Nonparametric Background Generation

Yazhou Liu, Hongxun Yao
School of Computer Science and Technology
Harbin Institute of Technology
Harbin 150001, P.R. China
{yzliu,yhx}@vilab.hit.edu.cn

Wen Gao, Xilin Chen, Debin Zhao
Institute of Computing Technology
Chinese Academy of Science
Beijing 100080, P.R China
{wgao,xlchen,dbzhao}@jdl.ac.cn

Abstract

A novel background generation method based on nonparametric background model is presented for background subtraction. We introduce a new model, named as effect components description *(ECD), to model the variation of the background, by which we can relate the best estimate of the background to the modes (local maxima) of the underlying distribution. Based on ECD, an effective background generation method,* most reliable background mode *(MRBM), is developed. The basic computational module of the method is an old pattern recognition procedure, the mean shift, which can be used recursively to find the nearest stationary point of the underlying density function. The advantages of this method are three-fold: first, backgrounds can be generated from image sequence with cluttered moving objects; second, backgrounds are very clear without blur effect; third, it is robust to noise and small vibration. Extensive experimental results illustrate its good performance.*

1. Introduction

Segmentation of moving objects in video sequence is a basic task in many computer vision and video analysis applications, for instance, video surveillance, indexing for multimedia, people detection and tracking, perceptual human-computer interface and "sprite" video coding. Accurate moving object segmentation will greatly improve the performance of object tracking and activity analysis. Background subtraction is one of the mostly adopted methods for moving object detection. Even though many promising methods have been presented in the literature [3, 7, 2], the fundamental problems for the precision of moving object detection are still far from being completely solved. The first problem is that the model should reflect the real background as accurately as possible, so that the system can detect the shape of moving object accurately. The second problem is that the background model should be sensitive enough to the changes of the background scene such as the start or stop of objects. If the problems mentioned above are not being solved properly, background subtraction causes the detection of false objects, often referred to as "ghosts".

According to different background modeling approaches, these methods can be further classified as parametric and nonparametric methods. For parametric background modeling methods, the most commonly used assumption is that the underlying distribution of the intensity value of a pixel is Gaussian or mixture of Gaussian. In [5], Chris Stauffer dealt with the motion segmentation problem based on an adaptive background subtraction method by modeling pixels as a mixture of Gaussian and used an online approximation to update the model. Several improvement on Gaussian mixture modeling have been made in [4]. In [7], a three-level Wallflower scheme was presented, which tried to solve many problems exist in background maintenance, such as light switch, foreground aperture and etc. In W4 [3], three values, maximum value (M), minimum value (N) and the largest interframe absolute difference (D), were stored for each pixel to model background.

Another class of common used background modeling methods is based on nonparametric techniques, as in [2, 6]. [2] build a nonparametric background model by kernel density estimation. For each pixel, observed intensity values are kept for estimating the underlying probability density function and new intensity value's probability can be calculated by this function. The model is robust and can handle situations where the background of the scene is cluttered and not completely static but contains small motions which are introduced by the moving tree branches and bushes.

The work presented in this paper falls into nonparametric category and has strong relations with Elgammal's work in [2]. But there are two essential differences. From rationale point of view, we introduce *effect components description* (ECD) to model the background variation and *most reliable background mode* (MRBM) to robustly estimate the background scene. From computation point of view, by ex-

ploiting the mean shift procedure, we can avoid the kernel density estimation procedure for calculating the probability of each newly observed intensity value, which is a somewhat time consuming operation. So in our method, only frame difference is needed for deciding the identity of the pixel. Consequently, the robustness and effectiveness of background subtraction can be improved.

The rest of this paper is organized as follows. In section 2, effect components description is presented to model the variation of the background. Section 3 explains most reliable background mode in detail. Section 4 and section 5 contain the experimental results and discussions of future extensions.

2. Effect Components Description

The camera type, environment and the objects to be captured may vary essentially with different applications. In order to model the background effectively, we start from the simplest condition, the *ideal condition*. For each position on the spatial lattice of image sequences, its intensity values along the temporal axes should be a constant C in the *ideal condition*, which means static scene taken by a stationary camera without moving objects and system noise. We refer to the scene taken in this condition as the *ideal background scene*. But in practical applications, this *ideal condition* can be seldom met. Therefore, the background pixels can be modeled as the sum of this ideal background scene and other effect components. We define this method as *effect components description* (ECD) of the background. The effect components include:

- **System noise** N_{-sys}: Introduced by image sensors and other hardware devices, if the environment is not too rigor, this value will not affect C essentially and will only cause modest deviation.
- **Moving object** M_{-obj}: Changes introduced by real moving objects or their shadows. Most of time, this component will cause great disturbance to C.
- **Moving background** M_{-bgd}: Changes caused by moving background regions, such as tree branches moving with wind, or rippling water.
- **Illumination** S_{-illum}: It denotes the illumination changes due to the change of the position of the sun in outdoor scene or the light's on or off in indoor scene.
- **Camera displacement** D_{-cam}: Pixel intensity variation caused by small displacement of the camera.

So the observed value V_{-obsv} can be modeled as:

$$V_{-obsv} = C + N_{-sys} + M_{-obj} + M_{-bgd} + S_{-illum} + D_{-cam} \quad (1)$$

Table 1. Classification of effect components

	long term	short term
constant deviation	S_{-illum},D_{-cam}	M_{-bgd}
random deviation	N_{-sys}	M_{-obj}

As a matter of fact, effect components above can be further classified according to different properties. The first property which should be emphasized here is the duration, by which we can classify these components as *long term effects* and *short term effects*. We divide the video stream into blocks of equal length along the temporal axes. *Long term* means that the effect of the component will last for several blocks or even the whole video stream, such as N_{-sys}, S_{-illum}, and D_{-cam}. For M_{-obj} and M_{-bgd}, they happen only occasionally and do not last for a long time, so we called them *short term effects*.

Another classification criterion is deviation property. We regard S_{-illum}, D_{-cam}, and M_{-bgd} as *constant deviation effects* (time invariant). Since their effects can be modeled as constant additions/subtractions or replacement of the ideal background value C for a relative long time period. Take S_{-illum} for example, if the scene is taken in the indoor environment, and then a light is switched on, in this case the S_{-illum} can be dealt with as a constant addition to C in the following frames. For N_{-sys} and M_{-obj}, they may have random values at different time. We call them *random deviation effects* (time variant). The analysis above can be summarized as in Table 1. One point should be make clear here is that this is not a strict classification, since it depends on the block size that we chose. But this will not affect our following analysis essentially.

Since S_{-illum} and D_{-cam} make long term constant deviation to the ideal background C, so we can integrate these components into the ideal background as $C' = C + S_{-illum} + D_{-cam}$. An intuitive explanation of this integration is that if the illumination has changed or camera has been moved, it is reasonable for us to think that the background (ideal background) has changed. So (1) can be further expressed as:

$$V_{-obsv} = C' + N_{-sys} + M_{-obj} + M_{-bgd} \quad (2)$$

Thus far the observation value V_{-obsv} can be modeled as the sum of the ideal background value C' and the effect components (N_{-sys}, M_{-obj}, M_{-bgd}). These effect components will cause different influence on C'.

- N_{-sys} take place over the whole video stream and cause modest deviation to C'. So most of the observed values will not deviate far from C'.
- M_{-obj} and M_{-bgd} happen only occasionally and may cause great deviation to C'. So only a minority of the observed values will be different from C' dramatically.

The observation is that the pixel values of a spatial location should keep stable with modest deviation for the most of the time (due to *long term random deviation* N_{-sys}) and significant deviation (due to *short term deviation* M_{-obj} and M_{bgd}) may occur only when a moving object passes this location. So the extreme values with significant deviations only form a minority of the observed values in a time period.

Our task is to find an estimate $\hat{C}'$ of the ideal background C'. From the analysis above, we can see that $\hat{C}'$ should be the center of the region where the majority of the observed values located. This task can be accomplished by mean shift procedure. Here we call the $\hat{C}'$ *most reliable background mode* (MRBM).

3. Mean Shift for MRBM

The basic computational model for locating the *most reliable background mode* (MRBM) is mean shift (refer to [1] for a more comprehensive description of mean shift). It's an elegant way to find the modes of the underlying density where the gradient is zero. The algorithm's outline can be seen in Figure.1. It consists of following steps:

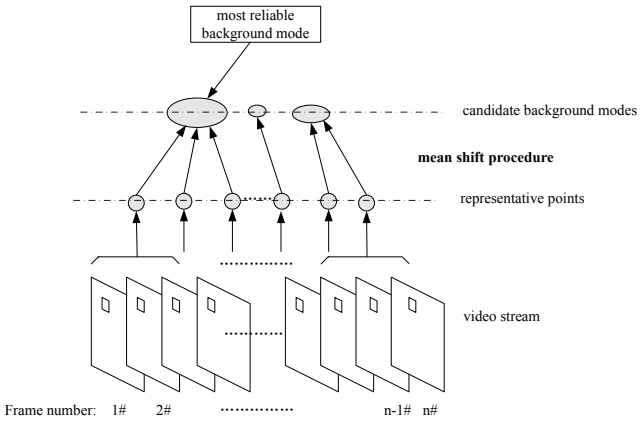


Figure 1. The outline of proposed MRBM algorithm

- **Sample selection**: We keep a sample $S = \{x_i\}, i = 1, \ldots, n$ for each pixel, where x_is are the intensity values of the pixel along the time axis and n is the sample size.
- **Representative points' selection**: To reduce the computational load, a set of $m \ll n$ points called representative sample set is selected or calculated from S. We denote this set by $P = \{p_i\}, i = 1, \ldots, m$. The entries of P can be the down sampling results or local means of the original sample points. In our experiments, we adopt the latter one.
- **Mean shift procedure**: m convergence points can be obtained by applying m mean shift procedures started from the representative sample points in P. Note that the computation of the mean shift vector is still based on the entire sample set S. Therefore the quality of the gradient density estimate is not diminished by the use of representative points.
- **Derive the candidate background modes**: Since some of the convergence points are very close or even identical to each other, these m convergence points can be clustered into $q \leq m$ classes. We can obtain q weighted cluster centers, $C = \{\{c_i, w_i\}\}, i = 1, \ldots, q$, where c_i is the intensity value and w_i is the weight of each cluster center. The number of points for each class is denoted by $l_i, i = 1, \ldots, q$, where $\sum_{i=1}^{q} l_i = m$. Therefore the weight of each class center can be defined as: $w_i = \frac{l_i}{m}, i = 1, \ldots, q$.
- **Obtain the MRBM**: $\hat{C}' = c_{i^*}$, where $i^* = \arg\max_i \{w_i\}$ and $\hat{C}'$ is the most reliable background mode as mentioned in section 2.

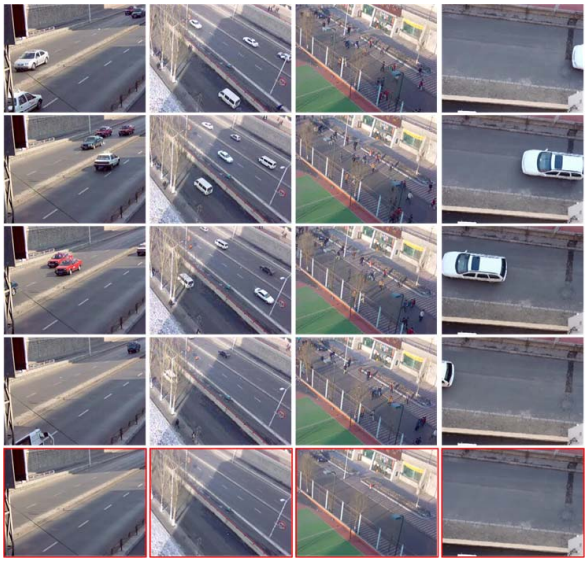
After applying the steps introduced above to all the pixels, we can generate the background scene B by these MRBMs.

4. Experiments

We evaluated our MRBM background generation algorithm in both indoor and outdoor environments. The video image size was 320×240 and the frame rate was 25 fps. In all experiments, the YUV color space was taken as feature space.

4.1. Background Generation

In many surveillance and tracking applications, it is desirable to generate a background image without moving objects. But most of the time, it's not very easy to obtain a video sequence without moving objects for training. Our algorithm MRBM can generate a very clear background image from a sequence with cluttered moving objects. Fig.2 shows some results of the generated backgrounds. In this experiment, the sample size is $n = 100$ and the length of the sequence is 360 frames. We keep 10 representative points for mean shift procedure. Frames 1, 33, 66, 99 of each video sequence are shown. The images on the bottom of Fig.2 are the backgrounds generated by our algorithm. Take Fig.2 (c) for example, this sequence is taken at a busy time from campus and there are always more than ten walking students in each frame. The generated background in Fig.2 is very clear and all the moving objects have been removed successfully.



(a) highway (b) highway (c) campus (d) road

Figure 2. Background images generated by MRBM. Frames 1, 33, 66, 99 of each sequence are shown.

4.2. Background Subtraction

Figure.3 shows the background subtraction results of our algorithm. Figure.3(a) are the observed frames and Figure.3(b) contain the background images generated by MRBM with 100 sample points. We present the difference images in Figure.3(c), in which we can see that the moving objects become prominent. Notice that the Figure.3(c) is just the difference image without thresholding or other image processing operations.

5. Conclusion

The contribution of this paper mainly consists of two aspects: first, we introduced *effect components description* (ECD) to model the variation of the background; second, based on ECD we developed a robust background generation method, *most reliable background mode* (MRBM). High quality background images can be generated by MRBM from a video sequence with cluttered moving objects. Several examples validate the method and show its efficiency.

Nevertheless, some aspects should be considered in future works. In this paper, only the cluster center with largest weight is considered, but as a matter of fact, other centers can still provide some useful information. So how to fuse this information to improve the description capacity of the background model will be addressed in our future work.

(a) current frames (b) background images (c) difference images

Figure 3. Background subtraction results

6. Acknowledgement

This research is partially sponsored by Natural Science Foundation of China under contract No.60332010, the Program for New Century Excellent Talents in University (NCET-05-0334) and ISVISION Technologies Co., Ltd.

References

[1] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(5):603 – 619, 2002.

[2] A. M. Elgammal, D. Harwood, and L. Davis. Non-parametric model for background subtraction. In *6th European Conference on Computer Vision*, Dublin, Ireland, 2000.

[3] I. Haritaoglu, D. Harwood, and L. S. Davis. W4: real-time surveillance of people and their activities. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):809–830, 2000.

[4] L. Li, W. Huang, I. Y. Gu, and Q. Tian. Foreground object detection in changing background based on color co-occurrence statistics. In *Sixth IEEE Workshop on Applications of Computer Vision*, pages 269 – 274, Orlando, Florida, 2002.

[5] C. Stauffer and W. E. L. Grimson. Learning patterns of activity using real-time tracking. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):747–757, 2000.

[6] D. Thirde and G. Jones. Hierarchical probabilistic models for video object segmentation and tracking. In *the 17th International Conference on Pattern Recognition, 2004 (CPR 2004)*, volume 1, pages 636 – 639, 2004.

[7] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers. Wallflower: Principles and practice of background maintenance. In *International Conference on Computer Vision*, Corfu, Greece, 1999.