International Journal of Engineering Inventions
e-ISSN: 2278-7461, p-ISSN: 2319-6491
Volume 2, Issue 10 (June 2013) PP: 58-61

Human Detection and Tracking System for Automatic Video Surveillance

Swathikiran S.[1], Sajith Sethu P.[2]
[1]PG Scholar, [2]Assistant Professor, Department of Electronics & Communication Engineering, SCT College of Engineering, Thiruvananthapuram-18

Abstract: *Detection and tracking of humans in a video has great potential applications. The paper proposes a novel scheme for performing automated human detection and tracking in videos. The method consist of two steps namely, motion detection and human – non human classification. Motion detection is done by performing background subtraction. The background model is generated using a mixture of Gaussian distributions. The human – non human classification is done by generating the feature vector from the local binary pattern of segmented regions. A radial basis function based support vector machine is used for classification. The proposed scheme was tested on some test videos and was able to detect and track human figures with acceptable accuracy.*

Keywords: *Background modeling, background subtraction, Gaussian mixture model, human detection, human tracking, local binary pattern, support vector machine*

I. Introduction

Object tracking is the process of tracing the location of an object of interest. The main application of object tracking comes in the field of automatic video surveillance. Nowadays almost all buildings and roads are provided with closed circuit television (CCTV) security cameras. The visuals from these CCTV cameras serve a great deal in monitoring human activities. Thus these become very much helpful in prevention of crime, theft, etc. The main problem associated with this is that a lot of human workforces are required for the monitoring of these video sequences. It is in this context that the need for an automatic surveillance system arises.

The conventionally used technique for motion detection in video sequences is to perform background subtraction [5] technique. The background subtraction technique subtracts a background mask from each of the video sequences and detects if there is any change in it. The simplest background mask is the image of the scene without any moving object in it. Another method is to choose the background mask as an average of first N number of frames of the video[8]. But these background masks will simply fail if there is any illumination change in the scene of interest. So as a solution to this problem, the background mask will be modeled [1]. For this, several techniques such as statistical modeling, fuzzy background modeling [6], neural network background modeling [7], background clustering[1], etc are present. Once the background is modeled and the foreground is computed by subtracting the background mask from the video frame, the next step is to determine whether the foreground contains any object of interest. For this an object recognition technique has to be used. In the human detection and tracking system, the object of interest is humans. Several techniques have been proposed for detecting a human figure in an image such as histogram of oriented gradients [9], partial least squares analysis [10], Haar-cascade classifier[11], etc.

The proposed method combines an adaptive statistical background modeling[2] technique and a local binary pattern[3] feature for the detection and locating of a human being in a video scene. Since an adaptive statistical background modeling technique is used to model the background mask, the proposed scheme is very much resistant to illumination changes. Also the moving objects in the scene are further classified as human or non human using the local binary pattern feature descriptor. The proposed scheme is also resilient to changes occurring in the background scene over time and thus it can successfully detect and track moving human figures in a video. The block diagram of the proposed scheme is given in Fig. 1

II. Background Subtraction for Motion Detection

The proposed method uses a background subtraction technique for locating the moving objects in the scene. Background subtraction uses a model of the background to find the foreground objects present in the video frame. The simplest technique uses a static background image as the background model. These models fail when the scene under consideration is of varying illumination. In order to cope with this problem, the proposed method uses a statistical background model as the background mask. The model is constructed from a mixture of Gaussian distributions. The method was originally proposed by Grimson and Stauffer[12] and later modified by KaewTraKulPong and Bowden[2].

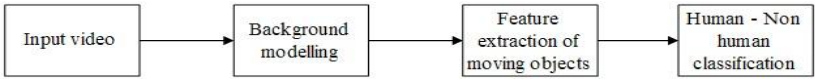

Figure 1**:** Basic block diagram of the human detection and tracking system

In this method, each pixel in the background is modelled by a mixture of a small number of Gaussian distributions. Generally, the number of Gaussian mixtures used ranges from 3 to 5. Each Gaussian distribution is used to model a single colour space. Consider the RGB colour space. The colour space consists of three colours namely, red, green and blue. Let the number of Gaussian distributions be 'K'. Now, each pixel 'X' in the frame will be modelled using these 'K' Gaussian distributions. At any time instant 'N', the probability that the pixel at a particular location has the value 'X_N' is given by

$$p(X_N) = \sum_{i=1}^{K} w_i \eta(X_N; \theta_i) \tag{1}$$

where, w_i is the weight associated with each of the Gaussian distribution and $\eta(X_N;\theta_i)$ represents the i^{th} Gaussian distribution which can be represented as

$$\eta(X_N; \theta_k = \frac{1}{(2\pi)^{D/2}\left|\sum_k\right|^{1/2}} \exp(-\frac{1}{2}(x-\mu_k)^T \sum_k^{-1}(x-\mu_k)) \tag{2}$$

where, μ_k and Σ_k represents the mean and covariance matrix of the k^{th} Gaussian component respectively.

Let $\sum_k = \sigma_k^2 I$ (3)

In each frame of the video, each pixel will be compared with the K Gaussian distributions. If the pixel matches with any of the distribution, the weight, mean and covariance of the corresponding Gaussian distribution will be updated using the following equations:

$$\hat{w}_k^{N+1} = (1-\alpha)\hat{w}_k^N + \alpha\hat{p}(\omega_k \mid x_{N+1}) \tag{4}$$

$$\hat{\mu}_k^{N+1} = (1-\alpha)\hat{\mu}_k^N + \rho x_{N+1} \tag{5}$$

$$\hat{\Sigma}_k^{N+1} = (1-\alpha)\hat{\Sigma}_k^N + (\rho x_{N+1} - \hat{\mu}_k^{N+1})(x_{N+1} - \hat{\mu}_k^{N+1})^T \tag{6}$$

$$\rho = \alpha\eta(x_{N+1}; \hat{\mu}_k^N, \hat{\Sigma}_k^N) \tag{7}$$

Where, ω_k is the k^{th} Gaussian component and α is the learning rate.

The first B number of distributions that are obtained when the K Gaussian distributions are arranged according to a fitness value given by w_k/σ_k will be selected as the required background model. The value B is given by

$$B = \arg\min_b (\sum_{j=1}^{b} w_j > T) \tag{8}$$

Where, the threshold T is the minimum amount of background present in the entire scene. If the pixel value does not match any of the distribution, the pixel will be marked as a foreground pixel.

III. Human Detection System

After performing the background subtraction technique mentioned above, the foreground containing the moving objects are segmented. The next step is to determine whether the foreground object is human or not.

This is required because the foreground may contain other moving objects such as vehicles, animals, or other moving objects such as leaves, trees, etc. The local binary pattern of the moving object is used for classifying the object as human or not. The classification is done using a support vector machine.

Local binary pattern (LBP) was introduced in 1996 by Ojala et al[3]. It was first used for face recognition applications. Later, several papers were published that used LBP in applications such as image retrieval, texture analysis, motion analysis, etc. In this paper, the LBP is used for detecting whether the object under consideration is human or not. Fig. 2 explains how the LBP of an image is generated. A local binary pattern corresponding to each of the pixel values present in the image will be generated and the original pixel value will be replaced with this number. The resulting image is called local binary pattern image. Consider Fig. 2. The pixel under consideration is the pixel at location (2,2) with value 200 In order to find the LBP of this pixel, the immediate neighbours of this pixel will be thresholded with 200 as the threshold value. This results in eight binary values around the pixel at (2,2) These binary values will be then converted to decimal form and the pixel at (2,2) will be replaced with this value. In this way the LBP of all the pixels will be calculated to obtain the LBP image. A modification to the LBP mentioned above is the uniform local binary pattern (ULBP). A binary sequence is said to be a uniform pattern if the number of transitions present in the pattern is exactly two

in number. Transitions can be a change from 0 to 1 or from 1 to 0. For example, the sequence 10001000 is not a uniform pattern since it contains four transitions. The sequence 00011100 is a uniform pattern since the number of transitions is only two. Consider the UBLP computation of an image. Suppose that the LBP of a pixel was found to be non-uniform. Then, the pixel value will not be replaced with the LBP value. Instead, the original pixel value will be retained. On the other hand, if the LBP was uniform, then the pixel value will be replaced with the LBP. The proposed scheme for human detection uses uniform local binary patterns for feature extraction and classification. Fig. 3 and Fig. 4 show the local binary pattern images of pedestrian image and car image respectively.

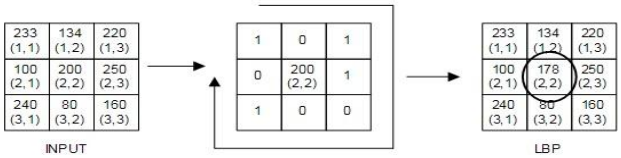

Figure 2: Local binary pattern computation

Once the ULBP of an image is computed, the next step is to extract the feature vector from the ULBP image. The feature is extracted as follows. First, the image will be divided into small blocks of size mxn. Then the histogram of each block will be calculated and concatenated together to obtain the feature vector. The similarity between these feature vectors are used for comparing images. The usage of uniform local binary pattern provides two major advantages. Firstly, for a 'p' bit binary sequence, the number of sequences with two transitions are given by p(p-1). Therefore, for a 8 bit sequence, the number of uniform sequences is given by 56.

This reduces the length of the histogram to a great extend. Secondly, the uniform local binary pattern detects only the important features such as line ends, edges, corners, etc.

IV. Experiments and Results

Two different experiments were conducted to determine the efficiency of the proposed scheme. The first experiment was done to check whether the local binary pattern is able to successfully classify human figures from other objects. The radial basis function SVM was used for the classification. The MIT CBCL pedestrian database #1 was used as the human training set and random images obtained from the internet was used as non human images. The SVM was trained using 100 human images and 100 non human images. The remaining images in the pedestrian database and other random images from the internet were used for testing the SVM. The pedestrian image was of size 128x64. After the computation of the uniform local binary pattern image the LBP image is divided into 8 blocks. Thus each block is of size 16x8. Then the histogram of each block is found and concatenated together to obtain the feature vector. The non human images found from the internet were first resized to 128x64 dimensions and then the features were found. The recognition accuracy was found to be 96.34%. This result shows that the local binary pattern is indeed a good feature for the classification of human and non-human objects.

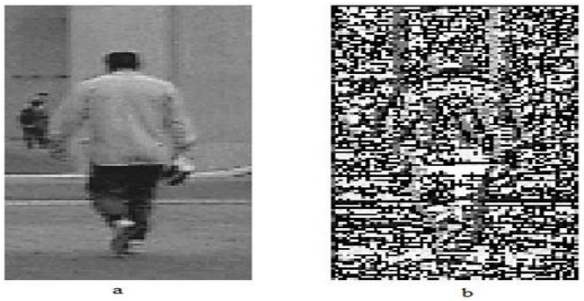

Figure 3: Local binary pattern image of pedestrian image. (a) Original image. (b) LBP image of (a)

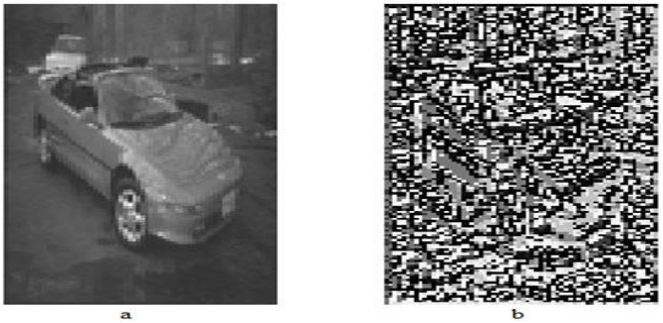

Figure 4: Local binary pattern image of car image. (a) Original image. (b) LBP image of (a)

The second experiment was conducted to determine the efficacy of the proposed scheme as a human detection and tracking system. The testing was done on videos taken using Sony Cyber-shot DSC-W710 at a frame rate of 30fps. The MIT CBSL pedestrian #1 database was used for training the SVM. The background model of the scene was generated using a Gaussian mixture model with the following parameters: number of Gaussian mixtures used, K=5; initial learning rate, ρ=0.005; threshold, T=0.7; variance of Gaussian distribution, σ=900. The proposed scheme was tested on three different videos. Video1 contained only a single human. Video 2 contained multiple humans and video3 contained humans as well as vehicles. The moving objects in the video frame were obtained after the background subtraction step. Then each of the moving objects were separated and resized to the dimension 128x64. Then the feature vectors were extracted. Recognition accuracy was computed by counting the correctly classified frames in the entire video. The recognition result is shown in TABLE 1.

Table 1:Results of experiment 2 showing the performance of the proposed system

Test Video	Total number of frames	Correctly classified frames	Accuracy (%)
Video1	540	500	92.59
Video2	510	448	87.84
Video3	600	514	85.67

V. Conclusion

The paper proposes a novel method for detecting and tracking humans in a video. The proposed scheme consists of a background subtraction technique for locating the moving objects within a scene and a local binary pattern based feature for classifying the moving objects as human or not. The background of the scene was modelled using a Gaussian mixture model and was used for performing background subtraction. The local binary pattern of the moving objects was used as the feature vector for human – non human classification.

The classification was done using a support vector machine (SVM). The proposed human detection and tracking system was tested on a number of test videos. The experimental result shows that the proposed scheme can be used for the detection and tracking of humans in CCTV camera videos for automatic surveillance.

REFERENCES

[1] T. Bouwmans, "Recent Advanced Statistical Background Modeling for Foreground Detection: A Systematic Survey", Recent Patents on Computer Science, Volume 4, No. 3, pages147-176, September 2011.

[2] P. Kaewtrakulpong, R. Bowden, “An Improved Adaptive Background Mixture Model for Realtime Tracking with Shadow Detection”, In Proc. 2nd European Workshop on Advanced Video Based Surveillance Systems, September 2001.

[3] T. Ojala, M. Pietikäinen, and D. Harwood, "A Comparative Study of Texture Measures with Classification Based on Feature Distributions", Pattern Recognition, vol. 29, 1996 pp. 51-59.

[4] Center for Biological and Computational Learning (CBCL) at MIT (2009) CBCL Pedestrian Database #1. http://cbcl.mit.edu/software-datasets/PedestrianData.html

[5] Fettke, M.; Sammut, K.; Naylor, M.; Fangpo He, "Evaluation of motion detection techniques for video surveillance," Information, Decision and Control, 2002. Final Program and Abstracts, vol., no., pp.247,252, 11-13 Feb. 2002

[6] F. El Baf, T. Bouwmans, B. Vachon, “Type-2 Fuzzy Mixture of Gaussians Model: Application to Background Modeling”, International Symposium on Visual Computing, ISVC 2008, pages 772-781, Las Vegas, USA, December 2008.

[7] D. Culibrk, O. Marques, D. Socek, H. Kalva, B. Furht, “A Neural Network Approach to Bayesian Background Modeling for Video Object Segmentation”, International Conference on Computer Vision Theory and Applications, VISAPP 2006, Setubal, Portugal, February 2006.

[8] B. Lee, M. Hedley, “Background Estimation for Video Surveillance”, Image and Vision Computing New Zealand 2002, IVCNZ 2002, pages 315-320, 2002.

[9] Dalal, N.; Triggs, B., "Histograms of oriented gradients for human detection," Computer Vision and Pattern Recognition, *2005*. CVPR 2005. IEEE Computer Society Conference on, vol.1, no., pp.886,893 vol. 1, 25-25 June 2005

[10] Schwartz, W.R.; Kembhavi, A.; Harwood, D.; Davis, L.S., "Human detection using partial least squares analysis," Computer Vision, 2009 IEEE 12th International Conference on , vol., no., pp.24,31, Sept. 29 2009-Oct. 2 2009

[11] Van-Dung Hoang; Vavilin, A.; Kang-Hyun Jo, "Fast human detection based on parallelogram haar-like features," IECON 2012 - 38th Annual Conference on IEEE Industrial Electronics Society , vol., no., pp.4220,4225, 25-28 Oct. 2012

[12] Stauffer, C. and Grimson, W.E.L,Adaptive Background Mixture Models for Real-Time Tracking, Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, Vol. 2 (06 August 1999), pp. 2246-252 Vol. 2.