PERIODICA POLYTECHNICA SER. TRANSP. ENG. VOL. 34, NO. 1–2, PP. 109–117 (2006)

A MIXTURE OF DISTRIBUTIONS BACKGROUND MODEL FOR TRAFFIC VIDEO SURVEILLANCE

Tamás BÉCSI and Tamás PÉTER

Department of Control and Transport Automation
Budapest University of Technology and Economics
Bertalan L. u. 2., H-1111 Budapest, Hungary
Phone: +36 1 4631044, +36 1 4633656
e-mail: becsi@kaut.kka.bme.hu;peter@sultan.kma.bme.hu

Received: March 27, 2005

Abstract

Video surveillance gathers more and more ground in the area of traffic measurement. Such computer vision applications use a static camera and some kinds of a detection algorithm for identifying moving objects. Background estimation plays an important part in these applications playing the part of generating the ideal background of the video stream, enabling the application to easily subtract the recent frame, and separate the foreground pixels from the background. The paper takes a short survey on the recently applied background estimation models, and introduces a distribution mixture model.

Keywords: computer vision, background modelling.

1. Introduction

During the research at the area of modelling Intelligent Transport Systems (ITS) our team decided to develop an own traffic measurement system. Our former projects, such as traffic simulation, incident detection [1] or traffic state estimation [8] need model verification, and video surveillance is suitable for traffic model identification and verification.

Video surveillance gathers more and more ground in the area of traffic measurement. Such computer vision applications use a static camera and some kinds of a detection algorithm for identifying moving objects. The conventional approach is the so-called 'Background Subtraction' algorithm. The algorithm compares each (or the selected) video frames to an ideal reference background model. Those pixels or regions differing significantly from the model are marked as foreground. The process does not end at this point, because there is a need of an algorithm that sorts the real foreground, from the noise, such as falling leaves, snow, rain, camera noise etc…

Background estimation plays an important role in these applications – playing the role of generating the ideal background of the video stream, enabling the application to easily subtract the recent frame, and separate the foreground pixels from the background.

Multiple background subtraction algorithms can be found in the literature, but the problem of identifying moving objects is unsolved. The problems of the algorithm come from two ways: technological and algorithmical.

The technological part of the problems comes from its real-time need: one not only has to develop a proper working algorithm, but also must stay under the computational and memory limitations.

The algorithmical problems come from the fact that traffic surveillance is a typical outdoor application, where multiple disturbances may occur. Change in illumination as time goes by, weather effects such as snow, fog or rain indicates that no static background model can be applied; the model must response to these changes. The speed of the moving objects are different, cars stop and become part of the background, etc… An algorithm that does not answer to these threats is not good enough for outdoor surveillance, so adaptiveness and the rate of adaptation are one of the main attributes of the models. [2, 3, 11]

2. A Background Subtraction Review

Background subtraction the most often uses the steps described below: [4].

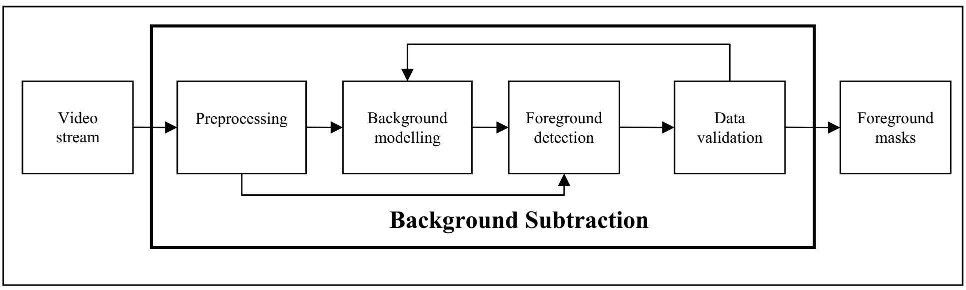

Fig. 1. Flow diagram of the background subtraction algorithm

2.1. Preprocessing

As a result of the outdoor type of surveillance multiple disturbances may occur that the algorithm must handle. There are five main tasks that the preprocessing stage must handle: [8, 12]

- The preprocessing stage of the algorithm handles the low-level disturbance. Image filtering is the best way to maintain camera noise. Single linear smoothing algorithms can be effective ways to deal with these problems.

- When the camera is moving, a certain kind of an image-relocation is necessary.
- In order to fulfill the requirements of the real-time needs frame-rate and frame-size reduction is commonly used.
- To speed up the application and to cover only the ‘interesting’ part of the frame, a mask can be used. The pixels/regions covering sidewalks or buildings can be deleted, or marked for the application, so there is no need to care about them.
- Finally, the preprocessing stage must produce the proper data format that the application needs. Color applications become more popular in the literature as the outcome of the realization that color images can carry much more information in low luminosity or low-contrast areas. But there are drawbacks of using color, the complexity of the application should grow up to four times as the simple luminance (grayscale) model.

2.2. Modelling the Background

The most important part of the application is the background-modelling algorithm. This algorithm produces the ideal background. Later, the subtraction algorithm uses the output frame of this part for determining the foreground pixels of the frame. It is a hard task to create a well-working background model, because it has to be robust against environmental changes, but also has to be sensitive enough to identify moving objects, so the rate of adaptation is the key point of success.

The literature divides adaptive techniques into two well-separated groups: recursive and non-recursive techniques having both advantages and disadvantages:

2.2.1. Non-Recursive Techniques

The non-recursive techniques are using the recent history of the video, storing multiple images in a buffer, and calculating the whole background at each stage. These sliding-window approaches need very high memory, but a very adaptive one, because the current background does not depend on the far history of the video. The memory requirements grow extremely high in slow-moving traffic, for the mass covers the area at most of the time. The milestones of this technique are described below:

Frame Differencing Frame differencing uses only two frames: the recent frame, and the frame before. The model simply divides the two frames, and provides the result as the foreground mask that needs only simple finalization. Though this is the simplest algorithm, it has many problems: Interior pixels of a slow-moving object are marked as background (known as the aperture problem), and pixels behind the moving object cast foreground (ghost effect).

Median Method Median filtering method calculates the median of each pixel using the frames stored in the buffer. The presumption is that the background is mostly visible, and the other (foreground) pixels are differing from each other. Problems occur when one tries to use this method with color frames, and the computational requirements are quite high: the complexity grows exponentially as the buffer grows.

Predictive Filtering Predictive filtering also uses a quite high buffer and uses linear estimation for determining background. The complexity of this method prevents it from real-time use.

Non-Parametric Model The non-parametric model is trying to calculate the Gaussian density function for each pixel according to its previous history. The model is working well, but it has the disadvantage of very high computational requirements.

2.2.2. Recursive Techniques

The main difference between recursive and non-recursive techniques is that recursive methods do not maintain a buffer for storing frame history; they are trying to update the background model using only the recent video frame. This holds advantages and disadvantages too; the history of the frames does not have direct effect on the recent background, but has indirect effect through the inner dynamics of the model. If the inner dynamics is not good enough, the model may diverge from the expected result. The list below describes some of the fundamental types of recursive techniques: [5, 7, 10, 6]

Fig. 2. Ghost Effect and the Aperture problem in frame differencing

Approximated Median Filter One of the most genuine ideas of background modelling was the development of the 'AMF' filter, which holds the advantages of median filtering, but achieves the median background using minimal effort. The method has the second least computational requirements after the simple frame differencing, maintaining only one background layer and simply modifying each pixels luminance by one in the direction of its difference from the same pixel of the recent frame. It is easy to understand that the method converges to the luminance value, where half of the luminance values is lower, the other half is higher, and that is the median of the pixels distribution.

Kalman-filtering The general 'Kalman-filter' is a widely used technique for estimating non-measurable state variables of a linear dynamic system under energy-limited disturbance (noise). The state space representation of the filter is:

$$x_n = \mathbf{A}x_{n-1} + \mathbf{K}(u_{n-1} - \mathbf{HA}x_{n-1}), \tag{1}$$

Where **A** is the parameter matrix of the inner dynamics, **K** is the Kalman gain matrix, and **H** is the measurement matrix.

Mixture of Gaussians Method The idea of using the 'MoG' method is that the background is not a single Gaussian distribution of luminance values, but a mixture of multiple Gaussians. The method tracks each pixel, and updates the Gaussian distributions' weight, mean and deviance values depending on the fit rate of the pixel. The number of the distributions used is from three to six depending on the limitations. The presumption is the mean of the distribution having the highest weight is the searched luminance of the background; other techniques are using some kind of a probable luminance calculation method.

2.3. *Foreground Detection*

After dealing with background estimation, the subtraction engine has to identify the foreground pixels. The obvious solution is directly subtracting the current frame and the background image:

$$|P_{est}(x, y) - P_{cur}(x, y)| \geq T \tag{2}$$

Where P_{est}(x,y) is the pixel of the background model; P_{cur}(x,y) is the corresponding pixel of the current frame and T is the threshold. If the above inequality is true the pixel (x,y) is marked as foreground.

Other two simple mechanisms are to normalize the difference to relative difference (3), or introduce the statistic parameters of the differences (4).

$$\frac{|P_{est}(x, y) - P_{cur}(x, y)|}{P_{est}(x, y)} \geq T \tag{3}$$

$$\frac{|P_{est}(x, y) - P_{cur}(x, y) - \mu_d|}{\sigma_d} \geq T \tag{4}$$

Where μ_d and σ_d are the mean a deviation values for the pixel differences.

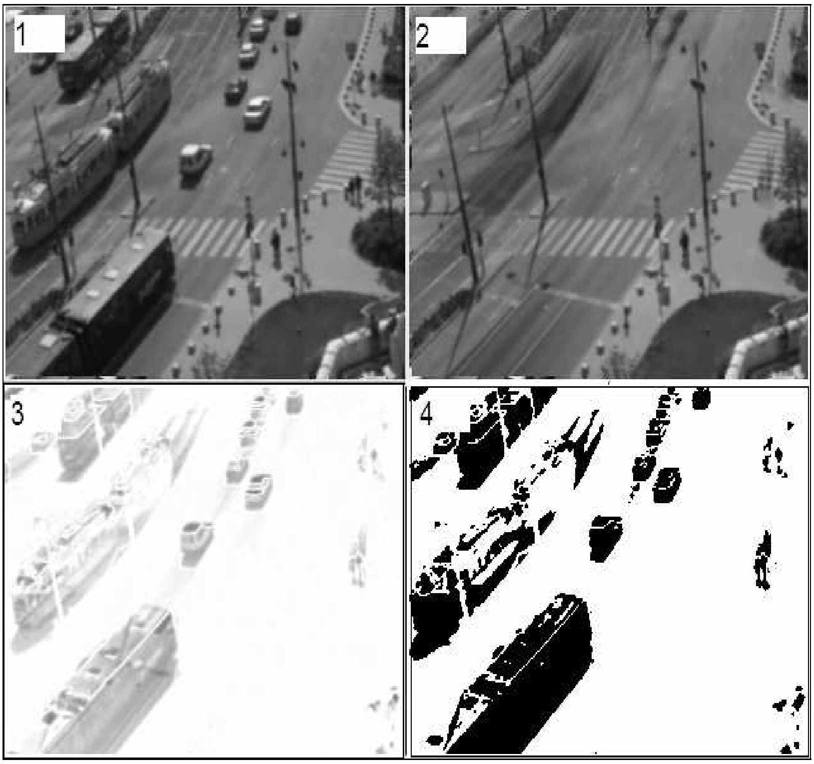

Fig. 3. AMF method; 1: Recent Frame, 2: Background Model, 3: Simple Difference, 4: Simple Noise Gate

3. The Mixture of Distributions Model

The idea of developing such model came from the recognition of the fact that the MoG model is the best performing model from those described above. However, there are disadvantages of the model which make it computationally ineffective. The calculation of the multiple distributions takes too much processing effort, and it is not always required to achieve eligible results.

3.1. Step 1, Initialization

At first, the algorithm initializes the distributions. This means that the process clears all distributions. By the arrival of the first video frame, the algorithm sets the first distributions luminance value to the luminance of the corresponding pixel of the

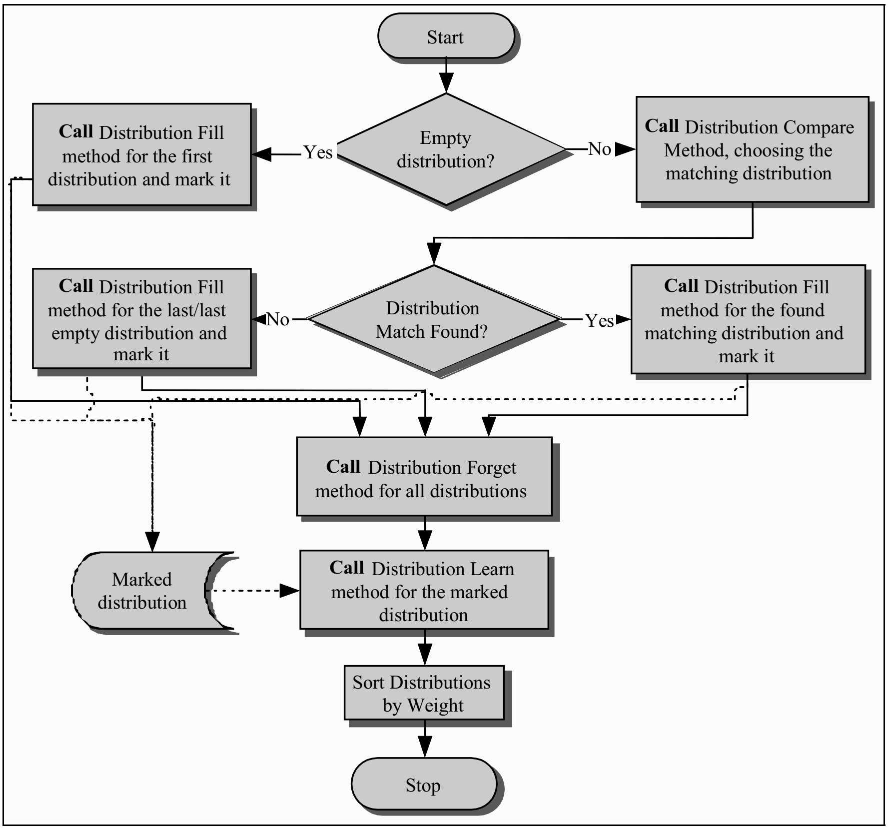

Fig. 4. Flow Chart of the Multi Distribution Model

video frame, and the distribution gets a minimal weight and a big width. At this time, the 'predicted' background is just the same as the first frame.

3.2. Step 2, Process Flow

After initialization, the process compares all incoming pixels with the distributions luminance values in order of their weights. Match is found, if the distance between the pixel's and the distribution's luminance is smaller than the distribution's width.

Fig. 2 shows the result of the method. The quality of the background is much better than the AMF's one and simple differencing and noise gate can produce optimal output. In case of a match, the algorithm increases the distribution's weight, decreases its width, and corrugates the luminance just as in the AMF method, by stepping a unit towards the pixel's luminance.

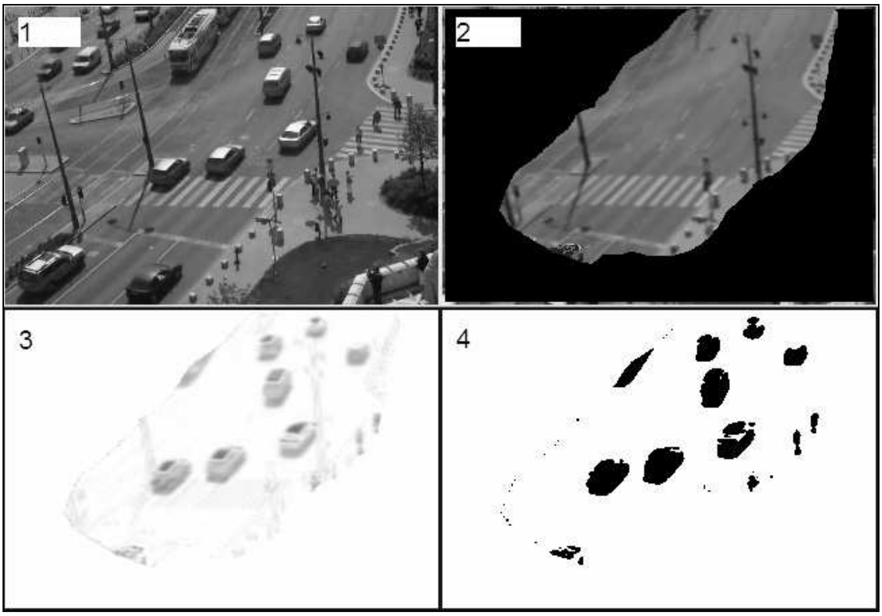

Fig. 5. MoD method; 1: Recent Frame, 2: Background Model, 3: Simple Difference, 4: Simple Noise Gate

If there is no match, the process terminates the weakest distribution (the one with the smallest weight), and replaces it with the current pixel, giving some initial small weight, and big width.

The distributions not involved in the match use the forgetting method, decreasing their weights, and increasing their widths.

At last, the process sorts the distributions by their weight, and the set of each pixel's strongest (highest weight) distributions realizes the estimated background. *Fig. 1* shows the flow chart of the process.

The developed MoD method combines the advantages of the MoG and the AMF methods, by using the multi-layer feature of the MoG, and the fast adaptiveness of the AMF.

By this hybridization, the outcome is an algorithm, which is sufficiently accurate just like the MoG, because it misses the shadow effect raised by a moving object in AMF, but faster, because of the simple AMF-like learning method.

4. Conclusions and Future Work

The paper introduced a method of background estimation which looks well applicable for traffic surveillance. Its results were shown using simple differencing, to easily compare with other methods. For more exact comparison, some measurable properties should be used at the future, e.g. pixel match percentage to optimal background, and detection percentage of foreground pixels using similar algorithms.

The future applications of the designed algorithm and application will be to evaluate and measure driver behaviours and the identification of microscopic traffic models. Currently we are at the final stage of this development.

References

[1] BÉCSI, T. – PÉTER, T., An Adaptive Approach to Modeling Traffic Flow and Incident Detection on Highways, *Proc. of Interacademia 2004*, pp. 57–65, Budapest, 2004.

[2] BONNESON, J.– ABBAS M., Intersection Video Detection Manual, FHWA, 2002, tti.tamu.edu/documents/4285-2.pdf

[3] BONNESON, J.– ABBAS, M., Intersection Video Detection Field Handbook, FHWA, 2002, tti.tamu.edu/documents/4285-3.pdf

[4] CHEUNG, S. – KAMATH, C., Robust Techniques for Background Subtraction in Urban Traffic Video, Video Communications and Image Processing, Volume 5308, pp. 881–892, SPIE Electronic Imaging, San Jose, 2004.

[5] CHO, J. H. – KIM, S. D., Object Detection Using Multi-resolution Mosaic in Image Sequences, Signal Processing: Image Communication 20, 2005, pp. 233–253.

[6] ELGAMMAL, A. – DURAISWAMI, R. – HARWOOD, D. – DAVIS, L. S., Background and Foreground Modeling Using Nonparametric Kernel Density Estimation for Visual Surveillance, *Proc. of The IEEE*, **90**, pp. 1151–1163, 2002.

[7] HARVILLE, M. – GORDON, G. – WOODFILL, J., Adaptive Video Background Modeling Using Color and Depth, *International Conference on Image Processing, 2001. Proceedings.*, **3**, 7-10, pp. 90–93.

[8] KULCSÁR, B. – BÉCSI, T. – VARGA, I., Estimation of Dynamic Origin Destionation MAtrix of Traffic Systems, *Periodica Polytechnica Ser. Transp. Eng.* **33**, No.1-2, pp. 3–14, 2005.

[9] PATIN, F., An Introduction To Digital Image Processing, 2003 http://www.programmersheaven.com/articles/patin/ImageProc.pdf,

[10] STAUFFER C. -GRIMSON W.E.L., Adaptive background mixture models for real-time tracking", CVPR99, Fort Colins, 1999.

[11] VEERARAGHAVAN, H.– MASOUD, O.– PAPANIKOLOPOULOS N. P., Managing Suburban Intersections Through Sensing, University of Minnesota - ITS Institute, 2002 http://www.cts.umn.edu/pdf/CTS_02-07.pdf

[12] WALTZ, F.M. – MILLER J. W.V., An Efficient Algorithm for Gaussian Blur Using Finite-state Machines, *SPIE Conf. on Machine Vision Systems for Inspection and Metrology VII*, 1998.