ISSN: 2278 – 909X
International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)
Volume 6, Issue 6, June 2017

Video Object Action Detection by Combining Gaussian Mixture Model with EBPNN

Ankita Shrivastava, Rajender Singh Yadav

***Abstract*— As Video is playing important role in various works so different researchers are working for getting automatic information. Here work focus on the video object detection and identify the action of the object as well. Foreground identification was done by using histogram feature and Gaussian model. Here EBPNN (error back propagation neural network) was trained by the vector obtained from the GMM where this trained neural network identify and classify the action of the object as well. Experiment was done on real dataset and compares with existing action detection methods. Results shows that proposed work reduce the execution time and increase the video object detection pixel localization parameter.**

***Index Terms*— Artificial Neural Network, Histogram Feature, Human Action detection, Digital Image processing.**

I. INTRODUCTION

As computer vision has provide different application out of those human action detection is highly important in these days[8]. Here this human detection automatically identify the action perform by the human like running, moving, kicking, etc. So this help in monitoring the sensitive area where it automatically generate alarm for some kind of unfair activities.
The potential applications of human action detection include film and television content analysis, video index and summarization, real-time active object monitoring for video surveillance, and on-line pedestrian detection for smart vehicles.

However, human action detection remains a challenging problem. First issue in this work is appearance and body shape of the detecting object is different for various angles of the observer. As different objects have their own apparel that make it difficult to understand.

Secondly issue, is highly varying background with illumination make it different for judging the movement of the object [7]. One more point in this issue is the moving camera which make variable image for the same object and background. Third issue is that detecting human do not repeat action in same manner either it change the velocity or angle of movement which make it difficult to judge.

Manuscript received June, 2017

***Ankita Shrivastava**,Electronics and Communication,Gyan Ganga College of Technology Jabalpur, Jabalpur, India,9827279702*
***Rajender Singh Yadav**, Electronics and Communication,Gyan Ganga College of Technology Jabalpur,Jabalpur,India,8818902100,*

Fourth issue include multiple objects or human being in the same scene where each perform its own activity then dis-occlusion and occlusion of the objects occurs which make it more difficult to judge the shape of the object with there action [6].

So researcher faces this problem of how to extract and characterize behavior from some video having multiple movements with multiple objects. The other is how to learn an efficient classifier to recognize a given behavior in a new context. With respect to those mentioned difficulties, the main challenge is to find a set of features that characterize behaviors well and account for most of those scenarios.

II. RELATED WORK

In IEEE 2012, paper titled Spatio-Temporal Traffic Scene Modeling for Object Motion Detection [1] gives the overview on Approach for activity observation utilizing Bayesian combination strategy where background demonstrating and Gaussian definition is estimated by kernel intensity formulation. Having positive perspective requires less computational time and Works well with quickly and gradually changing foundation yet having negative angle as Object's element indistinguishable to that of foundation are nullified.

In IEEE 2013, paper titled An Improved Moving Objects Detection Algorithm [2] gives the review on Enhanced three phase differential technique consolidated with canny edge discovery to increase finish data identified with moving target having positive angle Ghosting impact is dispensed with and Algorithm beats the vacant technique and edge erasure issues of standard three-phase differential strategy yet having negative perspective as The outcome is not perfect in the environment with solid light and clear shadow additionally Results corrupt for dynamic foundation.

In IEEE 2014 ,paper titled Moving Object Detection Based on Temporal Information[3] gives the outline on Makes utilization of temporal data for era of movement saliency which is then trailed by greatest entropy and fuzzy developing strategy to recognize moving target having positive angle No earlier learning of the background dummy is required and Robust to gentle foundation movements and camera butterflies, No client association for parameter tuning is required and Efficiently manages the bothers of the foundation yet having negative perspective as Shadow is resolved alongside moving item which might be misclassified as question itself.

Wang et al. in [4], all in all utilized thick courses and movement outer region to develop descriptors, i.e., a cross between the method of this part and those of going before part.

Viswanath et al. in [6] start an account technique to see noticeable movement relies on upon the possibility of " discernibleness" from the gainful pixels, when the casing arrangement is imply as a direct dynamical association. The gathering of gainful pixels with most elevated saliency is moreover used to reproduction the all encompassing elements of the remarkable region. The pixel saliency guide is fortify by two region saliency maps, which are figured relies on upon the similarity of progression of the distinctive spatiotemporal fixes in the video with the notable district flow, in a worldwide and in addition a nearby sense. The resultant work is tried on a place of requesting arrangement and assessed to cutting edge system to show case its better introduction on defense of its figuring effectiveness and capacity to notice remarkable movement.

Problem Identification

In [12] identification of different objects was done by large displacement optical flow LDOF which may not correctly identify object body uniquely. The large displacement optical flow (LDOF) tracker was adopt with a sampling step-size of 5 pixels to extract dense trajectories from it. The reason of adopting the LDOF tracker is that it tracks objects with fast motions and large displacements more reliably than conventional trackers. In this work merging of various detected object lead to misguide the detection of object. Instead of SVM classification method other technique need to be introduce, as SVM not classify the objects of different size. More feature inclusion in the work can increase the efficiency of the action detection. Execution time can be reduced by using alternating unsupervised, non-Gaussian.

V. PROPOSED WORK

PreProcessing

As video is the collection of frames which is called as frame. Here frames are display in fix rate. This rate should be greater than 16 frame per second. As human cannot judge one by one display of frames if rate is more than 16 frame per second. As contents of the consecutive frames are mostly same but change in object position is new information of the frame 9, 10]. So reading of video means conversion of video in sequence of frames of RGB format.

Feature Extraction

Original Video: As shown in block diagram first block will read original video. So collection of frames in sequence of matrix is term as reading of video. Now all frames are in form of two dimensional matrix and complete video is of three dimensional matrix. Here matrix is of three dimension first is for row, other for column, and third dimension for the frame sequence.

The histogram is alludes as discrete function where Histogram identifies with the recurrence of occurring of each dark level in that picture, that shows it informed us how the estimations of individual pixel in a picture are evacuated. Histogram is given as:

$$h(rk)=nk/N$$

Where and are intensity level and number of pixels in image with intensity respectively.

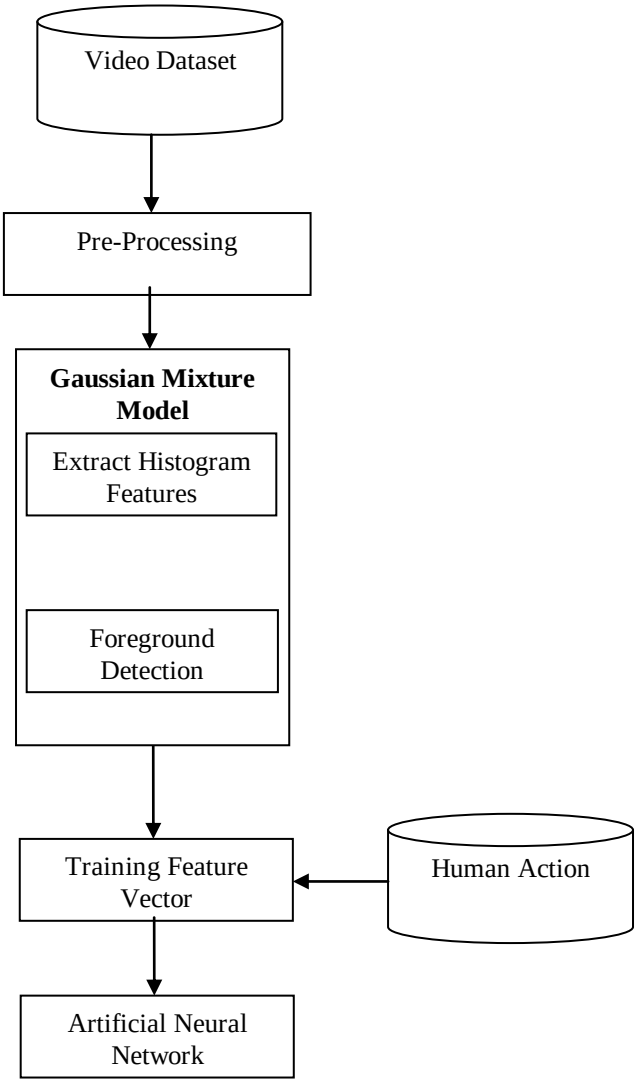

Fig.1. Block diagram of proposed model.

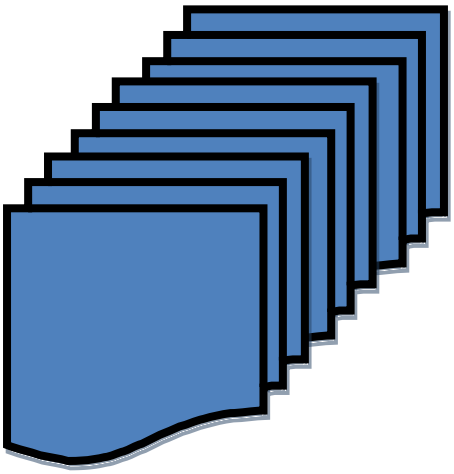

Fig. 3 Represent video frames collection.

Gaussian Mixture Model (Foreground Detection)

Staufer and Grimson [5] have proposed a versatile parametric GMM to decrease the impact of little tedious movements like trees, edges and brightening variety. A pixel I at position x and time t is demonstrated as a blend of K Gaussian conveyances. The present pixel esteem takes after the likelihood conveyance given by

$$P(I_{t,x}) = \sum_{i=1}^{k} w_{(t-1,x,i)} * \eta(I_{t,x}, \mu_{(t-1,x,i)}, \sigma^2_{(t-1,x,i)})$$

where η is the Gaussian probability density function.

$$\eta(I\ ,\mu\ ,\sigma^2) = \frac{1}{\sqrt{2\Pi\sigma^2}} \exp\left(-\frac{(I-\mu^2)}{2\sigma^2}\right)$$

While w, $\mu\ ,\sigma^2$ is the weights, mean value esteem and change of the ith Gaussian in the blend at time t - 1. For keeping up the Gaussian blend display, the parameters w, $\mu\ ,\sigma^2$ should be refreshed in view of the new pixel $I_{t,x}$. A pixel is said to be coordinated, if $I_{t,x}$ lies within σ standard deviations of a Gaussian. In our case σ^2 lies between 1 and 5. On the off chance that one of the K Gaussian is coordinated, the coordinated Gaussian is refreshed as following:

$$\mu_{(t,x,i)} = (1-\rho)\mu_{(t-1,x,i)} + \rho(I_{t,x})$$
$$\sigma^2_{(t,x,i)} = (1-\rho)\sigma^2_{(t-1,x,i)} + \rho(I_{(t,x)}, \mu_{t,x,i})^T (I_{(t,x)}, \mu_{t,x,i})$$

where

$$\rho = \alpha\eta(I_{t,x} \mid \mu_{t-1,x,i}, \sigma_{t-1,x,i})$$

is a learning rate that controls how fast μand σ^2 converges to new observations.

The weight of the K Gaussian is adjusted as follows:

$$w_{t,x,i} = (1-\alpha)w_{t-1,x,i} + \alpha(M_{t,i})$$

Where $M_{t,i} = 1$ is set for the matched Gaussian and $M_{t,i} = 0$ for the others. The learning rate _ is used to update the weight and its value ranges between 0 and 1.

If none of the K Gaussian component matches the current pixel value, the least weighted component is replaced by a distribution with the current value as its mean, a high variance, and a low value of weight parameter is chosen. Thereafter, the weights are normalized.

The K circulations are sorted in dropping request by w/σ This requesting moves the most plausible foundation with high weight and low fluctuation at the top. The main B Gaussian dispersion which surpass certain limit T are held for the foundation conveyances. In the event that a little estimation of T is picked, the foundation model is uni-modular and is multi-modular, if higher estimation of T is picked. On the off chance that a pixel It,x does not matches with any of the foundation part, then the pixel is set apart as forefront.

Training of Error Back Propagation Neural Network (EBPNN):

In this step after background separation pixel position of the foreground are store in feature vector where both x, y co-ordinates are store. This can be understands as the pixel position of the foreground from each frame are store in a fix size vector which act as the shape of the object.

- Let us assume a four layer neural network.
- Now consider i as the input layer of the network. While j is consider as the hidden layer of the network. Finally k is consider as the output layer of the network.
- If w_{ij} represents a weight of the between nodes of different consecutive layers.
- So the output of the neural network is depend on the below equation:

$$Yj = \frac{1}{1+e^{-X_j}}$$

where, $X_j = \sum x_i \,.\, w_{ij} - \theta_j$, $1 \le i \le n$; n is the number of inputs to node j, and θ_j is threshold for node j

- The error of output neuron k after the activation of the network on the n-th training example (x(n), d(n)) is:

$$e_k(n) = d_k(n) - y_k(n)$$

- The network error is the sum of the squared errors of the output neurons:

$$E(n) = \sum e_k^2(n)$$

- The total mean squared error is the average of the network errors of the training examples.

$$E_{AV} = \frac{1}{N}\sum_{n}^{N} E(n)$$

- The Back propagation weight update rule is based on the gradient descent method:
  - It takes a step in the direction yielding the maximum decrease of the network error E.
  - This direction is the opposite of the gradient of E.
- Iteration of the Backprop algorithm is usually terminated when the sum of squares of errors of the output values for all training data in an epoch is less than some threshold such as 0.01

$$w_{ij} = w_{ij} + \Delta w_{ij} \qquad \Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}}$$

Proposed Algorithm: Neural Network Algorithm

Input: V //Training video
Output: TNN // Trained Neural Network

1. V ←Pre_processing(V)
2. Loop 1:n // n : number of frames in the video
3. Feature[n]←Gaussian_Mixture_Model(F)
4. C←Class[F] // Set class as per the frame action
5. Loop 1:itr // itr: Iterations
6. TNN←EBPNN(F[n], C)
7. EndLoop
8. EndLoop

VI. EXPERIMENT AND RESULTS

This section presents the experimental evaluation of the proposed object detection work of video. All algorithms and utility measures were implemented using the MATLAB tool. The tests were performed on an 2.27 GHz Intel Core i3 machine, equipped with 4 GB of RAM, and running under Windows 7 Professional. Experiment done on the video that are of different environment.

Dataset

Action Dataset were taken over homogeneous backgrounds with a static camera with *25*fps frame rate. The sequences were down sampled to the spatial resolution of *160x120* pixels and have a length of four seconds in average.

Evaluation Parameters

Execution time :This is the time taken by the algorithm to detect the action in the video or it can also be said in terms of the total time taken by the system, which includes the video reading time and execution time completely.

Pixel Localization: This parameter of evaluation is defined as the capacity of system to locate pixels. When a video is read and processed for object detection, at the time of execution the pixels are plotted according to the actions of object.

Video Frame	Number of Frames	Action
	**155**	**Hand waving**
	**140**	**Boxing**
	**200**	**Walking**

Table 1 Represent different video with actions.

RESULTS:

Actions	Execution time in seconds	
	Previous Work [12]	Proposed work
Boxing	**27.8502**	**5.56931**
Jogging	**32.6674**	**7.31273**
Hand waving	**24.36**	**3.0060**

Table 2. Comparison of proposed and previous work on execution time.

Table 2 shows that proposed work has highly decrease the execution time. Here due to use of neural network for testing detection of action is quit feasible. While for training different type of video having various actions can be used.

Actions	Action localization pixels	
	Previous Work [12]	Proposed work
Boxing	1280	388
Jogging	1586	851
Hand waving	847	468

Table 3. Comparison of proposed and previous work on action localization pixels.

Table 3 shows that proposed work has highly decrease the action localization pixels. It has been obtained that proposed work run on various environment video but consistently values are higher than previous method in [12].

Actions	Human Action	
	Previous Work [12]	Proposed work
Boxing	Standing	Boxing
Jogging	Walking	Jogging
Hand waving	Standing	Waves

Table 4. Comparison of proposed and previous work on action localization pixels.

Table 4 shows that proposed work has accurately detect the actions what kind of movement is done in video. It has been obtained that proposed work run on various environment video but consistently values are higher than previous method in [12].

VII. Conclusions

Video object action detection was done in this work. The key idea is to separate the foregoing and background pixels in the frame by using histogram feature with Gaussian model. Output of the Gaussian mixture act as the training vector of the neural network. Values obtained from different evaluation parameters shows that proposed work was better as compare to previous work of SVM. Results shows that multiple actions are detect from the same trained neural network for different action of various environment. Future work will involve the spatial information while dynamics of the video sequence was considered.

REFERENCES

[1]. Jiuyuehao, Chao Li, Zuwhan Kim, And Zhang Xiong Spatio-Temporal Traffic Scene Modeling For Object Motion Detection, Ieee, Intelligent Transportation Systems, 2012. 9.

[2]. Liu Gangl , Ningshangkun ,You Yugan ,Wen Guanglei And Zhengsiguo, An Improved Moving Objects Detection Algorithm, In Proceedings Of The 2013 Ieee International Conference On Wavelet Analysis And Pattern Recognition, Pp. 96-102, 14-17 July, 2013.

[3]. Zhihu Wang, Kai Liao, Jiulongxiong, And Qi Zhang, Moving Object Detection Based On Temporal Information, Ieee Signal Processing Letters, Vol. 21, No. 11, Pp. 1404-1407, November 2014.

[4]. Wang, H., Kl¨Aser, A., Schmid, C., And Liu, C.-L. (2013). Dense Trajectories And Motion Boundary Descriptors For Action Recognition. International Journal Of Computer Vision, 103(1):60–79.

[5]. C. Stauffer And W. E. L. Grimson, "Adaptive Background Mixture Models For Real-Time Tracking," In Computer Vision And Pattern Recognition. IEEE Computer Society, 1999, Pp. 2246–2252.

[6]. Viswanath Gopalakrishnan, Deepu Rajan, And Yiqun Hu. A Linear Dynamical System Framework For Salient Motion Detection Ieee Transactions On Circuits And Systems For Video Technology, Vol. 22, NO. 5, MAY 2012 683.

[7]. W.T. Lee And H. T. Chen, "Histogram-Based Liu Gangl , Ningshangkun ,You Yugan ,Wen Guanglei And Zhengsiguo, An Improved Moving Objects Detection Algorithm, In Proceedings Of The 2013 Ieee International Conference On Wavelet Analysis And Pattern Recognition, Pp. 96-102, 14-17 July, 2013.

[8]. Idrees, H., Warner, N., and Shah, M. (2014). Tracking In Dense Crowds Using Prominence And Neighborhood Motion Concurrence. Image And Vision Computing, 32(1):14–26.

[9]. Zhong Zhou, Member, IEEE, Feng Shi, and Wei Wu. "Learning spatial And Temporal Extents Of human Actions for Action Detection". DOI 10.1109/TMM.2015.2404779, IEEE Transactions on Multimedia.

[10]. Interest Point Detectors," In Proceedings Of The IEEE Conference On Computer Vision And Pattern Recognition, 2009, Pp. 1590-1596.

[11]. Rupesh Kumar Rout A Survey On Object Detection and Tracking Algorithms Department Of Computer Science And Engineering National Institute Of Technology Rourkela Rourkela 769 008, India.

[12]. Jiuyuehao, Chao Li, Zuwhan Kim, and Zhang Xiong. Spatio-Temporal Traffic Scene Modeling For Object Motion Detection, IEEE, Intelligent Transportation Systems, 2012.