

Robust Background Modeling with Kernel Density Estimation

<http://dx.doi.org/10.3991/ijoe.v11i8.4880>

Man Hua, Yanling Li and Yinhui Luo

Civil Aviation Flight University of China, GuangHan, Sichuan, China

Abstract—Modeling background and segmenting moving objects are significant techniques for video surveillance and other video processing applications. In this paper, we proposed a novel adaptive approach for modeling background and segmenting moving objects with a non-parametric kernel density estimation. Unlike previous approaches to object detection that detect objects by global thresholds, we used a local threshold to reflect temporal persistence. With a combination of global thresholds and local thresholds, the proposed approach can handle scenes containing gradual illumination variations and noise and has no bootstrapping limitations. Experimental results on different types of videos demonstrate the utility and performance of the proposed approach.

Index Terms—Adaptive, Background Modeling, Thresholding, Kernel Density Estimation.

I. INTRODUCTION

Moving object detection and segmentation is an important research topic in the field of computer vision that has been widely used in the areas of video surveillance and video compression. At present, there are three methods for motion detection: optical flow method, frame difference method, and background subtraction method. Because the detecting speed and effects are more ideal for the background subtraction method, it has attracted a lot of research interest in recent years. The main idea of background subtraction is the use of a background subtraction frame for subtracting the background difference and using the threshold to identify the two values of difference. As a result, we find the moving target template. Therefore, the effective background modeling and the setting of the threshold are the key points of background subtraction.

Recently, there has been a large amount of work addressing the issues of background model representation and adaptation. A robust background modeling is used to represent each pixel of the background image over time by a mixture of Gaussians [1]. This approach was first proposed by Stauffer and Grimson [2,3] and has become a standard background updating procedure for comparison. Instead of modeling the feature vectors of each pixel by a mixture of several Gaussians [4], Elgammal proposed evaluating the probability of a background pixel using a nonparametric kernel density estimation (KDE) based on very recent historical samples in the image sequence [5]. Mittal improve this KDE-based background model by introducing variable bandwidth kernels and optical flow [6].

The main limitation of most traditional statistical solutions is their need for a series of training frames absent of moving objects [7]. However, in some situations, e.g., public areas, it is difficult or impossible to control the area

being monitored. In such cases, it may be necessary to train the model using a sequence that contains foreground objects. Another limitation of these methods is that most schemes determine the foreground threshold experimentally [8].

In this paper, we propose a method that overcomes these limitations. Our aim for such a framework is

- 1) A reference background image that contains no moving objects may be not required.
- 2) Adaptive thresholding to make the system adaptable to scene changes and illumination variation.

This paper is organized as follows: section two introduces the adaptive global and local thresholding method based on KDE; the third section describes the background update strategy; the fourth section gives the results of the experiment; and finally, conclusions are given.

II. ADPTIVE THRESHOLDING

A. Modelling the dissimilarity measure statistics

Embedding MS Visio drawings causes problems while transforming the document into PDF format. It is better to export them to GIF (graphics, screenshots) or JPEG (photos) format.

In this paper, we propose a novel adaptive thresholding scheme that uses two different types of adaptation. First, we performed a statistics-based threshold detection and then spatial cues were used to verify the threshold and perform adjustment according to the spatial continuous of foreground.

If we estimate the density of $p(x_i)$, the possibility of each pixel belonging to the background, we can get the KDE graph.

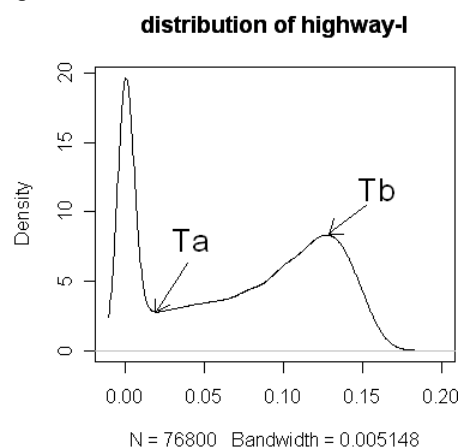


Figure 1. Kernel density of possibility of each pixel belongs to the background in an image. The training sample is 40 frames, take 50-th frame for example.

We can notice that the distribution of $p(x_t)$ is usually bimodal. If the input pixel is significantly different from the corresponding background estimate, the possibility this pixel belongs to the background, represented by $p(x_t)$, has small value. While the bigger $p(x_t)$ value represents the close similarity of background. The foreground distribution usually centers around zero and has a relatively smaller deviation and sharper crest than the background, and the background deviation depends on the variations, such as illumination or animated texture. We can use the first trough from zero as the foreground threshold, as any $p(x_t)$ left of this trough can be safely classified as foreground. As mentioned before, the second crest is from the background distribution. Any possibility larger than this point can be concluded as background.

Then we can determine two thresholds from the KDE graph, Ta and Tb , as background and background threshold accordingly.

If we treat the KDE graph as a histogram H and the bin index (i) of the histogram associated with the value of $p(x_t)$, the $H(i)$ denotes the density of $p(x_t)$. The Ta can be found by the first match of the condition below:

$$\begin{aligned} H(Ta) - H(Ta - 1) < 0 \\ H(Ta) - H(Ta + 1) < 0 \end{aligned} \quad Ta > 0 \quad (1)$$

While Tb can be determined by finding the largest peak from Ta by:

$$\begin{aligned} Tb = \text{argmax}(H(Tb)), \\ H(Tb) - H(Tb - 1) > 0, H(Tb) - H(Tb + 1) > 0, Ta < Tb \end{aligned} \quad (2)$$

In this manner, any $p(x_t)$ below Ta will be considered as foreground and larger than Tb will be taken as background. The value between Ta and Tb need further information for classification.

So far, only the temporal intensity distribution is considered in the background model, but the spatial clues also play an important role in foreground detection.

B. Robust background modeling

As it is usually hard to have prior knowledge of the scene, a non-parametric approach able to handle arbitrary densities is more suitable. A particular nonparametric technique that estimates the underlying density avoids having to store the complete data and is quite general in the kernel density estimation technique. In this technique, the underlying PDF is estimated as:

$$p(x_t) = \sum_{i=1}^N W(x_i) \cdot K(x_i - x_t) \quad (3)$$

where $K(x_i - x_t)$ is a kernel function, which is usually taken to be a density function, and $W(x_i)$ is a re-weighting function that can be adjusted to control the roles of different data points in the sample. N is the number of sample. If we choose our kernel function to be Gaussian, then the density can be estimated as:

$$p(x_t) = \sum_{i=1}^N W(x_i) \cdot \frac{1}{h\sqrt{2\pi}} \exp\left(-\frac{(x_i - x_t)^2}{2h^2}\right) \quad (4)$$

where x_t is a color/intensity feature and h is the bandwidth that controls the smoothness of the estimate. Often the re-weighting function $W(x_i)$ is required to be non-negative and sum up to 1.

In most of the previous works, uniform weights are typically used [9]; the same influence of each pixel assumption is made. However, assuming that each pixel plays the same role in background may be flawed.

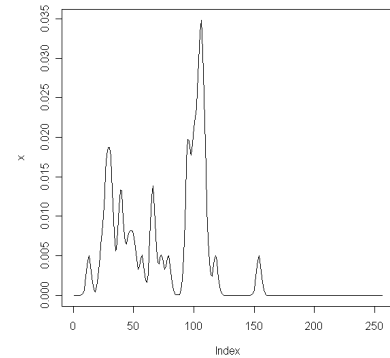
To obtain a reliable estimate, we used formula 5 to count the continual unchanged pixel values:

$$c_i(t) = \begin{cases} c_i(t)+1, & |g(t)-g(s)| < \delta \\ 0, & |g(t)-g(s)| > \delta \end{cases}, 1 < s < N \quad (5)$$

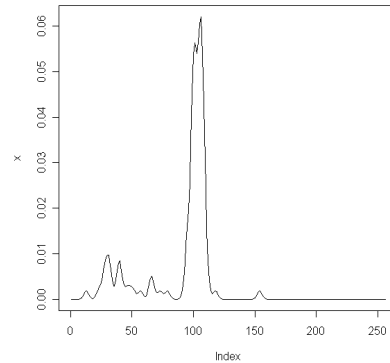
Where $g(t)$ is the intensity value on time t , and δ is a small threshold used to make a decision if a pixel value changes, the threshold is dependent on the noise of the image and normally can use a const less than 10. Then we define the weight function $W(x_i)$ as below:

$$W(x_i) = \frac{c_i(t)}{\sum_{i=1}^N c_i(t)} \quad (6)$$

As the background distribution is more temporally stationary than the foreground, the continual unchanged counter for a background is larger than the foreground. Consequently, more weight is assigned to a background pixel in the weighted kernel density estimation.



(a)



(b)

Figure 2. The KDE generated by weighted and not weighted with pixel

Figure 2 indicates the KDE generated by weighted(b) and without weighted(a); the temporal intensity distribution of the pixel without applied weight is multiple-modal, while the distribution is almost unimodal and centered at intensity 100 in weighted KDE. The interference with the foreground is suppressed. Here the background representation is drawn by estimating the probability density function of each pixel with a higher possibility of belonging to the background in the background model. The current pixel is declared as the foreground if it is unlikely to come from this background distribution, i.e. $p(\mathbf{x}_t)$ is smaller than some predefined thresholds. It is usually not easy to determine such a threshold; a popular threshold detection scheme is based on normalized statistics that consider the mean and the standard deviation of for all spatial locations. It can be adapted with noise and illumination variation.

III. BACKGROUND UPDATE STRATEGY

The background should be updated automatically when scenes change or illumination varies abruptly or gradually. Slow-moving objects may blend into the background model if the background model adapts too fast, and it will fail to identify the portion of a foreground object that has corrupted the background model. To overcome this problem, we checked if a pixel is stable enough by several continuous images to avoid the problem of blending the foreground into the background model. The checking is performed by:

$$|C_t(i) - C_{t-j}(i)| < \delta_c, 0 < j < L$$

Where $C_t(i)$ is the intensity value of pixel i at time t . δ_c is a small threshold defined to make decisions if two pixel values are unchanged, and L denotes the continuous unchanged image number. If the pixel value is left unchanged in several images, it will be updated into the background

If an abrupt scene change has been detected, the background model needs re-initialization; it usually occurs if a large percentage (above 80%) of the foreground is detected and in several continuous images [10].

IV. EXPERIMENTAL RESULTS

We compared our algorithm's output with that of several existing techniques used for background modeling, and present the compared results obtained with a typical mixture of Gaussian and KDE methods. Each of these methods is trained on 15 to 40 frames depending on the length of sequence. Also the parameter for the MoG method has been left unchanged in the implementation of OpenCv, and the threshold for KDE has been tuned to produce the best possible results for the sequences presented in Fig. 3. Tests were performed on several sequences representative of situations that might be commonly encountered in surveillance video. Here, we describe three typical scenes: VSSN06 (video 7) 390, towerl_set2, and highway I.

We have discovered that the adaptive thresholding methods have shown significantly better results for all the algorithms concerned.

To analyze the quantitative test results, multiple labeling standard tests in video VSSN06 were verified, and the method of MOG and FKDE were compared. The experimental results were measured by recall and precision, defined by [11], where recall is the ratio of the correct detections and total of the manual annotations (ground truth), and precision is the ratio of the correct number and total number. Figure 4 shows the quantitative experimental results of MOG, FKDE, and our method for the VSSN06-Video 7.

At the same time, we give the quantitative test results of video VSSN06, as shown in Table 1:

TABLE I.
 QUANTITATIVE RESULTS COMPARISON

	<i>Method/performance</i>	<i>Recall</i>	<i>precision</i>
MOG	vssn06_video4	0.87	0.12
	vssn06_video7	0.88	0.30
FKDE	<i>Method/performance</i>	<i>Recall</i>	<i>precision</i>
	vssn06_video4	0.93	0.18
	vssn06_video7	0.92	0.34
Our Method	<i>Method/performance</i>	<i>Recall</i>	<i>Precision</i>
	vssn06_video4	0.89	0.27
	vssn06_video7	0.93	0.37

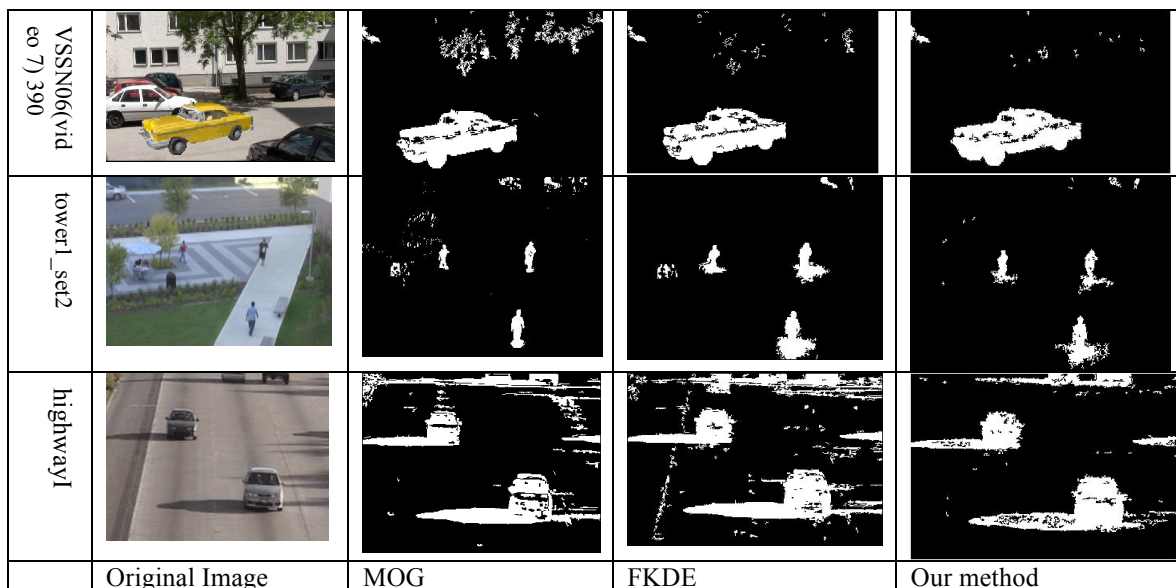


Figure 3. The test results comparison of different algorithm

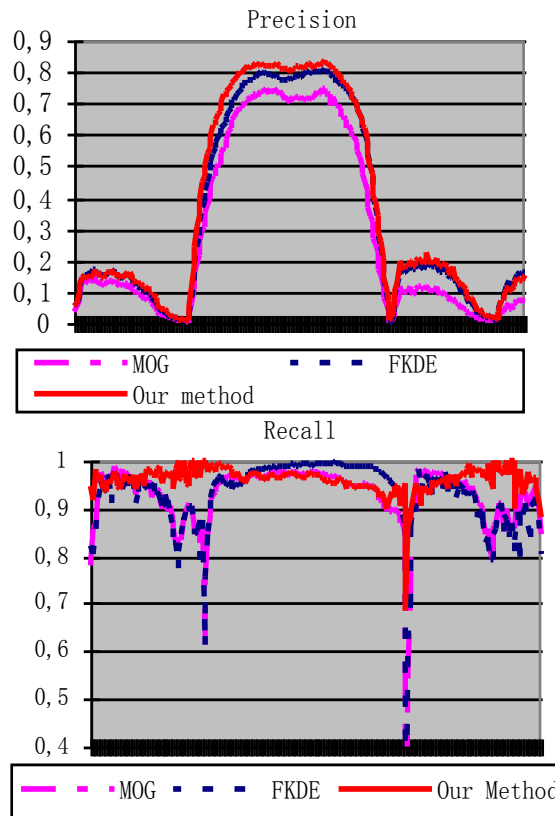


Figure 4. The test results comparison of VSSN06 video 7

As indicated in Table 1 and Figure 4, the precision and recall of our method are better than in MOG and FKDE. The average scores of our method for precision and recall are 2 and 4 percent more than FKDE and 7 and 5 percent more than MOG, respectively. While the foreground accounts for a small area of the screen, the average scores of the whole video are low. Also, the small amount of false foreground information makes the low scores.

From the point of executive speed, our algorithm can reach a detection speed of 15 frames per second likely to FKDE[5]. The reason is that the calculation complexity of the two methods both originate from the background of the Gauss score and number of clusters.

V. CONCLUSIONS

In this paper, we proposed a moving object detection algorithm with robust background modeling. The algorithm based on KDE uses a method combined with global and local thresholds that can effectively solve the problem of threshold setting through different laws of different pixels, improving the detection accuracy and avoiding the phenomenon of mixing. At the same time, the algorithm based on the data distribution can adjust the parameters automatically in a wide range of parameter variations. A large number of video experiments have been done, and the results show that this method can correctly segment the foreground object motion and demonstrate the robust-

ness of the method. How to eliminate the inner shadow and target "hole" will be the next step of work.

REFERENCES

- [1] Friedman N, Russell S. Image segmentation in video sequences: a probabilistic approach. In: Proceedings of Thirteenth Conference on Uncertainty in Artificial Intelligence. Providence, Rhode Island, USA, Morgan Kaufmann Publishers, 1997. 175~181.
- [2] Gutchess D, Trajkovic M, Cohen-Solal E, Lyons D, Jain AK. A background model initialization algorithm for video surveillance. In: Proc. of the 8th IEEE Int'l Conf. on Computer Vision. Vancouver, 2001. 733~740. <http://dx.doi.org/10.1109/iccv.2001.937598>
- [3] Hou ZQ, Han CZ. A background reconstruction algorithm based on pixel intensity classification. Journal of Software, 2005, 16(9):1568~1576
- [4] Lee D S. Effective Gaussian mixture learning for video background subtraction. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(5): 827~832. <http://dx.doi.org/10.1109/TPAMI.2005.102>
- [5] Elgammal A, Duraiswami R, Harwood D, Davis L S. Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. In: Proceedings of IEEE. 2002. 1151~1163 <http://dx.doi.org/10.1109/jproc.2002.801448>
- [6] A. Mittal and N. Paragios, "Motion-Based Background Subtraction Using Adaptive Kernel Density Estimation," Proc. Int'l Conf. Computer Vision and Pattern Recognition, pp. 302-309, 2004. <http://dx.doi.org/10.1109/cvpr.2004.1315179>
- [7] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. S. Davis, "Real-time foreground-background segmentation using codebook Model," Real-Time Imag., vol. 11, pp. 172~185, 2005. <http://dx.doi.org/10.1016/j.rti.2004.12.004>
- [8] XU Dong-Bin, HUANG Lei etc, Adaptive Kernel Density Estimation for Motion Detection, ACTA AUTOMATICA SINICA, 2008, 10
- [9] Sheather S J, Jones M C. A reliable data-based bandwidth selection method for kernel density estimation. Journal of the Royal Statistical Society Series B, 1991, 53 (3): 683 - 690.
- [10] Haritaoglu I, Harwood D, Davis L. : Real-Time surveillance of people and their activities. IEEE Trans. on PAMI, 2000, 22(8): 809~830. <http://dx.doi.org/10.1109/34.868683>
- [11] Maddalena, L. Petrosino, A. A Self-Organizing Approach to Background Subtraction for Visual Surveillance Applications. IEEE Trans, Image Processing, 2008, 17(7): 1168~1177 <http://dx.doi.org/10.1109/TIP.2008.924285>

AUTHORS

F. A. Author, Man Hua is with the School of Computer Science, Civil Aviation Flight University of China, Guanghan, Sichuan, CO 618307 China (e-mail: hua.man@163.com).

S. B. Author, Yanling Li is with the School of Computer Science, Civil Aviation Flight University of China, Guanghan, Sichuan, CO 618307 China (e-mail: lyi7652613@126.com).

T. C. Author, Yinhui Luo is with the School of Computer Science, Civil Aviation Flight University of China, Guanghan, Sichuan, CO 618307 China (e-mail: hua.man@163.com).

This work is supported by the Joint Fund of the National Natural Science Foundation of China and the Civil Aviation Administration of China (No. U1233105). Submitted, 09 March 2015. Published as resubmitted by the authors on 09 April 2015.