Reliable Background Prediction Using Approximated GMM

Tomosuke Maeda
Advanced Course of Electric and Electronic Engineering, National Institute of Technology, Tokyo College, Tokyo, Japan
ae13708@tokyo-ct.ac.jp

Tomohiko Ohtsuka
Department of Electronic Engineering, National Institute of Technology, Tokyo College, Tokyo, Japan
tootsuka@tokyo-ct.ac.jp

Abstract

Our study proposes a new reliable background prediction for object detection in a frame sequence. Our method generates the approximated Gaussian Mixture Model (GMM) from the standard GMM by eliminating moving objects that can be easily detected based on frame differences. This reduces the computational time taken to predict the background image by averaging the intensity of each pixel of approximated GMM. However, the computational time costs more to fit each GMM parameter using an EM algorithm. In addition, this method achieves a reliable background prediction. This is possible because the precision of the background prediction is higher than other conventional approaches. Using the proposed background subtraction method, our experimental results indicate that the precision and recall levels obtained were approximately 20% higher than other levels that were obtained using conventional approaches.

1. Introduction

Computational barriers had limited the complexity of real-time video processing applications in the past. Consequently, most systems were too slow to be useful, or could be realized only by restricting them to highly controlled situations. Recently, the advent of faster computers has allowed researchers to consider more complex, robust models for the real-time analysis of streaming data. These new methods allow researchers to model real-world processes in variable conditions.

In the field of video surveillance and monitoring, a robust system must be independent of camera placement. In addition, it should be independent of objects in the visual field, overlapping of these objects, lighting movement in cluttered areas, shadows, changes in lighting, moving scene element effects, slow moving objects, or objects being introduced, or removed from the scene. In general, traditional approaches that are based on background subtraction methods fail in these situations. Hence, our goal is to create a robust, adaptive background prediction method that is adequately flexible to manage lighting variations, multiple object movements, and other arbitrary changes in the experimental scene.

Background modeling is a technique that is used to model the background changes that occur in each observation scene. This technique can detect the foreground region without supervised learning. Hence, it is widely used as an effective methodology in image analysis. The background is generated as a model, based on a statistical analysis of the observed static visual field. The background subtraction method extracts moving objects in order to detect the difference between the current frame and the background model. However, there are limitations to establishing high-quality background models because changes are not caused only by the movement of objects.

In previous studies, typical background models were based on statistical approaches [1-3]. However, other approaches could be used to predict the background. These approaches are based on the intensities of a specified pixel, its peripheral pixels [4-5], the gradient of the pixel intensity, and the temporal information between frames [6-7].

In this study, we propose a new background prediction approach. This approach uses adaptive frame accumulation, driven by motion detection, to generate a robust and reliable background. To generate an adaptive averaged background, only the static pixel intensities are accumulated. This can be performed only if the frame difference at the corresponding location is sufficiently small. The adaptive averaged background is defined as the adaptive averaging frame.

2. Related Works

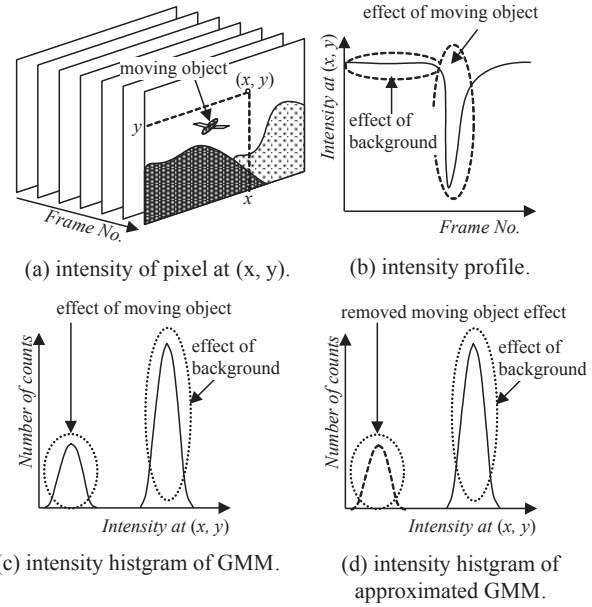
There are several reports on robust background modeling for background subtraction using sequenced scenes from past to present, in order to achieve high performance and low computational cost. For example, there are conventional approaches such as spatial approaches [4-5], temporal approaches [6-7], and statistical approaches that are based on the analysis of the pixel intensities [1-3] appearance frequency.

These conventional approaches [1-7], represent the pixel intensity distribution at the specified location using the Gaussian mixture model, as shown in Fig. 1. The pixel intensity is almost constant, when there is no moving object across the specified location. However, the pixel intensity derivative becomes higher when the moving object passes through the specified location. When this occurs, the profile of the pixel intensity at the specified location becomes a curve, as shown in Fig. 1 (b). The pixel intensity distribution, as shown in Fig. 1 (b), can be approximately represented by using the *Gaussian Mixture Model* (*GMM*), as shown in Fig. 1 (c). The *GMM* Parameter fitting can be performed with the Expectation-Minimization (*EM*) algorithm.

Robustness can also be achieved by combining hybrid approaches. However, these approaches increase computational costs. Our approach is very different when compared to these related approaches. The proposed approach applies motion detection for frame accumulation.

3. Overview of the Proposed Approach

3.1. Concept of Approximated GMM



(a) intensity of pixel at (x, y). (b) intensity profile.

(c) intensity histgram of GMM. (d) intensity histgram of approximated GMM.

Figure 1. Basic Design of Approximated GMM

The target of this study is focused on the video surveillance for the outdoor scene. *GMM* is used to achieve a reliable background prediction for object detection. Fig. 1 (a)–(c) displays the basic *GMM* features, which can be represented as a linear combination of simple Gaussian models. However, this requires high computational costs to fit every parameter of each Gaussian model in the *GMM*. Our study represents the pixel intensity distribution as a small amplitude and single Gaussian model when a moving object passes through the specified location in the frame, as shown in Fig. 1 (a) and (b). Based on the frame difference, it can detect moving objects easily. This could reduce the calculation cost, if the smaller amplitude distribution, which is caused by a moving object, can be separated from the *GMM*. We propose a new *GMM* approximation, called approximated *GMM*, which can eliminate the moving object, using the frame difference. This is depicted in Fig. 1 (a). The *GMM* is transferred into the single Gaussian model based on frame differences as shown in Fig. 1 (d). It can remove the part of the *GMM* distribution that is caused by the moving objects, using the proposed approach. It is easier to fit parameters of the single Gaussian model than the fitting problem for *GMM*.

3.2. Algorithm Outline of Proposed Approach

The proposed approach is separated into three parts: *Motion Detection*, *Object Tracking,* and *Predicted Background Generation*. The *Motion Detection* procedure can detect "*motion*" for each pixel, based on the frame difference between each adjacent frame pair. The frame difference is defined as:

$$d(x,y,n) = f(x,y,n) - f(x,y,n-1) \qquad (1)$$

where $f(x, y, n)$ is the pixel intensity on the coordinate (x, y) at the frame number n. The motion mask $m(x, y, n)$ indicates whether the object at (x, y) is moving. The motion mask $m(x, y, n)$ is set to one for the "*motion*" component, when the frame difference $d(x, y, n)$ is larger than the specified threshold value T_{hm}. Else, the motion mask $m(x, y, n)$ is set to zero. That is:

$$m(x,y,n) = \begin{cases} 1 & (d(x,y,n) \geq T_{hm}) \\ 0 & (\text{Otherwise}) \end{cases} \qquad (2)$$

The *Object Tracking* procedure detects regions of the moving objects based on moving detection results. This identifies the area where the frame difference is larger than the specified threshold value, and registers it as the moving object. It can be registered as the moving object, which is stopped in the frame window, even if the frame difference of that region is lower than the threshold value. The registered moving object is distinguished from the background if the frame difference becomes zero.

The *Predicted Background Generation* procedure estimates the ground precision background for each frame. We have introduced the adaptive frame accumulation procedure $s(x, y, n)$. This extracts the averaged background of each frame, if the position (x, y) does not belong to moving objects. Because the frame contains the moving element at the specified location (x, y), where $m(x, y, n)$ is set to one, the background pixels can be extracted based on $m(x, y, n)$. This indicates that the frame $f(x, y, n)$ contains a static element at (x, y), when the value of $m(x, y, n)$ is equal to one. The adaptive frame accumulation is defined as:

$$s(x,y,n) = \sum_{i=0}^{N_{acc}} \{1 - m(x,y,n+i)\} \cdot f(x,y,n+i) \qquad (3)$$

where N_{acc} is the period for one adaptive frame accumulation. This value can also denote the number N_m as the number of non-moving pixels at the position (x, y), which ranges from $I = 0$ to N_{acc}. That is:

$$N_m = \sum_{i=0}^{N_{acc}} \{1 - m(x,y,n+i)\} \qquad (4)$$

The predicted background $b_{ave}(x, y, n)$ can be represented as the period of an adapted frame accumulation. That is:

$$b_{ave}(x,y,n) = s(x,y,n) / N_m \qquad (5)$$

The computational cost is rather large, when $b_{ave}(x, y, n)$ is updated at every frame. We also assume that the illumination changes are not so large for a short period. Based on these observations, $b_{ave}(x, y, n)$ is updated for every N_{up} frames in the proposed approach.

When position (x, y) belongs to moving objects, $b_{ave}(x, y, n)$ is assigned to the previous value of $b_{ave}(x, y, n\text{-}N_{acc})$.

4. Experimental Results

Several experiments on background subtraction were performed for real frame numbers to evaluate the quality of the predicted backgrounds using the proposed approach. The *precision*, *recall*, and *F measure* were introduced as evaluation measures to evaluate the performance of our approach. The *precision*, *recall*, and the *F measure* are defined as:

$$Precision = TP/(TP + FP) \qquad (6)$$

$$Recall = TP/(TP + FN) \qquad (7)$$

$$F\ measure = 2/(1/Precision + 1/Recall) \qquad (8)$$

where True Positive (TP) is the number of pixels detected in the foreground, False Positive (FP) is the number of pixels falsely detected in the foreground, and False Negative (FN) is the number of pixels falsely undetected in the foreground. This indicates that the quality of the background prediction is higher when the precision and recall are larger.

The public scene databases *SCENE 1,* which contains 5337 frames, is introduced to evaluate the background quality of our approach. Sample results for *SCENE* 1 are shown in Figs. 2, 3, and 4. The frame sequence includes faster moving people with some illumination changes. Examples of the experimental results of background subtraction for *SCENE* 1 are shown as backgrounds predicted by the proposed approach. Figs. 2 (a) and 4 (a), depicts the current frame. Figs. 2 (b) and 4 (b), depicts the ground precision of the background subtraction, which is generated manually. Figs. 2 (c) and 4 (c), depicts the background predicted by the proposed approach. This is generated from a specified number n_p from past frames in these experiments. In the experiments, each averaged background is updated every n_{up} frames. In these experiments, n_p = 60 and n_{up} = 15, which are set based on human expertise. Figs. 2 (d) and 4 (d), show the background subtraction results using the background predicted by the proposed approach. The results of the *precision*, *recall*, and *F measure* are summarized in Table 1.

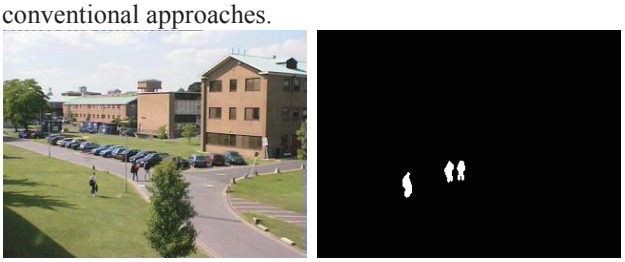
In cases where there were only fast moving persons in the foreground region, the *precision*, *recall*, and *F measures* were sufficient. These results are shown in Figs. 2, 3, and 4. Experimental results show that the background subtraction was able to detect moving objects in cases of fast moving objects. The background subtraction used the predicted background generated by our approach to detect these moving objects.

The *precision*, *recall*, and *F measure* shown in Figs. 2, 3, and 4 are adequate. However, the precision and F-measure are smaller than those of *SCENE* 1 because there are slower moving people in the foreground region, with larger illumination changes. Experimental results show that the background subtraction was able to detect moving objects in cases of fast moving objects with low illumination changes. The background subtraction used the averaged background generated by our approach to detect these moving objects.

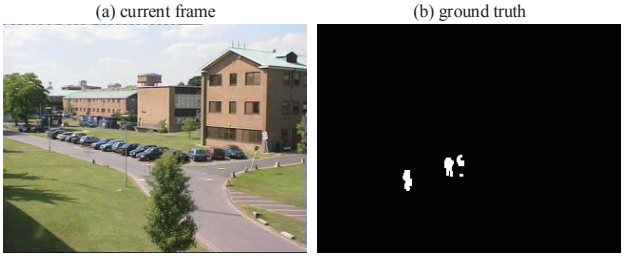
Table 2 summarizes the maximum value of the *F measure* for each of the scene databases that were generated by the conventional approaches, using the public scene database. The results can be compared with our approach. The *Stuttgart Artificial Background Subtraction Dataset* (*SABS*) [8] is applied to evaluate the performance. The Basic scene in *SABS,* contains a swaying tree, as the background fluctuates with small illumination changes. The Bootstrap scene in *SABS* does not contain training data. The Darkening scene in *SABS* contains a simple illumination change, which darkens gradually. The Light Switch scene in *SABS* contains a scene where lights in a store switch on or off. Table 2 proves that the proposed approach can achieve higher performance values for the background subtraction of each dataset.

5. Conclusion

In this study, we proposed a new background prediction approach where adaptive frame accumulation driven by motion detection can be used to generate a robust and reliable background. Only the static pixel intensities are accumulated to generate the adaptive averaged background, provided that the temporal frame difference at the corresponding location is sufficiently small. The adaptive averaged background was estimated as the adaptive averaging frame. Our experimental results showed that both the precision and recall levels obtained using our proposed background subtraction method were approximately 20% higher than those obtained using conventional approaches.

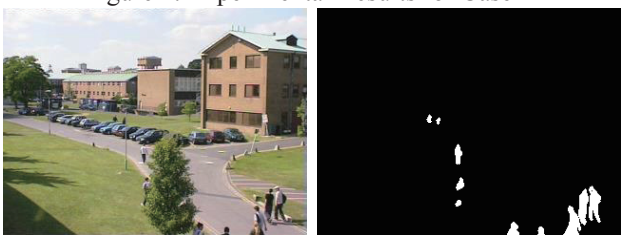


(a) current frame (b) ground truth

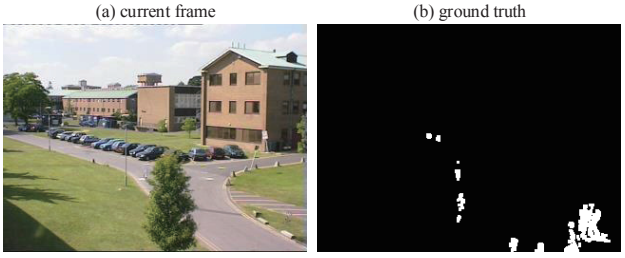


(c) predicted background by proposed approach (d) object detection result using predicted background

Figure 2. Experimental Results for Case 1



(a) current frame (b) ground truth



(c) predicted background by proposed approach (d) object detection result using predicted background

Figure 3. Experimental Results for Case 2

(a) current frame (b) ground truth

(c) predicted background by proposed approach (d) object detection result using predicted background

Figure 4. Experimental Results for Case 3

Table 1. Summarized Evaluation Results for SCENE 1

	Precision	Recall	F measure
Results of Fig. 2	0.87	0.72	0.79
Results of Fig. 3	0.76	0.81	0.78
Results of Fig. 4	0.89	0.48	0.62

Table 2. F measure Comparison for Scene Dataset SABS

Approach	Basic	Bootstrap	Darkening	Light Switch
McFrlane [9]	0.61	0.54	0.50	0.21
Stauffer [1]	0.80	0.64	0.40	0.22
Oliver [10]	0.64	-	0.30	0.20
McKenna [11]	0.52	0.30	0.48	0.31
Li [12]	0.77	0.68	0.70	0.32
Kim [13]	0.58	0.32	0.34	-
Zivkovic [14]	0.77	0.63	0.62	0.30
Maddalena [15]	0.77	0.50	0.66	0.21
Barnich [16]	0.76	0.69	0.68	0.27
Maeda[17]	0.76	0.70	0.74	0.35
Proposed	0.81	0.81	0.78	0.74

Acknowledgements

The authors are grateful to Prof. A. Piironen and Mr. N. Koskimaa at Helsinki Metropolia University of Applied Sciences for their valuable discussions. This work is supported by a Grant-in-Aid for Scientific Research (C) No. 25420395.

References

[1] C. Stauffer, W. E. L Grimson, “Adaptive Background Mixture Models for Real-Time Tracking,” Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, Vol. 2, pp. 246-252, 1999.

[2] Shimada, D. Arita, R. Taniguchi, “Dynamic Control of Adaptive Mixture-of-Gaussians Background Model,” Proceedings of IEEE International Conference on Advanced Video and Signal Based Surveillance, 2006.

[3] Elgammal, R. Duraiswami, David Harwood, “Background and Foreground Modeling using Non-parametric Kernel Density Estimation for Visual Surveillance,” Proceedings of the IEEE, Vol. 90, pp. 1151-1163, 2002.

[4] H. Marko, P. Matti, “A Texture-Based Method for Modeling the Background and Detecting Moving Objects,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 28, No. 4 pp. 657-662, 2006.

[5] S. Yoshinaga, A. Shimada, H. Nagahara, R. Taniguchi, “Object Detection Using Local Deference Patterns,” Proceeding Asian Conference on Computer Vision, 2010.

[6] Shimada, R. Taniguchi, “Hybrid Background Model using Spatial-Temporal LBP,” IEEE International Conference on Advanced Video and Signal Based Surveillance, 2009.

[7] S. Zhang, H. Yao, S. Lui, “Dynamic Background Modeling and Subtraction Using Spatio-Temporal Local Binary Patterns,” Proceedings of 15th IEEE International Conference on Image Processing, pp. 1556-1559, 2008.

[8] S. Brutzer, B. Hoferlin, Gunther Heidemann, “Evaluation of Background Subtraction Techniques for Video Surveillance,” Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, pp. 1937-1944, 2011.

[9] N. McFarlane, C. Schofield, “Segmentation and Tracking of Piglets in Images,” Machine Vision and Applications, Vol. 8, No. 3, pp. 187-193, 1995.

[10] N. Oliver, B. Rosario, A. Pentland, “A Bayesian Computer Vision System for Modeling Human Interactions,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 8, pp. 831-843, 2000.

[11] S. J. McKenna, S. Jabri, Z. Duric, A. Rosenfeld, H. Wechsler, “Tracking Groups of People,” Computer Vision and Image Understanding, Vol. 80, No. 1, pp. 42-56, 2000.

[12] L. Li, W. Huang, I. Gu, Q. Tian, “Foreground Object Detection from Videos Containing Complex Background,” Proceedings of International Conference on Multimedia, pp. 2-10, 2003.

[13] K. Kim, T. Chalidabhongse, D. Harwood, L. Davis, “Real-Time Foreground-Background Segmentation using Codebook Model,” Real-Time Imaging, Vol. 11, No. 3, pp. 172-175, 2005.

[14] Z. Zivkovic, F. van der Heijden, “Efficient Adaptive Density Estimation per Image Pixel for the Task of Background Subtraction,” Pattern Recognition Letters, Vol. 27, pp. 773-780, 2006.

[15] L. Maddalena, A. Petrosino, “A Self-Organizing Approach to Background Subtraction for Visual Surveillance Applications,” IEEE Transactions on Image Processing, Vol. 17, No. 7, pp. 1168-1177, 2008.

[16] Barnich, M. Van Droogenbroeck, “A Powerful Random Technique to Estimate the Background in Video Sequence,” Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing 2009, pp. 945-948, 2009.

[17] T. Maeda, T. Ohtsuka, H. Aoki, “Reliable Background Prediction by Approximate Gaussian Mixture Model Frame Differences for Background Subtraction”, 7th International Workshop on Image Media Quality and its Applications, pp. 84-87, 2014.