A Survey: Background Subtraction Techniques

Miss Helly M Desai, Mr. Vaibhav Gandhi

Abstract— Background subtraction is a widely used technique to detect a foreground image from its background. We present a study of different background subtraction methods and compare them. All review methods are compared based on their robustness, memory usage and computational effort they require. The overall evaluation shows that GMM and KDE gives the best performance in accuracy but by using different feature extraction algorithm like SURF algorithm we can improve the performance of the basic background subtraction methods.

Index Terms— Background subtraction, Gaussian mixture model, Kernel Density Estimation, SURF, Moving object detection, Frame difference, Codebook

———————————— ◆ ————————————

1 INTRODUCTION

BACKGROUND subtraction is a method which is basically used to detect a foreground object from the image. This is used in object tracking, target tracking, traffic analysis, and video appliances. Object detection is performed to check existence of an object in video and to locate that object. [19] During a video sequence spatial and temporal changes are monitored in order to track objects. Spatial and temporal changes include presence, size, shape, etc.

The main focus here is to track path of an object as it moves around a scene. There are some challenges in object tracking as described in [19]:

1. Loss of evidence caused by estimate of the 3D realm on 2D image,
2. Noise in an image,
3. Difficult object motion,
4. Imperfect and entire object occlusions,
5. Complex objects structures.

Today video surveillance system plays important role in safety and security area. It is used as a remote eye. It is useful in all the areas such as residence, malls, hospitals, airports to detect a real time moving object and to analyze that object.

There are different methods to detect a moving object.

1. Optical Flow method.
2. Consecutive frame difference.
3. Background subtraction.

In Optical Flow Method, automatic feature extraction has been done by using clustering so that features are extracted from the current image by using x-mean cluster and classify extracted features points based on their estimated motion parameters. The segmented region is labeled and labeling result characterize as moving object. Moreover it cannot be used for real time application without using some special hardware.

Consecutive frame subtraction is an easy method and also works good in dynamic environment but it is not extraction moving object completely. Gaussian mixture model and watershed are used in this type of method. Propose that where first the difference between two frames is calculated and then dividing it to moving area and background area.

Background subtraction is a method in which incoming frames are compared with the background model and the moving object will be detected. There is also a difficulty in such a situation where background is keep on changing.

One of the methods that is widely used to detect a moving object is background subtraction method. It is widely used for video security applications. The main reason to use this method is that it is simple, accurate and takes less computational time. Like other methods background subtraction method also have to face some challenges like system limitations and environmental changes. System limitation means that the platform on which application has been used and environmental changes means changes in illuminations, lights, shadows, colors similarity etc.

2 BASIC OF BACKGROUND SUBTRACTION METHOD

Background subtraction is used to detect foreground object by comparing two different frames and will find the difference and create a distance matrix. Basically it will compare the value of the difference with the threshold value. Now a threshold value is not predefined but it will calculate the threshold value by using first few frames that you have given. So the main scenario is that if the difference is greater than a threshold value than it is marked as a moving object otherwise it will take it as a background image.
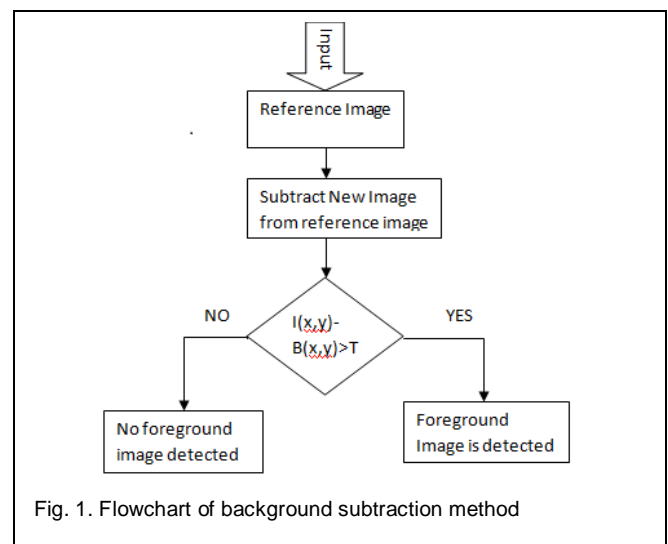

Fig. 1. Flowchart of background subtraction method

Now the challenges that have to be face during background subtraction is that background is changes frequently because of illumination changes, motion changes and changing in

background geometry. So a simple inter frame difference is quite weak solution to detect a moving object accurately.

There are many different background subtraction methods like frame difference, Gaussian mixture model, kernel density estimation, codebook. All the methods give different accuracy in different methods.

3   BACKGROUND SUBTRACTION ALGORITHM

Most BS techniques share a common denominator: they make the assumption that the observed video sequence I is made of a static background B in front of which moving objects are observed. With the assumption that every moving object is made of a color (or a color distribution) different from the one observed in B, numerous BS methods can be summarized by the following formula:

$$x_t(s) = 1 \text{ if } d(I_{s,t}, B_s) > \tau \text{ otherwise } 0.$$

Where τ is a threshold, X_t is the motion label field at time t (also called motion mask), d is the distance between $I_{s,t}$, the color at time t and pixel s, and B_s, the background model at pixel s. The main difference between several BS methods is how B is modeled and which distance metric d they use. In the following subsection, various BS techniques are presented as well as their respective distance measure.

3.1 Basic Motion Detection

The easiest way to model the background B is through a single grayscale/color image void of moving objects. This image can be a picture taken in absence of motion or estimated via a temporal median filter [5, 10], [17]. To handle brightness changes and background modification it is updated as follows:

$$B_{s,t+1} = (1-\alpha)B_{s,t} + \alpha.I_{s,t} \quad (2)$$

Where α is a constant whose value ranges between 0 and 1. With this simple background model, pixels corresponding to foreground moving objects can be detected by thresholding any of those distance functions:

$$
\begin{aligned}
d_0 &= \left| I_{s,t} - B_{s,t} \right| \\
d_1 &= \left| I_{s,t}^R - B_{s,t}^R \right| + \left| I_{s,t}^G - B_{s,t}^G \right| + \left| I_{s,t}^B - B_{s,t}^B \right| \\
d_2 &= (I_{s,t}^R - B_{s,t}^R)^2 + (I_{s,t}^G - B_{s,t}^G)^2 + (I_{s,t}^B - B_{s,t}^B)^2 \\
d_\infty &= max\{\left| I_{s,t}^R - B_{s,t}^R \right|, \left| I_{s,t}^G - B_{s,t}^G \right|, \left| I_{s,t}^B - B_{s,t}^B \right|\}
\end{aligned}
$$

Where R,G and B stand for the red, green and blue channels and d0 is a measure operating on grayscale images. Note that it is also possible to use the previous frame I_{t-1} as background image B [7]. With this configuration though, motion detection becomes an inter-frame change detection process which is robust to illumination changes but suffers from a severe aperture problem since only parts of the moving objects are detected.

3.2 Gaussian Mixture Model (GMM)

To account for backgrounds made of animated textures (such as waves on the water or trees shaken by the wind),

some authors proposed the use of multimodal PDFs. Stauffer and Grimson's method [15], for example, models every pixel with a mixture of K Gaussians. For this method, the probability of occurrence of a color at a given pixel s is given by:

$$P(I_{s,t}) = \sum_{i=1}^{k} \omega_{i,s,t} N(\mu_{i,s,t}, \sum_{i,s,t})$$

where N (μi,s,t , $\sum$i,s,t) is the ith Gaussian model and ωi,s,t its weight. Note that for computational purposes, as suggested by Stauffer and Grimson, the covariance matrix $\sum$i,s,t can be assumed to be diagonal, $\sum = \sigma^2$ Id. In their method, parameters of the matched component (i.e. the nearest Gaussian for which $I_{s,t}$ is within 2.5 standard deviations of its mean) are updated as follows :

$$
\begin{aligned}
\omega_{i,s,t} &= (1-\alpha)\omega_{i,s,t-1} + \alpha \\
\mu_{i,s,t} &= (1-\rho).\mu_{i,s,t-1} + \rho.I_{s,t} \\
\sigma_{i,s,t}^2 &= (1-\rho).\sigma_{i,s,t-1}^2 + \rho.d_2(I_{s,t}, \mu_{i,s,t})
\end{aligned}
$$

Where α is an user-defined learning rate, ρ is a second learning rate defined as $\rho = \alpha.N(\mu_{(i,s,t)}, \sum_{(i,s,t)})$ and d2 is the distance defined in equation above. Parameters μ and σ of unmatched distributions remain the same while their weight is reduced.

3.3 Kernel Density Estimation (KDE)

An unstructured approach can also be used to model a multimodal PDF. In this perspective, Elgammal et al. [6] proposed a Parzen-window estimate at each background pixel:

$$P(I_{s,t}) = \frac{1}{N} \sum_{i=t-N}^{t-1} K(I_{s,t} - I_{s,i})$$

A pixel is labeled as foreground if it is unlikely to come from this distribution, i.e. when P(Is,t) is smaller than a predefined threshold. Note that j can be fixed or pre-estimated following Elgammal et al.'s method [6]. Formal methods such as Mittal and Paragios's [14] which is based on "Variable Bandwidth Kernels".

3.3 Codebook

Another approach whose goal is to cope with multimodal backgrounds is the so-called codebook method by Kim et al. [13]. Based on a training sequence, the method assigns to each background pixel a series of key color values (called codewords) stored in a codebook. These codewords will take over particular color in a certain period of time. For instance, a pixel in a stable area may be summarized by only one codeword whereas a pixel located over a tree shaken by the wind could be, for example, summarized by three values: green for the foliage, blue for the sky, and brown for the bark. With the assumption that shadows correspond to brightness shifts and real foreground moving objects to chroma shifts, the original version of the method has been designed to eliminate false positives caused by illumination changes. This is done by performing a separate evaluation of color distortion

4 EXPERIMENTAL PROTOCOL

Our goal is to evaluate the ability of each method to correctly detect motion, a ground truth is available for all videos allowing the evaluation of true positives (TP), false positives (FP) and false negatives (FN) numbers. Those values are combined into a (Precision/Recall) couple defined as:

$$\text{Precision} = \big(TP/(TP+FP)\big) \quad \text{Recall} = \big(TP/(TP+FN)\big)$$

A good algorithm is one producing simultaneously a small number of FP and FN. The comparison between methods is made easier as we do not have to find the best threshold for each method over each video because different threshold values are used.

5 DISCUSSION

Here from the following table we can say that each method works with different efficiency in different environment. We can say that GMM works good in all environment comparatively with others. We can achieve good efficiency of basic method in multi model and noisy background by using some feature extracting methods. Here number of stars represents level of efficiency.

TABLE 1

COMPARISION OF METHODS

	Basic	GMM	KDE	Codebook
Static Background	***	***	***	***
Multimodel Background	*	***	***	***
Noisy Background	*	***	***	***
Computation Time	***	**	*	-
Memory Requirement	***	**	*	**

6 CONCLUSION

Here we can say that GMM and KDE are comparatively good methods for background subtraction but in real time applications they are not giving that much good result so if we use some feature extractions methods like SURF then we can improve efficiency of detecting a foreground object.

More over one more criteria also should be there that when background is also moving then what will be the detection rate of foreground object.

REFERENCES

[1] Y. Benezeth , B. Emile, C. Rosenberger. Comparative study on foreground detection algorithms for human detection. International Conference on Image and Graphics, pages 661–666, 2007.

[2] A.C. Bovik. Handbook of Image and Video Processing. Academic Press, Inc. Orlando, FL, USA, 2005.

[3] V. Cheng and N. Kehtarnavaz. A smart camera application: Dsp-based people detection and tracking. Journal of Electronic Imaging, 9(3):336–346, 2000.

[4] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati. Detecting moving objects, ghosts and shadows in video streams. Transactions on Pattern Analysis and Machine Intelligence, pages 1337–1342, 2003.

[5] R. Cucchiara, C. Grana, A. Prati, and R. Vezzani. Probabilistic posture classification for human-behavior analysis. in Transactions on Systems, Man and Cybernetics, Part A, 35:42– 54, 2005.

[6] A Elgammal, D. Harwood, and L. Davis. Non-parametric model for background subtraction. European Conference on Computer Vision, pages 751–767, 2000.

[7] S. Ghidary, Y. Nakata, T. Takamori, and M. Hattori. Human detection and localization at indoor environment by homerobot. International Conference on Systems, Man, and Cybernetics, 2:1360–1365, 2000.

[8] D. Hall, J. Nascimento, P. Ribeiro, E. Andrade, P. Moreno, S. Pesnel, T. List, R. Emonet, R.B. Fisher, J.S. Victor, and J.L. Crowley. Comparison of target detection algorithms using adaptive background models. International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, pages 113– 120, 2005.

[9] I. Haritaoglu, D. Harwood, and L.S. Davis. W 4: real-time surveillance of people and their activities. Pattern Analysis and Machine Intelligence, 22:809–830, 2000.

[10] J. Heikkila and O. Silven. A real-time system for monitoring of cyclists and pedestrians. Workshop on Visual Surveillance, pages 74–81, 2004.

[11] W. Hu, T. Tan, L. Wang, and S. Maybank. A survey on visual surveillance of object motion and behaviors. Systems, Man, and Cybernetics, Part C: Applications and Reviews, 34(3):334– 352, 2004.

[12] P. KaewTraKulPong and R. Bowden. An improved adaptive background mixture model for real-time tracking with shadow detection. Workshop on Advanced Video-based Surveillance Systems conference, 2001.

[13] K. Kim, T.H. Chalidabhongse, D. Harwood, and L. Davis. Real-time foreground-background segmentation using codebook model. Real-Time Imaging, 11:172–185, 2005

[14] A.Mittal and N. Paragios. Motion-based background subtraction using adaptive kernel density estimation. Proceedings of the international conference on Computer Vision and Pattern Recognition, 2004

[15] C. Stauffer and W.E.L. Grimson. Adaptive background mixture models for real-time tracking. International conference on Computer Vision and Pattern Recognition, 2, 1999.

[16] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfinder: Real-time tracking of the human body. Pattern Analysis and Machine Intelligence, 1997.

[17] Q. Zhou and J. Aggarwal. Tracking and classifying moving objects from video. Performance Evaluation of Tracking Systems Workshop, 2001.

[18] Z. Zivkovic. Improved adaptive gaussian mixture model for background subtraction. International Conference on Pattern Recognition, 2004.