Object Tracking Using Features Extracted From Compressed Domain

Miss. Pratiksha R.Bhalekar [#1], Ms. Vaishali Suryawanshi [*2]
[#]*M.E. Computer, Department of Computer Engineering, Mumbai University, India*
[*]*Assistant Professor, Department of Computer Engineering, Mumbai University, India*

***Abstract*—** *It is a challenging task to develop effective and efficient models for robust object tracking due to factors such as pose variation, illumination changes, occlusion, and movement obscure. Our methodology is Object tracking using features extracted from Compressed Domain. Features are extracted from the compressed domain with a Discrete Cosine Transform. We pack test of pictures of the frontal range target and the establishment using the same Discrete Cosine Transform. The system can be considered as generative because the target can be well represented theoretically with the features generated randomly. It is additionally discriminative since it utilizes these features to discrete the objective from the encompassing foundation. Calculating similarity measure utilizing Euclidian distance. Position tracking after is similarly done using Euclidian separation. These tracking results are compared with mean shift tracking.*

Keywords — *Object Tracking, Discrete Cosine Transform (DCT), Background Subtraction Method, Euclidian distance, Scale Invariant Feature Transform (SIFT)*

I. INTRODUCTION

Object tracking [1][2] is an important task within the field of computer vision. The multiplication of powerful PCs, the accessibility of brilliant and cheap camcorders, and the expanding requirement for computerized feature examination has created a lot of enthusiasm for object tracking algorithms. There are three key strides in video analysis: identification of intriguing moving objects, tracking of such protests from casing to frame to frame, and examination of object tracks to perceive their conduct In its simplest form, tracking can be defined as the problem of estimating the trajectory of an object in the image plane as it moves around a scene. In other words, a tracker assigns consistent labels to the tracked objects in different frames of a video. Additionally, depending on the tracking domain, a tracker can also provide object-centric information, such as orientation, area, or shape of an object [2].

Object tracking, in general, is a challenging problem. Tracking of objects [2] plays assumes a vital part in the security of regular man to a nation. This incorporates following of items like human, military vehicle, interlopers in the outskirt of each nations. Object tracking remains a challenging problem due to appearance change caused by pose, illumination, occlusion, and motion blur, among others. Real Time Compressive tracking [3] is challenging tasks to develop effective and efficient models for robust object tracking due to factors such as pose variation, illumination change, occlusion, and motion blur.

Object tracking [2][3], by definition, is to track an item (or different items) over an arrangement of pictures. Difficulties in tracking objects can arise due to sudden object movement,, changing appearance patterns of the object and the scene, non-unbending object structures, object-to-object and object-to-scene occlusions, and camera motion, illumination changes. What's more, another issue in Object following is Occlusion. Occlusion [2] can be classified into three categories: self-occlusion, interobject occlusion, and occlusion by the background scene structure. Self-occlusion occurs when one part of the object occludes another. This circumstance most frequently arises while tracking articulated objects. Interobject occlusion occurs when two objects being followed block one another.. Similarly, occlusion by the background occurs when a structure in the background occludes the tracked objects.

Tracking objects can be complex due to:

- Loss of data brought on by projection of 3D world on 2D picture
- Noise in images
- Complex object shapes / movement
- Non unbending or enunciated nature of objects
- Partial and full object occlusions
- Scene illumination changes
- Real-time processing requirements

Application areas includes-

- **Human Computer Interaction** (HCI) [4] becomes increasingly important which uses in object tracking.
- **Video tracking** [4] can be a time consuming process due to the amount of data that is contained in video. Adding further to the complexity is the possible need to use object recognition techniques for tracking, a challenging problem in its own right.

- **Medical imaging** [4] is a tracking application. Medical image processing tools are playing an increasingly important role in assisting the clinicians in diagnosis, therapy planning and image-guided interventions
- **Video surveillance** [4], has received growing attention in the last years as an important technology to achieve improved security, namely in large public environments (e.g., shopping malls, railway stations, airports, etc.)

II. PROPOSED WORK

Object tracking [1][2] is to track an object (or multiple objects) over a sequence of images. The aim of an object tracker is to create the direction of an object over time by locating its position in every frame of the video. Object tracker might likewise give the complete region in the image that is occupied by the object at every time instant. The tasks of detecting the object and setting up correspondence between the object instances across frames can either be performed separately or jointly. In the first case, object representation is performed by utilizing rectangular or circular format. Then feature selection is done and after that possible object regions in every frame are obtained by means of an object detection algorithm, and then the tracker corresponds objects across frames. In next case, Object tracking is performed.

General block diagram of proposed system as shown in figure 1.Input video frames are stored in database called test dataset.

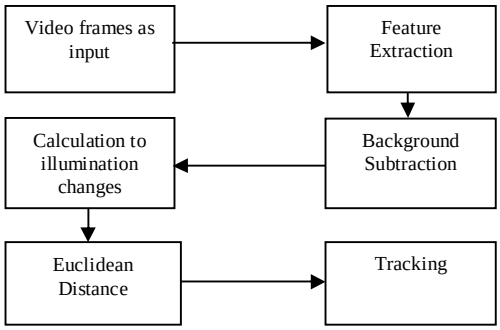


Fig.1. Block diagram of detailed Object tracking system

The system is using DCT (Discrete Cosine Transform Matrix) for extracting features of image.

i. Object Representation

In a following situation, an item can be characterized as anything that is of enthusiasm for further examination. For example, watercrafts on the ocean, angle inside an aquarium, vehicles on a street, planes noticeable all around, individuals strolling on a street, or rises in the water are a situated of items that may be essential to track in a particular space. Items can be spoken to by utilizing rectangular or elliptical format.

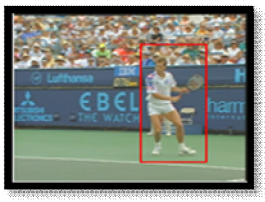


Fig. 2.Object Representation

ii. Feature Extraction

The feature extraction is the important step. Because based on the features extracted from the image, trackng procedure must be finished. Here the feature extraction calculations is Discrete Cosine Transform (DCT)[4][5].

Selecting the right features plays a critical role in tracking. When all is said in done, the most alluring property of a visual component is its uniqueness so that so that the objects can be easily distinguished in the feature space.

Discrete Cosine Transform

Discrete Cosine Transform (DCT) [4][5] to generate the feature vectors for the purpose of search and retrieval of database images. For the DCT change, we change over a HSV picture into dim level picture. For spatial localization, we then use the DC transformation. Each image is resized to *N*N* size. DCT is applied on the image to generate a feature vector as shown in figure 3.

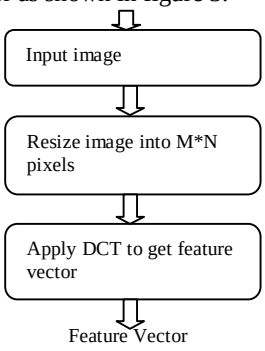


Fig.3. Flowchart for feature extraction

iii. Object Detection

Every tracking method requires an object detection mechanism either in every frame or when the object first shows up in the video. Background Subtraction Method is utilized.

Background Subtraction Method

Background subtraction (BS)[2][6] is a common and widely used technique for generating a foreground mask (namely, a double image containing the pixels belonging to moving objects in the scene) by using static cameras. Object detection can be accomplished by building a representation of the scene called the background model and then discovering deviations from the model for each incoming frame. Any significant change in an image region from the background model signifies a moving object. The pixels constituting the regions experiencing change are marked for further processing. Usually, a connected component algorithm is applied to obtain connected regions corresponding to the objects. This process is referred to as the *background subtraction.*

As the name proposes, BS [2] computes the frontal area veil performing a subtraction between the present casing and a foundation model, containing the static piece of the scene or, more in general, everything that can be considered as background given the characteristics of the observed scene.

Background modeling consists of two main steps:

- Background Initialization;
- Background Update.

In the first step, an initial model of the background is computed, while in the second step that model is updated in order to adapt to possible changes in the scene.

P [F (t)] =P [I (t)]-P [B] (1)

Where, F (t)] =Modified Frame,

I (t)] = Initial Frame,

P [B] = Background Frame

Euclidean Distance

In proposed approach, Feature similarity can be done by Euclidian Distance [7]. And it computes distance between location of Object in current Frame & location of object in Previous Frame**.** Euclidean distance or Euclidean metric is the "ordinary" distance between two points that one would measure with a ruler, and is given by the Pythagorean formula

$$d(q, p) \equiv \sqrt{\sum_{i=1}^{n} (q_i - p_i)} \qquad (2)$$

The similarity function defines a distance among object in current frame and object in previous frame.

III.FRAMEWORK

1. Take Input Video .Separate Frames from the Video.
2. In the first frame (this is reference frame).Select Object Using Rectangular Template.
3. Calculate features of the object in template using Feature Extraction technique- DCT. These features we have to track in subsequent frames.
4. To compute motion/Detection of object Background Subtraction method [3] is used.
5. Subtract each subsequent Frame from Previous Frame.
6. Eliminate small illumination Changes and obtain possible region in the image where the object may be present. For each of the region compute the features using DCT technique.
7. Compute Features of detected objects.
8. Compute spatial location of each region. (Row number, Column number)
9. Compute Euclidian distance for similarity measure.
10. Computes distance between location of Object in current Frame & location of object in Previous Frame
11. Minimum Distance Represent Object.

IV. EVALUATION PARAMETER

In this approach, in this methodology, a Frame-based Performance Evaluation measurement [8] is utilized. Set of performance evaluation metrics have implemented in order to quantitatively analyse the performance of our object detection and tracking system.

Frame-based Metrics

Frame-based metrics [8] are used to measure the performance of surveillance system on individual frames of a video sequence. This does not take into consider the response of the system in preserving the identity of the object over its lifespan. Each frame is exclusively tried to check if the number of objects as well as their sizes and locations match the corresponding ground truth information for that

particular frame. The results from individual frame statistics are then averaged over the whole sequence. This represents a bottom-up methodology.
Beginning with the first frame of the test sequence, frame based metrics [8] are computed for every frame in the sequence. From every frame in the video sequence, first a few true and false detection and tracking quantities are computed.

True Negative, TN: Number of edges where both ground truth and framework results agrees on the nonappearance of any object.

True Positive, TP: Number of edges where both ground truth and framework results agree on the presence of one or more objects, and the bounding box of at least one or more objects coincides among ground truth and tracker results.

False Negative, FN: Number of edges where ground truth contains at least one object, while system either does not contain any object or none of the framework's objects fall within the bounding box of any ground truth object.

False Positive, FP: Number of edges where system results contain at least one object, while ground truth either does not contain any object or none of the ground truth's objects fall within the bounding box of any system object.

In the above definitions, the two bouncing boxes are said to be incidental if the centroid of one of the boxes lies inside the other box. Likewise, aggregate ground truth TG is the aggregate number of frames for the ground truth objects and TF is the aggregate number of frames in the video sequence. Once the above defined quantities are calculated for all the frames in the test sequence, in the second step, the following metrics are computed:

Tracker Detection Rate (TRDR) = TP/TG (3)

False Alarm Rate (FAR) = FP/TP+FP (4)

Detection Rate = TP/TP+FN (5)

Specificity = TN/FP+TN (6)

Accuracy = TP+TN/TF (7)

Positive Prediction = TP/TP+FP (8)

Negative Prediction = TN/FN+TN (9)

False Negative Rate = FN/FN+TP (10)

False Positive Rate = FP/FP+TN (11)

V.COMPARISON

Mean Shift Tracking Using Gaussian Filter [9][10]

Begin from the position of the model in the current frame. Search in the model's neighbourhood in next frame Find best applicant by maximizing a similarity function .Repeat the same process in the next pair of frames.

Feature Extraction Using SIFT

The feature extraction is the important step. Because based on the features extracted from the image, trackng process has to be done. Here the feature extraction algorithms is Scale Invariant Feature Transform (SIFT).

Scale Invariant Feature Transform

Scale Invariant Feature Transform (SIFT)[10][11][12] is an approach for detecting and extracting local feature descriptors that are sensibly invariant to changes in illumination, scaling, rotation, image noise and small changes in viewpoint. This algorithm is initially proposed by David Lowe in 1999, and then further developed and improved.
Detection stages for SIFT features are as follows:

(1)**Scale-space extrema detection**: The first phase of computation searches overall scales and image locations. It is executed efficiently by means of a difference of- Gaussian function to identify potential interest points that are invariant to orientation and scale.

(2) **Keypoint localization**: At each competitor area, a detailed model is fit to determine scale and location. Keypoints are chosen on basis of measures of their stability.
Keypoint confinement: At every competitor area, an itemized model is fit to decide scale and area. Keypoints are chosen on premise of measures of their strength.

(3)**Orientation assignment**: One or more orientations are alloted to each keypoint location on basis of local image gradient directions. Every future operations are performed on image data that has been transformed relative to the assigned scale, orientation, and location for each feature, thereby providing invariance to these transformations.
(4) **Generation of keypoint descriptors**: The local image inclinations are measured at the selected scale in the region around each keypoint. These angles are transformed into a representation which admits significant levels of local change in illumination and shape distortion.

VI.DATASET

The training data set consists of videos of: coastguard, akiyo, Bus, Flower, Big Buck Bunny, Bridge, Stefan, foreman, grandma, waterfall, bridge-close.

VII.EXPERIMENTAL RESULTS AND ANALYSIS

7.1.Mean Shift Tracking [9][10] Result

First step is Selectng a video. In the first frame. Select object using rectangular template.

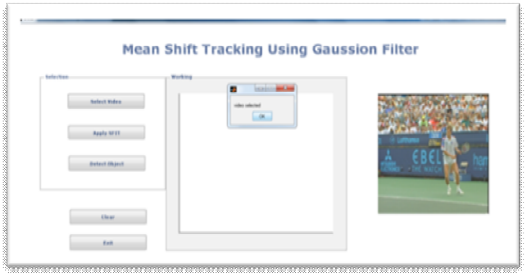


Fig.4.Selection of video

Calculating features of the object in template using Feature Extraction technique-SIFT

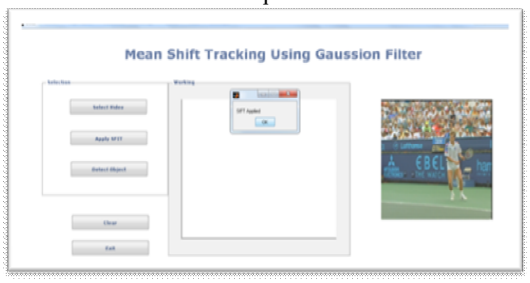


Fig.5.After Selection of video, SIFT applied

To compute motion/Detection of object

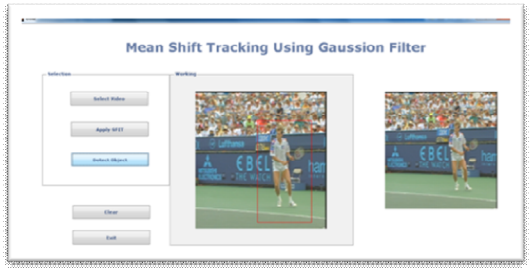
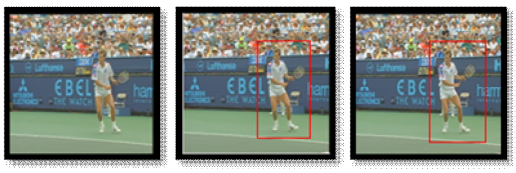


Fig.6. Detection of object



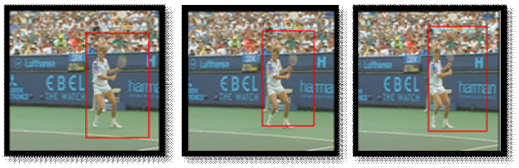


Fig.7 .After Detection, output which shows tracking Object

7.2.Object Tracking Using Features Extracted From Compressed Domain Result

To start with step is selecting a video. In the first casing (this is reference frame).Select Object Using Rectangular Template.

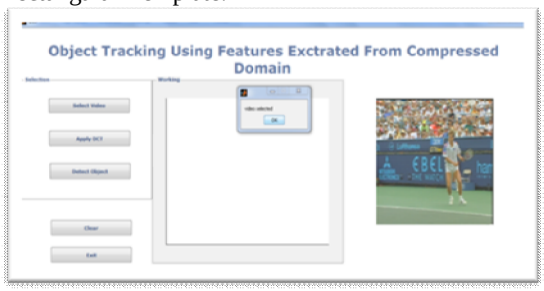


Fig.8.Selection of video

Calculating features of the object in template using Feature Extraction technique- DCT. These features we have to track in subsequent frames.

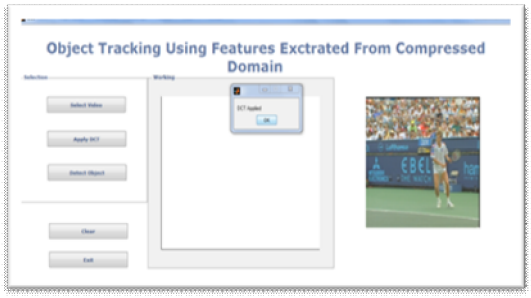


Fig.9.After Selection of video, DCT applied

To compute motion/Detection of object

Fig.10. Detection of object

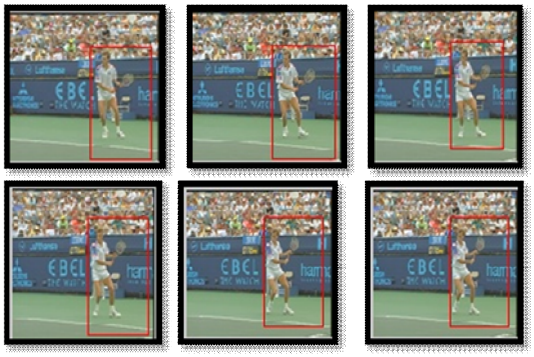


Figure 11 .After Detection, output which shows tracking Object

Analysis of Experimental results

Pose Variation: The object tracking algorithm get feature value.Therefore,this method can handle pose variation better than mean shift tracking using gaussion filter.

Occlusion: Occlusion means that there is something you want to see, but can't because of some property of your sensor setup, or some event. This method handles partial occlusion.

Time complexity: Time taken by object tracking using features extracted by compressed domain is lower than mean shift tracking. Mean shift tracking is very time consuming. And true positive rate is also low.

VII.CONCLUSION

This paper exhibits a methodology for object tracking using features extracted from compressed domain. This way to deal with create powerful and productive models for robust object tracking due to factors such as pose variation, illumination change, occlusion, and motion blur The paper is using highlight extraction strategy feature extraction technique (Discrete Cosine Transform) yields more features are extracted. Ths apprach handles small illumination changes and partial occlusion. This task is utilized as a part of Video tracking.

REFERENCES

[1] He, Yan, et al. "An improved real-time compressive tracking method." Proceedings of the Fifth International Conference on Internet Multimedia Computing and Service. ACM, 2013.

[2 Yilmaz, Alper, Omar Javed, and Mubarak Shah. "Object tracking: A survey."*Acm computing surveys (CSUR)* 38.4 (2006): 13.

[3] Zhang, Kaihua, Lei Zhang, and Ming-Hsuan Yang. "Real-time compressive tracking." Computer Vision–ECCV 2012. Springer Berlin Heidelberg, 2012. 864-877.

[4] Kekre, H. B., Tanuja K. Sarode, and M. S. Ugale. "An efficient image classifier using discrete cosine transform." Proceedings of the International Conference & Workshop on Emerging Trends in Technology. ACM, 2011.

[5] http://en.wikipedia.org/wiki/ Discrete Cosine Transform.

[6] Horprasert, Thanarat, David Harwood, and Larry S. Davis. "A statistical approach for real-time robust background subtraction and shadow detection." *IEEE ICCV*. Vol. 99. 1999.

[7] http://en.wikipedia.org/wiki/Euclidean_distance.

[8] Bashir, Faisal, and Fatih Porikli. "Performance evaluation of object detection and tracking systems." *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS)*. Vol. 5. 2006.

[9] COMANICIU, D. AND MEER, P. 1999. Mean shift analysis and applications. In *IEEE International Conferenceon Computer Vision (ICCV)*. Vol. 2. 1197–1203.

[10] http://en.wikipedia.org/wiki/Scale invariant_feature_transform [12] S.-B. Cho and J.-Y. Lee, "A human-oriented image retrieval system using interactive genetic algorithm," *IEEE Trans. Syst., Man, Cybern. A,Syst.,Humans*, vol. 32, no. 3, pp. 452–458, May 2002.

[11] G. David Lowe. Object recognition from local scale-invariant features. Proceedings ofthe International Conference on Computer Vision. 2. pp. 1150–1157,1997.

[12] Zhu, Chaoyang. "Video object tracking using SIFT and mean shift." (2011).

[13] Quast, Katharina, and André Kaup. "Shape adaptive mean shift object tracking using gaussian mixture models." *Analysis, Retrieval and Delivery of Multimedia Content*. Springer New York, 2013. 107-122.