# A Comparative Analysis of Motion Detection & Tracking System for Video Surveillance

MOIZ ABDUL HUSSAIN[1], GOVIND UKANDRAO KHARAT[2]
[1]Electronics & Telecommunication Engineering Department,
Anuradha Engineering College, Chikhli, India
[2]Principal
Sharad Chandra Pawar college of Engineering, Otur, India
[1]mymoiz2004@yahoo.co.in, [2]gukharat@gmail.com

*Abstract:* - In today's world, security of human being is the most active research area. Many different applications are being proposed to safeguard the public places. In this paper, we review the four different techniques of video surveillance system based on motion segmentation and tracking. The first system is based on dual frame differencing method followed by the morphological operations & Kalman filtering. The second technique is the use of visual background subtraction combined with illumination insensitive template matching algorithm. The third one is the optical flow used in combination of template matching. The final method is the design of AdaBoost classifier using sparse matrix & 45$^{o}$ rotated Haar features. This paper explores the different methods of visual tracking & their experimentation results to enhance the study in the field of image processing.

*Key-Words:* - Video Surveillance, Optical Flow, Frame Difference, AdaBoost Classifier.

## 1 Introduction

Ensuring the security at the public places is in general the most challenging task. In recent years [1, 2] there are numerous approaches proposed in this area. Today's security system acts like video storage anything that comes under the surveillance area is recorded in the computer system. Their basic concern is on the video compression and the frame retrieval. Such kind of surveillance system serves two important functions: First, assisting the human operator by providing the footage of the incident taken place for identifying the danger and second capturing the evidence for investigation purpose. Although these are the initial steps in surveillance system yet they could not be applied to detect crime in the real time and take necessary action against it.

It is obvious that human monitoring of the surveillance system is not efficient and reliable task. As it is well known fact that human being cannot cautiously and effectively monitor the area. Therefore there is a requirement of an involuntary monitoring and analysis system which is capable of detecting and tracking the subject. Also, the system can react according to the behaviour analysis of the individual in the scene. The involuntary visual surveillance technologies can be applied to develop the surveillance system that can be implemented in both real time threat monitoring and after crime investigation.

In this paper we begin with the survey of related techniques for video surveillance system. Then present the methods of motion segmentation and tracking. After that the result of the system are presented. Finally the paper ends with the conclusion.

## 2 Related Work

There is an enormous literature in the field of motion detection & object tracking. Here we discuss the relevant object tracking system.

Fan Yang *et al.* [3, 2014], proposed a super pixel tracking system in which the structural information is captured to make a tracker distinguish the target from the background. The method is also capable of dealing with occlusion and drift problems. An online object tracking method described by Wang *et al.* [4, 2013], with sparse prototypes, using principal component analysis (PCA) to train the appearance models. The tracking task is accomplished by online trained sparse prototypes. This algorithm is effective on motion blur & occlusion handling.

Olivier Barnich *et al.* [5, 2011], presented an innovative technique of adapting the pixel to the background based on its resemblance to the stored pixel. If the value of pixel matches with the stored pixel at the same location or in neighbourhood location the pixel is considered to be the part of background otherwise foreground. Peng Cui *et al.*

[6, 2012], proposed an action representation system using unsupervised categorization. In this video decomposition is done where videos are considered to be spatially distributed pixel and after replacing these pixels with their prototype multi action is recognized using duality between pixel clustering and action clustering.

Yuji Tachizaki *et al.* [7, 2009], proposed an object tracking system for motion JPEG movies. This technique makes the use of similarity between positive and negative values of DCT coefficients. Also the multiple blocks based processing decreases the wrong detection. This method generates good result in comparison with Gaussian Mixture Model. Tudor BARBU [8, 2012], proposed a human tracking system using static camera. In this system temporal differencing method along with morphological processing is used. This is then processed by human properties such as size, skin segments which is then matched by using correlation process. Tal Ben-Zvi et al. [9, 2012], presented a study of the signals approaching the target and classify the object intention as the suspicious based on tracking of the signals. Jianguo Lu et al. [10, 2010], presented the appearance model using the human body structure. This structure is then processed by Markov Chain Monte Carlo (MCMC) method to estimate the solution of target data problem. The method best suited the complex environment, occlusion & irregular target motion.

# 3 System Methodology

## A. Dual Frame Differencing Method

Human Motion detection & tracking algorithms have to deal with various challenges as described earlier. To overcome these problems the system, consists of three frame differences and morphological blob analysis to avoid false motion detection.

The proposed system works in three stages. At first stage gray scale frames are extracted from video & frame difference between frame Fn-2 and Fn-1 and second frame difference between the frame Fn-1 and current frame Fn are taken. It is different than other algorithms where only the difference between the current and previous frame is considered. In second phase, morphological operation is performed on the resultant frame difference to suppress the remaining errors. Then the background regions are extracted, holes are filled and small regions are removed. It gives, two background masks, after morphological operations which are then compared with the threshold calculated followed by the AND operation applied to eliminate the false motion detection. Finally, the Kalman filter [11] is used to remove the noise and other changes in pixel due to illumination or any other reason. Extracted foreground is then tracked by the rectangle around it. The system is depicted as figure 1 below.

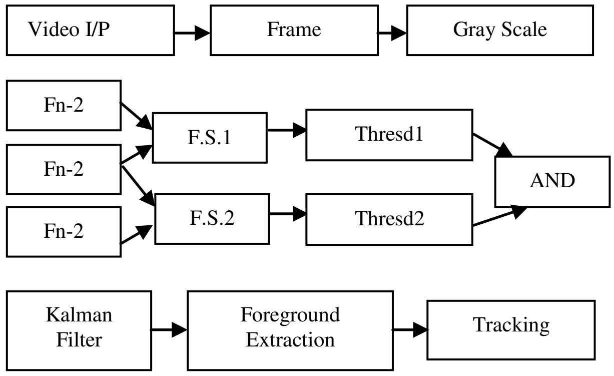



Figure 1: Dual Frame difference System Stages

## B. Visual background & Template Matching Method

A method for accurate detection & tracking of human motion in an image sequence by combining visual background extraction technique [5] for motion detection and illumination intensity minimization based on template matching technique for object tracking is proposed. The system will take care of illumination changes in real images caused due to automatic exposure alterations of the camera, adjustment of light cause irradiance, appearance of shadows or movement of the tracked objects. The objective of research is to provide a robust and improved method to find the moving objects in the video frame as well as to track them. The proposed method is effective in reducing the number of false alarms that may be triggered by a number of reasons.

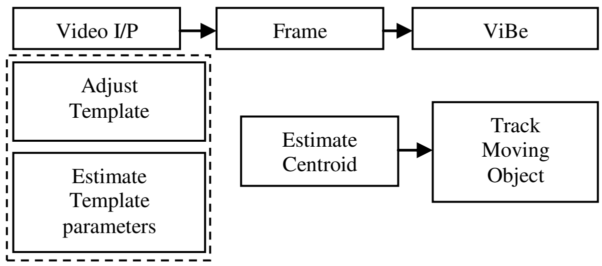



Figure 2: Visual Background & Template Matching System Stages.

Figure 2 shows the different stages involved in the system. First, image frames are taken out from video, and then, image motion is predicted using visual background extraction model to which morphological operation is applied to fill the gaps and nullify the small blobs present due to error. Extracted foreground is then tracked by rectangle for which centroid calculation is done. Simultaneously, illumination insensitivity hyperplane method [12] for template matching is also applied to perfectly track the object.

### C. Optical Flow Method
A framework for the detection and tracking of moving people using optical flow method is proposed. In order to detect foreground objects, first, optical flow algorithm is applied. This is then combined with the illumination insensitive template matching method to accurately track the object for visual surveillance system. Steps below shows the different stages involved in the proposed system. The algorithm to learn a system for motion detection & tracking can be briefly quoted as follows:

1. Extract the frames from the given video sequence.
2. Take last & current frame and calculate the optical flow in between the 2 frames.

$$Ix\,(x_i, y_i) \cdot dx\,(x_i, y_i) + Iy\,(x_i, y_i) \cdot dx\,(x_i, y_i) = -I_t\,(x_i, y_i) \quad \text{... (1)}$$

3. Compute the features as optical flow output.
4. Compute the value of ($\beta t$, $\Upsilon t$) for illumination insensitive iteration cycle. Contrast and brightness variations are represented by illumination compensation parameters $\beta$ and $\gamma$.

$$(\beta_t,\, \gamma_t) = \underset{(\beta,\gamma)}{argmin} \sum_{x\epsilon r}[\beta f(g(x,\mu(t)),t) + \gamma - f(x,t_0)]^2 \quad \text{... (2)}$$

5. Combine the value of step 3 and step 4 to track the given object in a video sequence.

### D. AdaBoost Classifier Using Sparse Matrix & 45$^o$ rotated Haar features
A simple yet effective and efficient tracking algorithm based on features extracted from a multiscale image feature space through data-independent basis is used. The model employs non-adaptive random projections that conserve the arrangement of the image feature space of targets.

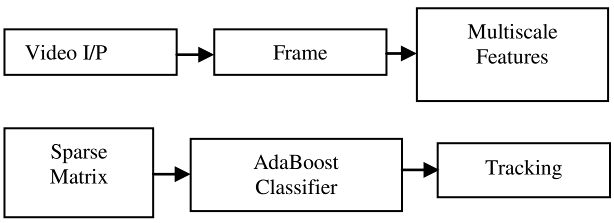



Figure 3: Visual Background & Template Matching System Stages.

A sparse measurement matrix is build to extract efficiently the features for the appearance model. Sample images are compressed of the foreground target and the background using the similar sparse measurement matrix.

The tracking task is formulated as a binary classification via an AdaBoost classifier with update in the dense domain. A coarse-to-fine search strategy is adopted to further reduce the computational difficulty in the recognition procedure. The compressive tracking algorithm runs fast and performs favourably against other state-of-the-art methods on challenging sequences in terms of efficiency, accuracy and robustness. Figure 3 shows the different stages involved in the proposed system. First, image frames are taken out from video, and then, image features are calculated by using sparse matrix. Extracted feature vectors are then given to the AdaBoost classifier for the tracking purpose.

## 4 Motion Database
Any algorithm must be carefully validated on appropriate data sets to show its potential for real-world applications. The data sets should be chosen to be representative for the applications under consideration. Experimentation work for the algorithms described is carried out on different databases. First, different publically available motion database namely "Weizmann database", "KTH database", "CMU Graphics Lab Motion Capture Database", "CRCV database" and different motion videos. Second, "Self database" developed by the author; this database contains 86 sequences having five classes of actions in indoor (namely Jump, Run, Walk, Side, Handwaving) & four classes of actions at outdoor (namely Jogging, Walking, Handwaving, Jumping) respectively performed by 19 different subjects

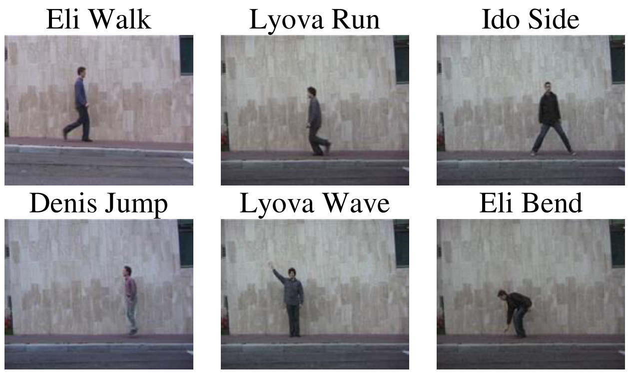



Figure 4: Images from the Weizmann data set.

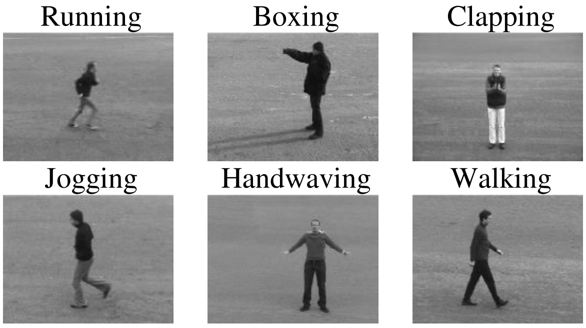



Figure 5: Images from the "KTH data set".

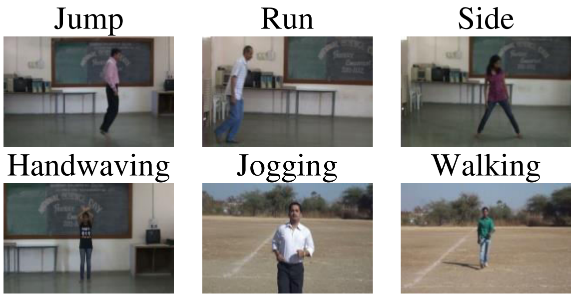



Figure 6: Images from the "Self data set".

## 5 Results & Discussions

The algorithms are implemented in MATLAB 7.10.0 (R2010a) on Windows 7.0 platform. The PC is equipped with Pentium corei3, 2.13 GHz processor and 4 GB RAM Memory. The experimental result of the dual frame background subtraction algorithm is shown in figure 7.

### A. Dual Frame Differencing Method

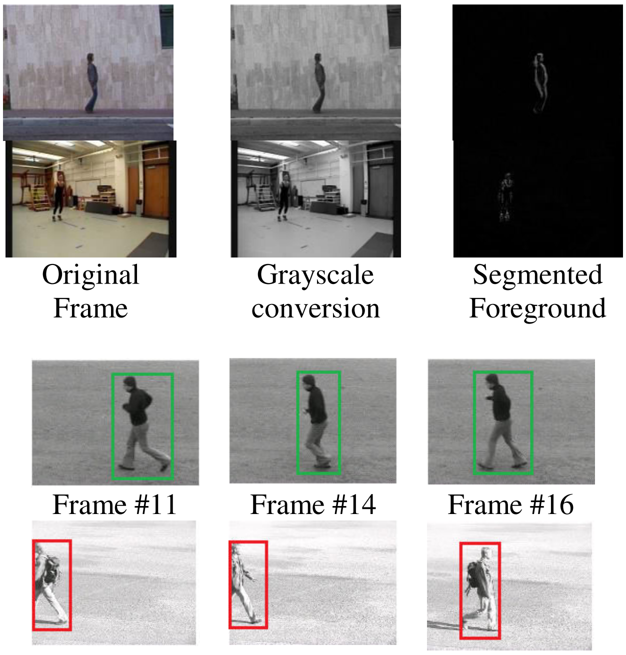



Figure 7: Experimental result of the background subtraction model on Database videos.

The computational difficulty for computing each difference is only $\tau$ ($|I_{t-} B_t|$) for the local context region of M x N pixels, thereby resulting in a fast method (100 FPS in MATLAB on an i3 machine). More importantly, the algorithm achieves robust results as shown in figures above. The summary of processing time and average speed of processing the video is given in the table 1 below.

Table 1: B.G. Subtraction model performance.

| S.N. | Sequence | Resolution (Pixels) | Frame Number | Overall Average Time (s) | Average Speed (frame/s) |
|---|---|---|---|---|---|
| 1 | Daria Jump | 180 x 144 | 67 | 48 | 1.373 |
| 2 | ADA Handwaving | 1280 x 720 | 30 | 12.209 | 2.457 |
| 3 | Girl | 320 x 240 | 20 | 3.688 | 5.422 |
| 4 | RDK Run | 1280 x 720 | 54 | 9.56 | 3.457 |

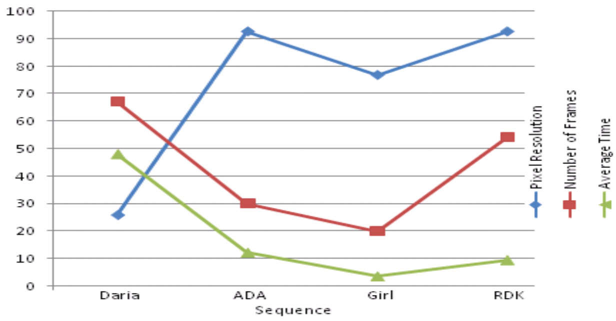



Figure 8: Graphical representation of performance.

The figure 8 shows the representation of overall average time for the frames to be processed. As seen in the graph increase in pixel resolution hardly affects the processing speed of the method. Moreover, this method takes less than 1s to compute the image difference, while the image scan takes less than 0 to 5s depending on the specific configuration.

Figure 9: Image from database sequence, grayscale conversion, "Adaptive mixture models" [13], "Haritaoglu" [14]. "Romain" [15], and described technique. (Read row-wise).

## B. Visual background & Template Matching Method

The Visual background & Template Matching Method results are compared with the other background subtraction processes as depicted in figure 9. In few numbers of video sequences the gradient mask is not accurately observed due to radiance variation & other problems.

Table 2: Tracking window variation for change in element σ performance.

| Sr. No. | Run | Scale Parameter σ | Target Scale |
|---|---|---|---|
| 1 | Run 1 | 0.25 | 0.9990 |
| 2 | Run 2 | 0.5 | 0.9994 |
| 3 | Run 3 | 0.75 | 0.9940 |
| 4 | Run 4 | 1 | 0.9866 |
| 5 | Run 5 | 1.25 | 0.9725 |
| 6 | Run 6 | 1.5 | 0.9687 |
| 7 | Run 7 | 1.75 | 0.9523 |
| 8 | Run 8 | 2 | 0.9412 |

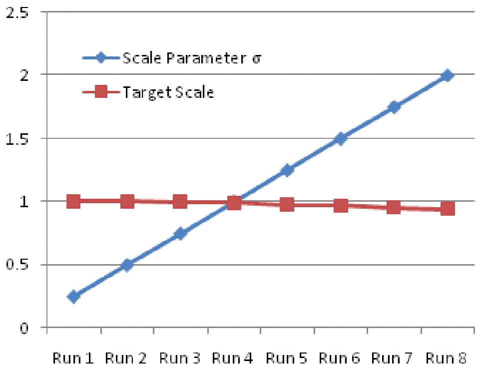


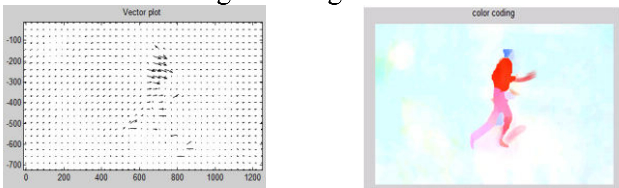

Figure 9: Graphical representation of table 2.

The target tracking efficiency is increased through the weighted context region of the last target location. Furthermore, the size of the object also changes with time depending on whether the object is approaching nearer to the camera or going away from the camera. Hence, the scale element σ should be updated accordingly. Table 2 below represents the variation of tracking rectangle when the element σ is changed for various values.

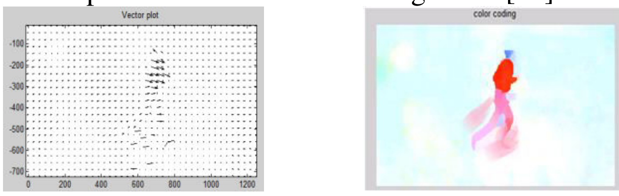
## C. Optical Flow Method

The experimental results of the described optical flow algorithm, Lucas & Kanade [16] method & Thomas Brox [17] techniques are shown in figure.

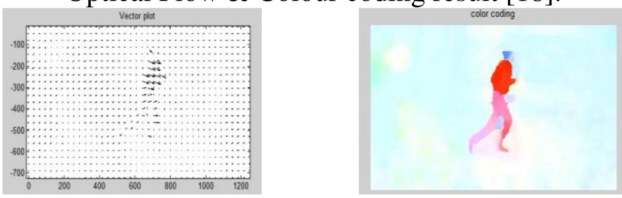
"Original Image Frames"




Optical Flow & Colour coding result [17].




Optical Flow & Colour coding result [18].




Optical Flow & Colour coding result described technique

Figure 9: Experimental result of the different optical flow techniques on Database videos.

Table 3: Execution Time, End Point Error (EPE) & Angular Error (AE) comparison for different methods.

| S.N. | Method | Time (s) | Avg. EPE | Avg. AE |
|---|---|---|---|---|
| 1 | Described method | 7.72 | 0.31 | 2.4 |
| 2 | Lucas Kanade | 18 | 0.86 | 4.9 |
| 3 | Thomas Brox | 21 | 0.63 | 5.3 |

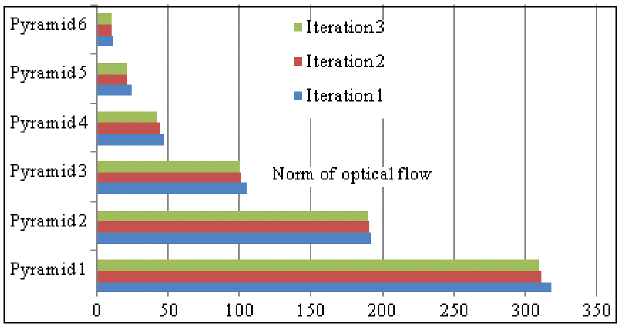



Figure 10: Graphical chart representation of norm of optical flow for different levels of pyramid for 3 Iterations (RDK sequence).

The optical flow performance is regularly measured in terms of angular error. This factor can be estimated by taking the product (dot) of the flow vectors & dividing this value with the product of their lengths, & then measuring the inverse cosine of this function. To attain the best average End-Point Error (EPE), regularization parameter $\lambda$ is selected amongst a set of various candidate values.

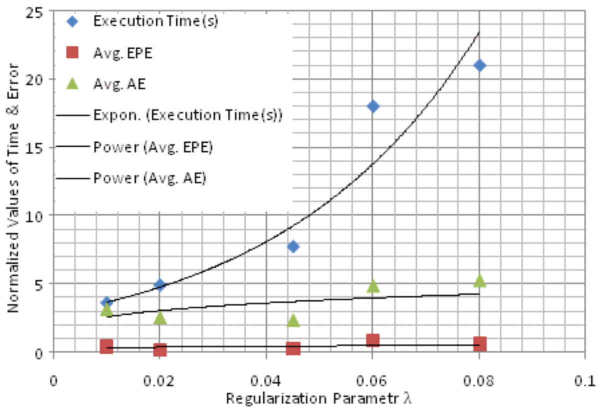



Figure 11: Scatter plot showing the variation of regularization parameter λ & its effect on Execution time & Errors (Average AE & EPE).

From the above chart of average angular error & end point error for the various values of λ, it is observed that error decreases gradually & then increases. The graph is in linear scale. From the chart it can be concluded that value of regularization parameter λ = 0.02 is a good choice. Various aspects of optical flow approaches & illumination insensitive tracker is verified & implemented using suitable algorithms. It is observed that the algorithm that integrates the illumination insensitivity tracker with optical flow model gives the best result in term of the error measurement & execution time.

## D. AdaBoost Classifier Using Sparse Matrix & 45⁰ rotated Haar features

45⁰ Tilted Haar features are simple & fast-to-estimate features. A characteristic of this set is defined by a filter bank that calculates the grey level variation among two defined areas i.e. black & white. In the Occlusion sequence, sparse matrix algorithm, L1 minimization [18], & fragment tracker [19] perform better, as shown in Figure 12, because these methods take partial overlap into account. The fragment tracker method is capable of handling occlusion via the part based representation with histograms. For the Occlusion sequence, sparse matrix tracker performs well particularly when partial occlusion occurs.

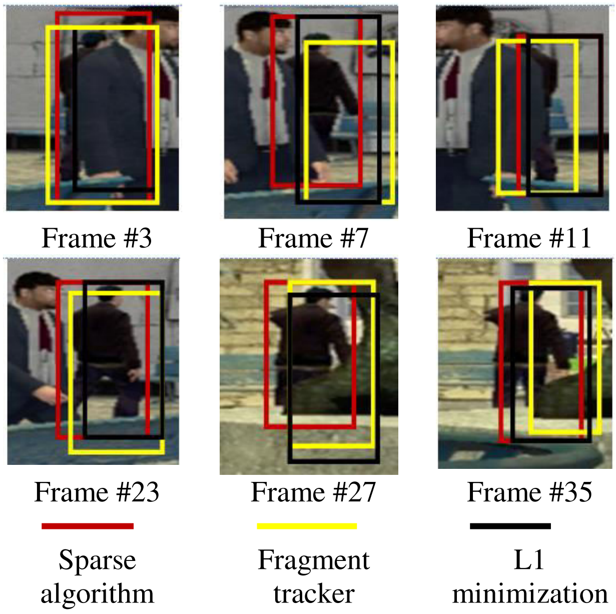



Figure 12: Tracking result of 3 different tracking model methods on Occlusion of two men walking.

When the object experience unexpected motion, it is difficult to guess its pixel position. Also, it is quite tough to properly update weight of the classification models. Performance evaluation of tracking objects in the scene normally involves estimating difference between the predicated & the ground truth locations.

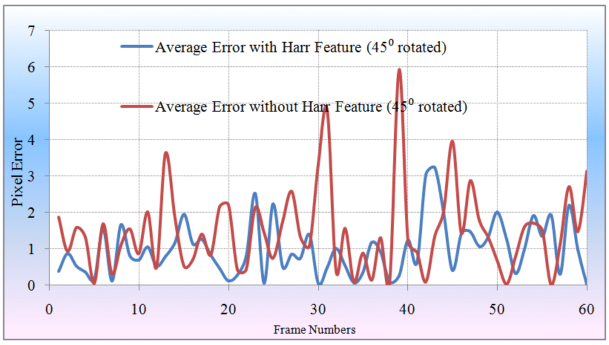



Figure 13: Error Estimation of targets (two men walking) tracked by different methods undergoing illumination, pose & Scale variations.

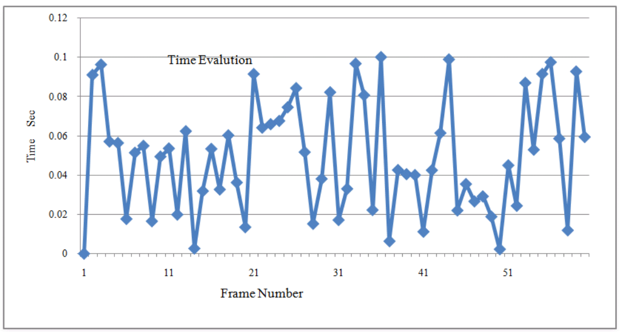
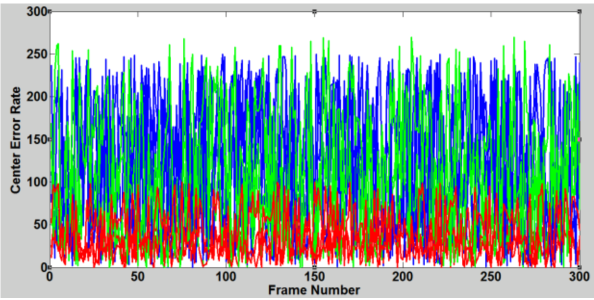



Figure 14: Frame processing Time of targets tracked.

Figure 13, reviews the output of average tracking errors. Sparse matrix measurement using AdaBoost classifier technique attains least errors. Figure 14 shows the algorithm processing time required for tracking the object as it takes the size and pose of the target object into account.

The average processing time per frame is 0.0921 second, that is, the sparse matrix algorithm handles more than 11 frames per second. It is the adequate speed for the real time video surveillance applications. The algorithm qualitative analysis can be estimated through another parameter known as centre error rate. CER shows that how much the tracking window is deviated from the centre of the foreground target. Figure 15 displays the graph of the average centre error of test sequences. From the graphical representation it is clear that the spare AdaBoost algorithm is efficient with most of the database videos.




Sparse algorithm L1 minimization
Fragment tracker

Figure 15: Centre Error Rate Vs Frame for three methods

## 6 Conclusion

In this paper, we have analyzed four methods of human motion segmentation & tracking. First method is the use of three frame subtraction which overcomes the disadvantage of simple frame differencing algorithm. This method achieves accurate segmentation of moving pixels. The second method is combination of visual background extraction and illumination insensitive template matching technique. Experimentation result of the model shows its effectiveness against motion changes due to camera movement, and cluttered background. Method also reduces false alarm triggering due to noise and illumination changes. Third method of optical flow parameters extraction in combination with illumination insensitive template matching method precisely tracks the object for visual surveillance system. The method is a solution to the problem of efficiently detection and tracking of moving people with ease and less processing time. The last method is the technique of sparse measurement matrix to extract the features for the appearance model & learn these features using AdaBoost classifier which is tested on large video database to show its effectiveness for the tracking & the low centre error rate. All the analyzed methods add to the study of human motion tracking system. Our future direction of study includes the study of surveillance system for the prediction of the speed & distance of the object in the scene. This would be useful in vehicle safety & sports activity analysis & applications.

*References:*

[1] Wei Zhong, Huchuan Lu et al., “Robust Object Tracking via Sparse Collaborative Appearance Model”, *IEEE Transactions on Image Processing*, Vol. 23, No. 5, May 2014.

[2] Furat N. Tawfeeq, “Real Time Motion Detection in Surveillance Camera Using MATLAB”, *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 3, Issue 9, September 2013.

[3] Fan Yang et al., “Robust Superpixel Tracking”, *IEEE Transactions on Image Processing,* Vol. 23, No. 4, April 2014.

[4] Dong Wang et al., “Online Object Tracking with Sparse Prototypes”, *IEEE Transactions on Image Processing,* Vol. 22, No. 1, January 2013.

[5] Olivier Barnich et al., “ViBe: A universal background subtraction algorithm for video sequences”, *IEEE Transactions on Image Processing,* 20(6):1709-1724, June 2011.

[6] Peng Cui et al., “A Matrix-Based Approach to Unsupervised Human Action Categorization”, *IEEE Transactions on Multimedia*, Vol. 14, No. 1, February 2012.

[7] Yuji Tachizaki et al., “A DCT Coefficient Sign-Based Background Model For Moving Objects Detection From Motion JPEG Coded Movies”, *International Symposium on Intelligent Signal Processing and Communication Systems* December 7-9, 2009.

[8] Tudor Barbu, “Novel Approach For Moving Human Detection And Tracking In Static Camera Video Sequences”, *Proceedings Of The Romanian Academy, Series A,* Volume 13, Number 3Pp. 269–277, 2012.

[9] Tal Ben-Zvi et al., “Intruder Detection: An Optimal Decision Analysis Strategy”, *IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews*, Vol. 42, No. 2, March 2012.

[10] Jianguo Lu et al., “A Detection-Aided Multi-target Tracking Algorithm”, *International*

*Conference on Machine Vision and Human-machine Interface*, 2010.

[11] Li Ying-hong, et al., “An intelligent tracking technology based on Kalman and mean shift algorithm”, *IEEE Second Intl. Conf. on Computer Modelling and Simulation*, 2010.

[12] Christoph Gr¨aßl et al., “Illumination Insensitive Template Matching with Hyperplanes”, *Pattern Recognition, 25th DAGM Symposium*, Magdeburg, Germany, pages 273-280, September 2003.

[13] C. Stauffer and W. Grimson, “Adaptive background mixture models for real-time tracking,” *in Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 2, pp. 246–252, 1999.

[14] I.Haritaoglu, D.Harwood, L.Davis, “A Fast Background Scene Modeling and Maintenance for Outdoor Surveillance”, *ICPR, pp*.179-183, Barcelona, 2000.

[15] Romain Marie, Alexis Potelle, “Dynamic Background Subtraction Using Moments”, *In Proceedings of the IEEE International Conference on Image Processing*, pp. 2369 - 2372, 2011.

[16] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. *European Conference on Computer Vision (ECCV)*, 2004.

[17] Jean-Yves Bouguet, “Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the algorithm”, *Intel Corporation Microprocessor Research Labs*, 2000.

[18] X. Mei and H. Ling, “Robust visual tracking using L1 minimization,” *In Proc. IEEE Int. Conf. Comput. Vision, Sep*.–Oct. pp. 1436–1443, 2009.

[19] A. Adam, E. Rivlin, and I. Shimshoni, “Robust fragments-based tracking using the integral histogram”, *In Proc. IEEE Conf. Computer Vision Pattern Recognition*, June. pp. 798–805, 2006.