International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 3 (2015) pp. 7631-7648

http://www.ripublication.com

Visual Tracking by Wavelet Moment, Shape and Histogram

Felix M Philip[1] and Dr.Rajeswari Mukesh[2]

[1] Research Scholar, Department of Electronics And Communication Engineering, Hindustan University, Chennai
felixphilip86@gmail.com
[2]Professor and Head, Department of Computer Science and Engineering, Hindustan University, Chennai

Abstract

Rapid development of the multimedia and the associated technologies urge the processing of a huge database of video clips. The processing efficiency lies on the search methodologies utilized in the video processing system. Usage of inappropriate search methodologies may make the processing system ineffective. Hence, effective object detection and tracking system is an essential pre-requisite for searching relevant videos from a huge collection of videos. This paper proposes object detection and tracking system making use of object segmentation and feature extraction. In object detection, initially the video is separated into different frames by use of shot segmentation and from each frame, object is segmented using background subtraction. After segmentation, the features are extracted which includes wavelet moment, histogram, shape and cologram. Subsequently, the object is tracked and saved from different frames. The proposed approach is evaluated using various video clips using error value and score value. Comparison is made with respect to the existing technique and we can observe that the proposed technique obtained lowererror value and higher score value shows the effectiveness of the technique. The highest score value obtained is about 0.91 when compared to the highest of 0.75 for the existing technique.

Keywords: Visual tracking, Wavelet moment, Shape, Histogram, Background separation, Shot segmentation

1. Introduction

Generally for human computer interaction (HCI) reliable visual tracking is seemed as one of the important task in real world [8]. But, developing robust visual tracking algorithm has open problems like what to track. Color trackers, for example, are distracted from their target by other objects of similar color, while edge-based trackers can be misled by clutter in the background [9]. In fact, more single visual cue method as, Kalman Filter(KF), Condensation, JPDAF and Mean Shift, etc are used earlier. But the single visual cue has various limitation such as, limited to a particular environment that is, typically, static, controlled and known a priori [8]. As a result, in real world the single visual cue is not a robust visual tracking method to deal with various scenes in real world. So in order to increase robustness of the single visual cue multiple combination cues are used [8]. Multiple cues have the capability to achieve more reliable visual tracking system [9]. It is clear that multi cues play an important role in human visual perception. Therefore, the integration of shape, texture, and color cues should result in better tracking performance against distractions [10].However, the key challenge is to optimally fuse multi-cues for developing reliable visual tracking systems. Dynamic Bayesian networks (DBNs) are a powerful, flexible statistical tool in machine learning and pattern recognition [8]. Multi-clue included shape feature and color feature [11].

Generally, Bottom-up and top-down approaches are two kinds of methodologies for the visual tracking problem [16].Also, due to the complex nature of human body, tracking human in video sequences is a very difficult task [18] and involves a number of hard issues such as occlusion, self-occlusion, cluttered background, and high-dimensional motion parameters [21].Occlusion reasoning for object tracking is one of the most challenging problems in visual surveillance[14].Mostly more single visual cue method as, Kalman Filter (KF), Condensation, JPDAF and Mean Shift, etc are used earlier in visual tracking[8]. In order to increase robustness of the single visual cue multiple combination cues are used. Multiple cues have the capability to achieve more reliable visual tracking system [9]. Therefore, the integration of shape, texture, and color cues should result in better tracking performance against distractions [10].

In visual tracking many representation schemes such as, for instance, spatial region template, non-parametric histogram, sparse features and mixture probabilistic model, etc were used already[13].For probabilistic tracking with multiple cues, most previous work focuses on incorporating reliability measures with dynamic Bayes networks (DBNs); usually Gaussian Mixture Models (GMMs) are used[9].The target appearance in visual tracking is generally represented by colour histogram. The colour of an object depends on illumination, viewpoint and camera parameters that tend to change during a long tracking. Moreover, the colour histograms omit the shape of the target, thus the static colour features are not always discriminative enough [21].So in visual contour tracking the edge detection is adopted to describe shape information by complementing the colour cue the effective feature [9]. It is noteworthy that the same as object detection and recognition, tracking performance is also subject to effective observation model severely [13]. The snake model to track cells in biological image sequences and proved the convergence of snake’s motion [17].Recognition and generation of natural speech as well as the visual perception of humans and human action are basic tasks in this area [12, 18]. The important, recovering noises and unexpected variation is important for robust object tracking, the major control points of shape information are defined to the boundary region of the moving object to guarantee the tracking performance [19, 20].

This paper proposes object detection and tracking system making use of object segmentation and feature extraction. The proposed technique consists of two modules, namely object detection module and object tracking module. In object detection module, the input video is initially split into frames by use of shot segmentation technique, which is subsequently object is detected by use of background subtraction. Afterwards, object parameters are found out and finally cropping is carried out. In object tracking module, the object is tracked by finding out successive positions of the object in successive frames by the use of checking conditions, expansion and feature extraction.

The rest of the paper is organized as follows: Section 2 gives the literature review. Section 3 gives the motivation behind the research and section 4 describes the proposed technique. Section 5 gives results and discussion. Conclusion is summed up in Section 6.

2. Literature Survey

Literature presents lot of works for visual tracking and in this section, some of work related to it is reviewed. Weiming Hu et al.[1] presented a framework for active contour-based visual tracking using level sets. The main components of our framework included contour-based tracking initialization, colour-based contour evolution, adaptive shape based contour evolution for non-periodic motions, dynamic shape-based contour evolution for periodic motions, and the handling of abrupt motions. For the initialization of contour based tracking, they developed an optical flow-based algorithm for automatically initializing contours at the first frame. For adaptive shape-based contour evolution, the global shape information and the local color information were combined to hierarchically evolve the contour, and a flexible shape updating model was constructed. For the handling of abrupt motions, particle swarm optimization was adopted to capture the global motion. HuiWanga et al. [2] proposed an active contour model and its corresponding algorithms with detailed implementation for image segmentation. In the proposed model, the local and global region fitting energies were described by the combination of the local and global Gaussian distributions with different means and variances, respectively. Then they presented an algorithm for implementing the proposed model directly. By adaptively updating the weighting coefficient and the time step with the contour evolution, algorithm was less sensitive to the initialization of the contour and can speed up the convergence rate. Experiment results demonstrated that the proposed model and its algorithms were effective with application to both the synthetic and real-world images.

Guillaume Caron et al. [3] presented a model-based pose estimation using a direct approach that takes into account the image as a whole. For this, they consider a similarity measure, the mutual information. Mutual information was a measure of the quantity of information shared by two signals. Exploiting this measure allowed the method to deal with different image modalities (real and synthetic). Furthermore, it handled occlusions and illumination changes. Results with synthetic (benchmark) and real image sequences, with static or mobile camera, demonstrated the robustness of the method and its ability to produce stable and precise pose estimations.Jihao Yin et al. [4] proposed object tracking using contour tracking. They tracked the global motion by weighted distance to subspace, which was adaptive to the complex environment variation by incremental learning, and then used contour model to track local deformation and evolve the contour to the edge points. The experimental results showed that their method can track object contour undergoing partially occlusion and shape deforming, which verify the effectiveness of the algorithm.

Samarjit Das et al. [5] developed a visual tracking algorithm that was able to track moving objects in the presence of illumination variations in the scene and that was robust to occlusions. They treated the illumination and motion (translation and scale) parameters as the unknown "state" sequence. The key idea was to importance sample on the motion states while approximating importance sampling by posterior mode tracking for estimating illumination. Experiments demonstrated the advantage of the proposed algorithm over existing PF-based approaches for various face and vehicle tracking. Luka Cehovin et al. [6] addressed the problem of tracking objects which undergo rapid and significant appearance changes. They proposed coupled-layer visual model that combined the target's global and local appearance by interlacing two layers. The local layer in model was a set of local patches that geometrically constrain the changes in the target's appearance. The addition of patches was constrained by the global layer that probabilistically models the target's global visual properties, such as colour, shape, and apparent local motion. The parameter analysis showed that tracker was stable over a range of parameter values.Richard J.D. Moore et al.[7] developed a vision-based tracking system, FicTrac (Fictive path Tracking software), for estimating the path an animal makes whilst rotating an air-supported sphere using only input from a standard camera and computer vision techniques. They have found that the accuracy and robustness of FicTrac outperforms a low-cost implementation of a standard optical mouse-based approach for generating fictive paths. FicTrac provided the experimenter with a simple yet adaptable system that can be combined with electrophysiological recording techniques to study the neural mechanisms of behavior in a variety of organisms, including walking vertebrates.

3. Motivation of the Research

Moving target detection in video image is the main substance of digital image processing and recognition; it is extensively used in robot navigation, intelligent visual surveillance system, medical image analysis, industrial inspection and video image analysis.The general difficulty of object detection in static imagesis a complex one as the object detection system is necessary todistinguish a particular class of objects from all others.Further, a robust object detection system should be able todistinguish objects in uneven illumination, objects which arerotated into the plane of the image, and objects that arepartially occluded or whose parts blend in with thebackground. For moving objects detection, higher density of featurecorrespondences can ensure lower missed detection rates andbetter localization accuracy while higher reliability of featurecorrespondences can ensure lower false detection rates.

The noises may be formed by not only the real environmental noise but also the fault of video acquisition systems and transmission channels. As a result, this will cause a considerable reduction of video quality, particularly for the high-level computer vision.To carry out video tracking an algorithm examines sequential video frames and outputs the movement of targets between the frames. There are a diversity of algorithms, each having strengths and weaknesses. Spatial segmentation can also create many problems when the algorithm is compared to outdoor sequences, whose background unlike most tests indoor sequences are not homogeneous. The progress of technology makes video acquisition devices better and less costly, in that way ever-increasing the number of applications that can successfully utilize digital video. Compared to still images, video sequences offer more information about how objects and scenarios modify over time. However, video needs more space for storage and wider bandwidth for transmission.

The actions of the camera, the projection of 3D world on2D image, non-rigid or expressed nature of objects, object desertion obscuration, clutter, fractional and full object occlusions, scene clarification changes and composite object shapes/motion these are the several problems aced in tracking objects in video sequences. Causing a loss of data the exterior look, pose, and the scale of the purpose may vary with the background changes. Hence, tracking algorithms under such conditions should be adaptive with composite scenes, robust with noisy images, and accomplished with real-time execution.

4. Proposed Visual Tracking by Wavelet Moment, Shape and Histogram

There has been a vast expansion in the multimedia field in the recent past which calls for effective retrieval and processing of videos from large databases. The effectiveness of such retrieval and processing largely depends on the searching and tracking techniques employed. In this paper, object detection and tracking system is proposed which makes use of back ground subtraction based segmentation and wavelet moment, shape and histogram based feature extraction. The proposed technique consists of two modules, namely object detection module and object tracking module. The block diagram of the proposed technique is given in figure 1.

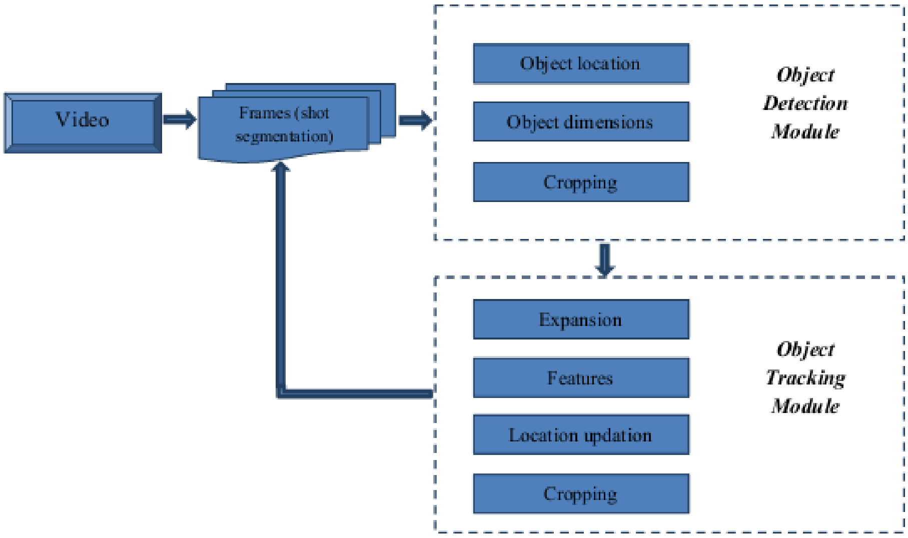


Figure 1: Block diagram of the proposed technique

4.1. Object Detection Module

In this module, the object is detected from the frame in consideration. The input video is initially split into frames by use of shot segmentation technique, which is subsequently object is detected by use of background subtraction. Afterwards, object parameters are found out and finally cropping is carried out.

Shot Segmentation

In most video applications like video retrieval, video watermarking andvideo summarization video retrieval, shot segmentation forms the primary step. In shot segmentation, video in consideration is initially split into non-overlapping units where each unit is called a shot. We can observe significant modifications in every shot which is actively contributed by the background noise and the objects in the videos. There are various video shot segmentation techniques in the literature.

We have employed Wavelet Transform based shot segmentation. The number of frames considered in each shot is dependent on the distance measure and the transformation. The initial two frames are separated intoset of blocks and DWT is employed foreach block of the frame. Wavelets can be defined as mathematical functions which has an average value of zero in a defined restricted interval. Suppose, the input is the sequence of n real numbers represented as $x_1, \cdots, x_n$, then the DWT of k sequences of z real numbers can be represented as:

$$\varpi_{kz} = \int_{-\infty}^{+\infty} x(\tau)\psi_{kz}(\tau)d\tau \tag{1}$$

Where,

$$\psi_{kz}(\tau) = \frac{1}{\sqrt{\alpha_{jk}}}\psi\left(\frac{\tau - \beta_{kz}}{a_{kz}}\right) \tag{2}$$

where ψ gives the mother wavelet, α gives the scaling parameter and β gives the shifting parameter respectively. Subsequently, the distance measure (Δ_{ij}) is calculated between the frames in consideration represented by Γ_i and Γ_j :

$$\Delta_{ij} = \sqrt{\sum_{j=1}^{m}(\Gamma_i - \Gamma_j)^2} \tag{3}$$

Once the distance is computed for the first two frames, similar methodology is carried out for the rest of consecutive frames. Finally, the frames within a shot can be recognizedby minimum distance. Minimum distance would imply maximum similarity between the frames. Let the input video sequence be represented by $G\,[j,k]$ and the shot segmented non-overlapping shots be represented by $S[j,k]$.

Background separation

Initially, first frame is considered and the object is found out from the considered frame. Mode based background separation is employed to find out the object location.Mode at any image location is defined by the pixel value that occurs most frequently at that location. Suppose the location be represented by lo_i and let the frequency of pixels $\{p_1, p_2, ..., p_k\}$ at lo_i be represented as $\{f_1, f_2, ..., f_k\}$. Then the mode at the location $\mathrm{mod}(lo_i)$ is represented as:

$$\mathrm{mod}(lo_i) = p_j \;\; where \;\; f_j \; is \; \mathrm{max}imum \;\; among \;\; f_i, 0 < i \leq k$$

These mode pixels are taken as the background pixels in the frame. As the background is motionless, the colour values of this pixel would approximately be the same during the entire analysis time. For the foreground, however, since the moving objects occupying the pixel may be in different colours and shapes at different times. Therefore, we assume that it should be the background pixel colour vector that occurs the most frequently in the frame. Hence the pixels other than the background pixels are taken as the object pixels and corresponding locations are found out. Suppose the background pixels be represented as $B = \{b_1, b_2, ..., b_m\}$, then object pixels can be represented as $\{O\} = \{P\} - \{B\}$ where $\{O\} = \{o_1, o_2, ..., o_{n-m}\}$. The corresponding object pixel locations are represented as $\{LO\} = \{lo_1, lo_2, ..., lo_{n-m}\}$.

After obtaining the object locations, compute the object dimensions so as to help in the cropping process. The dimensions found out are object centre, width and height. Initially, area constraint is checked as some of the found out object pixels may not represent the real object pixels and may be noise. Area constraint requires the object to be of minimum allowed area so that noise is avoided. Let the minimum allowed area be represented as Th_a , the area of the object pixel o_i be represented as A_i , then the criteria can be stated as:

$$if\,(A_i < Th_a),\, Then\;\; discard\, o_i$$

Subsequently, the dimensions are found out for each object o_i . Suppose the left and right edges be denoted by $E_R(i)\;\; and\;\; E_L(i)$ and top and bottom edges be represented by $E_T(i)\;\; and\;\; E_B(i)$. Object x centre ($OC_X(i)$) for the i^{th} object can be computed as:

$$OC_X(i) = \frac{|E_R(i)\;\; - E_L(i)|}{2}$$

Object y centre (OC_Y(i))for the i^{th} object can be computed as:

$$OC_Y(i) = \frac{|E_T(i)\;\; - E_B(i)|}{2}$$

Width (OW(i))for the i^{th} object can be computed as:

$$OW(i) = |E_R(i)\;\; - E_L(i)|$$

Height (OH(i))for the i^{th} object can be represented as:

$$OH(i) = |E_T(i)\;\; - E_B(i)|$$

By using these calculated dimensions, the image is cropped to have the object. Cropping can be defined as to cut the input image into smaller image leaving out unimportant areas of the image. That is after cropping, only important areas of the image are taken which in our case is the object. In our case, the image is resized to the object size containing only the object.

4.2. Object Tracking Module

In this module, the object is tracked by finding out successive positions of the object in successive frames. After detection of the object from the first frame, subsequently the object in the second frame is found out.

Initially, by taking the object location and dimensions of the first frame ($OC_X(i)$, $OC_X(i)$, OW(i) and OH(i)), crop the second frame image. Let the cropped image of the first image berepresented as C_1 and the cropped image of second frame using first frame object location be represented as C_2. The condition is such that if both are same, then keep the object location as same. This can be represented as:

$$\text{if } (C_1 = C_2),\ \textit{Keep the object location}$$

If the two are not same, the dimensions of the cropped image are extended from all sides. That is the cropped image is expanded through right side $E_R(i)$, left side $E_L(i)$, bottom side $E_B(i)$ and top side $E_T(i)$ to form new dimensions of $Ex_R(i)$, $Ex_L(i)$, $Ex_B(i)$ and $Ex_B(i)$. For this updated cropped image, extract the features which consist of density, correlogram, histogram and wavelet moment.

Density is often defined in Pixels per inch (PPI) or pixels per centimetre (PPCM) which as the name suggests gives the number of pixels in the cropped image per unit measure. Suppose the number of pixels contained in the cropped image is y and the centimetre covered is z , then PPCM density D is given by:

$$D = \frac{y}{z}$$

Correlogram gives the plot of auto-correlation of the cropped image values. It can be also be considered as the cross-correlation of image with itself. Autocorrelation can be defined by the formula (where the input image is represented by $g(x, y)$):

$$g(x,y) * g(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x^l, y^l)\ g(x + x^l, y + y^l) dx^l dy^l$$

Where $g(x, y)$ the two-dimensional brightness is function that defines the image, and $g(x^l, y^l)$ are the dummy variables of integration.

Third feature extracted is histogram. Histogram is a graphical representation of the distribution of pixel values. A histogram is a representation of tabulated frequencies, erected over discrete intervals, with an area proportional to the frequency of the pixel in the cropped image.

Image moments constitute an important feature extraction method which generates high discriminative features. Wavelet moment hails from orthogonal moment family that uses an orthogonal wavelet function as kernel. These moments combine the advantagesof the wavelet and moment analyses in order to construct improved moment descriptors. Wavelet moment for an image represented by $g(x, y)$ having dimension $m \times m$ can be given by:

$$W_{abq} = \sum_{x=1}^{m}\sum_{y=1}^{m}\psi_{ab}(r)\, e^{-jq\varpi} g(r,\varpi)$$

Where $r = \sqrt{x^2 + y^2}$, $\varpi = \tan arc(y/x)$ and $\psi_{ab}(r)$ is the mother wavelet basics. Hence from the expanded cropped image, we have found out the features. Let the extracted features of density, Correlogram, histogram and wavelet moments be represented as $den_i, cor_i, hist_i$ and $wavm_i$ for various locations. Among all the features, take the higher value based location of the density, histogram, wavelet moment and lower value based location for the correlogram. If all the taken locations are same, then take that location and crop the image of the second frame accordingly. That is, it can represented as:

$$if\ (Lo(\max den_i) = Lo(\min cor_i) = Lo(\max hist_i) = Lo(\max wavm_i)\ ), Then\ take\ the\ location$$

Where, Lo gives the location, $\max$ gives the maximum among all values and $\min$ gives the minimum among all values.

The process is continued for subsequent frames. In the subsequent step, third frame image is cropped using the object location of second frame and likewise. Generalising the process for all frames, the image in the $i+1^{th}$ frame is cropped using the object location of i^{th} frame. Save all the cropped region and object locations to find out the tracked object.

The pseudo- code of the proposed technique can be stated as:

START

Convert input video to frames using shot segmentation
For all frames
Read the current frame
Get the object locations using object subtraction to have $\{LO\} = \{lo_1, lo_2, ..., lo_{n-m}\}$
Calculate the object dimensions of $OC_X(i)$, $OC_X(i)$, $OW(\mathrm{i})$ *and* $OH(\mathrm{i})$
By using these parameters, crop the object in the frame
Take the subsequent frame and using the object location of current frame, crop the image of next frame
if $(\mathrm{C}_1 = C_2)$, , *keep the object location*
Else
Expand the cropped image dimensions.
Extract the cropped image features $den_i, cor_i, hist_i$ *and* $wavm_i$ $if\ (Lo(\max den_i) = Lo(\min cor_i) = Lo(\max hist_i) = Lo(\max wavm_i)\ )$,
Then take the location
Save all the cropped region and object locations to find out the tracked object

5. Results and Discussion

In this section, we discuss and analyse the results achieved by the proposed object detection and tracking technique. In section 5.1, database employed, experimental set up and evaluation metrics used are described. Section 5.2 gives the implementation screen shots and section 5.3 gives the performance analysis.

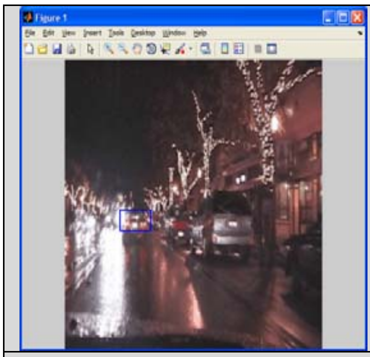
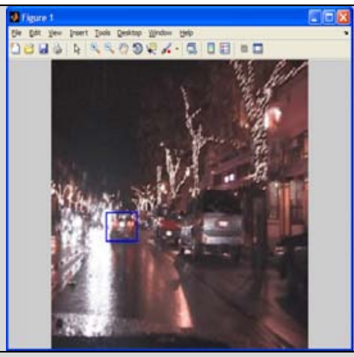
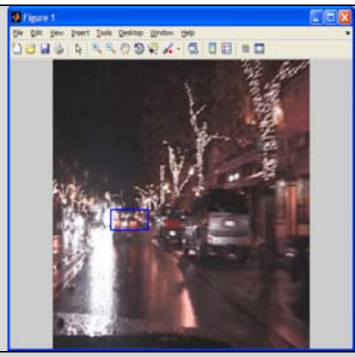
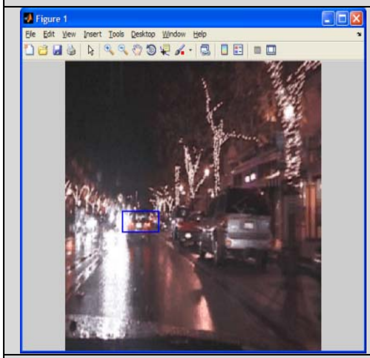
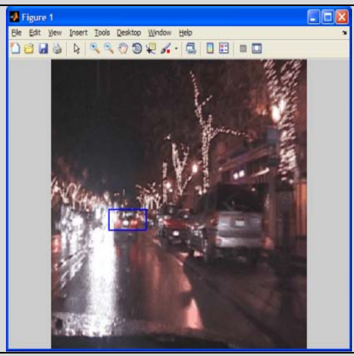
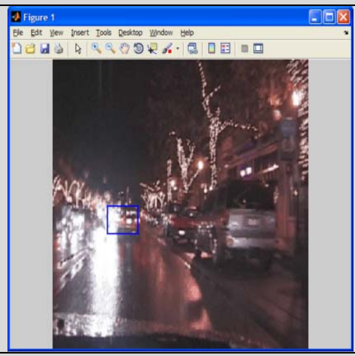
5.1. Database Employed, Experimental Set Up and Evaluation Metrics

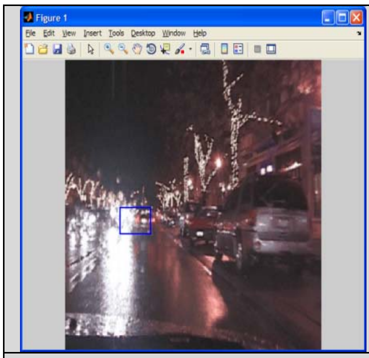
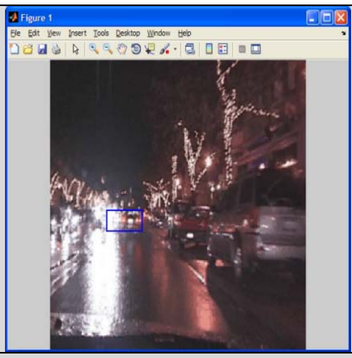
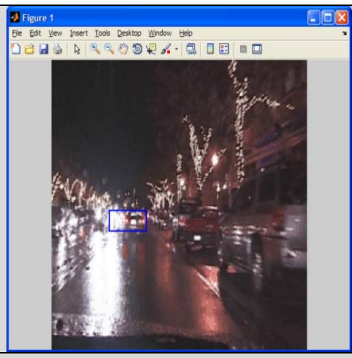
The database used includes many videos and for evaluation purpose, we have taken four videos namely two car videos, one face video and Akiyo video. The implementation is carried out using MATLAB which was simulated in a system having system specifications of 6 GB RAM and 3.2 GHz Intel i-7 processor. The evaluation metrics employed are error value and the score value. Error value has to be the minimum and score value has to be maximum for an effective technique. Error value is obtained as the difference of the original value and the obtained value.

5.2. Implementation Screen Shots

The implementation screen shots for car video 1 are given in table 1 given below.

Table 1: Implementation screen shots

		
Frame 1	***Frame 3***	***Frame 5***
		
Frame 8	***Frame 10***	***Frame 15***

		
Frame 20	***Frame 25***	***Frame 30***

5.3. Performance Analysis

In this module, the performance of the proposed technique is analysed and evaluated. The evaluation is made with the evaluation metrics of error value and score value. Comparison is also made with respect to the existing technique [1] so as to have an extensive analysis.

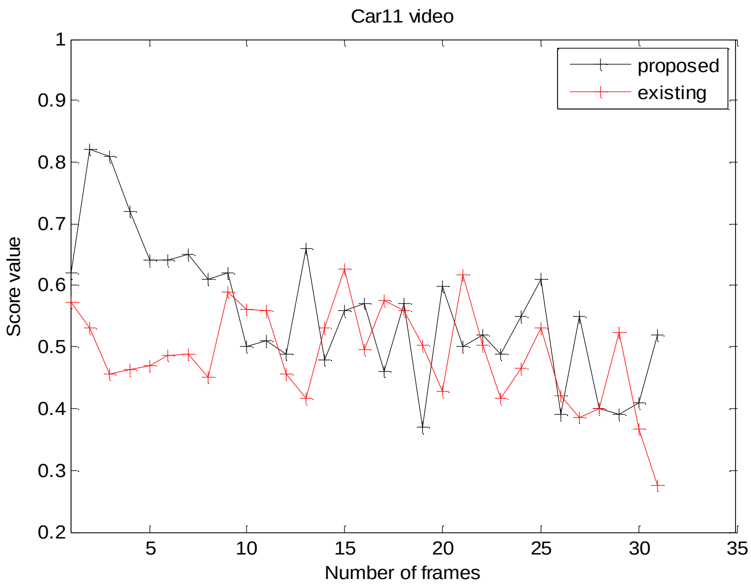


Figure 2:Score value for car1 video

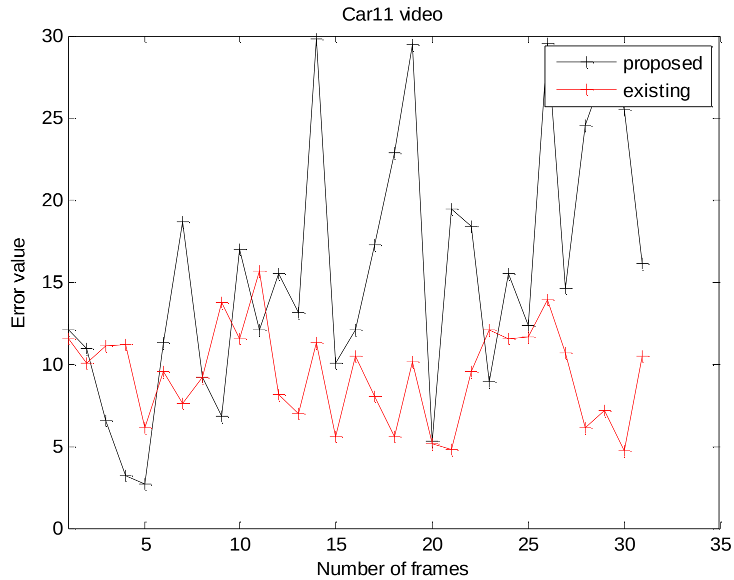


Figure 3: Error value for car1 video

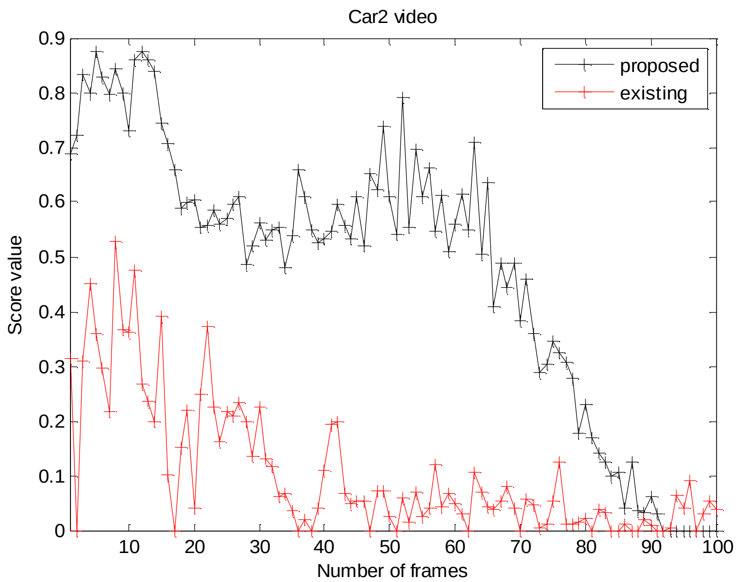


Figure 4:Score value for car 2 video

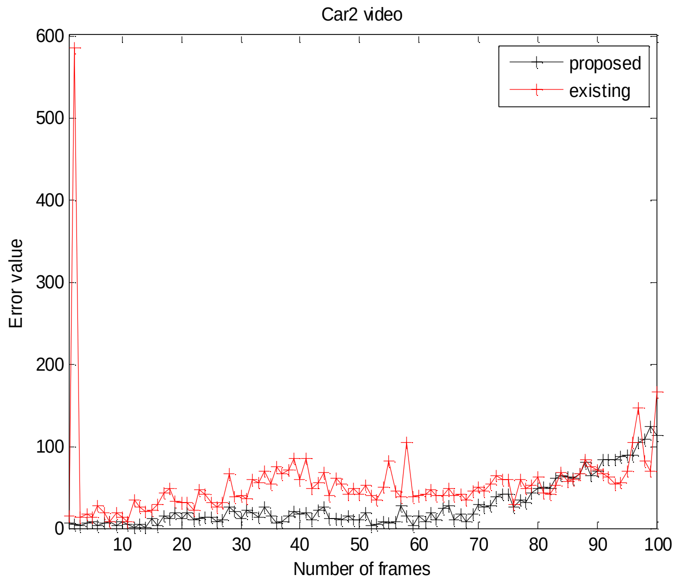


Figure 5: Error value for car2 video

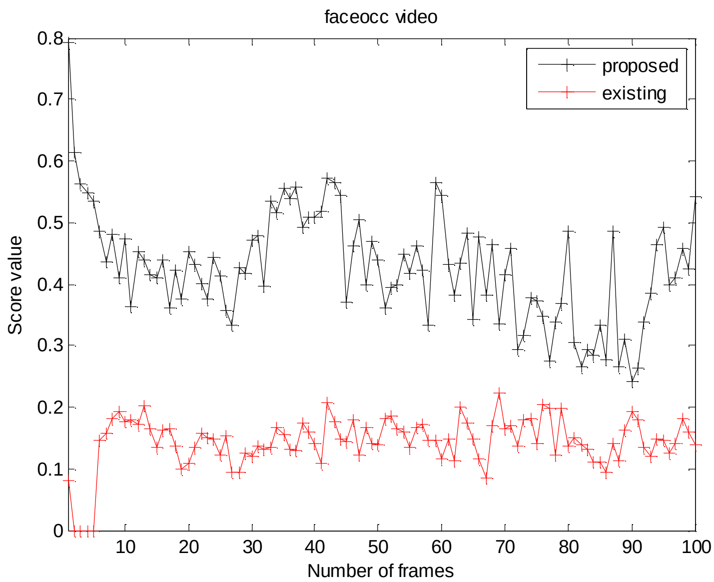


Figure 6: Score value for face video

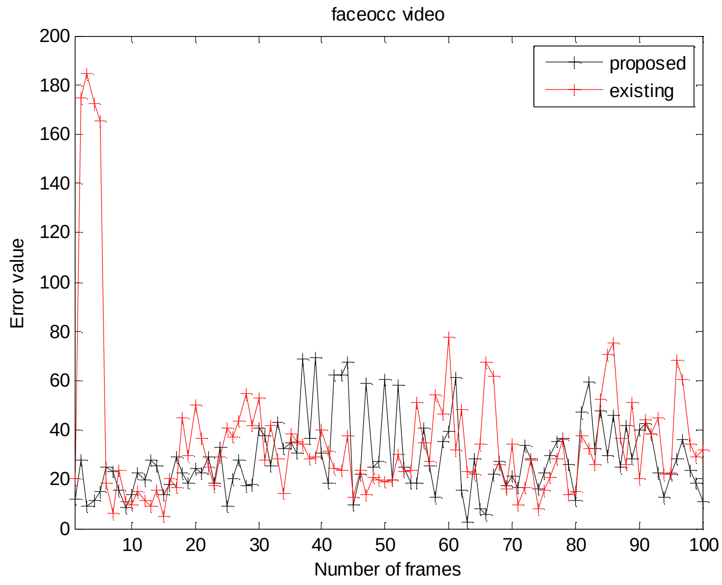


Figure 7: Error value for face video

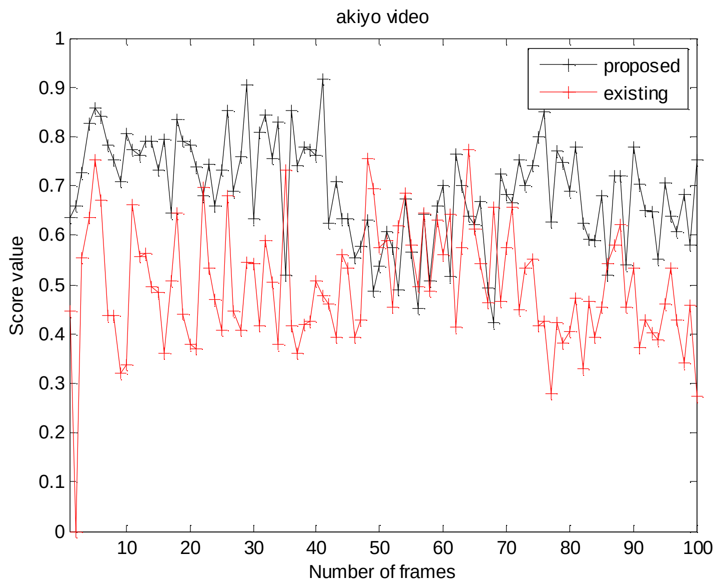


Figure 8: Score value for Akiyo video

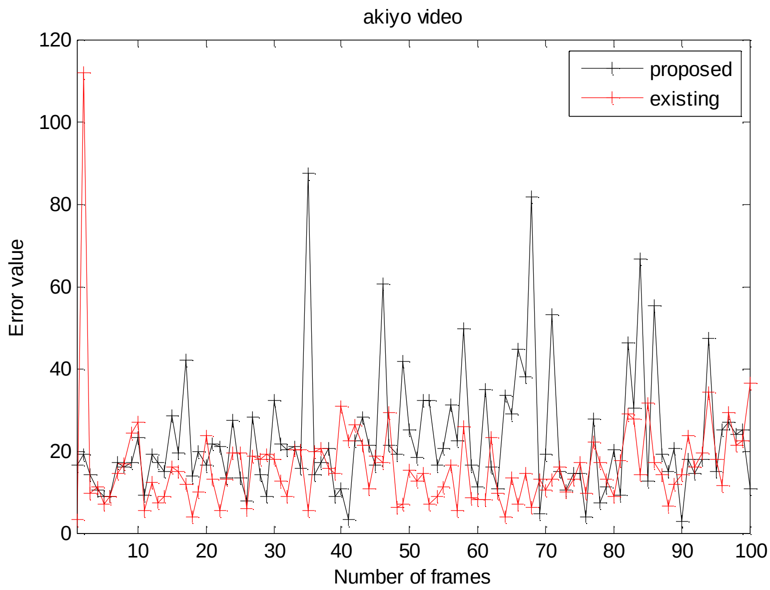


Figure 9: Error value for Akiyo video

Inferences from figures 2-9

- Figures give the comparative analysis by comparing our proposed technique with existing technique [1].
- The evaluation is made with the evaluation metrics of error value and score value.
- The results are taken by varying the number of frames.
- Figures 2,4,6 and 8 gives the score values. Figures 3,5,7 and 9 gives the error value.
- Comparison is also made with respect to the existing technique so as to have an extensive analysis.
- Figures 2 and 3 gives the car 1 video analysis, figures 4 and 5 gives the car 2 video analysis , figures 6 and 7 gives the face video analysis and figures 6 and 7 gives the Akiyo video analysis.
- From the figures, we can observe that the proposed technique achieved better score values attaining maximum values of 0.82, 0.87, 0.78 and 0.91 respectively. The existing maximum values for the images were 0.62, 0.53, 0.57 and 0.75 respectively.
- We can also observe smaller error value with respect to the existing technique.
- Low error value and high score value shows the effectiveness of the proposed technique.

6. Conclusion

This paper proposes object detection and tracking system making use of object segmentation and feature extraction. The proposed technique consists of two modules, namely object detection module and object tracking module.The proposed approach is evaluated using various video clips using error value and score value. Comparison is made with respect to the existing technique and we can observe that the proposed technique obtained lower error value and higher score value shows the effectiveness of the technique. The highest score value obtained is about 0.91 when compared to the highest of 0.75 for the existing technique.

Reference

[1] Weiming Hu, Xue Zhou, Wei Li, WenhanLuo, Xiaoqin Zhang, and Stephen Maybank, "Active Contour-Based Visual Tracking by Integrating Colors, Shapes,and Motions", IEEE Transactions On Image Processing, Vol. 22, No. 5, pp.1778-1792, 2013.

[2] HuiWanga,Ting-Zhu Huang, ZongbenXu and Yilun Wang, "An active contour model and its algorithms with local and global Gaussian distribution fitting energies", Information Sciences, Vol.263, pp.43–59, 2014.

[3] Guillaume Caron, Amaury Dame and Eric Marchand, "Direct model based visual tracking and pose estimation using mutual information", Image and Vision Computing, Image and Vision Computing, vol.32, pp.54–63, 2014.

[4] Jihao Yin, ChongyangFu and JiankunHub, "Using incremental subspace and contourtemplate for object tracking", Journal ofNetworkandComputerApplications, vol.35, pp.1740–1748, 2012.

[5] Samarjit Das, Amit Kale, and NamrataVaswani, "Particle Filter With a Mode Tracker for Visual Tracking Across Illumination Changes"IEEE Transactions On Image Processing, Vol. 21, No. 4,pp.2340-2347, 2012.

[6] Luka Cehovin, MatejKristan and Ales Leonardis, " Robust Visual Tracking Using an Adaptive Coupled-Layer Visual Model", IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 35, No. 4,pp. 941-957,2013.

[7] Richard J.D. Moore, Gavin J. Taylor, Angelique C. Paulk, Thomas Pearson, Bruno van Swinderen and Mandyam V. Srinivasan, "FicTrac: A visual method for tracking spherical motion and generating fictive animal paths", Journal of Neuroscience Methods, Vol.225, 106–119, 2014.

[8] Tao Wang, QianDiao, Yimin Zhang, Gang Song, Chunrong Lai and Gary Bradski, “A Dynamic Bayesian Network Approach to Multi-cue based Visual Tracking”, Proceedings of the 17th International Conference on Pattern Recognition, Vol.2, pp.167-170, 2004.

[9] ChunhuaShen, Anton van den Hengel, and Anthony Dick,” Probabilistic Multiple Cue Integration for Particle Filter Based Tracking”, Proceedings of7th Digital Image Computing: Techniques and Applications, 2003.

[10] Junqiu Wang and Yasushi Yagi, “Integrating Color and Shape-Texture Features for Adaptive Real-Time Object Tracking”, IEEE Transactions on Image Processing, Vol. 17, No. 2, 2008.

[11] Ge Yang and Hong Liu, "Visual Attention & Multi-cue Fusion based Human Motion Tracking Method", Sixth International Conference on Natural Computation, Vol.4, pp.2044 – 2054, 2010.
[12] Stefan Feyrer and Andreas Zell, "Tracking and Pursuing Persons with a Mobile Robot", In Proceedings of International Workshop on Recognition Analysis and Tracking of Faces and Gestures in Real Time Systems, pp.83-88, 1999.
[13] Hui Zhang, Baojun Zhao, Jianke Li and Jichao Jiao, "Visual Tracking for Arbitrary Shaped Object", 2nd International Conference on Future Computer and Communication, Vol.2, 2010.
[14] Weiming Hu, Xue Zhou, Min Hu, and Steve Maybank, "Occlusion Reasoning for Tracking Multiple People", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 19, No. 1,2009.
[15] Lei Xie, Xiu-Li Cai, Zhong-Wa Fu, Rong-Cwn Zhao and Dong-Me1 Jiang, "A Robust Hierarchical Lip Tracking Approach For Lip reading And Audio Visual Speech Recognition", Proceedings of International Conference on Machine Learning and Cybernetics, Vol.6,pp.3620-3624,2004.
[16] Ying Wu and Thomas S. Huang, "Robust Visual Tracking by Integrating Multiple Cues based on Co-inference Learning", International Journal of Computer vision, Vol.58, No.1,pp.55-71,2004.
[17] Kap-HoSeo, Tae-Yong Choi, and Ju-Jang Lee," Adaptive Color Snake Model for Real-Time Object Tracking", Proceedings of the IEEE International Conference on Robotics and Automation, Vol.1, pp.122-127, 2004.
[18] Yao WenJuan Liang YaLing Du MingHui, "A Real-time Lip Localization and Tacking for Lip Reading", 3rd International Conference on Advanced Computer Theory and Engineering,Vol.6,pp.363-366,2010.
[19] Sanghoon Kim , Sangmu Lee and Seungjong Kim, "Object Tracking of Mobile Robot using Moving Color and Shape Information for the aged walking", Proceedings of the Second International Conference on Future Generation Communication and Networking, Vol. 02, pp. 293-297, 2008.
[20] Peihua Li, Tianwen Zhang, "Visual Contour Tracking Based on Particle Filters",Image and Vision Computing,Vol.21,No.1,pp.111-123,2003.
[21] HuazhongNing, Liang Wang, Weiming Hu and Tieniu Tan, "Model-based Tracking of Human Walking in Monocular Image Sequences", Proceedings. 2002 IEEE Region 10 Conference on Computers, Communications, Control and Power Engineering, Vol.1, pp.537 – 540, 2002.
[22] Junqiu WANG and Yasushi YAGI, "Integrating Shape and Color Features for Adaptive Real-time Object Tracking", IEEE International conference on Robotics and Biomimetics, pp.1-6, 2006.