RECENT TRENDS IN MACHINE LEARNING FOR BACKGROUND MODELING AND DETECTING MOVING OBJECTS

Shobha.G[1] & N. Satish Kumar[2]
[1]Computer Science & Engineering Dept., R V College of Engineering, Bangalore, India.
[2]Computer Science & Engineering Dept., Jain Engineering College, India.

ABSTRACT

Background modeling is often used in the context of moving objects detection from static cameras. Numerous methods have been developed over the recent years and the most used are the statistical ones. This paper describes the current state-of-art in background modeling methods for moving object detection. We also propose a method for background modeling based on texture features and self organizing map through Artificial Neural Networks. This approach can handle video scenes containing moving background, illumination variation, and also include into the background model shadow cast by moving objects.

KEYWORDS: *Object detection, texture features, Neural Network.*

I. INTRODUCTION

A static camera observing a scene is a common case of a surveillance system. Detecting intruding objects is an essential step in analyzing the scene. An usually applicable assumption is that the images of the scene without the intruding objects exhibit some regular behavior that can be well described by a statistical model. If we have a statistical model of the scene, an intruding object can be detected by spotting the parts of the image that don't fit the model. This process is usually known as background subtraction.

Background subtraction is often one of the first tasks in machine learning applications. The performance of background subtraction model depends upon the background modeling technique used. Background modeling technique should be capable of dealing with movement through cluttered areas, objects overlapping in the visual field, shadows, lighting changes etc [1]. In this paper we explore the traditional approaches based on background modeling methods which typically fail in general situations. We also propose a robust, machine learning system that is flexible enough to handle variations, in lighting, moving scene clutter, multiple moving objects and other arbitrary changes to the observed scene. The detection of moving objects in video streams is the first relevant steps of information extraction, and can be used for recognition, classification, and activity analysis more efficiently, since only moving pixels need to be considered.

The basic approach of background modeling is averaging the images over time [2]. The background approximation formed is similar to the current static scene except where motion occurs. The main problem is its sensitivity to dynamic scene changes, and consequent need for the background model adaptation via background maintenance such as light changes, moving background, cast shadows and camouflage [3][4].

II. RELATED WORK

A very popular technique is to model each pixel in a video frame with a single Gaussian. This is a simple technique which calculates an average image on the scene, to subtract each new video frame from it and to threshold the result. Machine learning version of this algorithm updates the model parameters recursively by using the single adaptive filter. This model does not work well when the scene is dynamic. By using more than one Gaussian distribution per pixel i.e., Mixture of Gaussians (MoG) it is possible to handle such situations [5] [6].

Mixture of Gaussian algorithm have been proved to be a very successful one in outdoor scene but in indoor scene where the illumination changes are very drastic and very frequent, this algorithm gives a lot of noise or false positives along with foreground object after the background subtraction process. To overcome this, discriminative texture features was used to capture background statistics [7].

Texture analysis plays an important role in many image analysis applications. Even though color is an important cue in interpreting images, there are situations where color measurements just are not enough — nor even applicable. Texture measures can also cope better with varying illumination conditions, for instance in outdoor conditions. Therefore, they can be useful tools for high-level interpretation of natural scene image content. This approach used local binary pattern (LBP) texture operator which has shown excellent performance in many applications and has several properties that favor its usage in background modeling [8] [9].

The local binary pattern (LBP) operator was developed as a gray-scale invariant pattern measure adding complementary information to the "amount" of texture in images. It was first mentioned by Harwood et al. (1993), and introduced to the public by Ojala et al. (1996). Later, it has shown excellent performance in many comparative studies, in terms of both speed and discrimination performance. In a way, the approach is bringing together the separate statistical and structural approaches to texture analysis, opening a door for the analysis of both stochastic microtextures and deterministic macrotextures simultaneously. It also seems to have some correspondence with new psychophysical findings in the human visual system. Furthermore, being independent of any monotonic transformation of gray scale, the operator is perfectly suited for complementing color measurements — or to be complemented by an orthogonal measure of image contrast. The LBP operator can be made invariant against rotation, and it also supports multi-scale analysis [10] [11].

Although the moving object is detected by the above method, they leave behind "holes" where newly exposed background imagery differs from the known background model. While the background model eventually adapts to these "holes", they generate false alarms for a short period of time. The above method will include into the background model shadows cast by moving objects. Therefore it is necessary to construct an approach to motion detection based on background model that automatically adapts to changes in a self organizing manner and without a priori knowledge [12] [13] [14].

To overcome the above limitations, we propose an approach based on self organizing map (SOM) through artificial neural networks. SOM is used as a visualization tool for scene analysis [15]. In this approach training images are divided into sub images. LBP texture features for each block are used to train the SOM neural network. Training results into clusters, with each cluster containing similar data. This method classifies different regions of the background more accurately. For detection of the foreground object from the current frame, the current frame is divided into blocks; LBP texture features are calculated for each block and are fed into the trained SOM neural network. The LBP texture features of the moving object will not match to any of the clusters that are formed at the training phase and thus can be identified as the foreground. The proposed approach can handle scenes containing moving backgrounds, gradual illumination variations, can include into the background model shadows cast by moving object.

III. MODELLING THE BACKGROUND BY SELF-ORGANIZING THE TEXTURE FEATURES

The adopted artificial neural network is organized as 2-D flat grid of neurons or nodes [16]. Each node computes a function of the weighted linear combination of incoming inputs, where weights resemble the neural network learning. So each node represents a weight vector, obtained collecting the weights related to the incoming links. An incoming pattern is mapped to the node whose weight vector is "most similar" to the pattern, and weight vectors in a neighborhood of such node are updated. The whole set of weight vectors acts as a background model that is used for background subtraction in order to identify moving objects.

Our Experiment analysis consists of two stages

1. Initial Background Model
2. Subtraction and Update of the Background Model

3.1 Initial Background Model

This step consists of initializing the weight vectors of the network: The first image of the video sequence is good initial approximation of the background. An LBP texture feature [10] which is a binary pattern of 8 bit is used to initialize the weight vectors of the neural network. Let *lbp* be the LBP components of the generic pixel (x,y) of the first sequence frame I_0 and let C= $(c_1, c_2, c_3 \ldots\ldots c_n)$ be the weight vector for pixel (x,y).

3.2 Subtraction and updation of the background Model

After initialization, for the subsequent samples *lbp* texture features of each pixel are fed to the neural network. Each incoming pixle p_t of the *t*th sequence frame I_t is compared to the current pixel weight vector C to determine if there exists a weight vector that matches it. If a best matching vector c_m is found, it means that p_t belongs to the background and it is used as the pixel encoding approximation. If no matching weight vector exists, p_t will be either the shadow cast by some object, which should not be used to update the corresponding weight vectors. In this case p_t will be considered as a background or p_t is considered to be the foreground object

The above described background subtraction and update procedure for each pixel's texture feature is given in the following algorithm

Algorithm : self organizing texture features for background subtraction

Input: *lbp* of a pixel p_i in frame $I_{t,}$ t = 0, ….., Lastframe

Output: *lbp* of pixel p_t is background/foreground

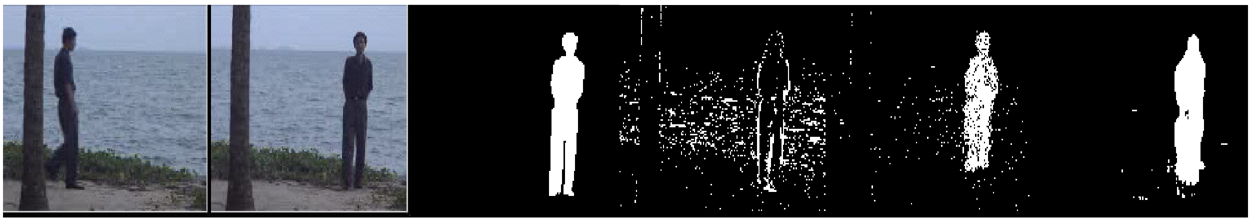
1. Initialize weight vector C with *lbp* components for pixel p_0 and store it into A
2. **for t = 1**, LastFrame
3. Find best match c_m in C to current sample $p_{t,}$ using Euclidean distance
4. **if** (c_m found) **then**
5. p_t is background
6. update A in the neighborhood of c_m
7. **else if**(p_t is shadow) **then**
8. p_t is background
9. **else**
10. p_t is foreground

The above algorithm consists of two phases

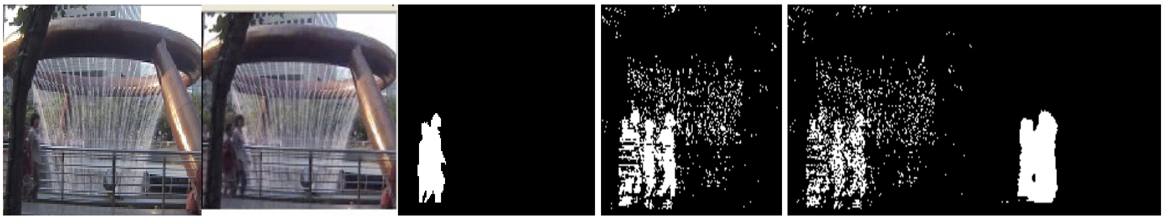
1. Neural network learning phase: This involves steps 1-6, which is executed for first few sequence frames F.
2. Neural network adaptation and background subtraction phase: This involves steps 2-10 executed on the LastFrame – F sequence frames. F depends on how many static initial frames are available for each sequence.

IV. EXPERIMENTAL RESULTS

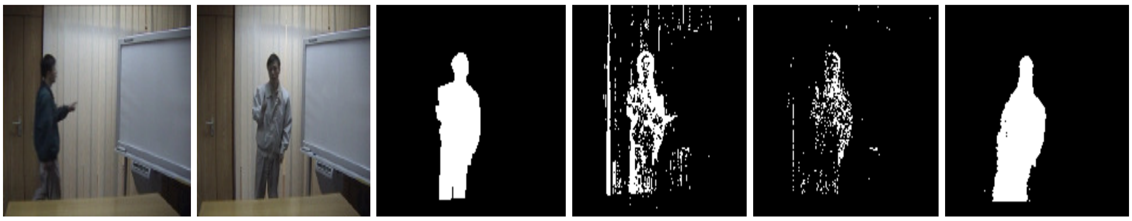
We evaluated the performance of the proposed algorithm with the single, mixture of Gaussians (MoG) and texture based method for modeling background objects for both indoor and outdoor environments.



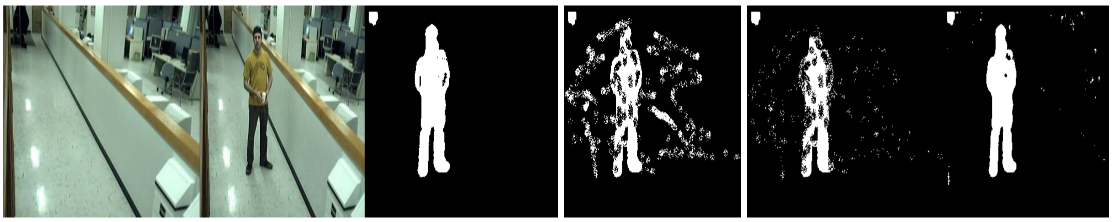
(a) (b) (c) (d) (e) (f)



(a) (b) (c) (d) (e) (f)



(a) (b) (c) (d) (e) (f)



(a) (b) (c) (d) (e) (f)

Figure 1. Comparison of experimental results of Background Subtraction of various algorithms
(a)Reference frame (b)current frame (c)ground truth (d)single Gaussian (e)mixture of Gaussian (f) LBP

From the above results it is observed that our proposed method works can handle video scenes containing moving background, illumination variation, and also include into the background model shadows cast by moving objects.

V. Conclusions

Providing a machine the ability to identify the foreground objects from the video sequences has long fascinated scientists, engineers and even the common man. In this paper, we have attempted to provide a survey on background modeling methods based on statistical learning. Performance analysis for each of the method is explored. We have also proposed a self-organizing method for modeling background. Unlike the existing method that use individual flow vectors as inputs, our method learns background texture in a self organizing manner; this makes the neural network structure much simpler. Experimental results, using different datasets, have demonstrated the effectiveness of the proposed approach, which proves also robust to moving backgrounds, gradual illumination changes and cast shadow. The only drawback of this approach is that it will not identify the foreground object if it is having the same texture and color as that of a background. This can be taken up for the future research work

ACKNOWLEDGMENT

I owe my sincere feelings of gratitude to ARMAMENT Research Board (ARB), DRDO, New Delhi for their supporting, guidance and suggestions which helped us a lot to write the paper.

REFERENCES

[1]. Chris Stauffer, W.E.L Grimson, " Adaptive background mixture models for rea-time tracking" The Artificial Intelligence Laboratory, Massachusetts institute of Technology,1999.
[2]. C.R. Wren, A. Azarbayejani, T. Darrell, and A.P. Pentland, "Pfinder: Real- Time Tracking of the Human Body," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 780-785, July 1997.
[3]. K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, "Wallflower: principles and practice of background maintenance," in Proc. 7th IEEE Conf. Computer Vision, 1999, vol. 1, pp. 255–261.
[4]. L. Maddalena and A. Petrosino, "A self-organizing approach to detection of moving patterns for real-time applications," in Proc. 2nd Int.Symp. Brain, Vision, and Artificial Intelligence, 2007, pp. 181–190, Lecture Notes Comput. Sci. 4729.
[5]. Friedman and S. Russell, "Image Segmentation in Video Sequences: A Probabilistic Approach," Proc. Conf. Uncertainty in Artificial Intelligence, pp. 175-181, 1997.
[6] C. Stauffer and W.E.L. Grimson, "Adaptive Background Mixture Models for Real-Time Tracking," Proc. IEEE CS Conf. Computer Vision and Pattern Recognition, vol. 2, pp. 246-252, 1999.
[7]. Marko Heikkilla, Matti Pietikainen "A Texture-Based Method for Modeling the Background and Detecting Moving Objects", IEEE Transactions on Pattern Analysis and Machine Learning , Vol. 28, NO4, April 2006.
[8]. M. Heikkila¨, M. Pietika¨inen, and J. Heikkila¨, "A Texture-Based Method for Detecting Moving Objects," Proc. British Machine Vision Conf., vol. 1, pp. 187- 196, 2004.
[9] T. Ojala, M. Pietika¨inen, and D. Harwood, "A Comparative Study of Texture Measures with Classification Based on Feature Distributions," Pattern Recognition, vol. 29, no. 1, pp. 51-59, 1996.
[10] T. Ojala, M. Pietika¨inen, and T. Ma¨enpa¨a¨, "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 971-987, July 2002.
[11] A. Prati, I. Mikic, M.M. Trivedi, and R. Cucchiara, "Detecting Moving Shadows: Algorithms and Evaluation," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 25, no. 7, pp. 918-923, July 2003.
[12] J. M. Ferryman, Ed., in Proc. 9th IEEE Int. Workshop on Performance Evaluation of Tracking and Surveillance, 2006.
[13] R. T. Collins, A. J. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, O. Hasegawa, P. Burt, and L.Wixson, "A system for video surveillance and monitoring," Tech. Rep. CMU-RI-TR-00-12, The Robotics Inst., Carnegie Mellon Univ.,
Pittsburgh, PA, 2000.
[14] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques," Int. J. Comput. Vis., vol. 12, no. 1, pp. 42–77, 1994.
[15] Lucia Maddalena and Alferdo Petrosino, "A Self-Organizing Approach to Background Substraction for visual Surveillance Applications", IEEE Tranasactions on Image Processing, Vol 17, NO 7, July 2008.
[16] T. Kohonen, *Self-Organization and Associative Memory*, 2nd ed. Berlin, Germany:
pringer-Verlag, 1988.

Authors:

Shobha G., Dean, PG Studies, (CSE & ISE) is associated with R.V. College of Engineering, Bangalore, since 1995.She has received her Master's degree from BITS, Pilani and Ph.D (CSE) from Mangalore University. Her research areas of interest are Database Management Systems, Data mining, Data warehousing, Image Processing and Information and Network Security.

N. Satish Kumar, Asst. Professor, CSE dept. Mahaveer Jain College of Engineering, Bangalore. He has received Master Degree (M.Tech) from VTU (R.V.C.E). His research areas are Image processing and computer vision.