International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

Video Object Tracking in Compressed Domain Using SIFT and STMRF

Kavita Borse[1], Rupali Nikhare[2]

[1]Department of Computer Engineering, Pillai College of Engineering, New Panvel, Navi Mumbai, India

[2]Assistant Professor, Department of Computer Engineering, Pillai College of Engineering, New Panvel, Navi Mumbai, India

Abstract: ***The compressed video object tracking are required for a following structure with both reasonable accuracy and complexity. The existing systems of video object tracking which doesn't provide good accuracy. Up-till now all systems assumes that video is already in the H.264/AVC format. All these types of errors are overcome in this video object tracking system. Here we are presenting video object tracking in compressed domain by using two algorithms Scale Invariant Feature Transform (SIFT) and Spatio-Temporal Markov Random Field (STMRF) algorithm. Here for tracking we use YUV video format but if input video is in avi format then we convert this video into YUV format. Scale Invariant Feature Transform (SIFT) perform object detection using frame difference method. SIFT tracking has steps such as scale space extrema, key point localization, orientation assignment,keypoint descriptor and match key point and so forth. STMRF utilizes just the movement vectors (MVs) and block coding modes from the compressed bit stream to perform tracking.it is start with, preprocessing the MVs are preprocessed through intra coded block motion estimation and motion compensation. Global Motion Estimation (GME) is utilized to take a short moments of camera and Global Motion Compensation (GMC) is utilized to remove Global motion which include intra-coded block processing. At every frame, the choice of whether a specific block belongs to object decided with the help of the ST-MRF, which is updated from frame to frame. STMRF gives preferred exactness over SIFT.***

Keywords: Scale Invariance Feature Transform (SIFT), Spatio-Temporal Markov Random Field (STMRF)

1. Introduction

Video object tracking assumes an essential part in the field of PC vision due to expanding powerful PCs, the accessibility of high caliber and economical camcorders. Because of this, there is a requirement for mechanized video examination. This created a lot of enthusiasm in object tracking. Really a video is arrangements of frames and these frames are shown at a high rate with the goal that human eyes can envision the congruity of its substance. So for object tracking methods can be applied to individual frames. In its basic way, tracking can be characterized as a strategy for tracking an object in progressive picture frames to decide movement with different items or objects. Here tracker gives labels to the track object in various frames of video. The initial step has identified the areas of interest. This is, for the most part, utilized as a part of numerous applications, for example, video monitoring, robotics and so forth. The each frame in a video sequence is isolated into two distinct arrangements of pixels. The principal set contains the pixels which relate to foreground objects and the second set corresponds to background pixels separately. Generally, foreground area objects are moving items like individual's person, vehicle, water crafts and ball and so on everything else is background. Tracking of an object are followed the fundamental three stages such as:

1) **Object Detection**
   Object detection is to choose the region of interest from the video sequence.
2) **Object Classification**
   Object classification includes the moving items, for example, vehicle, feathered creatures, swaying tree and so forth.
3) **Object Tracking**
   Tracking can be characterized as the estimating the way of an object in the picture plane when that object moves around a scene.

2. Problem Definition

Moving object tracking in compressed domain act an imperative part in some real-time applications, e.g. video ordering, video transcoding, video surveillance, and so on. Video object tracking is only the way of extracting an object from a video and constant track of its movement. The challenging assignment here is to give the best and exact outcome to the user by utilizing the different algorithms. Here basically two algorithms are implemented one is Scale Invariant Feature Transform (SIFT) and the second one is Spatio Temporal Markov Random Field (STMRF).This procedure requires yuv video, so a change of avi to yuv arrangement is additionally created in this framework. These algorithms try to give best and accurate outcome to the user by giving an efficient algorithm for object detection and tracking.

3. Architecture of Video object Tracking in The Compressed Domain

Figure below shows the systematic architecture for video object tracking in compressed domain using SIFT and STMRF algorithm.

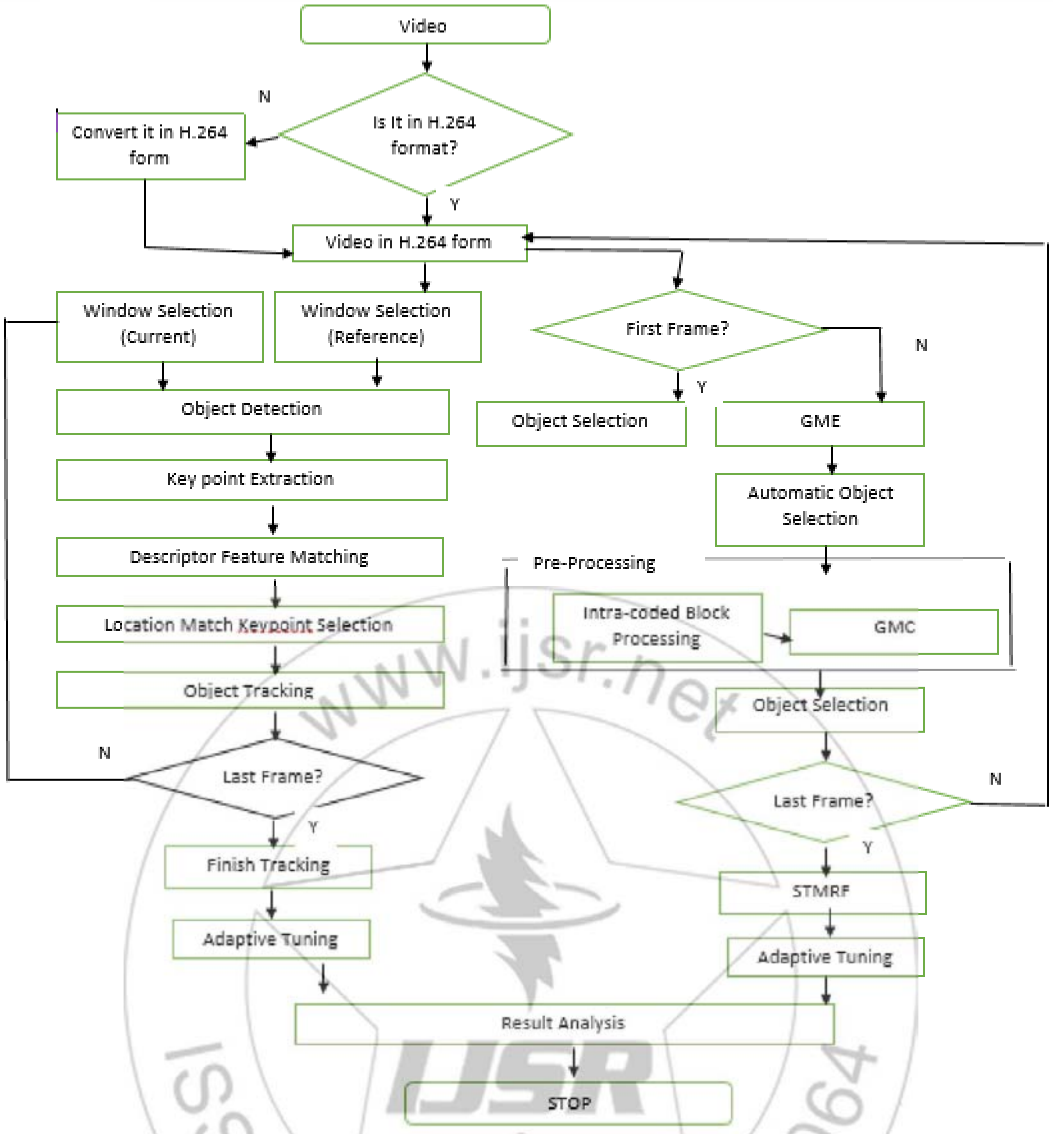

Figure 1: Architecture of video object tracking system

3.1 Input for the system

Accept a video as an input for processing. Video can be of either avi or yuv. If the video is in avi format then convert that video into yuv format. YUV is a compressed video format which is mostly used to perform an operation on video. After accepting video convert that video into frames because the video is a collection of frames.

3.2 Object Detection

Frame Differencing (FD) algorithm is mostly used to detect moving object and provide efficient result. These algorithms require reference frame, current frame and changes in illumination. This algorithm gives centroid of moving object and area of moving object as output. In case of recorded video and real time video, two consecutive frames are taken and then compared with each other. Objects which have changed their position with respect to previous frame are thus treated as moving objects.

$$D = I_n - I_{n-1} \quad \text{............................} \quad (3.1)$$

Where,

I_n = current frame

I_{n-1} = previous frame

3.3 Object tracking using SIFT

Scale Invariant Feature Transform (SIFT) depends on feature extraction and matching. The algorithm separates features of the tracking object which are invariant to changes in the object like illumination and orientation. Also the continue detection of object and match the features. It focuses on first extract the target model and then matches the feature point with candidate model. SIFT includes following steps:

1. Identifying the extrema (scale-space).
2. Localizing the keypoint.
3. Assigning the orientation.
4. Formulating the description of the keypoint.

SIFT Provide tracking of stable multi-objects by taking rectangular windows around the object of interest for both the present and reference frame. The essential thought is to decide the feature point and exclude those, whose areas are mismatched. SIFT produces the keypoint. A database holds their local features from the following back to back frame, keypoints are obtained in the same way. The coordinating keypoints are picked in view of Euclidian separate. The SIFT algorithm in detail[10]:

1) Load a video in the program.
2) Identify the moving item.
3) Select the window which bounds the object in the reference frame.
4) Apply SIFT algorithm to extract the feature point.
5) Perform step 2-4 on the current frame.
6) Match the descriptor feature.
7) Select location matched feature point.
8) Track the object.

Objects can be identified utilizing frame difference algorithm.

3.3.1 Keypoint Extraction

By Applying SIFT algorithm, key points are extracted in the reference frame and in the current frame.

[im,des,loc]=SIFT(image)

where,

im - incorporates pixel estimations of the test image
des - demonstrates descriptor vectors matrix
loc - orientation values location and scale.

The extracted feature point have both object and background. The keypoints identified with background are removed by utilizing

3.3.2 Keypoint Matching

Coordinating between keypoints extracted from object area of the reference and the current window is measured by comparing the distance of the closest neighbor to that of the second closest neighbor. The procedure to accomplish coordinating is as per the following:

1.Calculate the distance between m keypoints in the reference object area and n keypoints of the current object area by using dot production of the descriptor vectors.

$$d_{ij} = \cos^{-1}\left(des_{Ri} \cdot des_{Cj}\right) \text{................}$$

2. Distance d_{ij} are sorted and distance ratios between closest neighbor to that of second closest neighbor are calculated.

$$distRatio = \frac{\text{The closest distance}}{\text{the second closest distance}}$$

3.check the distance ratio if it is greater than 0.8 then all match are rejected, which 90% of the false match while discarding 5% of correct match.

$$\text{Match} = \begin{cases} Accept & if\ distRatio \leq 0.8 \\ Reject & if\ distRatio > 0.8 \end{cases}$$

The Keypoints accepted from the above matching are store and use for tracking.

3.3.3 Selection of location matched keypoints

Among the candidate keypoints accepted from distance matching of previous section there exist multiple candidates mismatched considering their location in both windows. To find these location-matched keypoints location difference d_{RC} between the corresponding candidates in both windows is calculated and if the location difference d_{RC} is less than or equal to the threshold dth the candidates are selected as a location-matched keypoints.

$$d_{RC} = \sqrt{(X_C - X_R)^2 + (Y_C - Y_R)^2}$$

Where,

(X_R, Y_R) location coordinate of the object in reference frame

(X_C, Y_C) location coordinate of the object in current frame.

$$\text{Location} - \text{matched} = \begin{cases} Yes & if\ d_{RC} \leq d_{th} \\ No & if\ d_{RC} > d_{th} \end{cases}$$

Select only location match point and display it.

3.4 Object tracking using STMRF

3.4.1 Preprocessing

Video object tracking makes utilization of two sorts of data from the compressed bit stream that is Motion Vector (MV) and block coding mode. This algorithm is used for texture data and texture data doesn't need the decoding. Four essential MB modes are characterized in the previously mentioned bit stream: 16 × 16, 16 × 8, 8 × 16, and 8 × 8, here the 8 × 8 mode can be again part into 8 × 4, 4 × 8, and 4×4 Modes. Being 4×4 is the smallest coding mode of partitioning in Advance Video Coding, to have a uniformly examined MV field, this strategy Delineate MVs to 4 × 4 mode. This is simple in inter coded block also here SKIP Block is also easy by just set MVs to zero.

3.4.2 STMRF Based Tracking

In MRF (Markov Random Field) tracking using spatial and temporal dimension tracking [5] of pre-recorded video is done. Initially, preprocessing of the video is done to remove noise by using Gaussian filter. GM [6] estimation is prepared to estimate the camera movement and GM compensation is done to remove GM which includes intra-coded block processing. After preprocessing tracking is performed using MRF method which uses spatial and time information. MRF [7] make use of markov property to track rigid object by using compactness behavior of the moving object which will not get dispersed in frame sequence. A block diagram of MRF Tracking method using spatial and temporal information is shown in Figure 1. In proposed MRF [8-9] based method compact space information for tracking moving rigid object in frame sequence, MV temporal continuity information is used find movement similarity between blocks in the frame in use by the objects. Video is taken as input and converted into frame; next frame is divided into 4X4 blocks. Blocks are labeled based on presence of object such as non-object and object blocks as 0 and 1 respectively. Suppose the block labeling is done for frame x and previous frame x-1 then tracking between the frame is done using motion information which is denoted by:

fx=(mx, bx)

where,

mx - MV for inter calculation between previous and current frame

bx(t) - block coding mode where t=(x,y) represent block position in the frame.

Labeling is done using MAP (Maximum A Posteriori) criterion where block is selected with maximum posterior probability:

P (wx | wx-1, fx)

Temporal continuity and spatial continuity is measured between previous frame labeling and candidate current frame labeling ψ which is the powerful cue for object tracking.

3.4.3 ST-MRF Optimization

The characteristics of a moving rigid object are relative similarity of motion within the region occupied by the object, spatial compactness i.e. not dispersed in different parts of the frame, and a continuous motion trajectory. Though the motion of flexible objects is not so characteristic it is in principle that these objects can be treated in a divide and conquer manner as a group of smaller sufficiently rigid objects. So this ST- MRF model is based on rigid object motion characteristics. The frame should be divided into smaller blocks (i.e. 4×4). Object blocks can be labeled 1 and non-object blocks can be labeled as 0. In past two algorithms Stochastic Relaxation (SR) and ICM were used. SR has some advantage in accuracy compared to ICM, but at a higher computational cost [10]. At first, the label of each block is initialized by projecting the previous frame labeling ω t−1 into the current frame. Then each block is relabeled with the label (0 or 1) that leads to the largest reduction in the energy function. Relabeling procedure is done until no further energy reduction is achieved. Normally, six iterations are needed to reach the local minimum. This is worth mentioning that because results are dependent on the initial labeling.

3.5 Result

This system is implemented in MATLAB 2012a and tested on various video sequences. Different sequences that represent typical situations critical for video surveillance systems because of its capacity in simulating various tracking conditions including illumination changes, position variations, occlusions, and distraction. A number of standard test sequences were used to evaluate the performance. Sequences were in the YUV 4:2:0 format, at two resolutions, CIF (352 × 288 pixels) and SIF (352 × 240 pixels), all at the frame rate of 30 fps.The first step is object selection.In this step user has to select the object from first frame.the output is shown in following figure.

Figure 1: Object Selection

After selecting the object now the tracking is done. The output of the object tracking using SIFT is shown below.

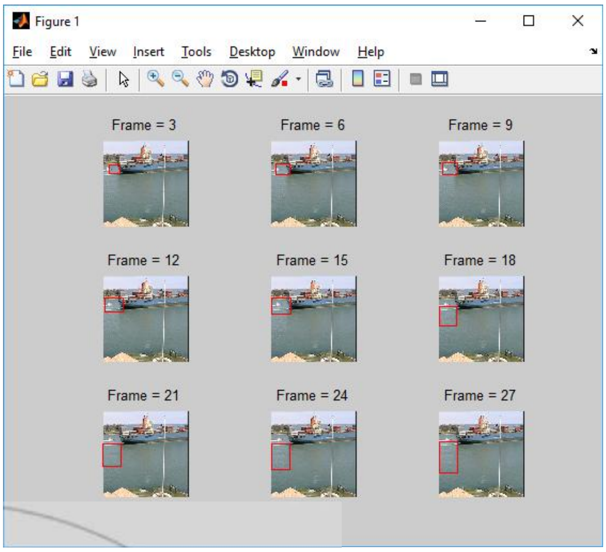

Figure 2 : Result of object tracking using SIFT

The output of the object tracking using STMRF is shown below.

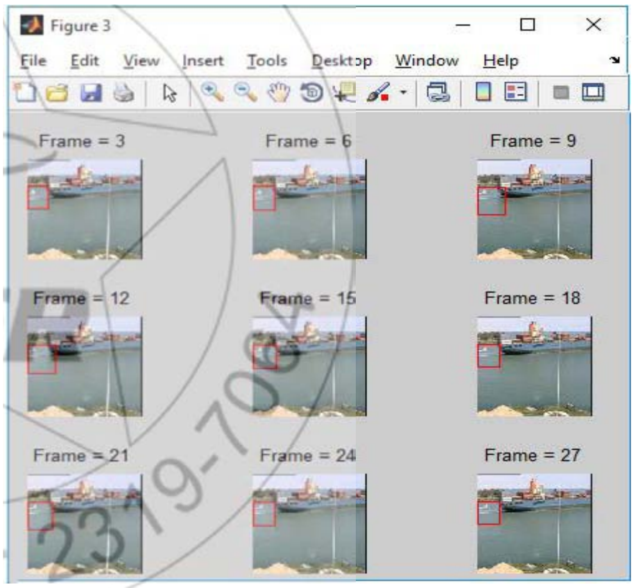

Figure 3: Result of object tracking using STMRF

If you observe clearly output of both tracking system.then you realize that STMRF object tracking is good than SIFT.

3.6 Analysis

Analysis contain the three graph precision, recall and f-measure.

3.6.1 Precision

Precision is calculated by the number of TP divided by the total number of labeled pixels, i.e., the sum of TP and FP.

$$\text{Precision} = \frac{TP}{TP+FP}$$

Where,

TP/True positive: detected regions that correspond to moving objects

FP/False positive: detected regions that do not correspond to a moving object,

3.6.2 Recall

Recall is defined as the number of TP divided by the total number of ground truth labels, i.e., the sum of TP and FN.

$$\text{Recall} = \frac{TP}{TP+FN}$$

Where,
TP / True Positive: detected regions that correspond to moving objects.
FN / False Negative**:** moving objects not detected

3.6.3 F-measure

F-measure is the harmonic mean of precision and recall.

$$\text{F-measure} = \frac{2.Precision.Recall}{Precision+Recall}$$

Here, for calculating performance we are processing on first 46 frames of video.

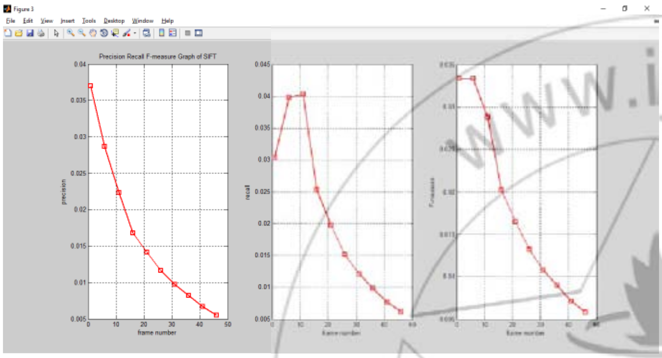

Figure 4 : Analysis graph of video object tracking using SIFT

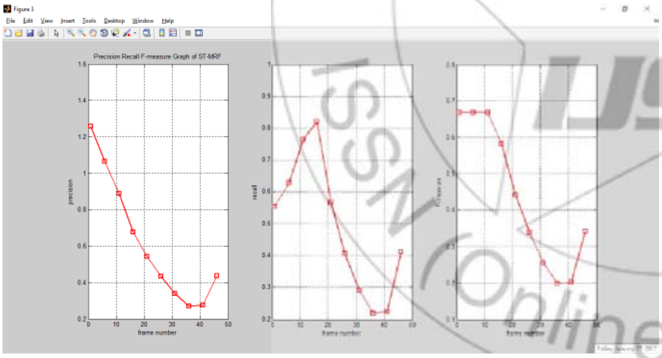

Figure 5: Analysis graph of video object tracking using STMRF

If you observe clearly then STMRF performance analysis is good than SIFT.

3.7 Conclusion

A robust and efficient video object tracking system is presented. The objective of this system is to track the object accurately in a video. The system is implemented by SIFT and STMRF method as compare to SIFT a STMRF algorithm shown accurate and effective object detection and tracking.

SIFT descriptor feature and its matching is widely used in object recognition and object tracking of computer vision fields because of its distinctive robust invariant characteristics. But according to distance ratio's value it's performance are not stable due to the false matched key points, and it is serious in case of small object tracking especially SIFT is a good for static images but for changing motion like video STMRF give best result.

References

[1] Hossein Khatoonabadi and Ivan V. Bajic, "Video Object Tracking in the Compressed Domain Using Spatio-Temporal Markov Random Fields", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 22, NO. 1, pp.300-313 JANUARY 2013
[2] W. Zeng, J. Du, W. Gao, and Q. Huang, "Robust moving object segmentation on H.264/AVC compressed video using the block-based MRF model", Real-Time Imaging, vol. 11, no. 4, pp. 290–299, Aug. 2005.
[3] Y.M. Chen, I. V. Bajic, and P. Saeedi, "Moving region segmentation from compressed video using global motion estimation and Markov random fields", IEEE Trans. Multimedia, vol. 13, no. 3, pp. 421–431, Jun. 2011.
[4] Z. Liu, Y. Lu, and Z. Zhang, "Real-time spatiotemporal segmentation of video objects in the H.264 compressed domain", J. Visual Commun. Image Represent, vol. 18, no. 3, pp. 275–290, 2007.
[5] C. Käs and H. Nicolas, "An approach to trajectory estimation of moving objects in the H.264 compressed domain", in Proc. 3rd Pacific Rim Symp. Adv. Image Video Technol., LNCS 5414, pp. 318–329, 2009.
[6] Bin Qi, Aishy Amer, "Robust and Fast Global Motion Estimation Oriented to Video Object Segmentation", IEEE hal-00348684, version 1, pp.1-12, 2009
[7] Chung-Tao Chu, Dimitris Anastassiou, Shih-Fu Chang, "Hierarchical global motion estimation/compensation in low bitrate video coding", IEEE International Symposium,Vol.2, pp.1149 - 1152, 1997.
[8] Samitha T, Prasanth C.R, Lekshmi P.R, Shanti K.P, "Study of H.264/AVC algorithm and its implementation in matlab",Vol.4,pp 53-68,2014.
[9] LUI Zhi, SHEN Li, ZHANG Zhao, "An efficient compressed domain moving object segmentation algorithm based on motion vector field", pp. 221-227
[10] Seok-Wun Ha,Yong-Ho Moon."Multiple object tracking using SIFT Feature and location Matching",International Journal of smart home,Vol.5,No. 4,October,2011.
[11] "YUV video sequence", [online] http://trace.eas.asu.edu/yuv/ (Accessed :12 feb 2015)

Author Profile

Ms. Kavita Borse is pursuing M.E in Computer Engineering from Pillai College of Engineering, New Panvel. She has 1 year teaching experience in Computer Science.

Prof. Rupali Nikhare is working as assistant professor in Computer Engineering department of Pillai College of Engineering, New Panvel. She obtain ME in Computer Engineering and BE in EXTC. She has 10 years teaching experience. Her area of interest is DSP, IP, CV, and Robotics.