Video Surveillance Based on SURF Extraction Using Hybrid BPNN

Parul Saxena, Department of Computer Applications, Madhav Institute of Technology & Science, Gwalior-474005
R. S. Jadon, Department of Computer Applications, Madhav Institute of Technology & Science, Gwalior-474005
gaurparul2007@mitsgwalior.in

Abstract— **Feature detection, rendering, and coordinating are basic segments of different computer vision applications; therefore they have gotten an impressive consideration in the most recent decades. A few element identifiers and descriptors have been proposed with a collection of explanations about an image (i.e., a particular characteristic). Matching between the existing database and the new real time images are determined by the fast speed up robust feature descriptor. This part presents fundamental documentation and numerical ideas for distinguishing and presenting video features. At that point, it talks about properties of faultless highlight and provides a description of various obtainable recognition and representation strategies. At last, the part examines the most utilized systems for execution assessment of location and depiction calculations. Background subtraction is a strategy to recognize object and well known utilized as a part of moving item detection. The point is to obtain a background and after that identify moving items by subtracting it and the present case. In this paper, Kalman filter methodology and Frame differentiate are used to recognize the articles. Kalman filter helps to track the position of an object. The Kalman filter has ideal accuracy over the Frame differentiate technique. The analysis comes to demonstrate that the Kalman filter is the best answer for acquiring high exactness, low asset necessities in a given video. The location of the object will appear in the results.**

Keywords— Video surveillance; Feature extraction and Feature Selection, SURF and hybrid BPNN **etc.**

I. INTRODUCTION

Object tracking is the way toward following the position and status of an object. Visual analysis system have served well in the field of video surveillance, militarily bearing, robot course, fake cognizance and restorative applications in the midst of the latest two decades. The real need for any vision-based system is its strength to the variability in the visual data appearance by powerful, uncontrolled condition. The main challenge is to track more than one object [1].The general following execution relies upon the exact removal and pinpointing the position of the moving things from the observation video. Video scrutiny is a more challenging task in today's environment. Tracking the capable of or having movement of thing toward which the action of a verb is directed has to pay attention of various researchers in the ground of computer vision and picture meting out. Surveillance is mainly used by governments for gathering information, for investigating and preventing the crime. Surveillance system is further divided into three: manual video supervision, Semi- automatic and fully automatic supervision.

In fully-automatic surveillance, the system will execute each task such as motion, tracking etc., without human interaction. Visual Surveillance (VS) involves the scrutiny and explanation of objects behavior in addition to object detection and tracking to observe the visual proceedings of the scene. Wide area surveillance control and scene analysis is the main task of IVS. Object tracking has got to deal with several illumination changes and well-acknowledged challenges. Mainly video analysis is categorized into three basic phases: moving entity detecting, discovery the route of the object from one case for enclosing a picture just before another case for enclosing a picture as well as detailed examination of something in order to get information about the entity track to know someone their presentation [2].

II. APPLICATION

Obtainable study on audio-based video experience discovery is quite less and yet to be additional discover. Audio-based event detection has been executed in the distinct field which is as follows:

- ***Surveillance***

In surveillance, although visual facial appearance may help in detecting events, sounds also may execute better. For example, sounds like gunshots, sudden screams may be required in surveillance. The approach originally to divide things into groups according to their type of specified audio surround addicted to speech and no speech procedures. Further usual and energized events are confidential using Gaussian Mixture Model (GMM). Four dissimilar audio characteristics such as Zero Crossing Rate (ZCR), Linear Prediction Coefficient (LPC), Linear Prediction Cepstral Coefficient (LPCC), and afterward frequently connected to customary action as a more formal articulation than do, yet as a rule inferring normal additional categorization into standard and energized procedures are used [3].

a. Meeting

Audio features like applause, cheers may be used to retrieve feedback from videos extracted from the meeting. It was proposed the involuntary recognition of social role played by an actor in small-group meeting, focusing on a) the consequence which is ***not*** associated with signs that have any original or primary intent of communication behaviors i.e. non linguistic, b) the relation time-consistency of the community

roles enact by a known person at some stage in the hours of a convention c) the behavior and mutual constraints in the middle of the roles played by the different people in a community encounter. Evaluation of model presentation between Support Vector Machine (SVM) and Hidden Markov Model (HMM) has been done. For investigating the complexity of segmenting a video into scenes, the approach use high-level audio info, in the type of audio events. Also, the procedure has also used for the creation of a lot of Scene Transition Graphs (STGs) that expand in sequence approaching from the similar speaker or writer can express certainty.

b. Sports

Extracting features in sports is an eye-catching, which not only requires visual features but also needs hearing cues. A novel framework has been explained for low-level collection of sports game (tennis) using audio track of a video soundtrack of the game. Gaussian permutation Model and a Hierarchical vocal communication model have been used to identify sequences of audio measures. The greatest entropy Markov representation to use to infer "match" events from these hearing events and multi-grams to be grateful for the segmentation of a sequence of match procedures into sequences of points in a tennis game. It has use Hidden Markov Model to classify TV broadcast video.

They have used TV programs such as basketball videos, commercials, news, football shows and weather reports for discrimination. Eight frame-based sound recording characteristic was utilized to satisfy the low-level sound recording quality and fourteen clip-based sound features was extract based on these frame-based skin to symbolize the high-level audio belongings. An erotic HMM is constructed for each of TV program. The greatest possibility technique is then utilized for test data to be classified using the models. Audio keywords were used to identify semantic proceedings in soccer video by the use of some gentle framework and these keywords are produced from low level audio keywords with the help of SVM in order to take out sports (soccer) highlight. It recognizes ball hits in table tennis by applying MFCC refined highlights and use SVM classifier. The results have been compared with data, which were energy features and their suggested Mel Frequency Cepstral Coefficient (MFCC) refined features.

c. Entertainment

It has been explored the use of audio terminology representation to detect exact audio events such as gunshots and explosion, in order to get more accurate and strong variation in multiple audio tracks that are present in Hollywood cinema. Each stationary audio segment has been described by one or more audio words obtained by performing product quantization to standard features. Automatic speech recognition technique use transcripts to automatically summarize videos. The full idea is divided into segments based on pause detection, the segment scores resultant and on the frequencies of the words and bi-grams, it contains. They proposed audio advantage that could suit the scene determination task. They have proved that the features affect the result more than clustering technique.

d. Consumer

The video has proposed an approach for audio-based semantic categorization for consumer video. Each cassette clip is represented as a progression of MFCC frames. Three clip level description, for example, Gaussian model, Gaussian mixture and probabilistic semantic analysis of a Gaussian component histogram was tested and used by Support Vector Machine classifiers based on the Kullback Leibler, Bhattacharyya or Mahalanob referred for classification. They have planned content-based video retrieval with the help of audio and visual cues combination. Adaptive video indexing technique is used to take out the visual feature that emphasizes spatiotemporal in sequence within video clips.

e. Other Applications

It presented a quick and accurate Motion Pictures Experts Group (MPEG) audio categorization algorithm based on sub-band data domain. Categorization task was carried out for four segments such as silent, music, speech and applause segment for the 1-second unit. Later Bayesian discrimination method for multivariate Gaussian distribution was used for the classification task.

III. TECHNIQUES

Human identification is an elegant surveillance method that is used to make difference between non stationary objects in any video. The success of detecting human motion depends on the accuracy of people identification. Recognition of human depends on two factors: object identification and object sorting. [4, 5]

3.1 Object detection

Object can be identified after partitioning motion in a video into multiple frames. This is the first step for tracking moving object. Some of the important methods for object tracking are optical flow, background subtraction, and spatial-temporal method. After that, they are framed in any of the more or less distinct parts into which something is or may be divided.

3.1.1 Background subtraction

In this method object is identified by subtracting it from the background of an image. The camera may be present, unadulterated or versatile in nature. This method is used for recognize non-static items with the help of difference between the present frame and the sample frame in pixel or block manner or by considering average of n frames as background image [2]. Sample frame is called background image. When there are dynamic changes on scene, a good model should be referred.

- ***Mixture of Gaussian model***

It presented a versatile Gaussian mixture is affected by illumination changes, extraneous events on dynamic scenes and rather than considering total pixels of an image as one, it consider every pixel at the same time as a mixture of Gaussian. After some time, novel pixel value was assigned to the MOG by means of live K-mean method. There are numerous strategies to improve the mixture of gaussian.

The requirement of preceding information about the foreground and background ratio is reduced by using an effective algorithm for mixture of gaussian. Without relinquishing surroundings quality, it controls the quantity of Gaussians adaptively for modify computational time. Kalman filter is utilized for adaptive background estimation.

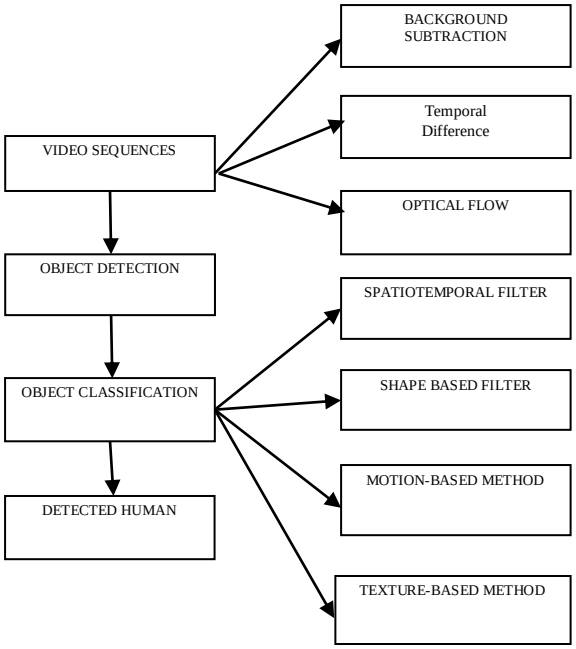


Fig.1 Flowchart of Human Recognition

Normalize consistent term identified with the properties of a result 5 sorts of orthogonal change (DCT, DFT, Haar change, solitary value deterioration and Hadamard change) are used to identify moving regions. Every point is demonstrated the same as gathering of local binary histograms with the intention of computed under a circular area covering a point.

- ***Nonparametric background model***

When we talk about parameters of specific environment then there is some problem in optimization. Hence, non parametric background model was introduced. This model points on features of statistical behavior of any image by segmenting the foreground from the background. Kernel method is used to present the color circulation for every background point without the need of argument estimate, but require more computational condition. It proposes a nonparametric frame for background subtraction, which was found strong in active texture image (e.g. moving leaves, spouting spring). Cluster feature which is known as FCH calculates the relationship among fuzzy histogram features and on live modify method to build background model. It will take more processing time but false rate of detection is low as compared to adaptive Gaussian mixture model.

- ***Temporal differencing***

This method contain 3 essential components: block alarm component, background component and object identification component. Block alarm finds the block whether it is a moving object or a background with the help of Laplacian method and allow the background component to process only those blocks which have bg points.

Next, the background demonstrating component utilizes for creating a high value background component with the help of two-stage training procedure and a way to recognize change in illumination. As the last step of this method, threshold values were used to find binary object detection mask with the help of object extraction module and is capable of utilizing their proposed edge preparing system. This execution assessment is refined by quantitative and qualitative procedures. General outcomes demonstrated the future strategy of achieving higher level of efficiency.

3.1.2 Optical flow

It is a method that examines movement in the video on the bases of comparing each point on object with image frame and vector based. Optical flow describes the motion of points between image frame under brightness constancy and spatial smoothness. Moving regions were detected using features of flow vectors.

One reason for using optical flow is, it can handle multiple cameras and object motions in crowded area and in dense motion. Even when camera is in motion it can recognize moving objects. Aside from their power to picture clamor, shading and illumination, the greater part of stream calculation strategies have high computational necessities and are sensitive to movement discontinuities. Due to the complexity of optical flow algorithm, it requires an advanced hardware and a high amount of frame for exact calculation.

3.1.3 SpatioTemporal filter

Movement identification in perspective of the spatiotemporal filter was examined in 3D data using the motion of a human in video. These strategies, for the most part, consider movement as a description of spatio-temporal distribution. They prepared a video which utilize a gaussian both spatial and derivative on the time axis. Due to this derivative operation on the temporal axis, it generate high acknowledgement at motion region. These responses help in generating thresholds to provide a binary movement mask and are efficient and easy to execute. They are used in conditions for less-confirmation or else low-quality film that is hard to extract different character, for example, frames. Spatio-transient movement strategies can better confine both space and time data of step movement with low computational complexity with basic need of execution. And hence, sensitive to change in time related with motion and noise.

3.1.4 Comparisons of detection techniques

Standard evaluation of object identification strategy is difficult as far as correctness and computational time. It is extremely hard to sum up; the exactness and computational time of various methods in every class by considering some characteristics because there are many procedures in every classification, and every strategy contain their own correctness and computational time. The general examples of these techniques are given in each category on the bases of many comparative surveys.

A few awesome systems counting the nonparametric procedure and the embedded version claims to require less calculation transparency. The hierarchical background model or block matching approach gives high correctness (approx. 5% to 15%) show contrast in connection to some excellent techniques contain MOG and less calculation time compared from MOG techniques so as use hybrid methods. As compared to background subtraction, optical flow strategy is good with moving camera and help in detection even there is crowd. But it requires high computational time.

3.1.5 Shape-based method

According to this methodology initially describe the contour features of moving area (points, blobs). Now a day's very difficult to distinguish a moving person from moving things because of appearances of the person body with respect to the difference in practical viewpoints. Despite of the shading or surfaces are the same for the objects, it propose basic and effective strategy for object finding by utilizing shades, contour data. It was found that distortion of people diagrams (or contour) under movement is used as selective characteristics to catch movement dynamic and utilize the DWT and DFT for person movement description and recognition. They projected a contour-based, part-template tree coordinating method to concurrent person identification and division merging local part-based and global shape-template based plans. This manner depended on the key concept of coordinating a hierarchical part template tree to pictures to identify people and estimate their postures. One of the drawbacks of the technique is can't catch the inner movement of the entity inside the section. Indeed, a background subtraction system doesn't generally dependably recoup exact region, particularly in powerful situations.

3.1.6 Motion-based method

This technique depends on the possibility that object motion features and examples are sufficient to recognize entity. Methodologies formulate utilization of the cyclic effects of the taken pictures for identify person without influencing from other items. A view-based method is used to identify human motions with the help of vector image template which contain two operators' i.e. binary motion picture and motion history picture.

There was s time-frequency based method for identifying and processing cyclic movement of person. This describes the individual movement based on space-time for descriptor, that shows movement going on multiple layers and dark with a Gaussian. An assertion could be executed within a nearby neighbor structure. As a result of handling a spatio-temporal association among an existing data of old labeled motion division, the majority for movement description of the doubt activity would be established.

3.1.7 Texture-based method

Intensity pattern of their neighboring pixels is based on this method like local binary pattern. LBP also provide multi-block local binary pattern for code intensity of the rectangular section. There is another procedure that utilizes high dimensional highlights of boundaries then recognizes individual body by SVM. Now procedure checks events for slope introduction in parts of a picture, for calculation on a thick grid of consistently spaced cells with covering neighborhood differentiate standardization for enhanced exactness.

3.1.8 Detection of non-moving person

In this main focus is on investigation of moving person detection because person movement highlights help in improved distinguishing people from different objects. The strategy for human identification from static pictures has various applications, for example, brilliant place and video surveillance in a crowded area from non movable pictures. This procedure assume human as a social event for ordinary human parts utilizing edge let characteristics, type of another kind of properties. Global shaped part and local shape method demonstrated hopeful outcomes. Learning based human detection framework and probability part detectors are used for human detection. Human identification with the help of sensor information is used with application of interest within the area of elderly worry support [6].

IV. LITERATURE SURVEY

Francisco Ortega-Zamoranoa et. al [2016] In this paper author compares the previous approaches in which computer vision require an expensive and high performance hardware with heavy computation. Due to this a different strategy is used to develop an inexpensive and easy to organize computer vision system for motion detection. This is accomplished by three means. Above all else, a reasonable and adaptable equipment stage is utilized. Furthermore, the movement location calculation is particularly custom-made to include a little computational load. Thirdly, a fixed point programming perspective is followed in developing the structure to further less the computational prerequisites [7].

I Bilik, J Tabrikian, et.al. [2015]This paper proposes a structure for accomplishing these non-coverings numerous camera calculations for such condition. Programmed object identification is the main task in a multi-camera inspection structure and background modeling (BM) is regularly used to remove predefined data, for example, object 's shape, geometry and so forth., for additionally preparing. Pixel-based

adaptive Gaussian mixture modeling (AGMM) is a standout amongst the most famous calculations for BM where object recognition is planned as an autonomous pixel discovery issue. It is invariant to bit by bit pixel alter, marginally moving background and regions. In any case, it as a rule yields unacceptable frontal area data (object cover) for object following because of sensor clamor and unseemly GM refresh rate, which will prompt gaps, unclosed shape and incorrect limit of the extricated object. Moreover, essential data of the object, for example, edge and shape are not used in such technique. In this manner, the execution of consequent operations, for example, object detection and recognition will be degraded [8].

Yifei Zhang, et.al [2013] In this paper, a novel activity scene display is investigated to learn logical connection amongst activities and scenes in practical recordings. With minimal earlier information on scene classifications, a generative probabilistic system is utilized for activity inference from background in view of visual words. The test comes about on a sensible video dataset approve the adequacy of the action scene method for motion identification from background settings. Extensive experiments were conducted on different feature extracted methods, and the results show the learned model has good robustness when the features are noisy [9].

Pedro Gil-Jimenez et. al [2009] In this paper, author defines that high illumination variations effect the output for variance estimation, due to which it decreases accuracy. The paper also suggests another technique and shows under such below conditions, the accuracy of the planned technique produce enhanced results whilst maintain the scenario with smaller change, hence maintaining the recognition of moving object of a video surveillance. It provide better result than standard procedures for noisy pixels with high illumination variability in real cases and even produce similar result in less illumination variability also. This paper helps to improve the performance of video surveillance under uncontrolled circumstances, such as outdoor scenarios [10].

Michael J.V. Leach et.al [2014] In this paper, the focus is on identifying human behavioral abnormalities in crowded surveillance condition. They concentrate on the unpredictability of distinguishing unassuming peculiar things in a behaviorally different surveillance scene. To achieve this objective it provides a method to improve behavior analysis. We found that in a crowded scene the info based social context allows the affinity to anticipate plain of self justifying groups. Scene setting reliably enhances the recognition of abnormality in both datasets [11].

Yiwen Wan et.al [2014] In this paper the real time highway system was discussed and contain information about vehicle's speed and volume and whenever requires generates incident signals. It additionally gives helpful fuse to existing surveillance condition with multiple levels of operations. Developments involve a novel 3-D Hungarian algorithm along with kalman filter (to project location of vehicles) compute for object analysis and hands-off instrument for camera alteration. Speed is judged after mapping with respect to evolutionary dynamic model. It also recognizes the stopped vehicles for more than Ts seconds, and triggers an alarm to traffic management for response when incident occurs [12]. Under different circumstances i.e. rain, low illustration or high illustration, it provide better result.

Enis Çetin et.al [2013] In this paper, it provides video handling systems for identification and investigation of uncontrolled fire. We know that human can detect fire easily even from long distance but machine cannot understand it. Traditional point sensors have transport delay whereas VFD reduces the recognition time in both inside and outside on the grounds because cameras can screen "volumes". It is feasible to cover a zone of 100 km^2 with the help of a single tilt-zoom camera set on a peak for wildfire detection. A further advantage of the VFD association is that it provides significant information about the size and direction of smoke circulation [13]. This system is used in high risk areas and in risky buildings. It will reduce the rate of false alarm.

Laila Alhimale et.al [2014] In this paper focus is on unintended falls which happen among senior citizens, particularly in indoor situations. A video based identification framework was used to provide security and watch the real activities of elder person. Neural network system was used to check a set of predefined situations that contain pray, standing, sitting, bending, and lying down in the fall identification [14]. This system is user friendly and provides more confidence among senior citizens to move at home or live alone without the fear of falls. Advantage of this system is it requires minimum maintenance and installation equipments with low cost.

V. PROPOSE WORK

In this paper, the techniques are SURF and hybrid BPNN. SURF is used to compare the georeferenced image with the real time captured images. It is used to detect interest of points and to create description for each point of interest. SURF performance is evaluated and tested with number of sample points in a region. SURF reduces computational time. The declared segment of the Kalman Filter Tracking capacity initially incorporates with few persistent factors. Determined factors took into account a sort of input frame that worked inside the requirements of the MATLAB function. In this manner, the calculation checks whether one of these constant factors is available before preparing each frame. Active factors are present just when the calculation first begins, i.e. at the point when the primary frame of information is handled. In such a case, starting conditions must be set for a few factors. Whatever is left of this region will be given to examining the particular starting conditions set for every factor of fault. In NN, whenever one element of NN failed other remains in working state because it has a property of working in parallel. NN need simple arithmetic operators such as addition and multiplication. It needs to be trained before the functions start

[14]. NN works in 2D images. It generates an error in the output layer when there is difference in the real system output and the output generated by computation. After this, error is fed back to the system to update its weights following some set of learning rules. These weights are initially assigned some random values, which are then updated during training process. In this we use hybrid NN because by default NN use activation function whose output value usually lies between 0 and 1 or -1 and 1, but hybrid NN use output values in points which increase conversion rate and also increase result accuracy.

1. Kalman Filter (KF)

Kalman filter is used for foreground object recognition for every pixel. A powerful Kalman filter structure is referred for the improvement of moving objects detection. It can work on dynamic videos also. The calculation of Kalman filter as appeared underneath reading the vehicle video and change video into multiple frames i.e. I (x, y)

$$I_k = F_k I_{k-1} + B_k U_k \quad (1)$$
$$P_k = F_k P_{k-1} F^T{}_k + Q_k \quad (2)$$
$$Y_k = Z_k - H_k I_k \quad (3)$$
$$S_K = H_k P_k H^T{}_k + R_k \quad (4)$$
$$K_k = P_k H^T{}_k S_k^{-1} \quad (5)$$
$$I_k = I_{k-1} + K_k Y_k \quad (6)$$

Assuming a little procedure change, let Q=1e-5. (Let Q=0 yet expecting a small but non-zero value provides us greater flexibility in "tuning" the filter as we will exhibit below.) Let's accept an experience of fact; we realize that the estimation of the random constant has a standard normal probability distribution, so we will "seed" our filter with the assumption that the constant i = 0. I_k and Z_k are the actual state and estimation vectors. I_k and Z_k are the approx. state and estimation vectors from eq. (1) and (3). H is the Jacobian network of partial derivatives of h as concerned with x.

$$H_{[i,j]} = \frac{\partial h_{[i]}}{\partial x_{[j]}}(\tilde{x}_k, 0)$$

The image I_k is the estimation of picture at k frame. P_k is the error covariance matrix. In this the present prediction is joined with present observation data to modify the state estimate. The two phases interchange, with the expectation advancing the state while the next scheduled perception and the modifications are combined with the observation. Sometimes, this isn't essential; if observations are not available for some reasons, the changes might be skipped and numerous prediction steps performed. Similarly, if various observations are present in the meantime, numerous changes might be performed with various observation matrices H_k. The equation for the modified estimate and covariance above is referred for the optical Kalman [15].

2. Object Detection Method

For a long period object recognition has been remaining a dynamic research area. The methodologies for this issue can be separated by a few ways. Some of them are area particular, i.e. they depend on the specific models' presumptions, and some of them are general, in light of the normal techniques for machine learning and autonomous of the specific functional issue. This segment is intended to demonstrate the short review of video object location systems and at the same time demonstrate the existing place of pictorial object identification. Object location can be isolated into two gatherings techniques: directed (as division) and unsupervised. On the opposite side, the strategies can be information-driven, i.e. the model is an element of a few informational index, or can be trade driven, where the possibility of the specialists is communicated in a few scientific show. The decision of the model should be taken by specific useful properties and various elements, for example, computational assets, memory utilization, efficiency of the model to the functional issue, and numerous others [16].

3. Back propagation neural network

BPNN is a feed-forward neural network consist of input, hidden and output layer and have weighted interconnections among them. The training of BPNN is supervised learning based on the output of network and minimum error. Architecture of BPNN is shown in figure(2), and delta learning rule is given below. Let us consider error at the k^{th} output node is given by:

$$e_k^p = o_k^p - d_k^p \quad (7)$$

where o_k^p and d_k^p is the neural network and desired output respectively, p denotes pattern. Let K and P be the total number of node and total pattern respectively, So MSE is given by:

$$E = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{2}\left(o_k^p - d_k^p\right)^2 = \frac{1}{K}\sum_{k=1}^{K}\left(e_k^p\right)^2 \quad (8)$$

In the updated of weights, after all the training patterns are feed as input then the cost function becomes:

$$E = \frac{1}{PK}\sum_{p=1}^{P}\sum_{k=1}^{K}\frac{1}{2}\left(e_k^p\right)^2 \quad (9)$$

For the output layer, differentiating E w.r.t. the weights w_{kj}, (stated in figure):

$$\frac{\partial E}{\partial w_{kj}} = \frac{\partial E}{\partial e_k}\cdot\frac{\partial e_k}{\partial o_k}\cdot\frac{\partial o_k}{\partial \propto_k}\cdot\frac{\partial \propto_k}{\partial w_{kj}} = e_k(1).\ \emptyset'(\propto_k) y_j \quad (10)$$

Similarly, the partial derivative w.r.t. the hidden layer weights v_{ji}:

$$\frac{\partial E}{\partial v_{ji}} = \sum_k \frac{\partial E}{\partial e_k}\cdot\frac{\partial e_k}{\partial \propto_k}\cdot\frac{\partial \propto_k}{\partial y_j}\cdot\frac{\partial y_j}{\partial \propto_j}\cdot\frac{\partial \propto_j}{\partial v_{ji}}$$

$$\frac{\partial E}{\partial v_{ji}} = -x_i\ \emptyset'_j(\propto_j)\sum_k e_k\ \emptyset'_k(\propto_k) w_{kj} \quad (11)$$

Where functions ϕ_j, ϕ_k are called activation functions.

The optimization of the error function are done via updating weights, the weight w_{kj} and v_{ji} is typically done by using steepest descent algorithm i.e. following adjustments are applied to the weights in the direction of steepest descent.

$$\Delta w_{kj} = \eta \frac{\partial E}{\partial w_{kj}} = \eta\, e_k\, \emptyset'_k(\propto_k) y_i \qquad (12)$$

$$\Delta v_{ji} = \eta \frac{\partial E}{\partial v_{ji}} = \eta\, x_i\, \emptyset'_j(\propto_j) \sum_k e_k\, \emptyset'_k(\propto_k).\, w_{kj} \qquad (13)$$

Where η is known as learning rate, if we add a momentum term into the above equations:

$$\Delta w_{kj}(t) = \eta\, \frac{\partial E(t)}{\partial w_{kj}} + \beta \Delta w_{kj}(t-1) \qquad (14)$$

$$\Delta v_{ji}(t) = \eta\, \frac{\partial E(t)}{\partial v_{ji}} + \beta \Delta v_{ji}(t-1) \qquad (15)$$

Where t denotes the number of iterations and β is positive constant s.t. $\beta \in [0, 1]$, known as momentum constant.

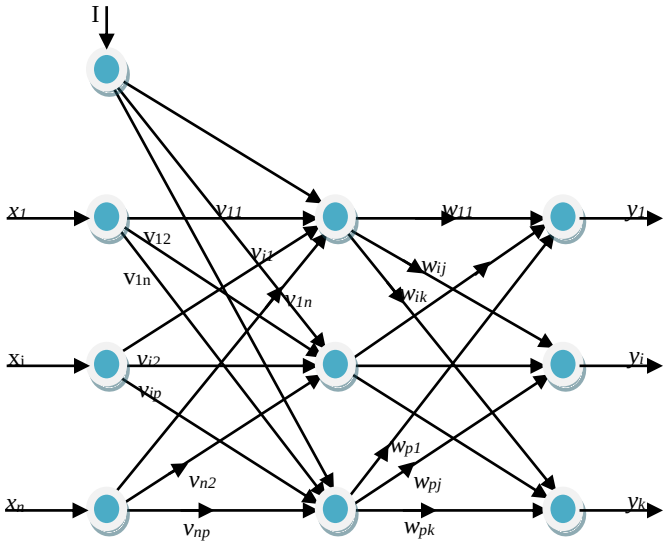


Figure 2: Architecture of BPNN

This calculation is best comprehended by arranging below 2 fundamental stages specifically:

Phase 1: Propagation:

i. Forward propagation: Input is nourished by the network to produce spread's yield enactments.
ii. Backward propagation: A feedback network is shaped by bolstering the yield as contribution to request to create a distinction amongst real and the objective the yields.

Phase 2: Weight update:

a) The gradient of weight is a product of the dissimilarity of effects and input creation.
b) Subtract a ratio (percentage) of the gradient from the weight.

$$f(x) = \frac{1}{1 + e^{-x}}$$

An expression for the sigmoidal function used where e is the natural logarithmic function and x can have any real value.

Firefly Algorithm

FA is a population based method developed by X.S. Yang [17] for solving optimization problems. In this paper MSE is consider as fitness function. Basically, FA is inspired by the bioluminescent behavior of firefly's i.e. short and regular flashing of light produced by fireflies. Lower intensity fireflies are easily attracted by the higher intensity firefly and the intensity of the flashing lights. This process of flashing lights can be formulated as matching with the fitness functions, MSE, which are given by equation (16) to be minimized. FA is based on certain assumptions given below:

1. all fireflies are unisex, so one firefly attracted to other firefly despite of their sex.
2. attractiveness are proportional to a firefly intensity. Thus for any two flashing fireflies, the less intensity one will move towards the higher intensity one.
3. intensity of every firefly indicate the quality of the solution.

$$\text{MSE} = \frac{\sum_{k=1}^{N} (\bar{f}\ (m,n) - f(m,n))^2}{N} \qquad (16)$$

where $\bar{f}$ (m, n) is target and f (m, n) is output.

The intensity (I) of firefly depends on the distance among fireflies i.e. intensity of fireflies is continuous decreases as the Euclidean distance between two fireflies say *r*, increases. The flashing light intensity I is inversely proportional to the square of distance *r* from source and expressed as:

$$I(r) = I_0\, e^{-\gamma r^2} \qquad (17)$$

where I_0 is the intensity of source, and γ (initial value 1) is absorption coefficient when one firefly attracts another with attraction coefficient β, which depends on distance between two firefly (*r*) is given by equation (18).

$$\beta(r) = \beta_0\, e^{-\gamma r^2} \text{ ; where } \beta_0 = 0.2, \text{ initial value.} \qquad (18)$$

The distance between i^{th} and j^{th} fireflies is:

$$r_{ij} = \left\| x_i - x_j \right\| = \sqrt{\sum_{k=1}^{D} \left(x_{ik} - x_{jk} \right)^2}$$

Where x_{ik} is the k^{th} element of i^{th} firefly position inside search space and D denotes the dimensionality. The i^{th} firefly attracts to j^{th} firefly as follows:

$$x_i = x_i + \beta_0 e^{-\gamma r_{ij}^2} (x_j - x_i) + \alpha N_i(0,1)$$

where x_i is the position of i^{th} firefly, term $\beta_0 e^{-\gamma r_{ij}^2} (x_j - x_i)$ represents attractiveness and $\alpha N_i(0,1)$ random movement of i^{th} firefly, where $\alpha \in (0,1]$.

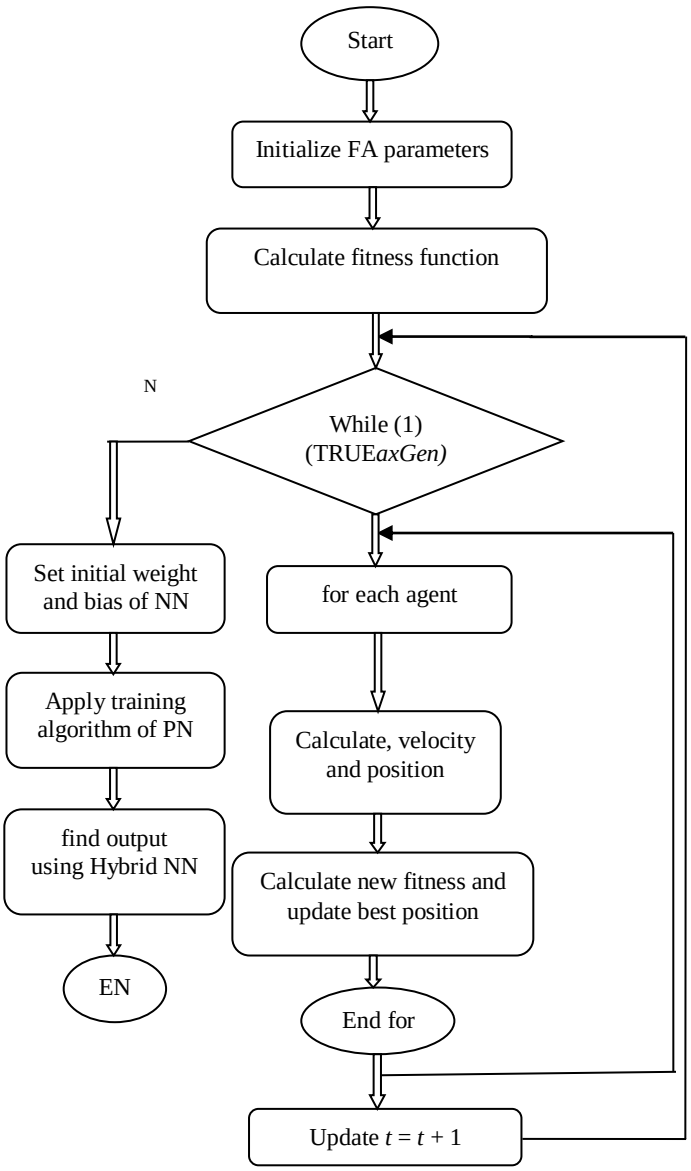


Figure 3: Flow chart of Hybrid BPNN

Hybrid BPNN

A nature inspired optimization method like FA is used to initialize weight of Hybrid BPNN for good solution and convergence rate. Instead of random weight initialization, assign initial weights to Hybrid BPNN using FA and train it by delta rule given in equation (14) and (15). FA is a global search method with adaptive factor alpha, has perceptible merits of avoiding local minima and slow convergence rate. Using this context, we represent a novel interest point detector and descriptor that compute and compare much faster as concern with discreteness, robustness and repeatability for real life object recognition application. Hybrid BPNN is trained with the training data and applies to testing data thus to find the best-fitting network. The flow chart of the Hybrid BPNN is shown in figure 3.

4. Feature Detectors

Feature detectors can be categorized into three classes: single-scale, multi-scale and affine invariant identifiers. More or less, single scale implies that there is just a single characteristic of the features or the object forms utilizing finder's inner parameters. The single-scale finders are invariant to picture changes, for example translation, rotation, changes in illumination and presence of noise. But, they are unfit to manage the scaling issue. Given two frames of a similar scene identified with a scale change, we have to decide if same interest points can be recognized or not. Accordingly, it is important to develop multi-scale identifiers that can fetch different features under scale changes.

5. *Speeded-Up Robust Features Descriptor (SURF)*

SURF is an interest point detector-descriptor technique. It is significantly quicker and more strong instead of SIFT. For the recognition phase, rather than depending on perfect Gaussian partial, the calculation depends on basic 2D filter; where it uses a scale-invariant blob locator under the observation of the measure of Hessian matrix for both scale determination and areas. It is based on second order derivations of an image intensity G(x, y). It is essential to compute the second order Gaussian derivatives in a productive path with the assistance of vital pictures utilizing an arrangement of box filter. The 9 × 9 box filter are approximations of a Gaussian second with standard deviation σ =1.2 and address the lowest scale for blob response maps. These approximations are shown by Gxx, Gyy, and Gxy. Subsequently, the estimated measure of Hessian can be communicated as

$$\det(H_{approx}) = G_{xx}G_{yy} - (wG_{xy})^2$$

Where w is a relative weight and it is used to modify the Hessian's determinant. The approximated determinant of the Hessian addresses the blob response in the image. These responses are secured in a blob response guide, local maxima are identified and filter by quadratic interpolation [18, 20].

6. Median Filter (MF)

In the proposed methodology the size of the window is permanent; in some cases, the median might be differing at the middle of the sorted pixel values. The proposed MF is planned to reduce the issue exist in the standard MF and removed by the adaptable MF. Moreover, with the standard middle technique, the window of k*k array of pixels is selected such that

$$k^2 = 2n + 1 \Rightarrow n = \frac{k^2 - 1}{2}$$

Where for whole number n>0, k=3, 5, 7 …. In the proposed arrangement of isolating, as in standard MF, the pixels are arranged and the median is selected from the sorted list of pixels present in the current window [19].

7. Features Matching

Features matching or image matching is a part of computer applications for picture recording, camera dimension and object identification, is the task of correlating two images of the same scene. Matching images is possible, by a set of interest points associated with the image descriptor present in image dataset. Once the features and their descriptors have been fetched from at least two pictures, the subsequent stage is to build up some preliminary feature matching between these pictures. The issue of picture matching can be defined as, assume that p is a point recognized by a detector in a picture related with their descriptor

$$\phi(p) = \{\phi_k(p) \mid k = 1, 2, \ldots, K\}$$

Where, for all K, the featurevector given by the k^{th} descriptor is

$$\phi_k(p) = (f_{1p}^k, f_{2p}^k, \ldots, f_{nkp}^k)$$

The point is to locate the best association q in a different picture from the arrangement of N interest points $Q = \{q_1, q_2 \ldots q_N\}$ by matching the feature vector $\phi_k(p)$ with those of the points in the set Q. A distance determine among the 2 interest point's descriptors $\phi_k(p)$ and $\phi_k(q)$ can be characterized as

$$d_k(p, q) = |\phi_k(p) - \phi_k(q)|$$

A match between a pair of interest points (p, q) is selected only if p is the best one for q in association with all other points in the first picture and q is the best one for p in association with all other points in the second picture. According to this, it is necessary to devise an effective algorithm to procedure as fast as possible. The nearest neighbor comparing in the feature space of the photo descriptors in Euclidean standard can be used for organizing vector-based highlights. Practically speaking, the ideal closest neighbor algorithm and its parameters rely upon the dataset features.

Besides, to smoother coordinating possibility for which the communication might be viewed as uncertain, the proportion among the distance to the nearest and the next nearest image descriptor is < threshold. As an exceptional case, for organizing HD features, 2 algorithms have been proposed: the randomized k-d forest and the FLANN [20].

8. Thresholding Methods and Otsu

According to gray level distribution, object in an image can be separated from its background by selecting some gray level threshold value. Gray images are converted into binary images and assign zero to all pixels whose value is below some threshold and assign 1 to all pixels whose value is above the threshold. In this condition, T is global threshold; f (x, y) is the gray estimation of point (x, y) and p(x, y) means some local property of the point, for instance, the ordinary gray estimation of the neighborhood concentrated on point (x, y) Based on this, there are two sorts of thresholding techniques.

a) Global thresholding: At this point when T depends just on f (x, y) (gray value) and the estimation of T only identifies with the character of pixels, then it is known as global thresholding.
b) Local thresholding: If edge T depends upon f (x, y) and p(x, y), this thresholding is known as local thresholding. This scheme segments entire image into multiple regions and for all these regions set some threshold T.

Otsu system is dependent on global thresholding i.e. on gray value only and it is very simple and effective. Otsu technique was proposed by Scholar Otsu in 1979. This technique requires dealing with a gray level histogram before running. In any case, in context of the 1D which just thinks about the gray level data, it doesn't give better segmentation outcomes. Because of this 2D Otsu calculation was proposed which work on both gray level edges of every pixel and in addition its spatial relationship data inside the area. That's why Otsu calculation gives better result when used in noisy images. There are many techniques to reduce computational time and maintain thresholding results. It proposed a fast recursive methodology that can capably diminish computational time. Otsu's procedure was one the decision strategies for general real world pictures as for consistency and shape measures. Be that as it may, Otsu's strategy utilizes a comprehensive look to assess the rule for augmenting the between-class change. As the number of classes of picture increases, Otsu's technique sets aside a lot of time for multilevel edge choice [21].

9. Background Subtraction

Background Subtraction otherwise called closer view recognition is a method in the fields of computer vision wherein a photo edge is isolated for handling. Background subtraction is done if the image we are accessing is the part of a video.

$$P [F (t)] = P [I (t)] - P [B]$$

The background is believed to be the frame at time t. This refinement picture would simply exhibit some power for the pixel territories which have altered affirmation two edges.

The edge is located in this refinement picture to enhance the subtraction.

$$[P [F (t)] – P [F (t- 1)] \& > \text{threshold}$$

This implies the distinction pictures pixels forces are edge. The exactness of this method is needy on rapidity of improvement in the scene.

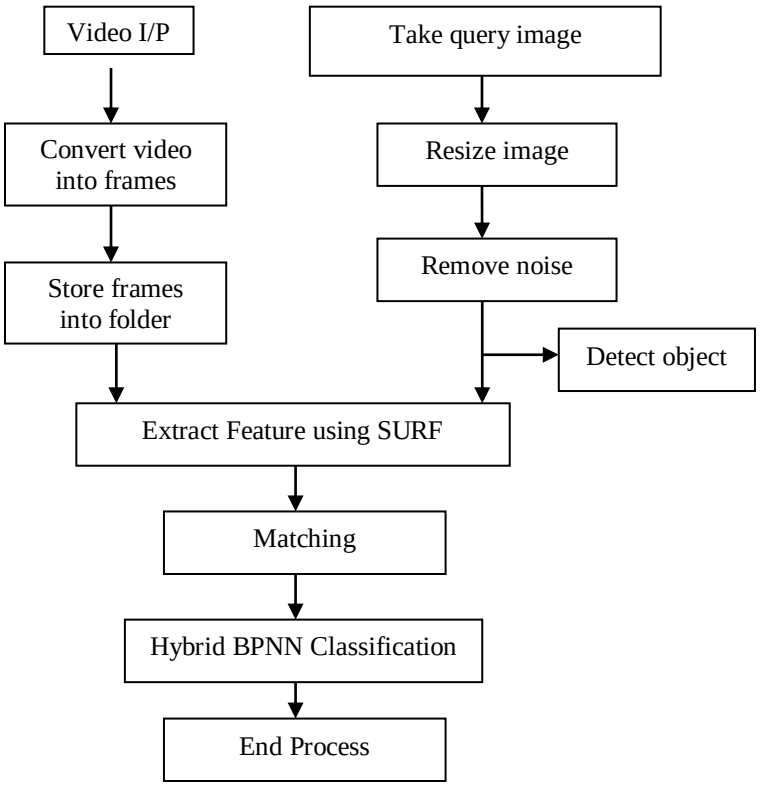


Fig. 4. Flowchart of propose work I

Propose Methodology I

Step1. First, browse a video.
Step2. Then, segment frames.
Step3. Store frames into folders after storing frames resize every frame.
Step4. Remove noise using median filter
Step5. Extract feature using SURF
Step6. Matching the features.
Step7. Apply Hybrid BPNN classification

Propose methodology II

Step1. Take a video.
Step2. Segment video into frames.
Step3. Check frame is gray or RGB. If frame is color then convert into gray.
Step4. Calculate threshold value using Ostu Method.
Step5. Extraction feature using SURF.
Step6. Detect object using Kalman filter
Step7. Calculate the value on parameter.
Step8. Stop.

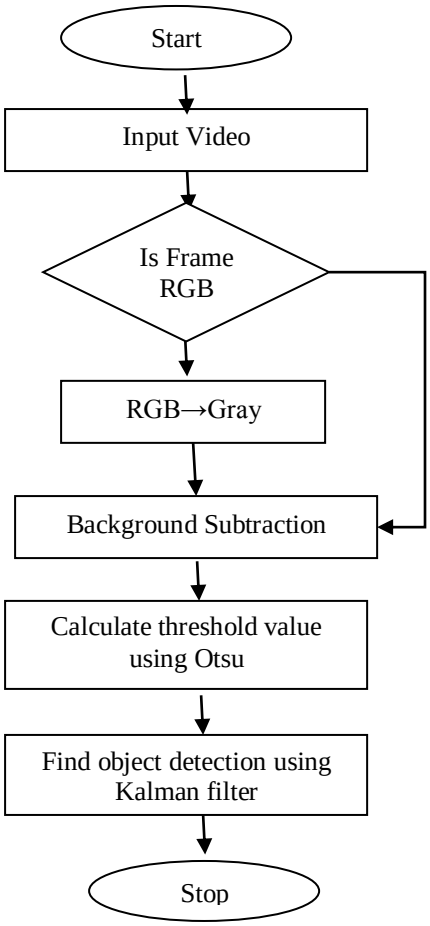


Fig. 5. Flowchart of propose work II

VI. RESULT ANALYSIS

For carrying out the training procedure a SURF and Hybrid BPNN is used. The training and testing of data have been done using the neural network tool in the MATLAB has been used. We have used Hybrid BPNN under the coordinated getting the hang of designing of neural network tool compartment to process our data, in which one-way associations work and no criticism component is incorporated.

1. First Result Analysis

Fig. 6. First convert rgb2gray

Fig. 7. Object detection

Fig. 8. Feature detection

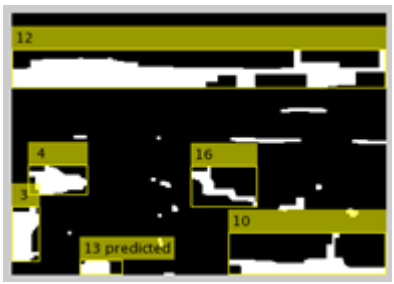


Fig.9.Predicted location using Kalman filter

1. Second Result Analysis

Fig. 10 Play original video and detect predict object location

Fig 11. Predicted location

Fig. 12. Convert RGB frame into Gray frame

Fig. 13. Subtract background

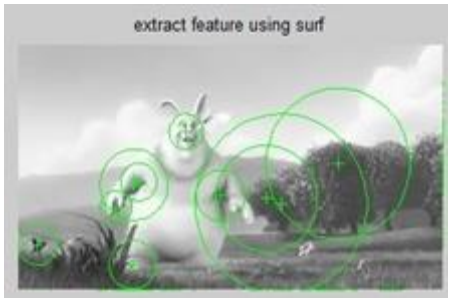


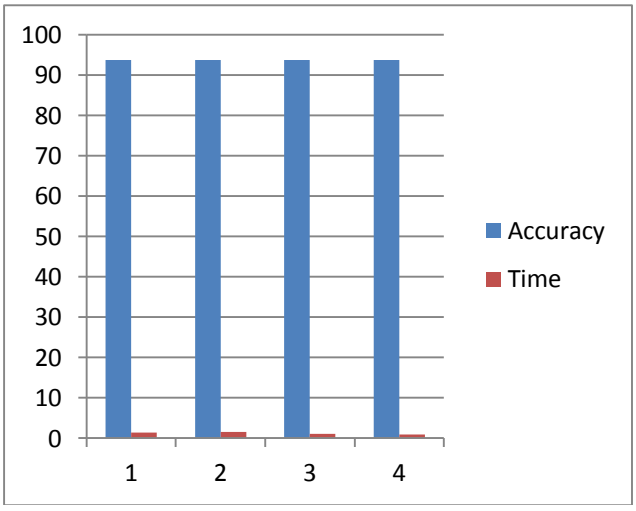
Fig. 14. Extract feature using SURF

Table.1 Threshold value

Threshold value
0.6235

Table.2 Accuracy and Time value

Accuracy	Time
93.75	1.4721
93.75	1.5782
93.75	1.0368
93.75	0.9740



Graph 1. Accuracy and Time value

Conclusion

The goal of this paper is to give a straight-forward, brief description for researchers about features recognition and extraction in research field. It presents the fundamental script and numerical ideas for recognizing and fetching images. It also focuses on different existing algorithm for detecting interest points. The most popular algorithms for example, SURF, Hybrid BPNN and Median filter,…etc are mentioned with their merits and demerits. We have proposed a technique that combines SURF with Kalman filter in order to predict the interest point of an object in long picture groupings. Utilizing Hybrid BPNN, we get an estimation of the object's area which is then sent as a perception to a Kalman filter. As a future upgrade, the creators propose the filters to be subsequently invigorated by sudden movement changes which imply a versatile Kalman filter. In this paper, Kalman filter has better accuracy then frame difference technique. Moreover, the divisions have been enhancing and the object identification more level. In future, work is required on practical filters to identify an object.

References

[1] Malik M. Khan, Tayyab W. Awan, Intaek Kim, and Youngsung Soh, "Tracking Occluded Objects Using Kalman Filter and Color Information". International Journal of Computer Theory and Engineering, Vol. 6, No. 5, October 2014.

[2] Shilpa, Prathap H.L, Sunitha M.R "A Survey on Moving Object Detection and Tracking Techniques" IJECS Volume 05 Issue 4 April 2016 Page No.16263-16269

[3] Rajeswari Natarajan and Chandrakala.S"Audio-Based Event Detection in Videos - a Comprehensive Survey" IJET Vol 6 No 4 Aug-Sep 2014 PP-1663-1674

[4] P Turaga, R Chellappa, VS Subramanian, O Udrea, Machine recognition of human activities: a survey. IEEE Trans. Circuits Syst. Video Technol. 18(11), 1473-1488 (2008)

[5] G Lavee, E Rivlin, M Rudzsky, Understanding video events: a survey of methods for automatic interpretation of semantic occurrences in video. IEEE Trans. Syst., Man, Cybern. C 39(5), 489-504 (2009)

[6] Manoranjan Paul, Shah M E Haque and Subrata Chakraborty"Human detection in surveillance videos and its applications - a review". EURASIP PP.8-16

[7] Francisco Ortega-Zamoranoa, Miguel A. Molina-Cabelloa, Ezequiel López-Rubioa, Esteban JPalomoa,b"Smart motion detection sensor based on video processing using self-organizing maps" (2016) pp476–489.

[8] I Bilik, J Tabrikian, MMSE-based filtering in presence of non-Gaussian system and measurement noise. IEEE Trans. Aerosp. Electron. Syst., 2010; 46: 1153–1170.

[9] Yifei Zhang, Wen Qu, Daling Wang" Action-Scene Model for Human Action Recognition from Videos" 2nd AASRI6 (2014) pp 111 – 117

[10] Pedro Gil-Jimenez, Hilario Gomez-Moreno, Javier Acevedo-Rodrguez, Saturnino Maldonado Bascon "Continuous variance estimation in video surveillance sequences with high illumination changes" Signal (2009) pp1412–1416

[11] Michael J.V. Leach, Ed.P. Sparks, Neil M. Robertson a "Contextual anomaly detection in crowded surveillance scenes" (2014) pp-71–79.

[12] Yiwen Wan, Yan Huang, Bill Buckles "Camera calibration and vehicle tracking: Highway traffic video analytics"2014, pp202-213

[13] Enis Çetin, Kosmas Dimitropoulos, Benedict Gouverneur, Nikos Grammalidis, Osman Günay, Y. Hakan Habiboglu, B. Ugur Töreyin ˇ d, Steven Verstockt e" Video fire detection – Review" 23 (2013)pp- 1827–1843.

[14] Laila Alhimale, Hussein Zedan, Ali Al-Bayatti "The implementation of an intelligent and video-based fall detection system using a neural network" 18 (2014)pp- 59–69.

[15] C.Srinivas Rao, P.Darwin, "Frame Difference And Kalman Filter Techniques For Detection Of Moving Vehicles In Video Surveillance". C.Srinivas Rao, P.Darwin / International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com Vol. 2, Issue 6, November- December 2012, pp.1168-1170

[16] 1T. NANTHINI, 2N. PUVIARASAN, "ROBUST VEHICLE DETECTION AND TRACKING". Proceedings of IISTEM International Conference, 15th October 2017, Pondicherry, India

[17] Yang, X. S. Cuckoo Search and Firefly Algorithm. Springer Press (2014).

[18] Herbert Bay, Tinne Tuytelaars and Luc Van Gool, "SURF: Speeded Up Robust Features". ScienceDirect- Computer Vision and Image Understanding, Volume 110, Issue 3, June 2008, Pages 346-359

[19] Kwame Osei Boateng1, Benjamin Weyori Asubam1,2 and David Sanka Laar, "Improving the Effectiveness of the Median Filter". International Journal of Electronics and Communication Engineering. ISSN 0974-2166 Volume 5, Number 1 (2012), pp. 85-97 © International Research Publication House

[20] M. Hassaballah, Aly Amin Abdelmgeid and Hammam A. Alshazly, "Image Features Detection, Description, and Matching". Springer International Publishing Switzerland 2016 A.I. Awad and M. Hassaballah (eds.), Image Feature Detectors and Descriptors, Studies in Computational Intelligence 630, DOI 10.1007/978-3-319-28854-3_2

[21] Miss Hetal J. Vala, Prof. Astha Baxi, A Review on Otsu Image Segmentation Algorithm". ISSN: 2278 – 1323 International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 2, Issue 2, February 2013