International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 11, November 2014, pg.316 – 324

SURVEY ARTICLE

A Survey on Action Recognition

A.Dhivya[1], S.Saranya Devi[2], M.Jayasudha[3]

[1,2]M.E Scholar, Department of Computer Science and Engineering

[3]Assistant Professor, Department of Computer Science and Engineering

Sri Eshwar College of Engineering, Coimbatore

divs392@gmail.com [1], ssarani022@gmail.com [2], jayasudha2080@gmail.com [3]

Abstract - This paper identifies various methods involved in recognizing the human action automatically for various applications. We focus on the various methods that can be applied for action recognition. The methods used are Background Subtraction Algorithm (BSA), Frame Differencing and Template Matching, BSA in security system, Cam-Shift Algorithm, Real time motion detection, Motion History Information (MHI) from compressed video. These methods are based on pixel based approach that gives single dot resolution in recognizing action (Movements).

Keywords: Action Recognition, Background Subtraction, Template Matching, Pixel Based, Single Dot Resolution

I. INTRODUCTION

Human action recognition has become more popular only because of its potential applications in the field of Bio-metrics, Security and also in human computer interaction. Although the work has been done, recognition in uncontrolled environment is still a challenging task. An active research topic in computer visions is portion detection, object classification, tracking with activity recognition and description of characteristics. Visual surveillance strategies are used from many long years ago to gather the information about the person and to monitor them along with the event detection. Video surveillance is used to detect the moving object to know the behavior of the person to recognize the action performed. Visual surveillance technologies, CCD cameras, thermal cameras and night vision device are the three most widely used devices in the field of surveillance.

The main intend of surveillance concept is to monitor and also to control the entire surveillance task. The aim of video surveillance is to build an automated controlling of camera in order to replace the number of persons required to monitor the camera for performing. Number of cameras required for security and surveillance process is becoming more than the operators required. The automated surveillance systems can be developed for both online and offline storage of video sequence and to analyze the information in that sequence. Surveillance system is very much useful in public places to provide better security from threats. This system is also used for private issues like home security and monitoring of patient in medical field and also in military appliances. This system also used in many traffic controlling activities and other public issues. Some of the areas where video surveillance system place a major role in many application are 1) Medical field 2) Military alliances 3)Extracting statistics for sport activities 4)Surveillance of forests for fire detection 5) Patrolling of highways and Railway for accident detection.

In a video stream monitoring process the recognition of activity is divided into three categories, which are single person with single object interaction, multiple people with multiple objects, and group behaviour. Here it is focused on visual surveillance in the direction of surrounding of model, objects segmentation, tracking and finally detecting action recognition. The following papers review about automatic recognizing of action immediately with the alert system at instant of time. The video focuses on to the static background from which portion to be tracked is specified using the some algorithms.

A. Background Subtraction Algorithm

The paper[1] describes the effective use of background subtraction algorithm for recognizing human action from static and different background images. Background subtraction is a widely used approach for detecting moving objects in videos from static cameras. The rationale in the approach is that of detecting the moving objects from the difference between the current frame and a reference frame,often called the “background image”, or “background model”. The background image must be a representation of the scene with no moving objects and must be kept regularly updated to adapt to the varying luminaries conditions and geometry settings. Models that are more complex have extended the concept of “background subtraction” beyond its literal meaning. In general, BS can be simplified as shown in the Equation below

$$\mathbf{f\ (x, y) = abs(Frame(x, y) - Background(x, y)}$$

Background Modeling:

The BM module designs a two phase background matching procedure using rapid matching followed by accurate matching in order to produce optimum background pixels for the background model.

- Initial Background Model.
- Optimum Background Modeling (OBM).

Matching:

The pixels are continuously checked by using following techniques.

- Rapid Matching.
- Stable Signal Trainer
- Accurate Matching.

By using anyone of these matching techniques between frames the action can be detected and the background can be updated automatically for better performance.

B. Frame Differencing and Template Matching

The paper[2] uses two techniques Frame Differencing and Template Matching for tracking movements and ensures that the change in orientation and position of object does not hinder the tracking system.

Frame Differencing:

Frame Differencing is a technique where the computer checks the difference between two video frames. If the pixels have changed there apparently was something changing in the image (consider frame). The Frame Differencing Algorithm is used for this purpose, which gives the position of object as output. This extracted position is then used to extract a rectangular image template (size is dynamic depending upon the dimension of object) from that region of the image (frame). The sequence of templates is generated as object changes its position.

Object detection using frame differencing:

The task to identify moving objects in a video sequence is critical and fundamental for a general object tracking system. For this approach Frame Differencing technique is applied to the consecutive frames , which identifies all the moving objects in consecutive frames. This basic technique employs the image subtraction Operator , which takes two images (or frames) as input and produces the output. This output is simply a third image produced after subtracting the second image pixel values from the first image pixel values. This subtraction is in a single pass. The general operation performed for this purpose is given by:

$$DIFF[i ,j] = I_1[i, j] – I_3[i, j]$$

DIFF[i, j] represents the difference image of two frames.

It seems to be a simple process but the challenge occurs is to develop a good frame differencing algorithm for object detection. These challenges can be of any type like

- Due to change in illumination the algorithm should be robust.
- The detection of non-stationary object (like wind, snow etc.) is to be removed.

To overcome such challenges we need to pre-process the *DIFF[i, j]* image. Pre-processing includes some mathematical morphological operations which results in an efficient difference image. *DIFF[i, j]* image is first converted into a binary image by using binary threshold and the resultant binary image is processed by morphological operations.This proposed algorithm for object detection is to achieve these challenges and provide a highly efficient algorithm to maintain such task of object tracking. This algorithm provides the position of the moving object.

Algorithm for object detection:

Assumption: All previous frames are stored in a memory buffer and the current frame in video is F_i

```
if DIFF[i, j] >= T then
Fbin[i, j] = 1 //for object
else
Fbin[i, j] = 0 //for background
```

This assumes that the interested parts are only light objects with a dark background. But for dark object having light background we use:

```
if DIFF[i, j] <= T then
Fbin = 1 //for object
else
Fbin = 0 //for background
```

Template matching:

The generated templates from each frame are passed on to the tracking module, which starts tracking the object with an input reference template. The module uses template-matching to search for the input template in the scene grabbed by the camera . A new template is generated if the object is lost while tracking due to change in its appearance and used further. Generations of such templates are dynamic which helps to track the object in a robust manner. The main objective of this study is to provide a better and enhanced method to find the moving objects in the continuous video frame as well as to track them dynamically using template matching of the desired object. The proposed method is effective in reducing the number of false alarms that may be triggered by a number of reasons such as bad weather or other natural calamity.

Tracking of object using template matching:

The limitation with this tracking module is that all the centroid information received by motion detection module for tracking of objects should always be in camera view. The next operation is to track the only interesting moving object irrespective of other moving objects. The object tracker module is used for this purpose, which keep track of interesting objects over time by locating the position of moving object in every frame of the video. The proposed algorithm has flexibility to perform both task, object detection and to track object instances across frames simultaneously. First of all tracking module will generate a template for all received centroid information and this template is used for the matching in next upcoming frame. By using mathematical correlation we match the template in next upcoming frame and the centroid of matched area is appended to make an estimation of trajectory and mean while the template is updated with new matched region .

C. Cam-shift Algorithm

The paper[3] discusses about five techniques for tracking moving object and detection of their actions that are mainly based on retargetting of missed window for each action detection. The discussed methods are explained in this paper.

Feature-based tracking

Feature-based tracking includes two processes: feature extraction and feature matching. For feature extraction, Moreover has realized operator detection using the feature point of image gray autocorrelation function; Canny proposed edge extraction algorithm ; Smith obtained the corner information using SUSAN operator; Jonathan et al. have proposed a motion compensation algorithm based on tangent distance . For feature matching, Polana and Tissainayagam have both achieved tracking based on the point feature; Nickels and Hutchinson proposed the tracking method based on SSD. Segen and Pingali adopted the angular point of motion profiles as tracking feature.; Yang et al recognized human activity based on motion trajectory features. For feature-based tracking methods , color feature has been widely used because it is easy to be extracted from the sequences of image or video, and robust for target object's shape changes and partial occlusions. For human movement, motion information analysis is the basic tracking mean, which is always been categorized into difference-based, background-estimation-based and motion-estimation-based the three methods. The key idea for them is to detect motion changes based on the correlation between adjacent frames of video sequences. One classic algorithm is optical flow.

Template-matching-based tracking:

This kind of method realizes targeting by computing the color or texture's similarity between template and image sequences, and the target could be located as where has the most similarity. Generally, human motion analysis is achieved by tracking and matching head, torso or four limbs. Such as, Essa completed face tracking; McKenna implemented multiplayer tracking with regional tracker and skin color model; Zhong proposed a deformable template to express a tracked object; Fieguth and Terzopoulos proposed to use the average color value of pixels within the rectangular target area as the template, then Caomaniciu used weighted color histogram distribution in circular area to express the target and adopted the mean shift iterative search strategy of mass center's displacement. Template-matching-based tracking method has advantages of simplicity and small calculation, so it has been widely used.

Shape-matching-based tracking:

Similar to template matching, shape matching method calculates the similarity between shape template and candidate target. Such as, Huttenloche constructed surfaces using Hausdorff distance, and the minimum point on the surface was the new location of the target. But the disadvantage of this kind method is that it can only deal with the translation between frames, and ineffective for the non-rigid movement tracking.

Model-based tracking:

When tracking rigid objects, like automobiles, aircraft, etc., model-based tracking algorithm also been adopted, such as Kalman filter and particle filter. Brodia and Chellappa used Kalman filter to track points in image sequences with noises; Beymer and Kinolige predicted the location and speed of tracked objects in X-Z dimension. Particle filter is one of the most representative methods based on Monte Carlo method, which occupies an important position in the statistical learning. For human motion analysis, Ju proposed cardboard model, whose parametric motion was constrained by articulated movement; Aguiar improved visual hull model of human body; Francesc et al. described a probabilistic integrated object recognition and tracking framework ,which improved traditional Bayesian approach; Antonio et al. also achieved human activity monitoring by local and global finite state machines ; Dimitrios et al. proposed a real-time behavior understanding based on a Bayesian filter supported by hidden Markov models , and the learning process has considered user's feedback. Zhou et al realized target tracking by combining the improved mean shift algorithm with the adaptive Kalman filter. The advantage of this kind model-based tracking method is that it can analyze and track the target even in the case of gesture changes, but its disadvantage is the accuracy of motion analysis depends on model's accuracy.

Active-contour-based tracking:

Active-contour was first proposed by Kass to track moving object, which is also called Snake model. Paragios and Vieren realized multi-target tracking based on active contour; Cootes and Taylor tracked hands with SmartSnake. The advantage of this kind method is small computation and suitable for deformable targets, and disadvantage is difficult for initialization. The above methods have their own characteristics and advantages. But there is not the most appropriate tracking algorithm to realize the monitoring for human motion, and this paper will solve the auto tracking problem oriented to upper-limb movement. In recent years, tracking algorithm based on camshaft (continuous the adaptive mean-shift) has gotten more and more attention because of its good performance in real-time and robustness. The algorithm is a nonparametric method, which detects moving target by clustering the color histogram. And the workflow is described as below:

a) Choose the initial position in given image or video, which is also the tracked target's search window;

b) Compute the color probability distribution of the search window as the color probability model;

c) Detect model's center mass by iterations, and save the spatial distribution areas and locations of the center mass;

d) Convert input image sequences of videos into color probability distribution images, and initialize the new search window using the distribution areas and locations gotten in step c) to obtain the new center mass;

e) Check whether the termination condition is met, which is the movement distance of center mass is less than a given threshold, or the iteration number reaches the set value; if yes, show the search result, and if not, return step b).

Cam-shaft algorithm not only needs to calculate the target's location, but also needs the orientation to achieve its adaptability. So, the following second moment needs to be used to obtain the orientation of probability distribution in two-dimensional space. This paper focuses on retargetting hence overcomes the limitation of cam shift algorithm by shrinking the window frame hence relocating the target.

D. *Real Time Motion Detection*

The paper[4] aims at processing the real time video captured by a Webcam to detect motion in the Scene using MATLAB 2012a, with keeping in mind that camera still recorded which means real time detection. The results show accuracy and efficiency in detecting motion.

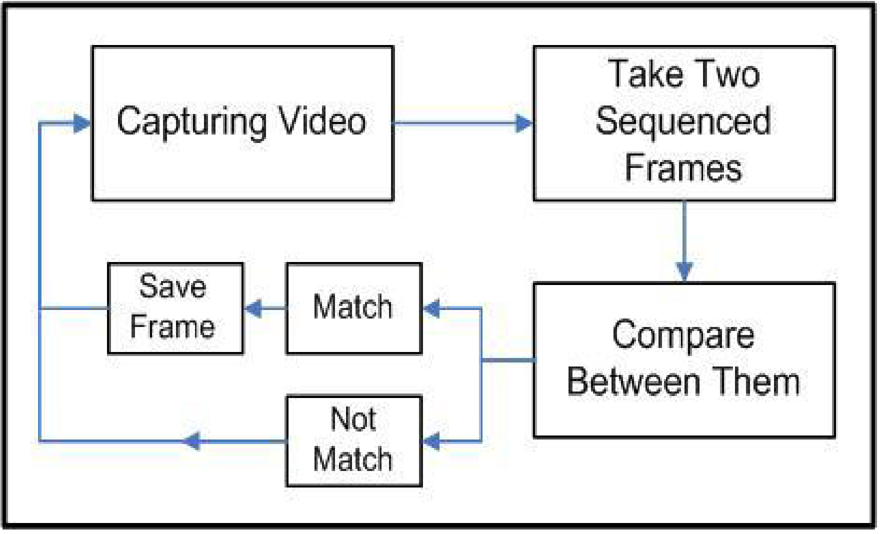

Fig. 1 System Flow Diagram

The system starts by clicking button “start preview” which will start to preview video captured from webcam attached to the system, if any motion occurred; then the system will start to capture images of that motion and save it in a previously selected place. The process of detection occurs by taking two sequenced frames (for example frame1 and frame2 after converting them to gray) and compares to find the mismatching between them according to a specific threshold which will be described later. The Graphical user interface will be like figure (2) which also contains two buttons to pause and resume the video capturing. The comparison starts between frame 1 and 2 then frame3 and 4 and so on, thus by using this method there is no reference image or background of detection which is compared sequentially with all video frames; so this proposal system can be applied in any place.

Noise Removal:

The main idea behind the motion detection algorithm is by filtering the frames based on algorithms. Most cameras produce images with a lot of noise, so motion is detected while there is no motion. Erosion filter is used to remove random noisy pixels in order to get only the regions where the actual motion occurs.

Error Threshold Value:

By choosing the appropriate value for threshold, the detection of motion will be more accurate and neglects all suspicious motions, ones we select a value (approximately between 0.5 and 2), then a computation is done. When two frames are added to the system, if the frame error (which is the ratio of mismatching) is less than error threshold, this means that there is no motion or there is but too small to detect. On the other hand if frame error is greater than error threshold, then there is a motion needs to be detected.

Thus the paper shows experiments for many threshold values for detecting human motion (slow and fast) with different error threshold values for the same motion; the results show that the appropriate value is between 1.3 and 1.5 especially if several motions (medium changes) occur in the scene; this will give less number of saved images that contain false or wrong detections. If the case and the threshold are not standard; we can specify a threshold according to our location that we want to monitor.

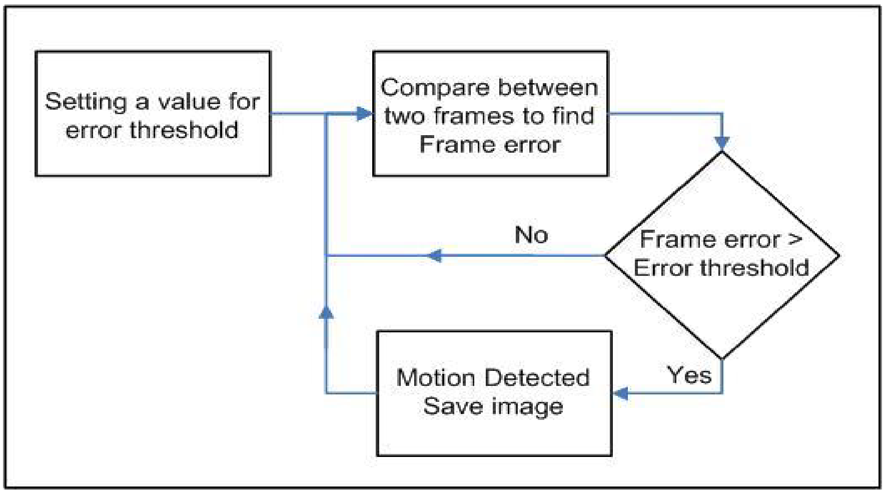
Setting a value for error threshold
Compare between two frames to find Frame error
Frame error > Error threshold
No
Yes
Motion Detected Save image

Fig. 2 Error Threshold Process

E. Motion History Information From Compressed Video

This paper we describe a system for recognition of various human actions from compressed video based on motion history information. We introduce the notion of quantifying the motion involved, through what we call Motion Flow History (MFH). The encoded motion information readily available in the compressed MPEG stream is used to construct the coarse Motion History Image (MHI) and the corresponding MFH. The features extracted from the static MHI and MFH compactly characterize the spatio-temporal and motion vector information of the action. Since the features are extracted from the partially decoded sparse motion data, the computational load is minimized to a great extent. The extracted features are used to train the KNN, Neural network, SVM and the Bayes classifiers for recognizing a set of seven human actions. The performance of each feature set with respect to various classifiers are analyzed.

State-space based approaches:

State-space approach uses time-series features obtained from the image sequences for recognition. The widely used state-space model for activity recognition is HMM due to its success in the speech community. The first attempt to use HMM for activity recognition is done by Yamato et al. where discrete HMMs are used for recognition of six tennis strokes. In their approach time sequential images expressing human actions are transformed to an image feature vector sequence by extracting mesh feature vector from eachimage. The mesh features are extracted from a binarized image obtained after subtracting the background image from original image by applying a suitable threshold. The drawbacks of this method are that it is sensitive to position displacement, noise, and also exhibits poor performance if the training and test subjects are different. The gesture recognition work by Darrell and Pentland uses time-warping technique for recognition which is closely related to HMM. On similar lines, dynamic time warping is used in Ref. [6] to match an input signal to a deterministic sequence of states. Starner and Pentland used HMMs to recognize a limited vocabulary of American Sign Language (ASL) sentences. Here, they used a view based approach with a single camera to extract two-dimensional (2D) features as input to HMMs.

In the work by Bregler ,this classification problem has been approached from a statistical view point. For each pixel in the image, the spatio-temporal image gradient and the color values are represented as random variables. Then the blob hypothesis is used wherein each blob is represented with a probability distribution over coherent motion, color and spatial support regions. Recently Ivanov and Bobick proposed a method, which combines statistical techniques used for detecting primitive component of an activity with syntactic recognition of process structure. In this approach the recognition problem is divided into two levels: i) The lower level detection of primitive components of activity followed by (ii) the syntactic recognition of the primitive features using a stochastic context-free grammar parsing mechanism. Another HMM based human activity recognition method is reported by Psarrou et al. Here the recognition is based on learning prior and continuous propagation of density models of behavior patterns. Ng et al. proposed a real-time gesture recognition system incorporating hand posture and hand motion. The recognition is done with HMM and recurrent neural networks (RNN). There are few works reported in literature which use neural networks for gesture recognition. Boehm et al. used Kohonen Feature Maps (KFM) for recognizing dynamic gestures. Oliver et al. proposed a system for modeling and recognizing human behaviors in a visual surveillance task. This system segments the moving objects from the background and a kalman filter tracks the object's features such as location, coarse shape, color and velocity. These features are used for modeling the behavior patterns through training HMMs and coupled HMMs (CHMM), which are used for classifying the perceived behaviors. Based on the above-mentioned work Madabhushi and Aggarwal presented a system for recognition of human action by tracking the head of the subject in an image sequence. The difference in centroids of the head over successive frames form their feature vector. The human actions are modeled based on the mean and covariance of the feature vector. Here detection and segmentation of the head is done manually.

Template matching based approaches:

One of the earlier works using this approach is found in the work done by Polana and Nelson ,where the flow information is used as feature. They compute the optical flow fields between consecutive frames and divide each frame into a spatial grid and sum the motion magnitude to get the high dimension feature. Here they assume that the human motion is periodic. The final recognition is performed using nearest neighbor algorithm. Davis and Bobick presented a real-time approach for representing human motion using compact MHIs in pixel domain. Here, the recognition of 18 aerobic exercises was achieved by statistically matching the higher order moment based feature extracted from the MHI. The limitations of the above method are related to the 'global image' feature calculations and specific label based recognition. To overcome these limitations the author extended the previous approach with a mechanism to compute dense local motion vector field directly from the MHI for describing the movement. For obtaining the dense motion, the MHI is represented at various pyramid levels to tackle multiple speeds of motion. These hierarchical MHIs are not directly created from the original MHI, but through the pyramid representation of the silhouette images. This indirect way of generating MHI pyramid increases the computational load. The resulting motion is characterized by a polar histogram of motion orientation. Rosales use these motion energy and MHIs for obtaining the spatial location and the temporal properties of human actions from raw video sequences. From these motion energy and MHIs, a set of Hu-moment features that are invariant to translation, rotation and scaling are generated. Using principal component analysis, the dimension of the Hu-moment space is reduced in a statistically optimal way. The recognition performances were evaluated for the following three classifiers namely KNN, Gaussian and mixtures of Gaussian. All the above mentioned techniques process the data in the pixel domain, which is computationally very expensive.

Thus the paper describes about the method for constructing coarse MHI and MFH from compressed MPEG video with minimal decoding. Various useful features are extracted from the motion representations for human action recognition. The KNN, Neural network (MLP) and SVM (RBFkernel) classifiers give the best classification accuracy of 98% and 1D projected and 2D polar features show consistent performance with all the classifiers. Since the data is handled at macroblock level, the computational cost is extremely less compared to the pixel domain processing.

II. CONCLUSION

This paper describes the comparison and analysis between various methods involved in the detection of human action recognition automatically. It also illustrates that there are many techniques that can be followed for detecting the target area for recognizing action, removing noise and compressing image. This kind of comparison reflects that the efficiency differs from each method. This paper shows that all the discussed methods are based on pixel based approach that gives better resolution of image.

REFERENCES

[1] Vishwanatha K & Murigendrayya M Hiremath, *Advanced Motion Detection Algorithm For Patient Monitoring Using Cell Phone With Video Display,* International Journal of Electronics Signals and Systems (IJESS) ISSN: 2231- 5969, Vol-1 Iss-4, 2012.

[2] N. Prabhakar, V. Vaithiyanathan, Akshaya Prakash Sharma, Anurag Singh and Pulkit Singha, *Object Tracking Using Frame Differencing and Template Matching,* Research Journal of Applied Sciences, Engineering and Technology 4(24): 5497-5501, 2012.

[3] Ru Wang, Chunjiang Zhao, Xinyu Guo, *Improved Cam-Shift Algorithm Based on Frame-Difference Method for Video's Auto Tracking,* International Journal of Digital Content Technology and its Applications(JDCTA Volume6,Number19,October 2012.

[4] Furat N. Tawfee **,** *Real Time Motion Detection in Surveillance Camera Using MATLAB,*International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 9, September 2013.

[5] R. Venkatesh Babua and K.R. Ramakrishnanb, *Recognition of human actions using motion history information extracted from the compressed video,* Image and Vision Computing 22 (2004) 597–607.

[6] A. Psarrou, S. Gong, M. Walter, *Recognition of human gestures and behaviour based on motion trajectories*, Image and Vision Computing 20 (5–6) (2002) 349–358.

[7] Shih-Chia Huang, *An Advanced Motion Detection Algorithm with Video Quality Analysis for Video Surveillance Systems*, IEEE Transactions On Circuits And Systems For Video Technology, Vol. 21, No. 1, January 2011.

[8] Naveen Kansal, Hardeep Singh Dhillon, *Advanced Coma Patient Monitoring System*, International Journal of Scientific & Engineering Research Volume 2, Issue 6, June- 2011 ISSN 2229-5518.

[9] M. F. Kazemi, A. H. Mazinan, A. Amir-Latifi, *A Knowledge-Based Objects Tracking Algorithm in Color Video Using Kalman Filter Approach*, International Conference on Information Retrieval & Knowledge Management , 13-15 March 2012.

[10] Ashish Kumar Sahu, *Abha Choubey, A Motion Detection Algorithm for Tracking of Real Time Video Surveillance*, International Journal of Computer Architecture and Mobilitym (ISSN 2319-9229) Volume 1-Issue 6, April 2013.

[11] Gottipati. Srinivas Babu, *Moving Object Detection Using Matlab*, International Journal of Engineering Research & Technology (IJERT), Vol. 1 Issue 6, August – 2012, ISSN: 2278-0181.

[12] Rao, I.S. ; Nandy, S. ; Murthy, P.H.S.T. ; Rao, V.M. ; Dept. Of ECE, GITAM Univ., Visakhapatnam, *India A real world system for detection and tracking*, International Conference on Advances in Recent Technologies in Communication and Computing, 2009 .

[13] Yao Shen, Parthasarathy Guturu, Thyagaraju Damarla, Bill P. Buckles, Senior ,and Kameswara Rao Namuduri, *Video Stabilization Using Principal Component Analysis and Scale Invariant Feature Transform in Particle Filter Framework*, IEEE Transactions on Consumer Electronics, Vol. 55, No. 3, AUGUST 2009.

[14] Francesc Serratosa, Rene Alquezar, Nicolas Amezquita, *A probabilistic integrated object recognition and tracking framework*, Expert Systems with Applications: An International Journal, vol.39, no.8, pp.7302-7318, 2012.

[15] Antonio Femandez-Caballero., Jose Carlos Castillo, Jose Maria Rodriguez-Sanchez, *Human activity monitoring by local and global finite state machines*, Expert Systems with Applications: An International Journal, vol.39, no.8, pp.6982-6993, 2012.