fundamentals of image processing", 1995, Delft University of Technology

[8] J. G. Skellam, "The frequency distribution of the difference between two Poisson variates belonging to different populations", Journal of the Royal Statistical Society: Series A, vol.109, no.3, pp.296, 1946

[9] E. R. Dougherty. Probability and statistics for the engineering, computing and physical sciences. Prentice Hall, 1990