Object Detection Using Local Difference Patterns

Satoshi Yoshinaga, Atsushi Shimada,
Hajime Nagahara, and Rin-ichiro Taniguchi

Graduate School of Information Science and Electrical Engineering,
Kyushu University, Fukuoka, Japan
{yoshinaga,atsushi,nagahara,rin}@limu.ait.kyushu-u.ac.jp

Abstract. We propose a new method of background modeling for object detection. Many background models have been previously proposed, and they are divided into two types: "*pixel-based* models" which model stochastic changes in the value of each pixel and "*spatial-based* models" which model a local texture around each pixel. Pixel-based models are effective for periodic changes of pixel values, but they cannot deal with sudden illumination changes. On the contrary, spatial-based models are effective for sudden illumination changes, but they cannot deal with periodic change of pixel values, which often vary the textures. To solve these problems, we propose a new probabilistic background model integrating pixel-based and spatial-based models by considering the illumination fluctuation in localized regions. Several experiments show the effectiveness of our approach.

1 Introduction

Background subtraction is one of the most widely used techniques to detect moving objects from image sequences. It enables us to detect objects by calculating subtraction of a background image from an observed image without any specific prior information about moving objects. However, when we use a simple background image in outdoor scenes, it will detect not only object regions but also a lot of noise regions. This is because it is very sensitive to changes in the pixel values caused by waving trees, fleeting clouds, illumination changes and so on. Therefore, many approaches to model these background changes have been proposed [1–9].

In general, the approaches of background modeling can be divided into two types: *"pixel-based approach"* and *"spatial-based approach"*. In the case of pixel-based background models, they commonly have a probability density function (PDF) for each pixel to represent the pixel value distribution observed in a video sequence. Stauffer et al. proposed a background estimation method, in which mixture-of-Gaussians is used to describe the background model [4], and Shimada et al. augmented this method by introducing a mechanism to change the number of Gaussians dynamically in each pixel [5]. Elgammal et al. employed Parzen density estimation to estimate the PDF of the pixel value non-parametrically

R. Kimmel, R. Klette, and A. Sugimoto (Eds.): ACCV 2010, Part IV, LNCS 6495, pp. 216–227, 2011.

[6]. These pixel-based models are effective for periodic changes of pixel values, which are caused by fleeting clouds, movement of tree branches or leaves, waves on water and so on. However, they cannot adapt for sudden illumination changes. This is because they construct their models based on statistical information of the pixel values observed in the past.

In the case of spatial-based background models, they model local textures in a localized region centered around each pixel to evaluate the similarity between the background image and the observed image [2, 3]. These models define several pairs of a target pixel and its neighbor pixels, and establish a background model using magnitude relations of pixel values of those pairs. Therefore, spatial-based background models are more robust than pixel-based one against sudden illumination changes, because there are little changes in magnitude relations of pixel values before and after a sudden illumination change. On the other hand, they cannot deal with periodic changes of pixel values, which are caused by the movement of tree branches or leaves and so on, since the textures change in such situations.

The *hybrid background models* are also proposed, in which both a pixel-based and spatial-based background models are utilized. Tanaka et al. proposed a hybrid background model [9], in which they combined the results of a pixel-based background model [8] and spatial-based one using “logical AND”. Their model is more robust than pixel-based or spatial-based ones, since it can utilize both properties by combining the results of two different models. However, objects should be detected accurately by both models, and mis-detection in either of the two models reduces the detection accuracy. Therefore, hybrid models require more sophisticated combinatorial algorithm for integrating the results of two different models with high accuracy.

In this paper, we propose a new probabilistic background model by integrating the methodology of both a pixel-based and a spatial-based approaches. Note that our approach, unlike previous works [9], does not combine two approaches in a naive way. We will give a detail explanation about our proposed method in Section2.

2 Probabilistic Background Model Considering Illumination Fluctuation in Localized Region

We propose a new probabilistic background model, as shown in Fig.1, by considering the illumination fluctuation in a localized region centered around each pixel.

2.1 Design of Local Difference Pattern

In our background model, the methodologies of both a pixel-based and a spatial-based approaches are naturally integrated. In the case of pixel-based model, the problem is that spatial information (e.g. texture) was not considered. On the

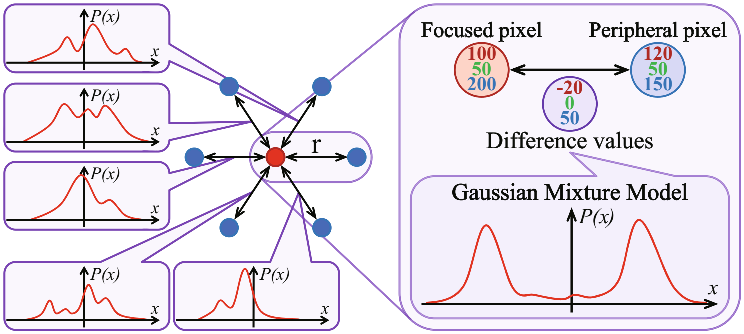


Fig. 1. Probabilistic background model using LDP: Our proposed model defines several pairs of focused pixel and its peripheral pixels in localized region which is a circular region with radius r. Eeach pair has a Gaussian Mixture Model (GMM) to model the distribution of the difference between the pixel values of them.

contrary, in the case of spatial-based model, local texture was represented in magnitude relations of pixel values in a background image and multiple hypotheses of the background can not be maintained, which causes a problem.

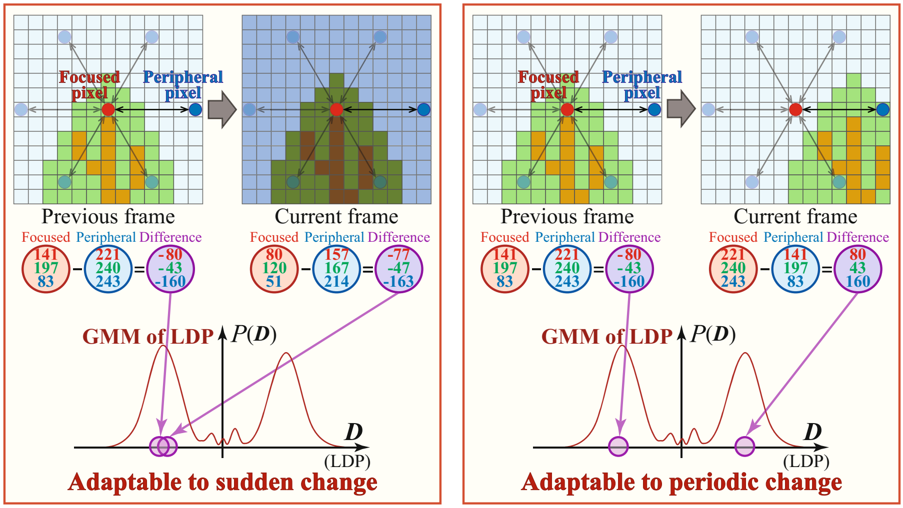
To solve these problems, we propose a new probabilistic background model integrating pixel-based and spatial-based models by considering the illumination fluctuation in localized regions, as shown in Fig.1. In the proposed model, we define several pairs of a focused pixel and its peripheral pixels, i.e., its surrounding pixels, in a localized region (Fig.1 is an example where the number of pair is 6), and we give each pair a Gaussian Mixture Model (GMM) to model the distribution of the difference between pixel values of each pair. Here, we call these pixel value differences in the localized region "*Local Difference Pattern*" (LDP).

The advantages of using LDP are as follows (see Fig.2). In most cases where sudden illumination changes occur, there are little changes in a LDP, since the pixel values in a localized region similarly increase and decrease in their values. Therefore, our proposed method can deal with sudden illumination changes as shown in Fig.2 (a). Furthermore, our proposed method can also deal with periodic changes of pixel values, since GMM represents multiple hypotheses of the background as shown in Fig.2 (b). Thus, our background model can utilize both properties of pixel-based and spatial-based model, without decreasing the accuracy.

2.2 Construction of Local Difference Pattern

A focused pixel in an observed image is represented by a vector $\boldsymbol{p}_c = (x_c, y_c)^T$. A directional vector $\boldsymbol{a}_j (j = 1, \ldots, N_{pair})$, which represents the direction of each reach or the direction of each peripheral pixel, is defined as follows.

$$\boldsymbol{a}_j = \left(\cos \frac{j-1}{N_{pair}} 2\pi, \sin \frac{j-1}{N_{pair}} 2\pi \right)^T \tag{1}$$



(a) Sudden illumination change (b) Periodic change of pixel value

Fig. 2. Adaptivities of the proposed model to background fluctuation: (a) shows the case that illumination suddenly changed (e.g. when sunlight is blocked out by clouds, etc.). LDP can absorb the effect of illumination changes, since it globally affects pixel values as a bias. (b) shows the case that texture is periodically changed (e.g. effect of movement of tree or grass, waves on water, etc.). GMM can also adapt to these kinds of deriodic changes, since it allows for multiple hypotheses of the background.

Each peripheral pixel $\boldsymbol{p}_j = (x_j, y_j)^T$, which is present on the circumference of a circle with radius r centered around a focused pixel $\boldsymbol{p}_c$, is represented by $\boldsymbol{p}_j = \boldsymbol{p}_c + r\boldsymbol{a}_j (j = 1, \ldots, N_{pair})$, where N_{pair} is the number of peripheral pixels.

We define N_{pair} pairs of a focused pixel $\boldsymbol{p}_c$ and its peripheral pixels $\boldsymbol{p}_j$. Then, a LDP observed at a focused pixel $\boldsymbol{p}_c$ at time t is defined by the difference between the pixel values of each pair, and we represent it by $\boldsymbol{D}^t = \{\boldsymbol{X}_1^t, \ldots, \boldsymbol{X}_j^t, \ldots, \boldsymbol{X}_{N_{pair}}^t\}$. Here, $\boldsymbol{X}_j^t = f(\boldsymbol{p}_c) - f(\boldsymbol{p}_j)$, where $f(\boldsymbol{p})$ is the d-dimensional vector representing the value of pixel $\boldsymbol{p}$ ($d = 3$ in case of RGB color images). Fig.1 shows an example of $N_{pair} = 6$.

In most cases where sudden illumination changes occur, there is little change in the LDP. Therefore, our proposed method based on the LDP can deal with sudden illumination changes.

2.3 Probabilistic Background Model Based on LDP

In our proposed method, we give each pair a GMM to represent the PDFs of a LDP. Here, we focus on the j-th pair of a LDP, and we model the difference between the pixel values of the j-th pair $\boldsymbol{X}_j^t$. Let $\{\boldsymbol{X}_j^1, \ldots, \boldsymbol{X}_j^t\}$ be the difference between the pixel values of the j-th pair observed until time t, then we can represent a PDF of them by a mixture of K Gaussian distributions. Then, the probability of observing the difference is

$$P(\boldsymbol{X}_j^t) = \sum_{k=1}^{K} w_{j,k}^t \eta(\boldsymbol{X}_j^t | \boldsymbol{\mu}_{j,k}^t, \boldsymbol{\Sigma}_{j,k}^t) \tag{2}$$

where j is a subscript representing the direction of the peripheral pixel based on the focused pixel, $w_{j,k}^t$, $\boldsymbol{\mu}_{j,k}^t$, $\boldsymbol{\Sigma}_{j,k}^t$ are a weight, the mean and the covariance matrix of the k-th Gaussian in the mixture at time t, and η is a Gaussian probability density function as follows.

$$\eta(\boldsymbol{X}_j^t | \boldsymbol{\mu}_j^t, \boldsymbol{\Sigma}_j^t) = \frac{1}{(2\pi)^{\frac{d}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\boldsymbol{X}_j^t - \boldsymbol{\mu}_j^t)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{X}_j^t - \boldsymbol{\mu}_j^t)\right) \tag{3}$$

K is determined by the available memory and computational power. Also, to reduce the computation cost, the covariance matrix is assumed to be of the form:

$$\boldsymbol{\Sigma}_{j,k}^t = \sigma_{j,k}^t \boldsymbol{I} \tag{4}$$

In the case of RGB color space, this means that the red, green, and blue pixel values are independent and have the same variances. While this is certainly not the case, the assumption allows us to avoid a costly matrix inversion at the expense of some accuracy.

Thus, a PDF of the difference between pixel values of a pair of a LDP observed until time t is characterized by a mixture of K Gaussian distributions. A new difference value will be represented by one of the components of the mixture model and used to update the model. We will describe the background model estimation process in 6 steps.

Step1. Every new difference value $\boldsymbol{X}_j^t$ is examined against the existing K Gaussian distributions until a match is found. Here, the match is defined as a difference within 2.5 standard deviations of distribution.

Step2. The prior weights $w_{j,k}^t$ of the K distributions of j-th GMM at time t are updated as follows

$$w_{j,k}^t = (1-\alpha) w_{j,k}^{t-1} + \alpha M_{j,k}^t \tag{5}$$

where α is the learning rate and $M_{j,k}^t$ is 1 for the matched distribution and 0 for the remaining distributions. After this process, these weights $w_{j,k}^t$ are renormalized.

Step3. The $\boldsymbol{\mu}_{j,k}^t$ and $\sigma_{j,k}^t$ parameters for unmatched distributions remain unchanged. The parameters of the distribution which matches the new observation are updated as follows

$$\boldsymbol{\mu}_{j,k}^t = (1-\rho) \boldsymbol{\mu}_{j,k}^{t-1} + \rho \boldsymbol{X}_j^t \tag{6}$$

$$\sigma_{j,k}^t = (1-\rho) \sigma_{j,k}^{t-1} + \rho (\boldsymbol{X}_j^t - \boldsymbol{\mu}_{j,k}^t)^T (\boldsymbol{X}_j^t - \boldsymbol{\mu}_{j,k}^t) \tag{7}$$

where ρ is the second learning rate and is defined as follows.

$$\rho = \alpha \eta(\boldsymbol{X}_j^t | \boldsymbol{\mu}_{j,k}^t, \boldsymbol{\Sigma}_{j,k}^t) \tag{8}$$

Step4. If none of the K distributions match the current difference value in **Step1**, a new Gaussian distribution is created as follows

$$w^t_{j,K+1} = W \tag{9}$$

$$\boldsymbol{\mu}^t_{j,K+1} = \boldsymbol{X}^t_j \tag{10}$$

$$\sigma^t_{j,K+1} = \sigma^t_{j,K} \tag{11}$$

where W is the initial weight value[1] for the new Gaussian. After this process, the weights are renormalized.

Step4-1: When the weight of the least probable distribution is smaller than a threshold, the distribution is deleted, and the remaining weights are renormalized.

Step4-2: When the difference between means of two Gaussians (the one is η_a and the other is η_b) is smaller than a threshold, these distributions are integrated into one Gaussian. The new wight, mean and variance of integrated Gaussian η_c are calculated as follows.

$$w^t_{j,c} = w^t_{j,a} + w^t_{j,b} \tag{12}$$

$$\boldsymbol{\mu}^t_{j,c} = \frac{w^t_{j,a}\boldsymbol{\mu}^t_{j,a} + w^t_{j,b}\boldsymbol{\mu}^t_{j,b}}{w^t_{j,a} + w^t_{j,b}} \tag{13}$$

$$\sigma^t_{j,c} = \frac{w^t_{j,a}\sigma^t_{j,a} + w^t_{j,b}\sigma^t_{j,b}}{w^t_{j,a} + w^t_{j,b}} \tag{14}$$

Step5. The Gaussians are ordered by the value of w/σ. This value increases as the distribution gains more evidence and as the variance decreases.

Step6. The first B distributions are chosen as the background model, and B is represented as follows

$$B_j = \operatorname*{argmin}_b \left(\sum_{k=1}^{b} w^t_{j,k} > T \right) \tag{15}$$

where T is a measure of the minimum portion of the data that should be accounted for by the background. If a small value for T is chosen, the background model is usually unimodal. If T is higher, a multi-modal distribution is created by repetitive background changes (e.g. fleeting clouds, the movement of tree branches or leaves, the waves on water, etc.) could result in multiple colors being included in the background model. This results in a transparency effect which allows the background to accept two or more separate colors.

[1] If W is higher, the distribution is chosen as the background model for a long time.

2.4 Object Detection Using LDP

Object detection based on the LDP using N_{pair} GMMs is defined by following equation.

$$f(x,y) = \begin{cases} background & \text{if } \sum_{j=1}^{N_{pair}} \phi(\boldsymbol{X}_j^t) > th \\ foreground & \text{otherwise} \end{cases} \tag{16}$$

In Equation 16, $\phi(\boldsymbol{X}_j^t)$ is a function which returns 1 or 0, according to whether the matched distribution found in **Step1** is one of the background models (described in **Step6**) or not. In addition, th is a threshold for determining whether a focused pixel $\boldsymbol{p}_c$ belongs to the background or the foreground.

3 Experimental Result

We conducted two kinds of experiments. First, we examined the parameters (r, N_{pair}) of LDP and decided one of the good parameters which was used in the following experiment. Second, we compared the accuracy with state-of-the-art methods. Due to space limitation, we'll report the result of PETS2001[2] dataset.

3.1 Preliminary Experiment for Adjusting Parameters

In our proposed method, we focus on a localized circular region of radius r centered around each pixel, and model a LDP using N_{pair} GMMs. Therefore, we investigated the parameters (r, N_{pair}) of LDP by several experimental analyses. Fig.3 shows that the variation of the accuracy across the parameters. Here, we employed Recall and Precision for the accuracy, and used manually-produced Ground Truth[3] dataset to evaluate them.

Fig.3 shows that there were little changes in the accuracy, when the number of pair N_{pair} was more than 4. Also it shows that the change of radius r caused little or no change in the accuracy, when N_{pair} was more than 4.

We evaluated the computational cost. The image size was 320×240 (pixel) and the PC had Core 2 Duo 2.8 GHz CPU and 4GB memory. Table 1 shows that computational cost increases in proportion to N_{pair}. Therefore, we have employed $N_{pair} = 6$ and $r = 10$ as an optimal parameter for PETS2001 dataset in terms of the balancing point of the accuracy and the computational cost.

We carried out preliminary experiments for using other data sets and confirmed that N_{pair} is not so changed depending on video contents. Therefore, N_{pair} was not critical for the performance. On the other hand, the radius r depends on the video content. However, it is easy to decide the parameter r from

[2] Benchmark data of International Workshop on Performance Evaluation of Tracking and Surveillance. Available from ftp://pets.rdg.ac.uk/PETS2001/

[3] The Ground Truth image denotes foreground regions which should be detected by background subtraction. We made Ground Truth for some Benchmark data including PETS2001 manually, and have published them to the web. Available from http://limu.ait.kyushu-u.ac.jp/dataset/

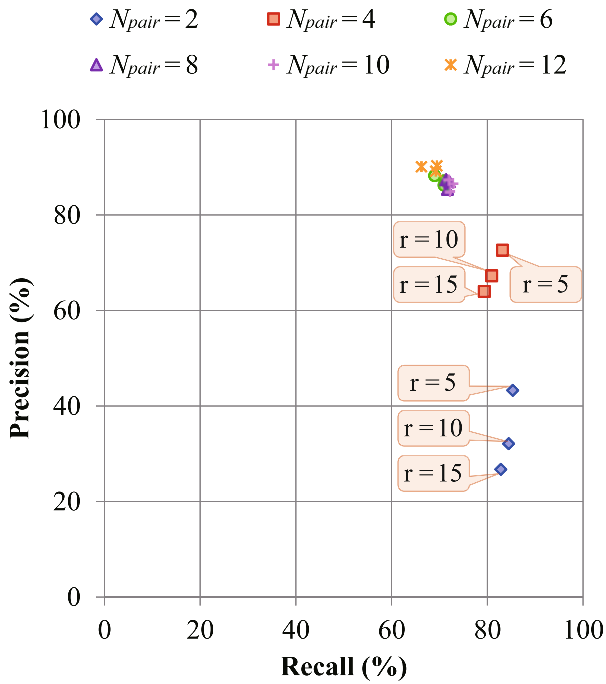


Fig. 3. The accuracy of object detection in relation to the parameters of LDP

a prior knowledge of the object sizes. In most cases (e.g. surveillance, security, etc.), we would predict the size of the objects, since the camera is stationary and observes similar objects in these applications. Hence, it does not lose a generality or effectiveness of the proposed method, although the parameter can be determined by the prior knowledge.

3.2 Object Detection Accuracy

We evaluated the accuracy of object detection based on Recall and Precision. According to experimental result in Section3.1, the parameters of LDP were

Table 1. Computational cost in relation to the parameter N_{pair}

Parameter	Average processing time (ms)
$N_{pair} = 2$	68.4
$N_{pair} = 4$	149.6
$N_{pair} = 6$	231.1
$N_{pair} = 8$	321.7
$N_{pair} = 10$	391.4
$N_{pair} = 12$	472.3

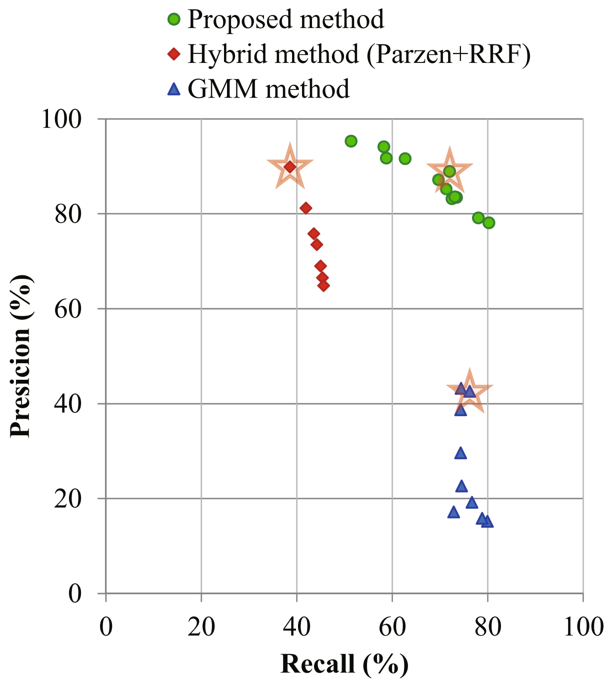


Fig. 4. The accuracy of object detection in relation to the changes of several parameters in each method: Star signs indicate the best performances on each method

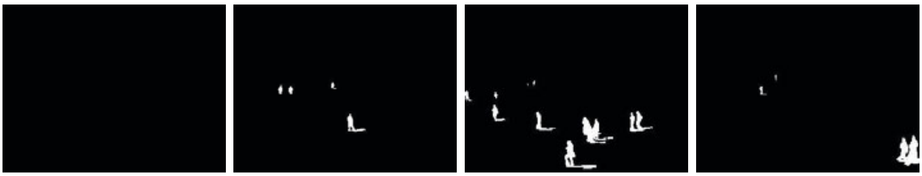
fixed $N_{pair} = 6$ and $r = 10$ respectively. However, our approach has some other parameters; the parameters of GMM (e.g. the initial weight W, threshold T, etc.), which affect the accuracy. Thus, we conducted some experiments with varying the parameters. We used GMM method [5] and Hybrid method (Parzen + RRF) [9] to compare the accuracy with our proposed method. These methods had also several parameters, therefore we performed some experiments with varying the parameters of them as well as our proposed method. Fig.4 and Table 2 show the accuracy of each method. The representative results of background subtraction are shown in Fig.5. In Table 2 and Fig.5, the states flagged with a star sign in Fig.4 have been used as the parameters of each method.

We can see that Precision of the Hybrid method was high but Recall of it was low. This was also confirmed form Fig.5 since there was little noise but

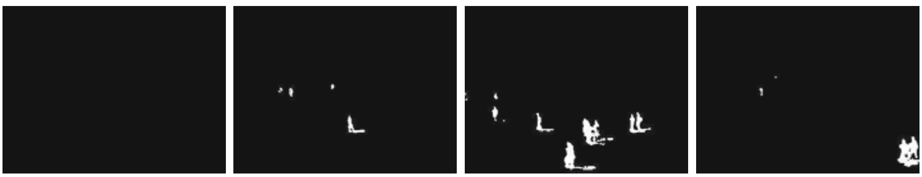
Table 2. The accuracy of object detection

	Recall	Precision
Proposed method	**72.0%**	**88.9%**
Hybrid method (Parzen + RRF) [9]	38.6%	89.9%
GMM method [5]	76.3%	42.6%

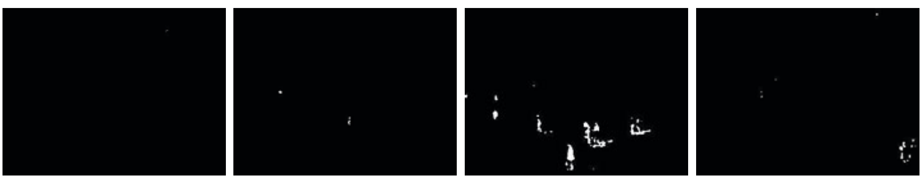
(a) Input image



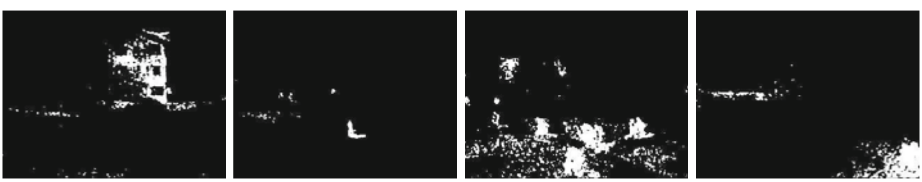
(b) Ground Truth



(c) Proposed method



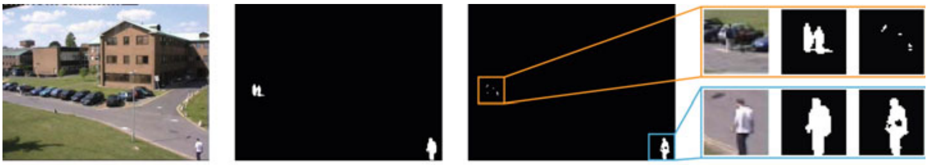
(d) Hybrid method (Parzen + RRF) [9]



(e) GMM method [5]

Fig. 5. Object detection by the Proposed method, Hybrid method and GMM method

object regions detected by the Hybrid method were abnormally-small. Fig.4 and Table 2 also show that Recall of the GMM method was high but Precision of it was low, and the GMM method detected not only objects but also many noises in Fig.5. On the other hand, both Recall and Precision of our proposed method were high, and object regions were accurately detected with little noise.



(a) Input image (b) Ground Truth (c) Proposed method (d) Examples of failure

Fig. 6. Problem with Our Proposed Method

3.3 Discussion

Totally, LDP gave use better results than state-of-the-art methods. However, we found out that following problems were caused by the characteristic of LDP that it used the difference values between focused pixel and peripheral pixels.

- **The objects with uniform texture:**
  We assumed that the LDP ignores global changes such as illumination changes. However, the LDP sometimes causes a problem in the case where "a object with uniform texture" appears on "the background with uniform texture". In this case, if radius r is smaller than foreground object, the local changes caused by the objects are mistakenly regarded as global changes. As the result, our proposed model fails to detect internal regions of the objects and False Negative increases (see blue-rectangle in Fig.6).
  As discussed in Section3.1, this problem can be avoided by selecting a suitable radius r from a prior knowledge of a scene. It would be desirable to decide the parameter automatically, which will be our future work.
- **Color similarity between object and background:**
  In the case where "an object has similar color with background" appears, the difference between the object and background becomes smaller. It is difficult to detect such objects in our method (see orange-rectangle in Fig.6).

4 Conclusion

In this paper, we have proposed a new probabilistic background model using several GMMs. We considered the illumination fluctuation in the localized region, and model LDP (the difference values between the values of focused pixel and its peripheral pixels, which is present on the circumference of circle centered around focused pixel) using GMMs. We could integrate pixel-based and spatial-based model themselves by using LDP, and background model using LDP could utilize both properties without decreasing the accuracy, unlike traditional model. In our experiment, we have got a good result where both Precision and Recall were superior to the traditional background subtraction methods.

Future works are summarized as follows.

- **Reduction of computational time**
  Our proposed method have the N_{pair} GMMs, therefore it is cost to update them, where N_{pair} is the number of peripheral pixels. In the case of

Image Size = 320 × 240 (pixel) and $N_{pair} = 6$, computational time was about $230ms$ using a PC with a Core 2 Duo 2.8GHz CPU and 4GB memory. It is not good for real-time processing, and therefore we should develop a mechanism to reduce the computational time.

– **Improvement of the accuracy of object detection**
  Our proposed method has some problems associated with objects, as described in section 3.3. Therefore, it is necessary to sophisticate our proposed method to cope with the objects described in section 3.3.

References

1. Toyama, K., Krumm, J., Brumitt, B., Meyers, B.: Wallflower: Principle and Practice of Background Maintenance. In: International Conference on Computer Vision, pp. 255–261 (1999)
2. Marko, H., Matti, P.: A Texture-Based Method for Modeling the Background and Detecting Moving Objects. IEEE Transactions on Pattern Analysis and Machine Intelligence 28, 657–662 (2006)
3. Satoh, Y., Kaneko, S., Niwa, Y., Yamamoto, K.: Robust object detection using a Radial Reach Filter (RRF). Systems and Computers in Japan 35, 63–73 (2004)
4. Stauffer, C., Grimson, W.: Adaptive background mixture models for real-time tracking. In: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), vol. 2, pp. 246–252 (1999)
5. Shimada, A., Arita, D., Ichiro Taniguchi, R.: Dynamic Control of Adaptive Mixture-of-Gaussians Background Model. In: CD-ROM Proceedings of IEEE International Conference on Advanced Video and Signal Based Surveillance (2006)
6. Elgammal, A., Duraiswami, R., Harwood, D., Davis, L.: Background and Foreground Modeling using Non-parametric Kernel Density Estimation for Visual Surveillance. Proceedings of the IEEE 90, 1151–1163 (2002)
7. Monari, E., Pasqual, C.: Fusion of Background Estimation Approaches for Motion Detection in Non-static Backgrounds. In: CD-ROM Proceedings of IEEE International Conference on Advanced Video and Signal Based Surveillance (2007)
8. Tanaka, T., Shimada, A., Arita, D., Ichiro Taniguchi, R.: A Fast Algorithm for Adaptive Background Model Construction Using Parzen Density Estimation. In: CD-ROM Proceedings of IEEE International Conference on Advanced Video and Signal Based Surveillance (2007)
9. Tanaka, T., Shimada, A., Ichiro Taniguchi, R., Yamashita, T., Arita, D.: Towards robust object detection: integrated background modeling based on spatio-temporal features. In: The Ninth Asian Conference on Computer Vision (2009)