Tracking Moving Objects in Video

G. Jemilda[1*], S. Baulkani[2], D. George Paul[1], J. Benjamin Rajan[1]

[1] Faculty of Computer Science and Engg., Jayaraj Annapackiam CSI College of Engg., Nazareth, Tamil Nadu, India.

[2] Faculty of Electronics and Communication Engg., Government College of Engg., Tirunelveli, Tamil Nadu, India.

* Corresponding author. Tel.: 919443979437; email: jemildajeba@yahoo.com

Manuscript submitted June 23, 2015; accepted March 24, 2016.
doi: 10.17706/jcp.12.3.221-229

Abstract: Object tracking is a well-studied problem in computer vision and has many practical applications. The problem and its difficulty depend upon several factors such as the knowledge about the target object, its quantity and type of parameters being tracked. Although there has been some success with building trackers for specific object classes such as human, face, mice etc. Tracking generic objects has remained challenging issue because an object can drastically change its appearance when deforming, rotating out of plane or when the illumination of the scene changes. Especially in the videos which are not clear in its original form i.e. the video which requires enhancement and its quality has to be improved. Here in the proposed work different algorithms are carried out to detect the generic objects such as the non-rigid objects and the deformable objects which are in occlusion, (i.e. in a cluttered environment) not only in the images which contain some sort of objects but also from the images which contains numerous objects which are unseen so as to improve the tracking efficiency. By doing this, the objects which are non-rigid and changing in nature can also be predicted in a perfect manner so that the computational complexity can be reduced, reliability, accuracy, and efficiency can be improved.

Key words: Deformable objects, generic objects, non-rigid objects, object tracking, occlusion.

1. Introduction

Object tracking depend on several factors such as the amount of prior knowledge about the target object and the number and type of parameters being tracked. Tracking generic object is a challenging task in video processing and computer vision because an object can drastically change appearance when deforming, rotating out of plan or when the illumination of the scene changes. The overall motion of an object can be described by the combination of a global motion and local deformations, the so-called deformation. The task of tracking is to locate and segment the object from the background at each frame of the image sequence. It has the ability to deal with the single object difficulties such as occlusion with background changing appearance, illumination, non rigid motion and the multi object difficulties such as occlusion between objects and object confusion. It has wide applications such as video surveillance, event detection, activity recognition, activity based human recognition, fault diagnosis, anomaly detection, robotics, autonomous navigation, dynamic scene analysis, path detection, human computer interaction, virtual reality, object navigation and others.

2. Review of Literature Survey

Object tracking depend on several factors such as the amount of prior knowledge about the target object and the number and type of parameters being tracked. Tracking generic object is a challenging task in video processing and computer vision because an object can drastically change appearance when deforming, rotating out of plan or when the illumination of the scene changes. The overall motion of an object can be described by the combination of a global motion and local deformations, the so – called deformation. The task of tracking is to locate and segment the object from the background at each frame of the image sequence. It has the ability to deal with the single object difficulties such as occlusion with background changing appearance, illumination, non rigid motion and the multi object difficulties such as occlusion between objects and object confusion. It has wide applications such as video surveillance, event detection, activity recognition, activity based human recognition, fault diagnosis, anomaly detection, robotics, autonomous navigation, dynamic scene analysis, path detection, human computer interaction, virtual reality, object navigation and others. Jehoon Lee *et al.* discussed the object tracking problem into two parts:

1) Segmentation process in which a weighted depth map for object extraction via the Bhattacharyya gradient flow is defined.
2) Filtering process in which global position of the object is estimated via particle filtering and active contours.

A reliable algorithm is designed to track deformable targets by combining particle filtering and geometric active contour models. The particle filtering is used for estimating the motion of the object and tracking of its deformations is achieved by active contours. Region based active contours driven by the Bhattacharyya gradient flow is used for object segmentation due to its robustness against noise and its ability to deal with cluttered environments. An on-line shape learning method based on Principal Component Analysis (PCA) is used which allows the tracker to detect the disappearance and the reappearance of a target. The limitations are, it will fail if the object reappears with an unexpected shape and it will lose the target, if the unexpected object occludes the target or appears with a similar shape to the tracked object [1].

Amir Salarpour *et al*. described all moving objects and used Kalman filter to track the vehicles. The two problems considered in tracking systems are:

1) Prediction
2) Correction

Prediction is used to predict the location of an object using Kalman filter. Mean-shift was combined with it to predict the search region. Correction is used to identify the object within designated region. The algorithm can be distinguished from its surroundings. The features such as color, distance and shape are used to represent the objects being tracked which play a major role in tracking performance. The color information alone is not sufficient because color varies due to light changes and road conditions. Robust tracking performance can be gained by shape information when color information becomes less reliable. The distance information is used to find the location of an object. The algorithm acts correctly in clutter scene and controls problem of multi object tracking such as appearance and disappearance of objects, occlusion and missing of a vehicle. The advantages of using this algorithm is it minimizes the computation time and increases the efficiency of traffic control systems [2].

Nicolas Papadakis *et al.* focused on the silhouette tracking which aims at extracting successive segmentations by introducing an energy involving a temporal consistency between visible and occluded parts of the objects using optical flow estimations, through a system of predictions. The predicted areas are separated into good and bad parts with respect to the final segmentation and the objects are represented with visible and occluded parts which handle partial and complete occlusions. The energy is minimized using a graph cuts optimization. The energy function which contains new terms allows tracking and

segmenting visible and occluded parts of an object [3].

Zhaoxiang Zhang *et al.* addressed the problem of model based object recognition. A local gradient based method is proposed to estimate the 12 shape parameters and three pose parameters. The shape parameters are set up as prior information and used for vehicle recognition. The pose parameters are needed to determine the vehicle pose and used for vehicle localization. Model based vehicle localization and recognition plays an important role in vehicle detection, tracking, recognition, high-level trajectory analysis and semantic interpretation. Previously, 2D geometric primitive features such as edge points, edge lines, vertices and conic sections are extracted and recognition is achieved by 2D – 3D correspondence and pose determination but they are time consuming and error prone. To achieve model – based localization, the 3D – model is projected into an image plane so that fitness between projections based on the distance error can be evaluated directly but they are not robust to noise and occlusion. So prior information for fitness evaluation are used instead of redundant examples and evolutionary computing instead of hypothesis – test based strategy to generate a large number of models based on the deformable model and to choose the best model and position by iterative evolution and the problem is solved in an optimization framework [4].

Lingfei Meng *et al.* narrated an object tracking using high resolution multi spectral satellite images with multi – angular look capability. In the moving object estimation step, the moving objects are identified on the time–series images. Detecting moving objects is calculated by frame difference. Then the target modeling step is done by extracting both spectral and spatial features. The target is defined as the interested object to be tracked. The spectral feature is defined as the target's Probability Density Function (PDF) and is estimated by discrete densities and m – bin histograms in which the background pixels are eliminated by reducing the background occlusion effect. The spatial feature is defined as the geometric area of the reference target surface in the window region and is estimated by pixel count. In the target matching step, the target is indicated as a sliding window over other image sequences. The Bhattacharyya distance and histogram intersection are used for spectral feature matching and pixel count similarity is used for spatial feature matching.

The main advantages are:

1) Better tracking accuracy due to the usage of both spectral and spatial features.
2) Applied in the research areas such as object velocity estimation and traffic control.

The limitations are:

1) Difficult to identify identical objects visually with small geometric size in the image sequences.
2) Tracking small objects in a denser scene is difficult [5].

Boris Babenko *et al*. studied the problem of tracking an object in a video by giving its location and the classifier is trained to separate the object from the background. They also stated that the generic object tracking is a challenging one because an object can change appearance when deforming, rotating out of plane or when the illumination of the scene changes. Also, object tracking depends on several factors such as the amount of prior knowledge about the target object and the number and type of parameters being tracked. But this paper focuses on the problem of tracking an arbitrary object with no prior knowledge other than its location.

The three components of a tracking system used are:

1) An appearance model which can evaluate the likelihood at particular location
2) A motion model which relates the locations of the object over time
3) A search strategy for finding the most likely location in the current frame.

The goal of this paper is to develop a more robust way of updating an adaptive appearance model and to handle partial occlusions without drift. When an object undergoes partial occlusion, instead of supervised learning, multiple instance learning avoids drift problems and therefore it is more robust.

The disadvantage of this paper is if an object is completely occluded for a long period of time or if the object leaves the scene completely, object tracking cannot be performed [6].

Badri Narayan Subudhi *et al.* proposes an algorithm which includes two schemes for moving object detection and tracking

1) Spatio-temporal spatial segmentation
2) Temporal segmentation

To obtain a Spatio-temporal spatial segmentation, a compound Markov Random Field (MRF) model and Maximum-A-posteriori Probability (MAP) estimation techniques are used. They are used to determine the boundary of the regions in the scene accurately with faster execution time.

For temporal segmentation, label frame difference based Change Detection Mask (CDM) is used. The proposed approach provides a better spatial segmentation compared to the other JSEG, edgeless and edge based methods. It is used to determine the foreground and the background parts and the effect of silhouette is reduced.

Moving object detection in a video is the process of identifying different object regions which are moving with respect to the background. It is used to find speed/ velocity, acceleration and position of the object at different time.

It was achieved by two different ways:

1) Motion detection/change detection is the process of identifying changed and unchanged regions when the camera is fixed and the objects are moving.
2) Motion estimation is the process of estimating the positions of the moving object and here both the objects and the camera may move.

So the change information based scheme has less computational burden and gives more accuracy and hence it is more viable for real time implementation [7].

Mosalam Ebrahimi *et al*. is concerned with the components of the object recognition pipeline namely feature detection, feature description and feature tracking. It leads to simpler pipeline and easier to implement. To speed up feature detectors and feature tracking, adaptive sampling concept is used to combine tracking and recognition resources. It uses less memory.

Feature based visual object recognition is a dominant approach in computer vision. In a conventional feature-based pipeline, local features are detected, a local descriptor is computed and descriptors are queried which is then used by a classifier.

The most popular feature detectors and descriptors is the Scale Invariant Feature Transform (SIFT). It is computed in two main stages:

1) Using a detector followed by a descriptor
2) Using a filter

The performance benefits of these two visual methods are:

1) To speed up FAST corner detector, adaptive sampling can be used
2) To speed up feature tracking and object recognition, an adaptive descriptor based on Binary Robust Independent Elementary Features (BRIEFs) can be used.

The two main advantages are it is simple and generic. It does not prevent us from using other algorithmic or implementation wise optimization methods [8].

Lingfei Meng *et al.* illustrated the Sparse Representation (SR) based tracking method to track the objects robustly due to the factors such as pose variation, illumination change, occlusion, background clutter and partial or full occlusion which is the effective one. To solve the problems like high computational cost due to trivial occlusion template and difficult in adopting object features due to raw template object representation, Kernel Sparse Representation (KSR) based tracking algorithm is used.

Multiple object features such as spatial color histogram and spatial gradient – orientation histogram are used by multi – kernel fusion.

The visual object tracking is formulated as a SR problem in which each object is represented by a set of templates. To tackle the occlusion and corruption problems, a set of trivial templates are used.

To decrease computational cost, an accelerated proximal gradient approach is used. To increase tracking accuracy, multi – kernel fusion based SR is proposed.

The main advantages of KSR with multi – kernel fusion are:

1) It is less sensitive to partial occlusion, illumination variation and object deformation.
2) It speed up atleast 4 trivial templates as compared with the SR based methods.
3) It achieves a complementary effect in object representation.

The limitations of this method are;

1) It fails to track the object whose appearance is very similar to its local background.
2) It cannot track the object which is long – term entirely occluded.
3) It cannot achieve superior results for fast moving object.
4) It is hard to achieve real – time speed [9].

Longyin Wen *et al.* proposed a robust spatio – temporal context model based tracker to complete the tracking task in unconstrained environments. The temporal appearance context model captures the historical appearance of the target to prevent the tracker from drifting to the background in a long – term tracking. The spatial appearance context model integrates contributors around the target to build a supporting field. Using this method, more stable and effective tracking is achieved.

Spatio – Temporal context model based Tracker (STT) consists of temporal and spatial context models. For temporal context model, a subspace learning model is proposed to represent the target with low – dimensional feature vectors. For spatial context model, local contextual information is viewed by considering the relationships between the target and its surrounding contributors by boosting method [10].

3. Methodology

In this paper our work is divided into 6 modules and is shown in Fig. 1: 1) Video Choosing; 2) Frame Splitting; 3) Extracting Features; 4) Training Objects; 5) Tracking Objects; 6) Creating video.

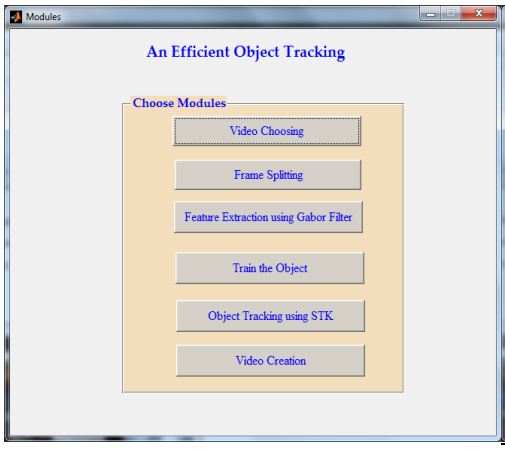

Fig. 1. Modules.

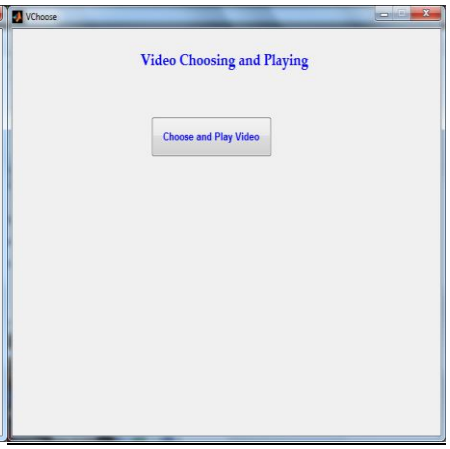

Fig. 2. Module 1.

3.1. Video Choosing

Video Choosing module is used to choose an AVI file from the particular path. To choose the AVI file the open file dialog box is used. In the open file dialog box, the AVI which is to be processed is chosen and then

open button is pressed. The open file dialog box gets the name and path of the selected AVI file. And then the AVI is read by using matlab command. After opening the file, the video can be played using the command implay. Fig. 2, Fig. 3 and Fig. 4 represents how to choose and play the video file.

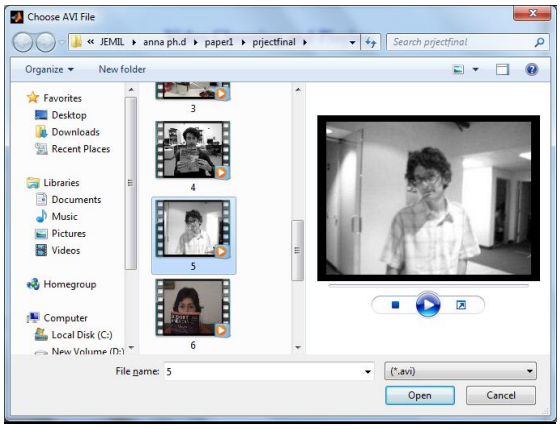

Fig. 3. Choosing AVI file.

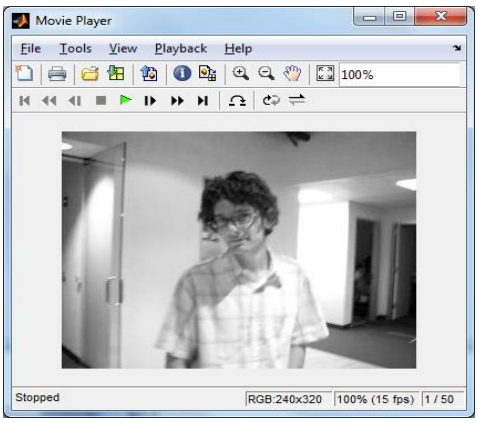

Fig. 4. Playing AVI file.

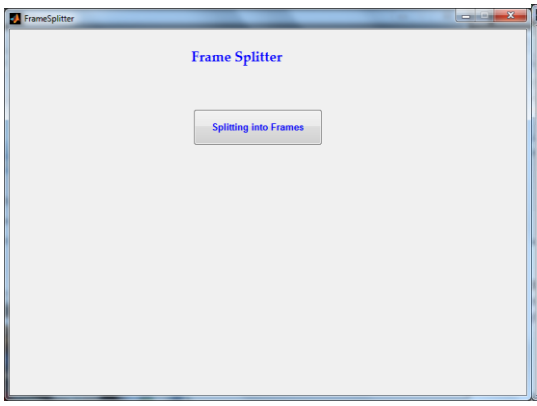

Fig. 5. Module 2.

Fig. 6. Splitted frames.

3.2. Frame Splitting

Frame Splitting module is used to split the chosen video file into number of frames. A video file consists of frames. These frames when appear before in a rate more than the perception of vision, gives a sensation of an object moving before, by looking just at the screen on which frames are appearing at high rate. The frames are the fundamental entity of a video file. So the frames are splitted by using aviread command. And then these frames are stored with an extension .jpg. The full information about the avi file is got by the command aviinfo. Some of the splitted frames are displayed in the axes continually. Fig. 5 and Fig. 6 shows how the video file is splitted into frames.

3.3. Extracting Features

Feature Extraction module is done using gabor filter. Gabor filters are band pass filters which are used in image processing for feature extraction. The impulse response of these filters is created by multiplying a Gaussian envelope function with a complex oscillation. In the spatial domain, a 2D Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave. Because of the multiplication-convolution property, the Fourier transform of a Gabor filter's impulse response is the convolution of the Fourier transform of the harmonic function and the Fourier transform of the Gaussian function. The filter has a real and an imaginary component representing orthogonal directions. The two components may be formed into a complex number or used individually. By applying various frequencies, the features are extracted. Then

convolution is applied and the mean value for the frames and testing feature values are calculated and displayed. Fig. 7 shows the extracted feature values.

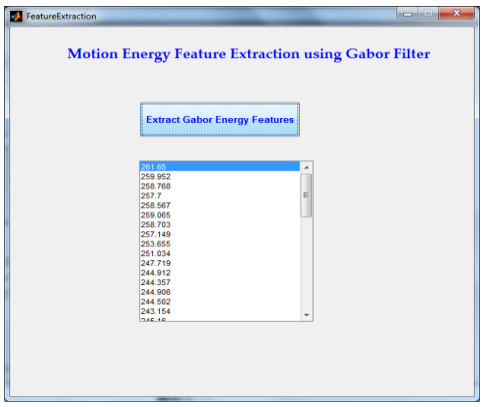

Fig. 7. Module 3.

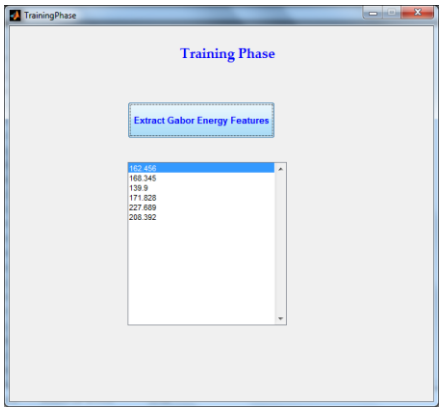

Fig. 8. Module 4.

3.4. Training

In training module, the objects or images are trained and stored in a database. Training is the process of creating an object detector or classifier to detect or recognize a specific object of interest. The training process utilizes:

- Positive images of the object of interest at different scales and orientations
- Negative images of backgrounds typically associated with the object of interest
- Non-objects similar in appearance to the object of interest
- Apply gabor filter and extract features from the training object images.

The feature values for the trained objects or images are also displayed and are shown in the Fig. 8.

3.5. Tracking

In tracking module, the object is tracked by applying spatio – temporal optimization in which multiple kernel learning technique is used. Here the support vector coefficients are calculated by projected gradient descent. It is used for object detection by detecting a set of features in a reference image, extracting feature descriptors, and matching features between the reference image and an input. This method of object detection can detect reference objects despite scale and orientation changes and is robust to partial occlusions. The tracked object frames are displayed in Fig. 9.

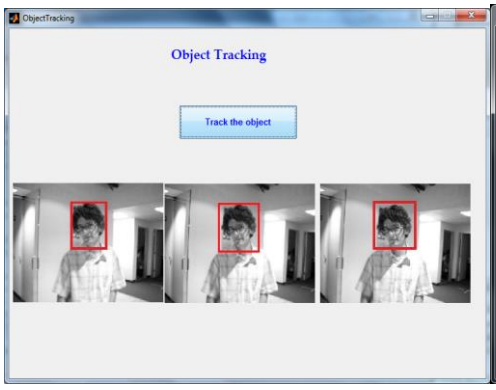

Fig. 9. Module 5.

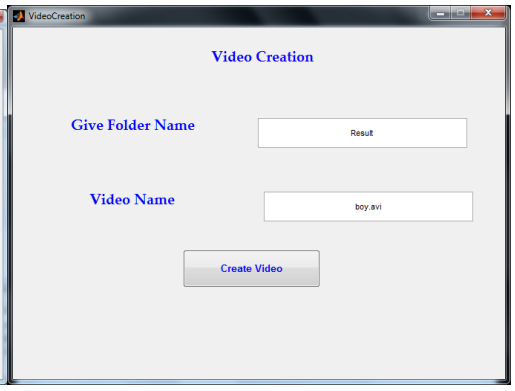

Fig. 10. Module 6.

3.6. Creating Video

This module is used to create an avi file from the tracked object frames. It can also be viewed. Fig. 10 and

Fig. 11 shows how and where the video was created again. The video is created using matlab. To create an Audio/Video Interleaved (AVI) file from MATLAB® graphics animations or from still images, follow these steps:

1) Create a VideoWriter object by calling the VideoWriter function. For example:

myVideo = VideoWriter ('myfile.avi');

By default, VideoWriter creates an AVI file using Motion JPEG compression. To create an uncompressed file, specify the Uncompressed AVI profile as follows:

uncompressedVideo = VideoWriter ('myfile.avi', 'Uncompressed AVI');

2) Optionally, adjust the frame rate (number of frames to display per second) or the quality setting (a percentage from 0 through 100). For example:

myVideo.FrameRate = 15; % Default 30

myVideo.Quality = 50; % Default 75

3) Open the file:

Open (myVideo);

4) Write frames, still images, or an existing MATLAB movie to the file by calling writeVideo. For example, suppose that you have created a MATLAB movie called myMovie write your movie to a file:

writeVideo (myVideo, myMovie);

Alternatively, writeVideo accepts single frames or arrays of still images as the second input argument. For more information, see the write Video reference page.

5) Close the file:

Close (myVideo);

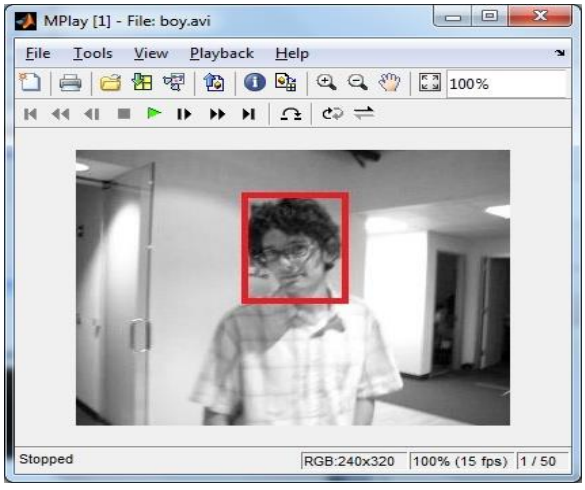

Fig. 11. Created video.

4. Conclusion

Here the moving objects in the videos are tracked and it is highlighted to show that the object is moving in the video. This paper can detect and recognize an object in an image by training an object classifier using spatio – temporal optimization with multiple kernel learning algorithms that create classifiers based on training data from different object classes. The classifier accepts image data and assigns the appropriate

object or class label. Before applying the classifier, the features are extracted. A feature is an interesting part of an image, such as a corner, blob, edge, or line. Feature extraction enables to derive a set of feature vectors, also called descriptors, from a set of detected features. Here Gabor filter is used as the feature extractors.

References

[1] Lee, J., Shawn, L., & Allen, T. (2011). Object tracking and target reacquisition based on 3-D range data for moving vehicles. *IEEE Transactions on Image Processing, 20(10)*, 2912-2924.

[2] Amir, S., Arezoo, S., Mahmoud, F., & Mir, H. D. (2011). Vehicle tracking using Kalman filter and features. *An International Journal on Signal and Image Processing, 2(2)*, 1-7.

[3] Nicolas, P., & Aurelie, B. (2011). Tracking with occulsions via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(1)*, 144-157.

[4] Zhang, Z. X., Tan, T. N., Huang, K. Q., & Wang, Y. H. (2011). Three-dimensional deformable-model-based localization and recognition of road vehicles. *IEEE Transactions on Image Processing, 21(1)*, 1-13.

[5] L. F., Meng, & Karekes, J. P. (2012). Object tracking using high resolution satellite imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 5(1)*, 146-152.

[6] Boris, B., Yang, M.-H., & Belongie, S. (2011). Robust object tracking with online multiple instance learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(8)*, 983-990.

[7] Badri, N. S., Pradipta, K. N., & Ashish, G. (2011). A change information based fast algorithm for video object detection and tracking. *IEEE Transactions on Circuits and Systems for Video Technology, 21(7)*, 144-157.

[8] Mosalam, E., & Walterio, W. M.-C. (2011). Adaptive sampling for feature detection, tracking and recognition on mobile platforms. *IEEE Transactions on Circuits and Systems for Video Technology, 21(10)*, 1467-1475.

[9] Wang, L. F., Yan, H. P., Lv, K., & Pan, C. H. (2011). Visual tracking via kernel sparse representation with multi-kernel fusion. *IEEE Transactions on Circuits and Systems for Video Technology, 21(10)*, 1-9.

[10] L. Y., Wen, Zhao, W. C., Zhen, L., Dong, E., & Li, S. Z. (2014). Robust online learned spatio-temporal context model for visual tracking. *IEEE Transactions Image Processing, 23(2)*, 785-796.

G. Jemilda received the B.E. degree in computer science and engineering from Dr. Sivanthi Aditanar College of Engineering, Tiruchendur in 1999, and the M.Tech. degree in computer science and engineering from Dr. M.G.R. University, Chennai in 2006. She has been working as a faculty in computer science and engineering in Jayaraj Annapackiam CSI College of Engineering, Nazareth-628 617 since January 2007. She is currently a research scholar in information and communication engineering in Anna University, Chennai. Her current research interests include image processing, data structure and computer networks. She has already published 6 papers in internationally reputed journals.

S. Baulkani received the B.E degree in electronics and communication engineering from Madurai Kamaraj University, Madurai and M.E. degree in computer science and engineering from Bharathiyar University, Coimbatore in 1986 and 1998 respectively. She received her Ph.D. degree in information and communication engineering from Anna University, Chennai in 2009. Presently, she is working as an associate professor in the Department of Electronics and Communication Engineering, Government College of Engineering, Tirunelveli, Tamil Nadu, India. Her areas of interest are digital image processing, network security, web mining and soft computing.