Video Object Segmentation and Tracking Using GMM and GMM-RBF Method for Surveillance System

Prasad I. Bhosle
M.E. Student, Computer Science and Engineering
Padm. Dr.V.B.Kolte College of Engineering, Malkapur
Maharashtra, India
prasadbhosle88@gmail.com

Lokesh Bijole
Assistant Professor, Computer Science and Engineering
Padm. Dr.V.B.Kolte College of Engineering, Malkapur
Maharashtra, India
lokeshmit5588@gmail.com

Abstract:- Now a day's computer vision has been applied to every organisation. Such that the all in security systems, computers are widely used regarding to this the security purpose every organisation are used different monitoring system i.e. surveillance system, suspicious monitoring system etc. Object tracking and explanation is the definitive purpose of many video processing systems. The two critical, low-level computer vision tasks that have been undertaken in this work are: Foreground-Background Segmentation and Object Tracking. In surveillance system cameras capture the footage for tracking suspicious movement in organisation, in this condition the videos prepare with the help of surveillance cameras the most difficult task is to tracking the object from the video and make the another image so that image should be vague to identification. Generally the surveillance system work We use a stochastic model of the background and also adapt the model through time. This adaptive nature is essential for long-term surveillance applications, particularly when the background composition or intensity distribution changes with time. In such cases, concept of a static reference background would no longer make sense.

*****

I. INTRODUCTION

Now a day's computer vision has been applied to every organisation. Such that the all in security systems, computers are widely used regarding to this for the security purpose every organisation are used different monitoring system i.e. surveillance system, suspicious monitoring system etc. [1]. In practical application, since the camera moves and rotates, it needs to track objects in a dynamical background. In this situation the visual object tracking is one of the most important parts of the video processing. In video processing system there are various problems occurred due to some insufficient configuration of hardware as well as software such as How to select the initial target objects in video footage automatically and establish objects motion model, and how to update object and background models at each frame are the key in real-time Video object tracking with an active camera, some cases due to limitations of hardware of a CCTV cameras the video there are too much noise in the footage, video footage the most difficult task is to track the object or segment because in the footage the object in moving condition that's why the captured image of the particular object is unable to visualise in cleared form due to its background [1].

To resolve this problem the various algorithms are published for tracking and segment the video objects efficiently such as Threshold decision and diffusion distance, partial list square analysis, CAMSHIFT, etc. [2]. All these algorithms are propose the various methods to tracking object form input image by solving their difficulties regarding object tracking. Such as Continuously Adaptive Mean Shift algorithm (CAMSHIFT) [2] is a popular algorithm for visual tracking, providing speed and robustness with minimal training and computational cost. An adaptive robust object tracking algorithm based on active camera is proposed. At the other algorithm Partial least squares analysis is a statistical method for modelling relations between sets of variables via some latent quantities. In PLS analysis, the observed data is assumed to be generated by a process driven by a small number of latent variables. In Threshold Decision and Diffusion Distance algorithm video object segmentation and tracking framework for smart visual surveillance cameras is proposed with two major contributions [3]. First, we propose a robust threshold decision algorithm for video object segmentation with a multi background model. The proposed algorithm can determine an appropriate optimal threshold value for our proposed multi background model; hence, it can enable good performance for conditions with dynamic backgrounds without threshold tuning by developers. Second, we propose a video object tracking framework based on a particle filter with diffusion distance (DD) for measuring colour histogram similarity and motion clues from video object segmentation. For better tracking of no rigid objects, we include colour histogram in our object model as it is more stable for no rigid moving objects. [3].

For video object tracking, the data association of segmentation blobs is highly dependent on the quality of segmentation results. Gradient descent-based methods search for the most likely object candidate regions with gradient descent optimization techniques. However, they suffer from local minima problems, and it is difficult for them to address objects that have large motions. In the Kalman filter is employed to predict object motion and track objects; however, it may fail for objects that have random motions. Particle filter is a more robust methodology for object tracking, and it can address large and random motions more effectively; however, the features employed for object modelling and the distance measurements used to decide the weights of the particles,

which are essential for making these algorithms effective, have to be appropriately selected and designed. For object modelling, colour, gradient, edge, texture, and motion are usually the features employed. However, several defects may appear when these features are used as the object model. For example, models with colour features may fail to address appearance variations due to changes in the illumination of the environment.

II. LITERATURE SURVEY

To implement this object tracking system are used various algorithms. In this system object tracking is most important field in surveillance systems. In surveillance system cameras capture the footage for tracking suspicious movement in organisation, in this condition the videos prepare with the help of surveillance cameras the most difficult task is to tracking the object from the video and make the another image so that image should be vague to identification. Generally the surveillance system work in client server architecture, at the client side, video is captured by surveillance cameras [4].Such cameras can be either analogue or digital. Digital camera has become more and more popular, mainly because the captured video by digital surveillance cameras is easier to track and analyse with object detection and content analysis tools. The captured video is sent to the server for further processing. At the server side, video data is used for object detection as well as tracking.

The advances in the development of these algorithms would lead to breakthroughs in applications that use visual surveillance. Such as monitoring of banks, department stores, Airports, museums, stations, private properties, parking lots for crime prevention, detection patrolling of highways, Railways for accident detection [5]. Measuring traffic flow, pedestrian congestion and athletic performance Compiling consumer demographics in shopping canters and amusement parks Extracting statistics from sport activities Counting endangered species Logging routine maintenance tasks at nuclear and industrial facilities Artistic performance evaluation and self-learning Law enforcement: Measuring speed of vehicles Detecting red light crossings and unnecessary lane occupation Military security: Patrolling national borders Measuring flow of refugees Monitoring peace treaties Providing secure regions around bases Detecting the natural phenomenon fire besides normal object motion would be an advantage of a visual surveillance system, thus, the presented system is able to detect fire in indoor and outdoor environments. Conventional point smoke and fire detectors typically detect the presence of certain particles generated by smoke and fire by ionization or photometry. An important weakness of point detectors is that they are distance limited and fail in open or large spaces. The strength of using video in fire detection is the ability to serve large and open spaces. Current fire and flame detection algorithms are based on the use of colour and simple motion information in video. In addition to detecting fire and flame colour moving regions, the method presented in this thesis analyses the motion patterns, the temporal periodicity and spatial variance of high-frequency [5].

The object tracking and segmentation from the video are the most difficult task because in running video there are various objects are in moving condition, while operator is going to track that object, the captured image become so noisy due to its pixel rate of that image. To overcome this problem till now various algorithms are used. Most of the algorithms are success to tracking and segment the video object but most of the algorithms are only track the object but not segment the background of the image. Another problem to use these algorithm is that, the tracked object stored in the database at the administrator of that system so that it can't be track that object next time until the particular algorithm apply on the video. There must be one provision in this that tracked object must be learned by the system so that the next time object will track and detect automatically [5].

This kind of provision will be implemented in artificial neural network, such as in character recognition system the neural network plays vital role, in this system the input character must be learned by the neural network so that the next time the same kind of character is given to the system at that time the system will recognised very fast. The artificial neural network works in approximation so that the every input will mostly recognise perfectly. As we use neural network in surveillance system the most of the object will detect correctly and it reduces the time of the administrator [5].

III. PROPOSED SYSTEM

The increasing rate of multimedia data and transmission facility induces some problem of data loss and delay of delivery. Now in the process of video object detection background updating is important factor for analysis. For the background updating used segmentation process and segmentation used clustering technique. Now in our dissertation used RBF neural network model for segmentation process and reduces the loss of frame and video data during object tracking process. The basic processing elements of neural networks are called artificial neurons, or simply neurons or nodes. In a simplified mathematical model of the neuron, the effects of the synapses are represented by connection weights that modulate the effect of the associated input signals, and the nonlinear characteristic exhibited by neurons is represented by a transfer function. The neuron impulse is then computed as the weighted sum of the input signals, transformed by the transfer function. The learning capability of an artificial neuron is achieved by adjusting the weights in accordance to the chosen learning algorithm. The basic architecture consists of three types of neuron layers: input, hidden, and output layers. In feed-forward networks, the signal flow is from input to output units, strictly in a feed-forward direction. The data processing can extend over multiple (layers of) units, but no feedback connections are present [6]. Recurrent networks contain feedback connections. Contrary to feed-forward networks, the dynamical properties of the network are important. In some cases, the activation values of the units undergo a relaxation process such that the network will evolve to a stable state in which these activations do not change anymore. In other applications, the changes of the activation values of the output neurons are significant, such that the dynamical behaviour constitutes the output of the network [6].

In this dissertation Proposed improved technique for video segmentation based on Gaussian mixture model and RBF neural network. In process of improved segmentation technique is proposed and we have used GMM technique. The Gaussian mixture model is kernel based segmentation, here kernel play a role of hyper plane. The size and efficiency of hyper plane decide the efficiency of segmentation. In segmentation process Linear discriminate analysis suffered two types of problem in region segmentation one is core point problem and another is outlier of feature point, in single class segmentation. In the process of video segmentation the lower content of visual feature such as color texture and dimensions. The feature extractor process extracts the feature of video database and store in the form of matrix. The processing of feature mapping convert into vector form for processing of fused property of classifier. The optimizations of feature selection process is improved the segmentation rate of GMM model. The process of segmentation discuss in this section is GMM, FGMM and kernel based FGMM classifier. The kernel based FGMM classifier used Gaussian kernel for segmentation analysis. The selection process of feature in fused method is used genetic algorithm [6].

A. Gaussian Mixture Model (GMM)

The adaptive Gaussian mixture model is originally designed by Fisher for taxonomic segmentation. GMM searches for those vectors in the underlying space that best discriminate among classes (rather than those that best describe the data). More formally, given a number of independent features relative to which the data is described, GMM creates a linear combination of these which yields the largest mean differences between the desired classes. It tries to find an optimal reducing-dimensionality linear projection that maximizes the scatter of all projected samples. However, for segmentation, the between class scatter should be maximized, while the within-class scatter should be minimized [9].

If a data set is categorized, it makes sense to use the class information to build a more desirable projection space to improve discrimination while reducing the dimensionality of the feature space. GMM is an example of a class specific method, in the sense that it tries to "shape" the scatter in order to make it more favourable for segmentation [10]. This method seeks the projections that maximize the ratio of the between-class scatter to the within-class scatter in the projection space.

Let the between-class scatter matrix be defined as

$$S_b = \sum_{i=1}^{j} P(i)(\mu_i - \mu)(\mu_i - \mu)^T \quad \ldots\ldots\ldots (1)$$

and the within-class scatter matrix be defined as

$$S_w = \sum_{i=1}^{j} P(i)E[(\mu_i - \mu)(\mu_i - \mu)^T]_{\cdot x \in class\ i} \quad \ldots\ldots\ldots (2)$$

Where P(i) denotes the empirical probability of the class i, E[.] is the expectation function, x is the feature vector of a video sample, μi is the mean feature vector of video class i, μ is the mean feature vector of the video set, J is the total number of the classes in the whole video set. GMM chooses the optimal projection Wopt such that:

$$W_{opt} = \text{argmax}\frac{W^T S_b W}{W^T S_w W} = [w_1 w_2 \ldots\ldots w_m] \quad \ldots\ldots\ldots (3)$$

Here {Wk|k=1,2,…,m} is the set of generalized eigenvectors of Sb and Sw corresponding to the m largest generalized Eigen values {λk|k=1,2,…,m}, i.e.

$$w_k = \lambda_k S_w w_k \quad \ldots\ldots\ldots (4)$$

Problems arise when dealing with high dimensional data. The main difficulty in this case lies in the fact that the within-class scatter matrix is almost always singular; therefore the standard algorithm cannot be used. Another disadvantage is a high computational complexity of solving Sw, Sb when working with the high dimensional input space. To solve this propose the use of an intermediate space. Principal Component Analysis is performed to reduce the dimensionality of the data, followed by applying GMM to this data in the reduced dimension.

B. Radial Basis Function

A Radial Basis Function (RBF) neural network has an input layer, a hidden layer and an output layer. The neurons in the hidden layer contain Gaussian transfer functions whose outputs are inversely proportional to the distance from the centre of the neuron.

An RBF network positions one or more RBF neurons in the space described by the predictor variables (x,y in this example). This space has as many dimensions as there are predictor variables. The Euclidean distance is computed from the point being evaluated (e.g., the triangle in this figure) to the center of each neuron, and a radial basis function (RBF) (also called a kernel function) is applied to the distance to compute the weight (influence) for each neuron. The radial basis function is so named because the radius distance is the argument to the function [12].

Weight = RBF (distance)

The further a neuron is from the point being evaluated, the less influence it has.

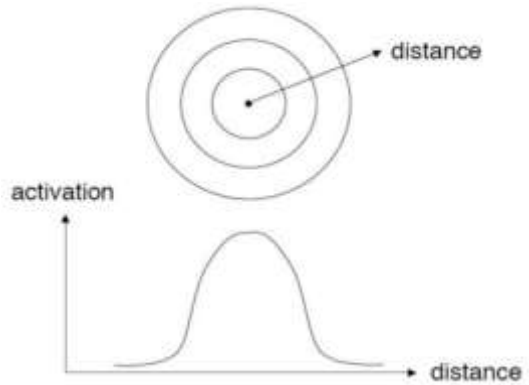

Fig.3.1 RBF computes the weights

Different types of radial basis functions could be used, but the most common is the Gaussian function:

 ISSN: 2321-8169
3588 - 3594

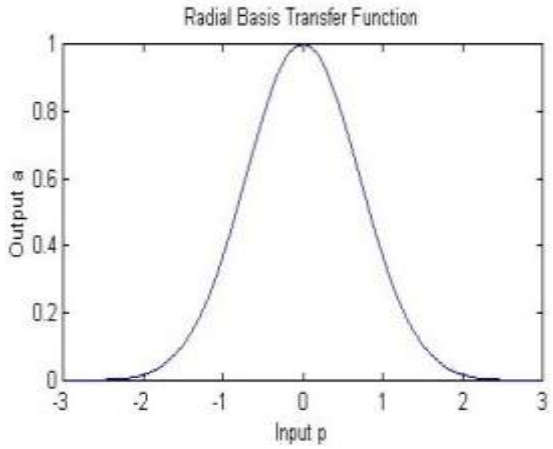

Fig3.2 Radial Basis Function

If there is more than one predictor variable, then the RBF function has as many dimensions as there are variables. The following picture illustrates three neurons in a space with two predictor variables, X and Y. Z is the value coming out of the RBF functions:

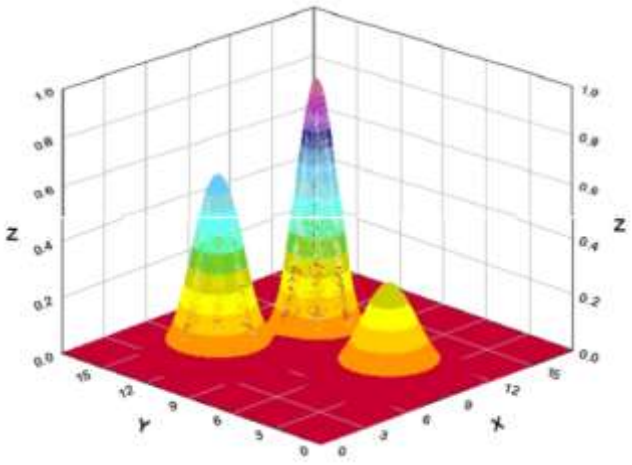

Fig.3.3 RBF three dimension values

The best predicted value for the new point is found by summing the output values of the RBF functions multiplied by weights computed for each neuron.

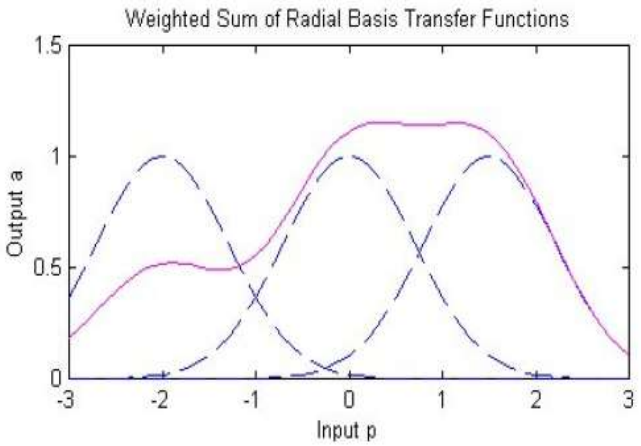

Fig. 3.4 RBF functions multiplied by weights computed for each neuron.

C. *Training RBF Networks*

The following parameters are determined by the training process:

1. The number of neurons in the hidden layer.
2. The coordinates of the centre of each hidden-layer RBF function.
3. The radius (spread) of each RBF function in each dimension.
4. The weights applied to the RBF function outputs as they are passed to the summation layer [11].

Various methods have been used to train RBF networks. One approach first uses K-means clustering to find cluster centers which are then used as the centers for the RBF functions. However, K-means clustering is a computationally intensive procedure, and it often does not generate the optimal number of centers. Another approach is to use a random subset of the training points as the centers [11].

D. *Proposed Algorithm*

The proposed algorithm is a combination of RBF kernel for feature separation of video feature fused technique. GMM weight key is a vector value given by the data set. The GMM value passes as a vector for finding a near distance between superior video feature separations. After finding a superior video feature separation the nearest distance divide into two classes, one class take a higher odder value and another class gain lower value for feature selection process. The process of selection of class also reduces the passes of data set. Dividing. After finding a class of lower and higher of given GMM value, compare the value of distance wet vector. Here distance weight vector work as a fitness function for selection process of genetic algorithm. Here we present steps of process of algorithm step by step and finally draw a flow chart of complete process.

Steps of algorithm

1. Select video data set
2. Put value of GMM and kernel of GMM
3. Start segmentation process of selected featured in video dataset
4. Generate process of bagg-off feature
5. Generate frequent feature set
6. Compute the distance with equilateral distance formula
7. Generate distance vector value for selection process
8. Compare the value of distance vector with population set
9. If value of GMM greater than vector value
10. Processed for encoded of data
11. Encoding format is binary
12. After encoding offspring are performed
13. Set the value of probability for mutation and the value of probability is 0.006.
14. Set of segmented video are generated.
15. If video is not segmented go to selection process
16. Else optimized segmented video is generated.

 ISSN: 2321-8169
3588 - 3594

17. Exit

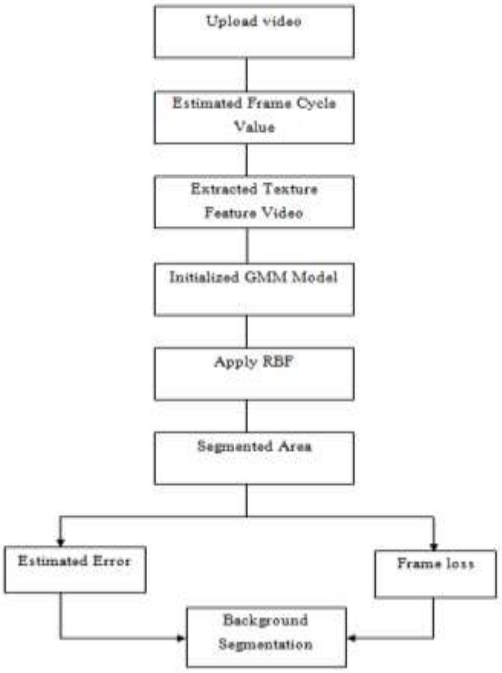

Fig. 3.5 Video segmentation based on RBF network

IV. RESULT ANALYSIS

A. Datasets

To evaluate the performance and to verify the robustness of the proposed method for automated multi object tracking in visible band, family dataset have been used. This dataset contains sequences taken in a real world public environment and it includes busy scenarios. Video sequences include people walking with their luggage as single or as a part of a larger group and multiple occlusions occur in all scenarios. These videos were captured with digital video (DV) cameras, in phase alternate line (PAL) standard with a 720 x 576 resolution and 25 frames per second and compressed as Joint Photographic Experts Group (JPEG) image format. To prove the proposed method is not affected from hot objects in the scene, such as heating system, radiators and so forth, method has also been tested in such environments. As there were no public datasets for abandoned object detection and only a few limited set for object tracking having both modalities, we captured our own dataset for various scenarios.

Table 4.1 Datasets for various scenarios

Dataset	Video format	Number of frames	Number of living object	Description of dataset
A1	.3gp	2700	10	Birthday party video
A2	.avi	4313	Party video number of busy object	Children's dance
A3	.mp4	1164	Various animals in nature	Animals Life
A4	.mpeg	1116	04	Boxing Match

Table 4.2 gives the computational value estimated by segmentation area of "birthday party" video for frame loss.

Table 4.2 Estimated result of "birthday party" video

Method	Segmented Area (%)	Frame Loss (%)
GMM	**80**	**30.75**
GMM-RBF	**85**	**29.62**

Chart 4.1 shows that comparative result analysis of frame loss and estimated value of segmented are using GMM and GMM-RBF model of birthday party video.

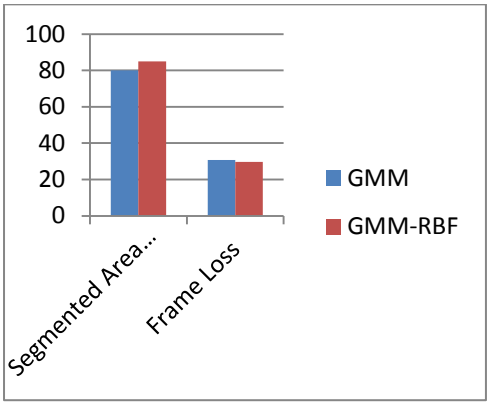

Chart4.1 Results of birthday party video.

Table 4.3 gives the computational value estimated by segmentation area of "children dance" video for frame loss

Table 4.3 Estimated result of "children dance" video

Method	Segmented Area (%)	Frame Loss (%)
GMM	**90**	**37.66**
GMM-RBF	**95**	**28.21**

Chart 4.2 shows that comparative result analysis of frame loss and estimated value of segmented are using GMM and GMM-RBF model of Children dance video.

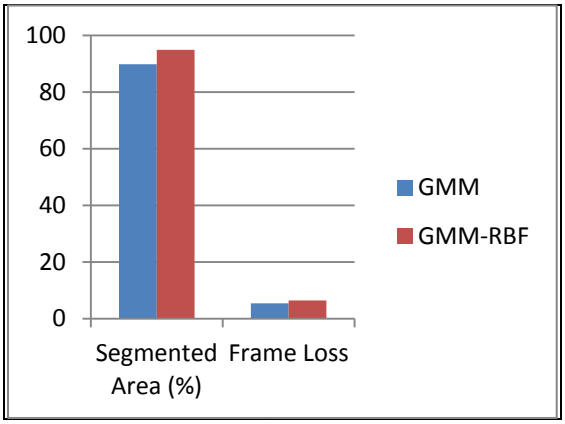

Chart4.2 Results of children dance video.

Table 4.4 Estimated result of "Animal life" video

Method	Segmented Area (%)	Frame Loss (%)
GMM	89.81	5.49
GMM-RBF	94.81	6.49

 ISSN: 2321-8169
3588 - 3594

Chart 4.3 shows that comparative result analysis of frame loss and estimated value of segmented are using GMM and GMM-RBF model of Animal life video.

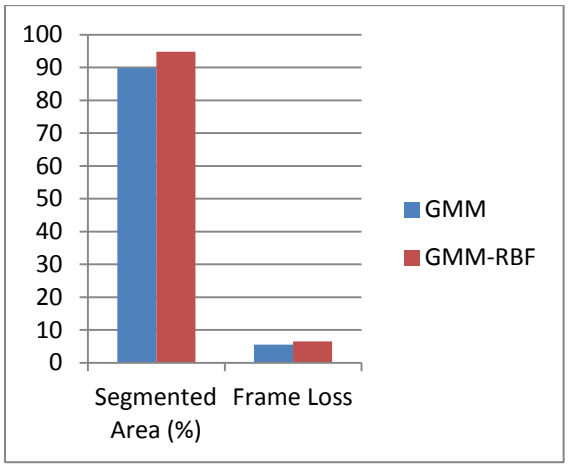

Chart4.3 Results of animal life video.

Table 4.5 gives the computational value estimated by segmented area and frame loss for "Boxing" video

Table 4.5 Estimated result of "Boxing" video

Method	Segmented Area (%)	Frame Loss (%)
GMM	81.79	11.54
GMM-RBF	86.57	10.62

Chart 4.4 shows that comparative result analysis of frame loss and estimated value of segmented are using GMM and GMM-RBF model of Boxing video.

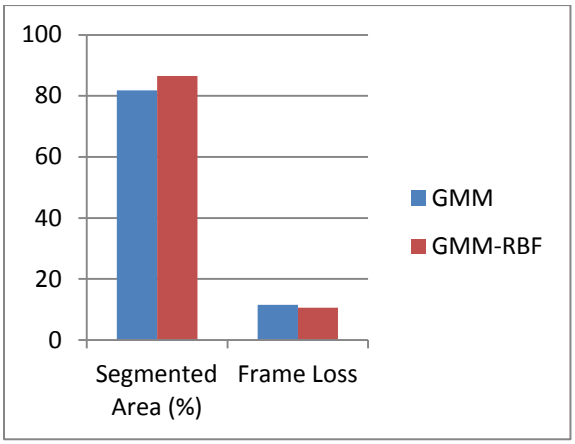

Chart4.4 Results of boxing video.

B. Comparative Result Analysis

As per the estimated the results of different videos are different formats which show the variations in the results so comparative studies can useful to show the compatibility of proposed system. Table 4.6 shows the comparative results of various input videos.

Table 4.6 Comparative Results

Video Name	Video format	Computed Result in	Method	
			GMM	GMM-RBF
Birthday party	.3gp	Segmented Area (%)	80	85
		Frame Loss (%)	30.75	29.62
Children dance	.avi	Segmented Area (%)	90	95
		Frame Loss (%)	37.66	28.21
Animal Life	.mp4	Segmented Area (%)	89.81	94.81
		Frame Loss (%)	5.49	6.49
Boxing	.mpeg	Segmented Area (%)	81.79	86.57
		Frame Loss (%)	11.54	10.62

Chart 4.5 shows the result of all formats of video for Comparative studies, in this all the computed results are shown as per their parameters.

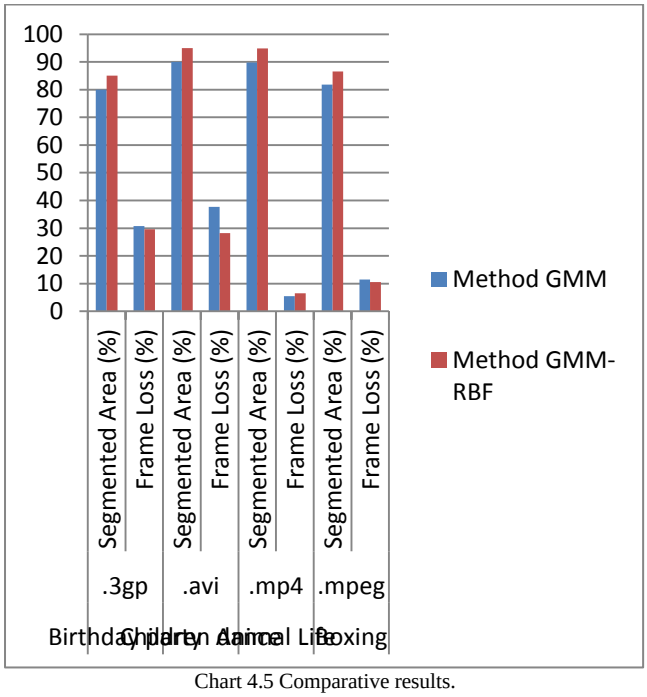

Chart 4.5 Comparative results.

V. CONCLUSION

This dissertation has presented the state of development of surveillance systems, including a review of current image

processing techniques that are used in different modules that constitute part of surveillance systems. As far as this image processing tasks is concern it has identified research areas that need to be investigated further such as adaptation, data fusion and tracking methods in surveillance system. Thus the artificial neural network is best concept to develop surveillance system.

This dissertation proposed a novel method for video segmentation and background removal for video object tracking. The proposed method is very efficient in compression of frame loss and segmentation area for video object tracking. The proposed method comes along with wavelet filter and RBF neural network. So the complexity of method is increase in terms of segmented area and frame loss minimisation. The performance can be further improved by fusing multi-modal information such as by applying the vehicle classification result to constraining the size of the object and vice versa. Future work will include applying the algorithm to a larger number of data and performing comparative studies on various applications with various vision- and other sensor-based approaches.

REFERENCES

[1] Lee L, Romano R, Stein G. "*Monitoring activities from multiple video streams: Establishing a common coordinate frame*", IEEE Transactions on PAMI, Cornell University, 2000, pp. 758–767.

[2] R.Stolkin, I. Florescu, G. Kamberov, "*An adaptive background model for CAMSHIFT tracking with a moving camera*."Center for Maritime Systems (Vol.1) (Hobken: Stevens Institute of Technology), 2000, pp. 50.

[3] Shao-Yi Chien, Wei-Kai Chan, "*Video Object Segmentation and Tracking Framework With Improved Threshold Decision and Diffusion Distance*" IEEE Transactions On Circuits And Systems For Video Technology, Vol. 23, No. 6, June 2013.

[4] Ananya Pathak, kandarpasharma, "*Human Detection And Surveillance System Using Real Time Video Feed And Ann*" IRNet Transactions on Electrical and Electronics Engineering (ITEEE) ISSN 2319 – 2577, Vol-1, Iss-2, 2012

[5] David Houcque. "*Introduction to Matlab ForEngineering Students*" Northwestern University, version 1.2, August 2005.

[6] J. Zupan, J. Gasteiger, *"Neural Networks for Chemists: An Introduction, VCWeinheim,"* IEEE Trans. on Image Processing, vol. 21, no. 5,may 2012

[7] Krumm, J., Harris, S., Meyers, B., Brumit, B., Hale, M., and Shafer, *"Multi-camera multi-person tracking for easy living"*. Third IEEE Int.Workshop on Visual Surveillance, Ireland, 2000, pp. 8–11

[8] H. Z. Ning, L. Wang, W. M. Hu, and T. N. Tan, *"Articulated model based people tracking using motion models", Proc. Int. Conf. Multi-Model Interfaces* , pp.115 -120 2002.

[9] S. Hare, A. Saffari, and P. Torr, *"STRUCK: Structured Output Tracking with Kernels"* in Proc. of the International Conference on Computer Vision (ICCV), Barcelona, Spain, 2011,

[10] R. T. Collins, Y. Liu, and M. Leordeanu, *"Online selection of discriminative tracking features"* IEEE Trans. Pattern Anal. Mach. Intell., vol. 27,no. 10, Oct. 2005, pp. 1631–1643.

[11] Kulkarni, A.D.: "*Artificial Neural Networks for Image Understanding"*. VNR Com- puter Library. Van Nostrand Reinhold, New York (1994)

[12] Vedaldi, A., Gulshan, V., Varma, M., & Zisserman, A.*"Mul-tiple kernels for object detection"* In International conference on computer vision 2009.

[13] A. Gilat. "*MATLAB: An introduction with Applications*." John Wiley and Sons, 2004.

[14] K. R. Coombes, B. R. Hunt, R. L. Lipsman, J. E. Osborn, and G. J. Stuck. "*Di®erential Equations with MATLAB"*. John Wiley and Sons, 2000.

[15] Weber, M.E., and Stone, M.L.: *"Low altitude wind shear detection using airport surveillance radars"*, IEEE Aerosp. Electron. Syst. Mag.,1995, 10, (6), pp. 3–9