Background modeling by combining joint intensity histogram with time-sequential data

Yasuyo Kita
Information Technology Research Institute
National Institute of Advanced Industrial Science and Technology (AIST)
Tsukuba, Japan
y.kita@aist.go.jp

***Abstract*—In this paper, a method for detecting changes from time-sequential images of outside scenes which are taken with several minutes interval is proposed. Recently, statistical background intensity model per pixel using Gaussian mixture model (GMM) has shown its effectiveness for detecting changes from video streams. However, when the time interval between consecutive images is long, enough number of frames can not be sampled for building useful GMM. To robustly build a pixel-wise background model at time t_0 from small number of fore and aft frames, we propose to use the joint intensity histogram of the images at time t_0 and $t_0 + 1$, $H(I_{t_0}, I_{t_o+1})$. Under "background dominance" condition, background probability distribution for each intensity level at t_0 can be estimated from $H(I_{t_0}, I_{t_o+1})$. By taking this background probability distribution per intensity as a prior probability, GMM which models the variation in each pixel is robustly calculated even from several frames. Experimental results using actual field monitoring images have shown the advantage of the proposed method.**

***Keywords*-change detection; background model; joint intensity histogram; outside scenes**

I. Introduction

Change detection from camera surveillance images is one of the most important subjects recently. In the case of monitoring outside scenes, large variations in environmental lighting cause big problems. When images are taken consecutively with small time interval, as like video surveillance, statistical model of each pixel value has shown good results to judge whether the pixel is background or not [1][2][3][4]. However, there are also several situations where images can be taken only with longer time interval. Field surveillance where electric supply is limited is one of such examples [5][6], where images are taken every several minutes. In such a situation, it becomes difficult to derive reliable background model per pixel because of too small number of frames under the same lighting condition.

As an extreme end of decreasing the number of referred frames, a method to detect changes from just two images using the joint intensity histogram (JIH) of the images has been proposed[7]. The joint intensity histogram is a two-dimensional (2D) histogram of combinatorial intensity pairs, $(I_1(\mathbf{x}), I_2(\mathbf{x}))$, where $I_{1(2)}$ and $\mathbf{x}$ represent the intensity level (0-255) of each image and the positions of pixels on images, respectively. In the method, clusters on the JIH are used to determine presumable intensity changes of background pixels. By selecting pixels showing different changes from the estimated background changes, object appearance/disappearance has been successfully detected. However, there is a risk to miss important changes if the method fails in detecting "background clusters" from the JIH.

Under "background dominance" assumption, the JIH of images I_1 and I_2, $H(I_1, I_2)$, can be thought to approximate the distribution of background pixels. Then, a cross-section of $H(I_1, I_2)$ along the plane of $I_1 = I0(Constant)$ gives background probability distribution of I_2 for each intensity ($I0$) of I_1. So it should be a promising direction to evolve this background model per intensity into a background model for each pixel by using time sequential values of the pixel. In other words, we use the JIH as a prior probability when calculating the GMM which models the variation in a particular pixel and weight the time sequential data of the pixel by this prior probability. To simplify the problem without changing its essentials, we deal with grey-level images as input data in this paper.

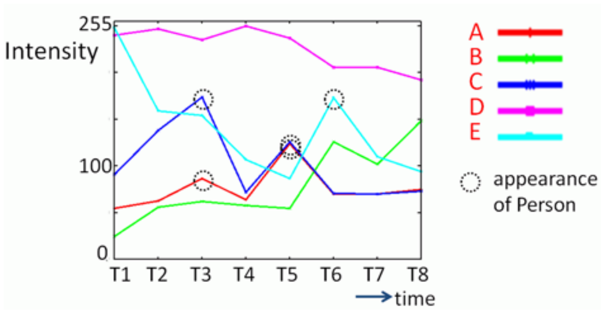
II. Method

A. Problem to solve

The application we consider in this paper is to make index images of 24 hour field surveillance images[5] by selecting only frames with appearance/disappearance of some objects, mainly for the purpose of watching illegal waste-disposal events, tracing of agricultural treatments and so on. In this application, field of view is relatively wide compared to supposed changes to detect, so the number of background pixels are much larger than that of the changes ("background-dominance" condition). Fig. 1(a) shows an example of sequential images, $T1 \sim T8$ in raster scan order, which are taken with five minute interval. Fig. 1(b) shows the intensity sequence of the pixels marked in Fig. 1(a); A,C: pixels corresponding to person appearances at $T3$, $T5$; B: a pixel corresponding to tree leaves; D: a pixel corresponding to a sunshine road; E: a pixel corresponding to person appearance at time $T6$.

1051-4651/10 $26.00 © 2010 IEEE
DOI 10.1109/ICPR.2010.248

(a)



(b)

Figure 1. Example of time sequential data; (a) image sequences of T1 to T8 in raster scan order; (b) intensity sequence of the five marked points.

Since intensity variations are fairly different depending on the position, background intensity model per pixel[1][2] is desirable. When lighting keeps changing, like gradually brightening in the morning, newly input data always becomes outliers if we build the background model only from past frames. Although it is one solution to predict the intensity transition as proposed in [4], reliable prediction is almost impossible with the long time interval we deal with. Since we do not necessarily need real-time processing for indexing, we use both fore and aft frames for modeling of variations in the intensity of each pixel.

In order to build the background model of one pixel at some specific time accurately, it is desired to use the data of the pixel at times as close to the time as possible. In the case of several minutes interval, only several frames can be used even if we allow to use the data in half of an hour from the interesting time. To compensate this lack of sampling data, we propose to leverage statistics of pixels with same intensity values in the image.

B. Algorithm

Fig.2 illustrates the concept of our method. Here,we consider change detection between time t_0 and t_0+1. First, the joint intensity histogram (JIH) of I_{t_0} and I_{t_0+1}, $H(I_{t_0}, I_{t_0+1})$, is calculated (Fig.2(a)), where I_t represents image at time t. For denoising, all joint intensity histogram used in this paper are smoothed by the 2D Gaussian filter of $\sigma_G = 3.0$. If we represent the cross-section of $H(I_{t_0}, I_{t_0+1})$ along the plane of $I_{t_o} = I_0$ as $h_{I_0}(v)(v = 0 \sim 255)$,

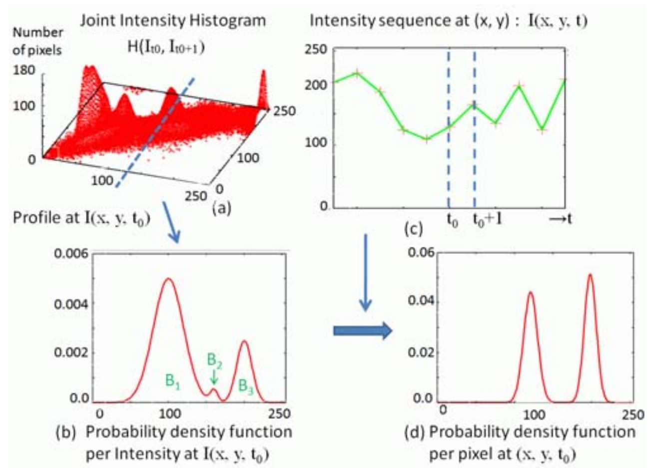


Figure 2. Concept of proposed method

normalized $h_{I_0}(v)$, $\bar{h}_{I_0}(v) = h_{I_0}(v)/\sum_{v=0}^{255} h_{I_0}(v)$, can be thought as the probability function of the intensity at time t_0+1 for the pixels whose intensity level is I_0 at time t_0. In Fig.2(b), $\bar{h}_{I_0}(v)$ in the case of $I_0 = I(x, y, t_0)$ is shown, where $I(x, y, t)$ represents the intensity of the pixel at (x, y) at time t. Under the "background dominance" condition, $\bar{h}_{I_0}(v)$ approximates background probability function at t_0+1 for intensity I_0 at t_0. That is, $\bar{h}_{I_0}(v)$ is the summation of background probability functions of the pixels with I_0 at t_0. In order to obtain the background probability function of pixel at (x, y), we propose to treat $\bar{h}_{I(x,y,t_0)}(v)$ as prior probability function of the pixel and sharpen it using time-sequential data at (x, y) as shown in Fig.2(b)(c)(d).

Based on this idea, Gaussian mixture model (GMM) which models the variation in the intensity of each pixel is calculated as follows. Here, we represent time series of intensity level at (x, y) as $X(t) = I(x, y, t)$ for conciseness.

1) Background probability function per intensity, $\bar{h}_{X(t_0)}(v)$, is obtained from the cross-section of $H(I_{t_0}, I_{t_0+1})$ at $I_{t_0} = X(t_0)$. If the function has clear plural peaks, it is partitioned as different background clusters, $B_1, B_2, .., B_K$, where K is the number of clusters [1]. The reason of this clustering is to separate different factors under the assumption that intensity variation caused by one factor should show single-peak distribution.
2) GMM is calculated with the following equations using fore and aft frames of t_0+1:

$$\mu_k = \frac{B}{A} \qquad \sigma_k = \frac{C}{A}$$

$$A = \left(\sum_{t=t_0-N_F}^{t_0+1+N_F} W_k(X(t))\right) - W_k(X(t_0+1))$$

[1]Owing to large sampling number, this segmentation can be done relatively stably, though there is room for further improvement. Resultant K is generally 1 to 5 in our experiments.

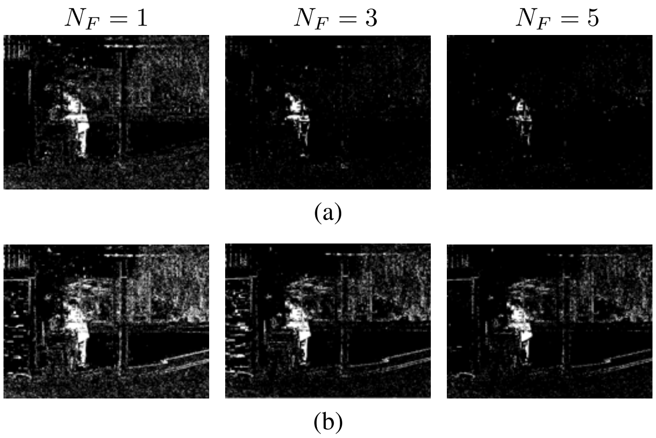


Figure 3. Effect of the number of referred frames N_F: (a) simple Gaussian model; (b) proposed model.

$$B = \left(\sum_{t=t_0-N_F}^{t_0+1+N_F} W_k(X(t))*X(t)\right) - W_k(X(t_0+1))*X(t_0+1)$$
$$C = \left(\sum_{t=t_0-N_F}^{t_0+1+N_F} W_k(X(t))*(X(t)-\mu_k)^2\right)$$
$$-W_k(X(t_0+1))*(X(t_0+1)-\mu_k)^2$$
$$\begin{cases} W_k(v)=h_{X(t_0)}(v) & if \quad v \subset B_k \\ W_k(v)=0 & else \end{cases}$$

where N_F represents the number of referred frames.
Then, the GMM is
$P(v) = \sum_1^K \eta(v, \mu_k, \sigma_k)$
where η is a Gaussian probability density function
$\eta(v, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(v-\mu)^2}{2\sigma^2}}$

If $X(t_0+1)$ matches with none of K distribution models, the pixel at (x, y, t_0+1) is detected as the pixel with change.

III. Experiments

We applied the proposed method to the image sequence of Fig. 1. Since it is almost impossible to build meaningful Gaussian mixture model from only several frames, we compared the proposed model with a single Gaussian model (SGM) which is simply calculated with ($X(t_0 - N_F)$,...$X(t_0)$,$X(t_0+2)$,...,$X(t_0+1+N_F)$).

Fig. 3 shows the effect of the number of referred frames, N_F, using $T3$ in Fig. 1. The white points illustrate detected pixels with change which is out of 3σ from each distribution model. The upper column is the results of SGM with N_F= 1,3,5 from left to right and the lower is the ones of the proposed model. Since SGM rapidly gets blurred as the referred period gets long, appearing person which has slight intensity difference from its background could not be well detected with $N_F = 3, 5$. On the other hand, in the results by the proposed model, the person area well remains even with N_F= 5.

The upper and lower rows of Fig. 4 show the results of T5 and T6 in Fig. 1 respectively. From left to right, the figures show the results of SGM, proposed model without the clustering process in the first procedure described in Section II.B (JIH1), and proposed model with the clustering process (JIH2). The reason of including the model without clustering, is to examine the pure effect of weighting with $\bar{h}_{X(t_0)}(v)$. In Table 1, the results comparing to manually given ground truth area (GT) are summed up. Here, True positive (TP) is (number of detected pixels in GT)/ (number of all pixels in GT), while False Negative(FN) is (number of detected pixels out of GT)/(number of all pixels out of GT). Here, SGMs were calculated with $N_F = 1$ because it gives their best results. Since the model obtained by the proposed method with $N_F = 1$ tends to be too sharp as shown in Fig. 5, $N_F = 2$ is used for this experiment. The reason of selecting $N_F = 2$ is that we want to use closer number to the one for SGM and the results with $N_F = 2$ are not largely different from larger number of N_F. 4σ was used for the proposed method to make its FN is almost the same with the result by SGM with 3σ. As you see, proposed models without the clustering process still give better results than SGM. This indicates that the weighting with the background probability function per intensity takes a great role. This point is clearly seen from comparison of the resultant background models at point C at T5:
Simple Gaussian model: $\eta(v, 105.7, 48.3)$
Proposed model: $\eta(v, 71.1, 2.7)$
Although the number of clustering of the background probability distribution for the intensity of this pixel is two ($K = 2$), derived pixel-wise model becomes a single

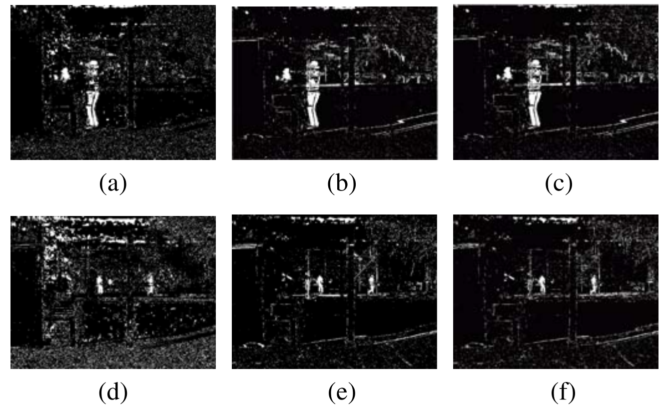
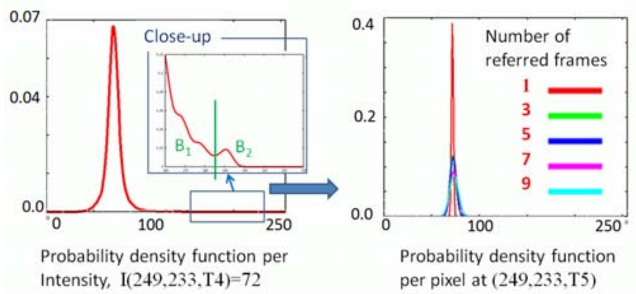


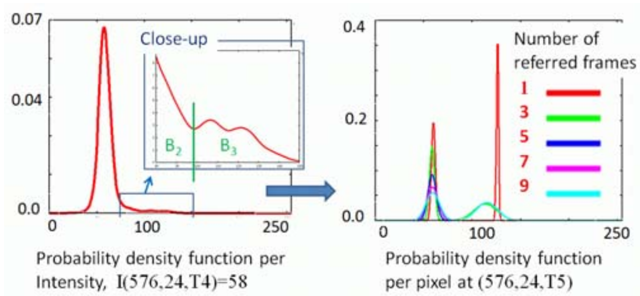
Figure 4. Comparison with simple Gaussian model: (a), (d) simple Gaussian model; (b),(e) proposed model without clustering (JIH1); (c),(f) background model with clustering (JIH2).

Table 1 True Positive and False Negative of the cases in Fig. 4

		SGM	JIH1	JIH2
T5	TP	0.532	0.643	0.641
	FN	0.069	0.050	0.051
T6	TP	0.712	0.723	0.741
	FN	0.127	0.052	0.054



(a) Point C



(b) Point B

Figure 5. Obtained background intensity model of point C and point B at T5 in Fig.1.

Gaussian distribution, because no data in the 2nd cluster are observed at this pixel during this observation period. The reason of the large difference between the two models is that intensity at T3, which is actually not the intensity of the background but the one caused by the person appearance at the time, is almost disregarded by the weighting effect of the proposed method.

The resultant background models for point B at T5 are:
Simple Gaussian model: $\eta(v, 82.0.31.2)$
Proposed model: $\eta(v, 58.4, 2.4) + \eta(v, 112.8, 11.9)$.
Here, the proposed model successfully represents swaying or shining effect of tree leaves with plural Gaussian distributions.

Fig. 5(a) and (b) show resultant background models at point C and point B with several N_F. The background models are fairly stable against the change of N_F at both positions.

IV. Conclusion

In this paper, we proposed a method to robustly model the variation of each pixel from small number of frames by using joint intensity histogram (JIH) as a prior probability. By weighting intensity sequence with the prior probability, the influence of outliers is effectively alleviated. As a result, Gaussian mixture models can be stably obtained from several frames including outliers. Although experiments up to the present are still limited, we think these preliminary results showed good perspective of the proposed strategy.

Acknowledgment

We are thankful to Dr. S. Ninomiya, Dr. M. Hirafuji and Dr. K. Tanaka of National Agricultural Research Center for data provision and discussions, and also to Dr. K. Sakaue and Dr. N. Kita for their support to this research.

References

[1] C. Stauffer and E. Grimson: "Learning Patterns of Activity Using Real-Time Tracking", *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 22, 9, pp. 747–757, 2000.

[2] A. Elgammal, R. Duraiswami, D. Harwood, L.S Davis: "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance", *Proceedings of IEEE*, Vol. 90, No. 7, pp. 1151–1163, 2002.

[3] M. Pic, L. Berthouze, and T. Kurita: "Adaptive background estimation: computing a pixel-wise learning rate from local confidence and global correlation values", *IEICE Trans. on Information and System*, Vol. E87-D, No. 1, pp. 50–57, 2004.

[4] T. Tanaka, A. Shimada, D. Arita, R. Taniguchi: "A fast algorithm for adaptive background model construction using parzen density estimation", In *Proc. of the IEEE Conf. on Advanced Video and Signal Based Surveillance*, pp.528–533, 2007.

[5] http://model.job.affrc.go.jp/FieldServerEn/default.htm

[6] T. Fukatsu and M. Hirafuji: "Field Monitoring Using Sensor-Nodes with a Web Server", *Journal of Robotics and Mechatronics*, Vol. 17, No.2, pp. 164–172, 2005.

[7] Y. Kita: "Change detection using joint intensity histogram", In *Proc. of 18th International Conference on Pattern Recognition*, pp.351–356, 2006.