Asian Journal of Information Technology 15 (5): 846-854, 2016
ISSN: 1682-3915

Capturing Moving Objects in Video Using Gabor and Local Spatial Context Model

G. Jemilda and S. Baulkani
[1]Faculty of Computer Science and Engineering,
Jayaraj Annapackiam CSI College of Engineering, Nazareth, Tamil Nadu, India
[2]Faculty of Electronics and Communication Engineering,
Government College of Engineering, Tirunelveli, Tamil Nadu, India

Abstract: Object tracking is the process of tracking a moving object in video over time using a camera. It is an important task within the field of computer vision. It has a wide variety of applications in computer vision such as video compression, video surveillance, vision-based control, human-computer interfaces, medical imaging, augmented reality and robotics. In this study, the object is tracked in video using the following steps. First, the input given is a video which is divided into frames. For each frame the features are extracted by Gabor filter which is used to identify the edges clearly. By this process, the object can be identified in the frame. In order to track the object in video, spatial context model is used. It checks the difference between the frames and keeps track of the object. The spatial correlation not only tracks the object but also reduces the time complexity. If there is not much difference between the first frame and the third frame then the same value will be on the sec frame. Thus, the sec frame will not be processed. The proposed method can produce an accurate result.

Key words: Tracking, gabor filter, spatial context model, video, object

INTRODUCTION

Object tracking is used to track an object over a sequence of images. In general, object tracking is a challenging problem. Difficulties in tracking objects can arise due to rapid object motion, changing appearance patterns of the object and the scene, non-rigid object structures, object-to-object and object-to-scene occlusions and camera motion.

Tracking is usually performed in the context of higher-level applications that require the location and/or shape of the object in every frame. The proliferation of high-powered computers, the availability of high quality and inexpensive video cameras and the increasing need for automated video analysis has generated a great deal of interest in object tracking algorithms.

There are three key steps in video analysis: detection of interesting moving objects, tracking of such objects from frame to frame and analysis of object tracks to recognize their behaviour. Additionally, it provides input to higher level vision tasks such as 3D reconstruction and 3D representation. It also plays an important role in video database such as content-based indexing and retrieval. Tracking moving objects in space is important for the maintenance of spatio-temporal continuity in everyday visual tasks. In Multiple Object Tracking (MOT) task, participants track a subset of moving objects with attention over an extended period of time. The ability to track multiple objects with attention is severely limited.

Literature review: Chien *et al*. (2002) investigatedan efficient video segmentation algorithm that can handle situations with any object motion, uncovered background and shadow effect. A background registration technique is used to construct the reliable background information from the video. A morphological gradient operation is used to filter out the shadow area while preserving the object shape and to achieve the real-time requirement for many multimedia communication systems. This method avoids the use of computation intensive operations.

Doghmane and Bouden (2007) inferred that the motion segmentation of image sequences is based on visual motion perception. They generally recognized that the analysis of moving objects proceeds in four stages: The first is the detection of variation in intensity over time in the environment. The sec is the segmentation of moving areas and objects. The third is the estimation of motion parameters. The fourth one is the 3D motion

Corresponding Author: Emilda Faculty of Computer Science and Engineering, Jayarajannapackiam Csi College of Engineering, Nazareth,

interpretation. They also dealt with detection and region-based segmentation methods. These methods help to estimate motion parameters. The comparative study using deterministic and stochastic modeling (images difference, maximum likelihood detector and Markov random field model) is used to detect the moving objects masks. The frame difference method is better in computing time and gives noisy masks of moving area but in the case of synthetic sequences without noise it is the better one. The likelihood detection and Markov model detection methods realized the best compromise in sensitivity to noise and cost of calculation. They take enough computing time and give the best and higher promise mask quality.

Prakash *et al.* (2012) proposed a novel object tracking method using Daubechies Complex Wavelet Transform (DaubCxWT). This transform is used to track the object from video sequences because of its approximate shift-invariance nature. Tracking of object in the first frame is done by computing the daubechies complex wavelet coefficients corresponding to the object of interest and then matching energy of these coefficients to the object neighbourhood in daubechies complex wavelet domain to perform tracking in the next consecutive frames. The proposed method needs only complex wavelet coefficients for tracking and hence, it is simple in implementation and tracks object efficiently.

Rosenberg and Werman (1998) studied a real time system for image registration and moving object detection. The algorithm is based on describing the displacement of a point as a probability distribution over a matrix of possible displacements. A small set of randomly selected points is used to compute the registration parameters. Moving object detection is based on the reliability of the probabilistic displacement of image points with the global image motion.

Cohen and Medioni (1999) described how to detect and track the object in video. The proposed method relies on a graph representation of moving objects which allows deriving and maintaining a template of each moving object by enforcing their temporal coherence. This inferred template along with the graph representation allows to find object trajectories as an optimal path in a graph. The proposed tracker allows dealing with partial occlusions, stopping and going motion in very challenging situations. This method shows the results on a number of different real sequences.

Roy *et al.* (2010) explained how to detecta moving object using a webcam in a commodity laptop without special hardware for high speed image processing. The moving object detection provides clear edges of a detected object. This method is used to reduce the average delay upto 45.5% and to decrease the memory consumption up to approximately 14%. Pan and Hu (2007) experimented a Content Adaptive Progressive Occlusion Analysis (CAPOA) algorithm to handle occlusionsrobustly. The context information and motion properties are taken as the reference target. The CAPOA algorithm makes much clear distinction between the target and the occluder. To perform Variant Mask Template Matching (VMTM), the non-occluded portion of the target is used to align the target from the erroneous location to its true location. The non-occluded portion of the target serves as a benchmark for the alignment of the target location. Using this technique, the object tracker is found to be much more robust against various types of occlusions.

Fatichah and Widyanto (2008) suggested that the human object detection is a very big issue because of the many applications of human object detection system, visual search engine and intelligent vehicles. In real application human object detection requires high accuracy and fast testing time. They also proposed that K-Boosting method considers quadratic kernel as base classifiers for boosting and it gives high accuracy and fast testing time.

Ercan *et al.* (2007) presented a sensor network approach for tracking a single object in a structured environment using multiple cameras. He tracked only the target object and treated others as occluders. The tracker is provided with complete information about the static occluders and some prior information about the moving occluders. One of the main contributions of his research is developing a systematic way to incorporate this information into the tracker formulation. Using this method, the number of cameras used, the number of occluders present and the accuracy of tracking are found out easily.

Xue and Ling (2011) introduced Sparse Representation (SR) into visual object tracking. Here, the object is represented by multiple templates. It utilizes only fixed object template. For ever changing object appearance, it performs poor. A set of trivial templates are used to find occlusion and corruption problems.

Ross *et al.* (2008) proposed Incremental Visual Tracker (IVT) algorithm which represents the object with a set of sub-space templates with the sequential appearance variations. To improve the efficiency, sub space is updated with multiple samples. Adam *et al.* (2006) integrated inner structure of the image and handled partial

occlusion. For significant appearance variations, fragment based tracker (Frag Track) algorithm is difficult to track objects. This algorithm fails quickly as the object is partially occluded due to optical flow because it is sensitive to occlusion.

Babenko *et al.* (2009) handled the unreliable labelled positive and negative data using Multiple Instance Learning (MIL) algorithm to soften the drift problem. They considered only the target appearance but not considered the relationship between the background and its target. Kalal *et al.* (2010) introduced detection module into tracking process using P-N algorithm. Here, the appearance model is corrected by the detector and the target is recaptured even if the target has moved out of view.

Kwon and Lee (2010) used Visual Tracking Decomposition (VTD) algorithm. This algorithm failed to track the target when abrupt motion happens. In such cases, the temporal information becomes unreliable but the spatial information is still discriminative. Zhong *et al.* (2012) introduced a Sparsity-based Collaborative appearance Model (SCM) which exploits both holistic and local information. They proved that the drifting is easy in significant pose variation and also the tracking accuracy is low. The result is not satisfied, if partial occlusion presents. This algorithm performs better in tracking the object after drastic illumination changes and performs well when the target undergoes out-of-plane rotation.

MATERIALS AND METHODOS

This part briefly explains about the methods used for tracking the objects. Firstly, the video is given as the input which is divided into frames. The frames are then filtered and classified using Gabor filter and weak classifier. Finally, the filtered frames and classified frames are combined to form a video which highlights the detected object. This process was shown in Fig. 1.

Input video: The input video is chosen as it consists of the objects to track. The objects in the video must be moving since, our study is to capture the moving objects in the video.

Divide video into image frames: Video technology is used for electronically capturing, recording, processing, storing, transmitting and reconstructing a sequence of

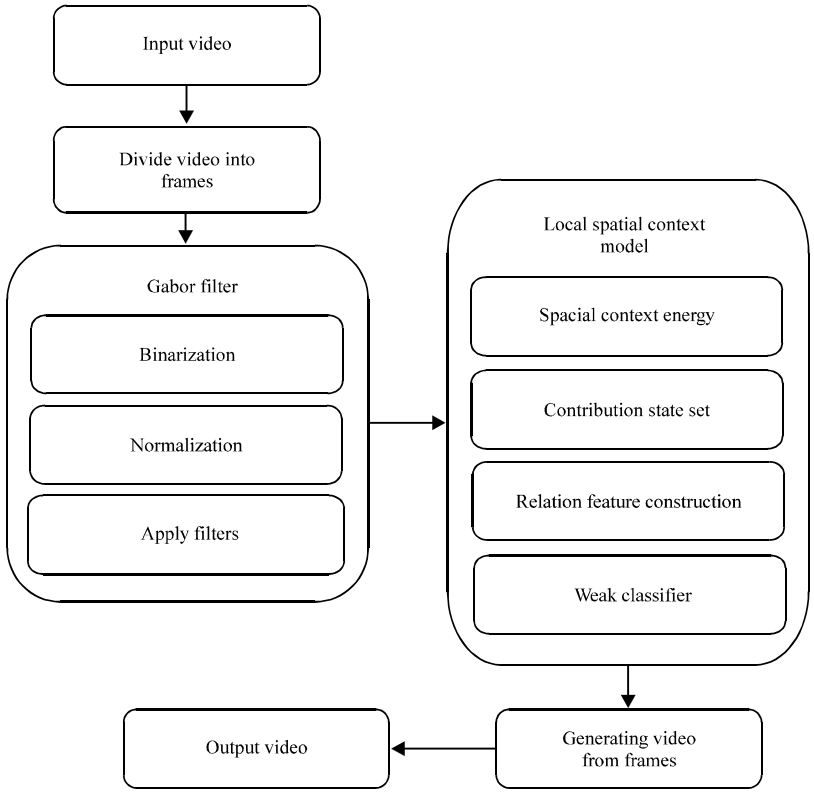

Fig. 1: Arechitecture diagram

still images representing scenes in motion. The number of still pictures per unit time of video ranges from six or eight frames per sec to 120 or more frames per sec. These frames are saved and processed consequently. Here, the video is divided into number of frames for example the video of upto 10 sec is break down into 80-85 frames.

Gabor filter for feature extraction: A Gabor filter is a linear filter used for edge detection. Frequency and orientation representations of Gabor filters are similar to those of the human visual system and they are helpful for extracting useful features from an image. The Gabor space is very useful in image processing applications particularly in recognition. Relations between activations for a specific spatial location are very distinctive between objects in an image. Further more important activations can be extracted from the Gabor space in order to create a sparse object representation. The output of Gabor filter is used as the recognition process. Each Gabor filter is defined as:

$$g(x,y:\gamma,\theta,\varphi,\sigma)=\exp\left(-\frac{x_1^2+\gamma^2y_1^2}{2\sigma^2}\right).\exp i.\left(\frac{2\pi x_1}{\gamma+\varphi}\right) \quad (1)$$

In this study, the Gabor filter is applied at orientations 0, $\pi/4$, $\pi/2$, $3\pi/4$ to an object. Then the response output is selected at some sampling points to form a feature vector. The number of pixels are counted that response maximally at θ_k and then a n-dimension vector is formed where n is the number of θ_k. This vector is defined as orientation map. Though orientation map can describe the orientation distribution of an image's edge at different angle, it also lost the position information of each pixel in the image. But, the position information is important. So, Gabor orientation map feature is not suitable for object recognition.

Gabor dominant orientation matrix is used as recognition feature. For an X×Y image V, the Gabor dominant orientation matrix m is having X×Y dimensions. The value of m (x, y) in dominant orientation matrix is valued by comparing the response output l_0, l_1,...l_{n-1} of image V convolved by the set of Gabor filters. If V(x, y) obtain the maximum response output at an orientation θ_k, then m (x, y) is assigned the value k. Thus, each element in matrix m is valued between 0 and n-1. So, an X×Y dimension vector is got. The following are the steps to extract dominant orientation matrix as recognition feature.

Binarization: Binarize the low resolution gray character and find the circumscribing frame of the character.

Normalization: Extend the circumscribing frame to a 32×32 normalized image.

Apply filters: Apply Gabor filters whose orientation is $\theta = 0, \pi/18,..., 17\times\pi/18$ to the normalized imageand then obtain 18 response outputs l_0 l_1,...l_{17}. Provide the Gabor dominant orientation matrix m where m is defined as: M (x,y) = k where max {I(x,y,)} for all k.

Local spatial context model

Spatial context energy: In multiple instance boosting, each selected weak classifier corresponds to each weak correlation. The selected correlations are combined together to evaluate the spatial context energy (namely the spatial energy function) of a candidate state. The spatial context energy is expressed as:

$$U\left(Z_t\middle|f^r(.),o_t\right)\propto-\sum_j g_j^t\left(X^t\right) \quad (2)$$

Where:

$f^r(.)$ = The contributor state set
Z_t = The target candidate state
O_t = The corresponding observation
$g_j^t(x^t)$ = The jth selected weak classifier at time t
X^t = The corresponding relation feature of the candidate state

Contributor state set f': The patch at the key point is calledcontributor and the key points around the target is generated in the rectangle centered atthe target center with the width $r_e\times w$ and height $r_e\times h_e$ where r_e is the enlargement factor. Set the enlargement factor as $r_e\in(0.5, 1.6)\times w$ and h are the width and height of the target in the current frame. If the extracted candidate key points are more than the required ones, randomly select some of them to be the final key points and use them to generate the contributors but if they are inadequate, randomly generate some other points in the rectangle to supplement them.

Relation feature construction: To incorporate the structure information of the target, partition the regions of the target and contributors into a predefined number of blocks. Let, $N = n_1\times n_2$ be the predefined number of partitioned blocks where n_1 and n_2 are the partitioned numbers of blocks in the row and column respectively. The structure information is

integrated by modeling the relationships between blocks. The weak relation function between two blocks is defined as:

$$d_f\left(b_p(.),b_q(.)\right)=\sum_{(i,j)\in b_p(.)} 1(i,j)-\sum_{(i,j)\in b_p(.)} 1(i,j) \quad (3)$$

Where:

I (i,j) = The image I with
I×j = The pixel value
b_p and b_q = Two blocks

The structure information comes from two parts: one is the mutual-pairwise features between the corresponding blocks of the target and the contributors and the other one is the self-pairwise features between the inner blocks of the target itself. Specifically, the self-pairwise and mutual-pairwise feature pools are constructed as:

$$F_s=\left\{d_f\left(b_i(z),\left(b_j(z)\frac{|i,j=1,...N}{i\neq j}\right)\right\}\right. \quad (4)$$

$$F_m=\left\{d_f\left(b_i(z),\left(b_j\left(f_k^r(z)\right)\right)\right|\begin{matrix}i,j=1,...N\\k=1,...m_c\end{matrix}\right\} \quad (5)$$

Where F_s be the self-pairwise feature pool and F_m be the mutual-pairwise feature pool.

Weak classifier: The Gaussian Mixture Model (GMM) is used to estimate the posterior probability of the weak classifier that is:

$$P\left(x_j|y\right)=\sum_{i=1}^{k}\omega_{ij}(y)\eta\left(x_j,\mu_{ij}(y),\sigma_{ij}(y)\right) \quad (6)$$

Where:

k = The number of Gaussian models
ω_{ij} (y), μ_{ij} and σ_{ij} (y) = The weight, mean and variance of the ith Gaussian model of the sample with label y

The probability density function of the gaussian distributionis:

$$\eta(x,\mu,\sigma)=\frac{1}{\sqrt{2\mu\sigma}}\exp\left(-\frac{\left(x-\mu^2\right)}{2\sigma^2}\right) \quad (7)$$

The positive and negative samples have equalprior probability in the task, i.e., P(y = +1) = P(y = -1). So, itis easy to get the continuous Bayesian weak classifier based on the GMM, i.e., Gaussian models to get the matching measure is given as:

$$0H_{kj}=\omega_{kj}(+1)\left(\prod_{i|y^{(i)}=1}\eta\left(u_j^{(i)},\mu_{kj}(+1),\sigma_{kj}(+1)\right)\right) \quad (8)$$

Where is the jth dimension of u (i). Let be the symbol indicating whether the kth Gaussian model matches the j th dimension of the feature. The mean and variance of the matched Gaussian model will be updated as:

$$\mu_{kj}(+1)=(1-\lambda)\mu_{kj}(+1)+\frac{\lambda 1}{n}\sum_{i|y^{(i)}=1}uj^{(i)} \quad (9)$$

$$\sigma_{kj}^2(+1)=(1-\lambda)\sigma_{kj}^2(+1)+\lambda \left(\frac{1}{n}\sum i\left|y(i)=1\left(x_j^{(i)}-\mu_{kj}(+1)^2\right)\right.\right)^{\frac{1}{2}} \quad (10)$$

Where, λ is the updating step. Otherwise, the mean and variance of the unmatched ones will not be updated. Finally, all the weights are updated and the updating rule of the negative samples is similarly defined.

Generating video from frames: To generate a video from a set of sequences or set of frames, start with the number zero. This will work as long as the sequence is unbroken once it starts. If there are gaps due to the stills, renumbering may be necessary to fill the gaps. If the frame rate of the resulting video is 3 frames per sec then each still can be seen for a short period of time. The rescaling of the picture is necessary to obtain the desired resolution so that the size of the resulting video is managed.

Output video: The filtered and classified frames are combined to form the output video.

Experimental images: The data sets used for experimentation is shown below. Only the frames which are taken from the video are shown. Figure 2 shows various movements of the moving object and Fig. 3 shows various movements of occlusion in the object. The captured moving object is shown in Fig. 4 and the tracked occluded object is shown in Fig. 5. Figure 6 and 7 show the frames of various movements of multiple object and the detected multiple objects.

Fig. 2: Various movements of the moving object

Fig. 3: Various movements of occlusion in the object

Fig. 4: Captured moving object

Fig. 5: Tracked occluded object

Fig. 6: Various movements of multipleobject

Fig. 7: Detected multiple objects

RESULTS AND DISCUSSION

To evaluate the performance of the proposed system, compare with the state-of-the-art tracking algorithms, i.e., IVT [9], SR based (L_1) [8], PN [12], VTD [13], MIL [11], frag track [10] and SCM [14]. The challenges of these video include abrupt motion, occlusions, pose variations, scale variations and complex backgrounds. To examine the effectiveness, it is compared with various performance metrics. The various performance metrics used to compare the effectiveness are Average Center Error in Pixels (ACEP), overlap rate and computational time taken.

Average center error in pixels: It shows the error rate for each frame:

$$\text{Average center error} = \text{area}\left(R_T \cup R_G\right) - \text{area}\left(R_T \cap R_G\right) \quad (11)$$

Where:
R_T = Tracking result of each frame
R_G = Corresponding ground truth

Overlap rate: It shows the overlapping rate of each frame:

$$\text{Overlap rate} = \frac{\text{area}\left(R_T \cap R_G\right)}{\text{area}\left(R_T \cap R_G\right)} \quad (12)$$

Where:
R_T = Tracking result of each frame
R_G = Corresponding ground truth

Computational time: It is used to find the moving object identification time in each frame:

$$\text{Computational time} = \text{End time} - \text{Start time} \quad (13)$$

Where:
End time = Total time taken for execution
Start time = Starting time

To analyse the performance of the proposed system, it is compared with the above mentioned performance metrics. The performance metrics are tabulated and given in Table 1-3.

To know the working of the proposed system effectively, the performance values are plotted in the graphs and are shown in Fig. 8 and 10.

Figure 8-10 show the various performance metrics with different methods for objects tracking in varying condition.

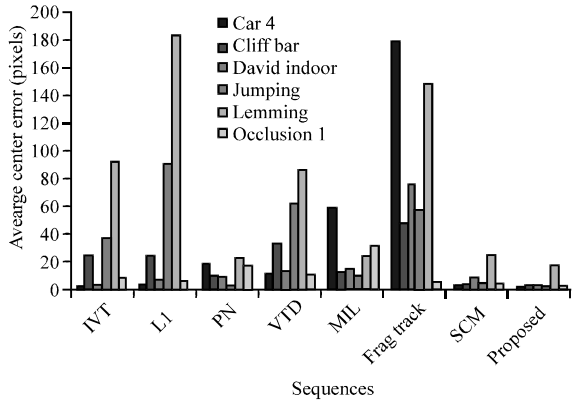

Fig. 8: Average center error

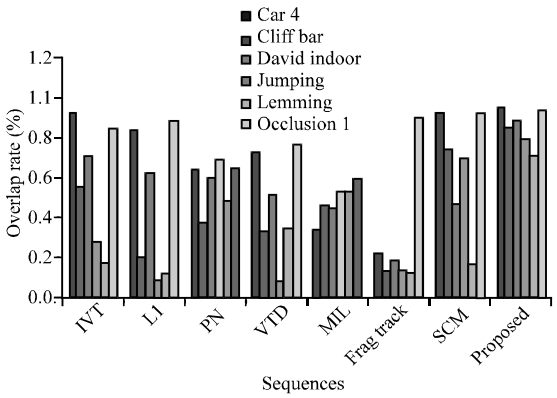

Fig. 9: Overlap rate

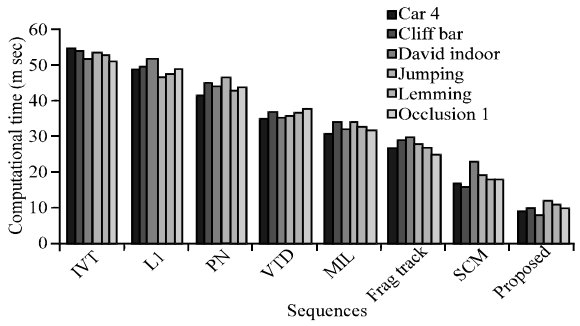

Fig. 10: Computational time

Table 1: Average center error (Pixels)

Sequences	IVT	L_1	PN	VTD	MIL	Frag track	SCM	Proposed
Car 4	2.9	4.1	18.8	12.3	60.1	179.8	3.1	2.2
Cliff bar	24.8	24.8	11.3	34.6	13.4	48.7	4.8	3.5
David indoor	3.6	7.6	9.7	13.6	16.2	76.7	9.2	2.8
Jumping	36.8	92.4	3.6	63.0	9.9	58.5	4.6	2.7
Lemming	93.4	184.9	23.2	86.9	25.6	149.1	25.0	17.5
Occlusion 1	9.2	6.5	17.5	11.1	32.3	5.6	4.4	3.2

Table 2: Overlap rate

Sequences	IVT	L_1	PN	VTD	MIL	Frag track	SCM	Proposed
Car 4	0.92	0.84	0.64	0.73	0.34	0.22	0.92	0.96
Cliff bar	0.56	0.20	0.38	0.33	0.46	0.13	0.74	0.85
David indoor	0.71	0.62	0.60	0.52	0.45	0.19	0.47	0.89
Jumping	0.28	0.09	0.69	0.08	0.53	0.14	0.70	0.79
Lemming	0.18	0.13	0.49	0.35	0.53	0.13	0.17	0.72
Occlusion 1	0.85	0.88	0.65	0.77	0.59	0.90	0.92	0.94

Table 3: Computational time

Sequences	IVT	L_1	PN	VTD	MIL	Frag track	SCM	Proposed
Car 4	55	49	42	35	31	27	17	9
Cliff bar	54	50	45	37	34	29	16	10
David indoor	52	52	44	35	32	30	23	8
Jumping	54	47	46	36	34	28	19	12
Lemming	53	48	43	37	33	27	18	11
Occlusion 1	51	49	44	38	32	25	18	10

The IVT, SR based (L_1), PN, VTD, MIL, frag track and SCM methods are difficult to keep tracking of the object after occlusion but our proposed approach achieves lowest tracking error and highest overlap rate. In comparison, our method succeeds.

CONCLUSION

Here, the moving objects in the videos are tracked and it is highlighted to show that the object is moving in the video. First, the video is divided into frames. These frames are then given into the Gabor filter which binarizes and normalizes the frames. These normalized frames are then given into the local spatial context model which forms a contribution set and classifies the frames with the weak classifier. The classified frames are then combined to form a video. In the output video, the moving object is highlighted. To evaluate the performance of the proposed method, we compared the results with various methods. From the experimental results, our proposed method shows better result for identifying the moving object in the video file.

REFERENCES

Adam, A., E. Rivlin and I. Shimshoni, 2006. Robust fragments-based tracking using the integral histogram. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Volume 1, June 17-22, 2006, New York, USA., pp: 798-805.

Babenko, B., M.H. Yang and S. Belongie, 2009. Robust object tracking with online multiple instance learning. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 20-25, 2009, Miami, FL., USA., pp: 983-990.

Chien, S.Y., S.Y. Ma and L.G. Chen, 2002. Efficient moving object segmentation algorithm using background registration technique. IEEE Trans. Circ. Syst. Video Technol., 12: 577-586.

Cohen, I. and G. Medioni, 1999. Detecting and tracking moving objects for video surveillance. Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, Volume 2, June 23-25, 1999, Fort Collins, pp: 319-325.

Doghmane, N. and T. Bouden, 2007. Basic methods for motion detection in images sequence. Asia J. Inform. Technol., 6: 296-302.

Ercan, A.O., A. El Gamal and L.J. Guibas, 2007. Object tracking in the presence of occlusions via a camera network. Proceedings of the 6th International Conference on Information Processing in Sensor Networks, April 25-27, 2007, Cambridge, MA., USA., pp: 509-518.

Fatichah, C. and M.R. Widyanto, 2008. Boosting with kernel base classifiers for human object detection. Asia J. Inform. Technol., 7: 183-190.

Kalal, Z., J. Matas and K. Mikolajczyk, 2010. P-N learning: Bootstrapping binary classifiers by structural constraints. Proceedings of the 23rd IEEE Conference on Computer Vision and Pattern Recognition, June 13-18, 2010, San Francisco, CA., pp: 49-56.

Kwon, J. and K.M. Lee, 2010. Visual tracking decomposition. Proceedings of the Conference on Computer Vision and Pattern Recognition, June 13-18, 2010, San Francisco, CA., USA., pp: 1269-1276.

Pan, J. and B. Hu, 2007. Robust occlusion handling in object tracking. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 17-22, 2007, Minneapolis, MN., USA., pp: 1-8.

Prakash, O., M. Khare, C. Mani and A.K. Singh, 2012. Moving object tracking in video sequences based on energy of Daubechies complex wavelet transform. Proceedings of the National Conference on Communication Technologies and its Impact on Next Generation Computing, October 20, 2012, Ghaziabad, UP, India, pp: 6-10.

Rosenberg, Y. and M. Werman, 1998. Real-time object tracking from a moving video camera: A software approach on a PC. Proceedings of the 4th IEEE Workshop on Applications of Computer Vision, October 19-21, 1998, Princeton, NJ., USA., pp: 1-2.

Ross, D.A., J. Lim, R.S. Lin and M.H. Yang, 2008. Incremental learning for robust visual tracking. Int. J. Comput. Vision, 7: 125-141.

Roy, A., S. Shinde and K.D. Kang, 2010. An approach for efficient real time moving object detection. Proceedings of the International Conference on Embedded Systems and Applications, July 12-15, 2010, Las Vegas, NV., USA., pp: 1-6.

Xue, M. and H.B. Ling, 2011. Robust visual tracking and vehicle classification via sparse representation. IEEE Trans. Pattern Anal. Mach. Intell., 33: 2259-2272.

Zhong, W., H. Lu and M.H. Yang, 2012. Robust object tracking via sparsity-based collaborative model. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 16-21, 2012, Providence, Rhode Island, pp: 1838-1845.