# Occlusion Aware Online Deformable Object Tracking Based On Structure-Aware Hyper-Graph

[1]M.Gayathri
Assistant Professor, Dept. of ECE
K.Ramakrishnan College of Technology
Trichy, India

[2]B.Revathi,[3]J.Selvin Prabha, [4]M.Rajeswari
UG Student, Dept. of ECE
K.Ramakrishnan College of Technology
Trichy, India

***Abstract* – Recent advances in online visual tracking focus on designing part-based model to handle the deformation and occlusion challenges. This paper describes a new and efficient method for online deformable object tracking. Different from most existing methods, this paper exploits higher order structural dependences of different parts of the tracking target in multiple consecutive frames. In this project, we present a robust tracking method by exploiting a fragment-based appearance model with consideration of both temporal continuity and discontinuity information. In the first stage, by adopting the estimated occlusion state as a prior, the optimal state of the tracked object can be obtained by solving an optimization problem, where the objective function is designed based on the classification score, occlusion prior, and temporal continuity information. In the second stage, we propose a discriminative occlusion model, which exploits both foreground and background information to detect the possible occlusion, and also models the consistency of occlusion labels among different frames. The experimental result of the proposed method shows considerable improvement in performance over the state-of-the-art tracking methods.**

***Index Terms*— Online Tracking, Deformable Object Tracking, Classification Score, Spatial Temporal Consistency, Occlusion Model Map, And Temporal Continuity Information.**

## I. INTRODUCTION

Online visual tracking is an important step toward fully automatic understanding of videos, which finds wide applications in video surveillance, behavior analysis, human computer interaction, to name a few. In spite of significant progress in the recent years, online tracking a deformable object accurately remains a difficult problem. Challenges in online visual tracking originate from large variations of the tracking target in appearance, shape, and motion, as well as the occlusions caused by other objects or background. Currently, the predominant approaches in online visual tracking aim to obtain a bounding box of the tracking target. Many methods are built on models which focus on capturing appearance variation of the tracking target in the bounding box, which are discussed in related works. In this paper, we propose a new online deformable object tracking method based on a structure-aware hyper-graph, spatial temporal consistency and occlusion model map, which can effectively incorporate higher-order dependencies among more than two consecutive frames, which is discussed in proposed work.

## II. RELATED WORK

Optical flow estimation is classically marked by the requirement of dense sampling in time. While coarse-to-fine warping schemes have somehow relaxed this constraint, there is an inherent dependency between the scale of structures and the velocity that can be estimated. This particularly renders the estimation of detailed human motion problematic, as small body parts can move very fast. In [1] Thomas Brox and Jitendra Malik presented a way to approach this problem by integrating rich descriptors into the variational optical flow setting. This way we can estimate a dense optical flow field with almost the same high accuracy as known from variational optical flow, while reaching out to new domains of motion analysis where the requirement of dense sampling in time is no longer satisfied.

In paper [5] Luka Cehovin, Matej Kristan and Ales Leonardis addresses the problem of tracking objects which undergo rapid and significant appearance changes. We propose a novel coupled-layer visual model that combines the target's global and local appearance by interlacing two layers. The local layer in this model is a set of local patches that geometrically constrain the changes in the target's appearance. This layer probabilistically adapts to the target's geometric deformation, while its structure is updated by removing and adding the local patches. The addition of these patches is constrained by the global layer that probabilistically models the target's global visual properties, such as color, shape, and apparent local motion. The global visual properties are updated during tracking using the stable patches from the local layer. By this coupled constraint paradigm between the adaptation of the global and the local layer, we achieve a more robust tracking through significant appearance changes.The experimental results on challenging sequences confirm that our tracker outperforms the related trackers in many cases by having a smaller failure rate as well as better accuracy. Furthermore, the parameter analysis shows that our tracker is stable over a range of parameter values.

In [7] Stefan Duffner and Christophe Garcia presented a novel algorithm for fast tracking of generic objects in videos. The algorithm uses two components: a detector that makes use of the generalized Hough transform with pixel-based descriptors, and a probabilistic segmentation method based on global models for foreground and background. These components are used for tracking in a combined way, and they adapt each other in a co-training manner. Through effective model adaptation and segmentation, the algorithm is able to track objects that undergo rigid and non-rigid deformations and considerable shape and appearance variations. Finally, the proposed models allow for an extremely efficient implementation, and thus tracking is very fast.

In [8] an efficient and scalable technique for spatiotemporal segmentation of long video sequences using a hierarchical graph-based algorithm was proposed by Matthias Grundmann, Vivek Kwatra, Mei Han and Irfan Essa. We begin by over-segmenting a volumetric video graph into space-time regions grouped by appearance. We then construct a “region graph” over the obtained segmentation and iteratively repeat this process over multiple levels to create a tree of spatio-temporal segmentations. This hierarchical approach generates high quality segmentations, which are temporally coherent with stable region boundaries, and allows subsequent applications to choose from varying levels of granularity. We further improve segmentation quality by using dense optical flow to guide temporal connections in the initial graph. We also propose two novel approaches to improve the scalability of our technique: (a) a parallel out-of-core algorithm that can process volumes much larger than an in-core algorithm, and (b) a clip-based processing algorithm that divides the video into overlapping clips in time, and segments them successively while enforcing consistency. We demonstrate hierarchical segmentations on video shots as long as 40 seconds, and even support a streaming mode for arbitrarily long videos, albeit without the ability to process them hierarchically.

In paper [11], the problem of tracking an object in a video given its location in the first frame and no other information was addressed by Boris Babenko, Ming-Hsuan Yang and Serge Belongie. Recently, a class of tracking techniques called “tracking by detection” has been shown to give promising results at real-time speeds. These methods train a discriminative classifier in an online manner to separate the object from the background. This classifier bootstraps itself by using the current tracker state to extract positive and negative examples from the current frame. Slight inaccuracies in the tracker can therefore lead to incorrectly labeled training examples, which degrade the classifier and can cause drift. In this paper, we show that using Multiple Instance Learning (MIL) instead of traditional supervised learning avoids these problems and can therefore lead to a more robust tracker with fewer parameter tweaks. We propose a novel online MIL algorithm for object tracking that achieves superior results with real-time performance. We present thorough experimental results (both qualitative and quantitative) on a number of challenging video clips.

## III. PROPOSED FRAMEWORK

A new online deformable object tracking method based on a structure-aware hyper-graph, which can effectively incorporate higher-order dependencies among more than two consecutive frames. As such, we refer to our method as structure aware tracker (SAT) subsequently. In our method, the tracking target is represented by multiple parts, which are ensembles of super-pixels that are similar in appearance and motion. To find such parts, we first apply the SLIC over-segmentation algorithm to generate super-pixels in each video frame, and then apply the graph cut algorithm to optimize an energy objective function to produce candidate parts.

Then, we construct a hyper-graph to capture the higher-order dependencies among candidate parts across multiple frames. Specifically, the nodes of the hyper-graph correspond to the candidate parts, and the hyper-edges encode the higher-order dependencies (as consistencies in both appearance and motion) of the candidate parts across multiple frames. We first segment each frame into super-pixels, and collect candidate parts in each frame of a frame buffer by the MRF based segmentation method.

After that, we construct a structure-aware hyper-graph, whose nodes correspond to the candidate parts in a frame buffer and hyper-edges correspond to the higher-order dependencies among the parts. We then group super-pixels into sub-graphs with appearance and motion-consistent target parts corresponding to the object across multiple frames. Assembling all parts belonging to the target, we find the precise location and boundary of the target and output the location of the target.

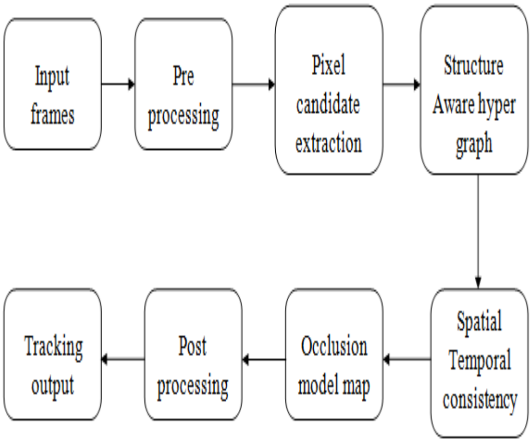

Input frames
Pre processing
Pixel candidate extraction
Structure Aware hyper graph
Spatial Temporal consistency
Occlusion model map
Post processing
Tracking output

Fig. 3.1: Block Diagram of the Proposed System

**Pre processing**

An input frame is dataset and the pre processing read all frames simultaneously that is frame buffer. Online updating is an important step to prevent drifting problem for online tracking. For better efficiency, the frame buffer is implemented as a queue. We add the head frame (EnQueue) and drop the tail frame (DeQueue) to collect a new frame buffer.

**Pixel candidate extraction**

Extraction using MRF based segmentation to accommodate the online nature of the tracking algorithm; we use a frame buffer to predict the location, appearance and boundary of the target. We then form the energy function to generate candidate parts. A Markov Random Field (MRF) is a graphical model of a joint probability distribution. It consists of an undirected graph in which the nodes represent random variables. The unary energy corresponding to likelihood of super-pixel p belonging to foreground (p = 1) or background (p = 0).

The binary energy encoding the consistency between pairs of spatially neighboring super-pixels, which encourages the target to be a collection of connected parts with similar appearance. Where N is the spatial neighborhood of super-pixels: two super-pixels p and q are in N if the Euclidean distance between their centers in the image plane satisfies, N is the number of pixels in each super-pixel, where ρ is the number of super-pixels in the searching window with width W and height H. The energy function is minimized with the graph cut algorithm, leading to a coarse labeling of each super-pixel as belonging to the target and the background.

**Structure-aware hyper-graph**

We refer to our method as structure aware tracker (SAT) subsequently. In our method, the tracking target is represented by multiple parts, which are ensembles of super-pixels that are similar in appearance and motion. To find such parts, we first apply the SLIC over-segmentation algorithm to generate super-pixels in each video frame, objective function to produce candidate parts.

Then, we construct a hyper-graph to capture the higher-order dependencies among candidate parts across multiple frames. Specifically, the nodes of the hyper-graph correspond to the candidate parts, and the hyper-edges encode the higher-order dependencies (as consistencies in both appearance and motion) of the candidate parts across multiple frames. We adopt a pair wise updating algorithm to extract parts belonging to the target. Finally, the target state (i.e., the center and scale of the target) is determined by comprehensively analyzing the searched part states.

Given the collected candidate parts, we construct the structure-aware hyper-graph encoding the dependencies among candidate parts. Specifically, each node in the node set V corresponds to a candidate part, and each hyper-edge in E represents the relations among the nodes The collected sequences are diverse with respect to object categories, camera viewpoints, sequence lengths and challenging levels. Different from 11 attributes for general object tracking in, our dataset includes videos reflecting typical challenges in tracking large deformation.

The non-rigid target occurs with local structural or significant deformation in shape. Severe occlusion the target is partially or fully occluded by other objects or background. Abnormal movement, the target moves abnormally, including fast motion, in-plane and out-of-plane rotation and other complex motions, illumination variation. The illumination in the target region is moderately to significantly change. Scale change. The scale of the target changes drastically, background clutter the background near the target has the similar appearance as the target.

### Hyper graph segmentation

Then the structure-aware hyper-graph segmentation using pair wise algorithm is updated by extracting candidate parts in the new frame buffer. Given the collected candidate parts, we construct the structure-aware hyper-graph encoding the dependencies among candidate parts. Specifically, each node in the node set corresponds to a candidate part, after constructing the hyper-graph; we search the dense subgraphs on it to determine the state of each target part. Image segmentation has come a long way. Using just a few simple grouping cues, one can now produce rather impressive segmentation on a large set of images. Behind this development, a major converging point is the use of graph based technique. Graph cut provides a clean, flexible formulation for image segmentation. It provides a convenient language to encode simple local segmentation cues, and a set of powerful computational mechanisms to extract global segmentation from these simple local (pairwise) pixel similarities. Computationally graph cuts can be very efficient.

### Sub graph

After constructing the hyper-graph, we search the dense subgraphs on it to determine the state of each target part. We call a subgraph of G “dense” if its nodes are inter-connected by a large set of hyper-edges with high weights.

*Appearance Weight*: The appearance weight is computed from HSV color features of vi and v j square distance of features of two nodes, and σ controls the sensitivity of the appearance term. The visual appearance of objects is given by the way in which they reflect and transmit light. The color of objects is determined by the parts of the spectrum of (incident white) light that are reflected or transmitted without being absorbed.

*Motion Weight*: The motion weight among nodes coupled in the hyper-edge provides an important cue for grouping nodes into subgraphs. Based on the assumption that target parts move smoothly in a short time interval, we compute motion weight based on fitting their motions using a simple linear model. Specifically, for each hyper-edge e, the parameters of the linear model are determined based on every node in by least squares fitting.

### Occlusion Model

In this work, the occlusion model is developed in a discriminative manner rather than a generative manner, i.e., models the posterior probability directly rather than model the likelihood and prior separately. To be specific, we introduce an occlusion score function, to model the occlusion labels of different fragments. By assuming that occlusion labels of different fragments are mutually independent Thus, the maximum of the objective function can be converted into the maximum of all sub-objective functions for each fragment.

### Post processing

The green and red rectangle represents sampled and optimal target state, respectively. The output from the algorithm is the target location. The plot rectangle box,

- Search window,
- Ground truth,
- Tracking position.

### Search Window

A search window is a way of locating matching macro blocks in a sequence of digital video frames for the purposes of motion estimation. The underlying supposition behind motion estimation is that the patterns corresponding to objects and background in a frame of video sequence move within the frame to form corresponding objects on the subsequent frame. This can be used to discover temporal redundancy in the video sequence, increasing the effectiveness of inter-frame video compression by defining the contents of a macro block by reference to the contents of a known macro block which is minimally different.

### Ground Truth

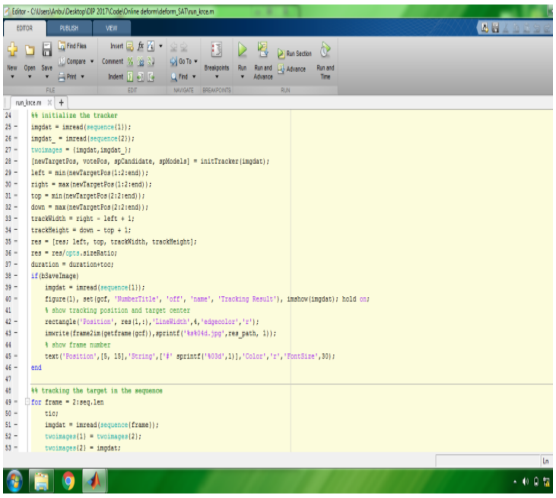
Ground truth means a set of measurements that is known to be much more accurate than measurements from the system you are testing.

### Tracking position

Generally a tracking system is used for the observing of persons or objects on the move and supplying a timely ordered sequence of respective location data to a model e.g. capable to serve for depicting the motion on a display capability.

## IV. SIMULATION RESULTS

1. Script


Editor - C:\Users\Anbu\Desktop\DIP 2017\Code\Online deform\deform_SAT\run_krce.m
EDITOR
PUBLISH
VIEW
run_krce.m

```
%% initialize the tracker
imgdat = imread(sequence{1});
imgdat_ = imread(sequence{2});
twoimages = {imgdat,imgdat_};
[newTargetPos, votePos, spCandidate, spModels] = initTracker(imgdat);
left = min(newTargetPos(1:2:end));
right = max(newTargetPos(1:2:end));
top = min(newTargetPos(2:2:end));
down = max(newTargetPos(2:2:end));
trackWidth = right - left + 1;
trackHeight = down - top + 1;
res = [res; left, top, trackWidth, trackHeight];
res = res/opts.sizeRatio;
duration = duration+toc;
if(bSaveImage)
    imgdat = imread(sequence{1});
    figure(1), set(gcf, 'NumberTitle', 'off', 'name', 'Tracking Result'), imshow(imgdat); hold on;
    % show tracking position and target center
    rectangle('Position', res(1,:),'LineWidth',4,'edgecolor','r');
    imwrite(frame2im(getframe(gcf)),sprintf('%s%04d.jpg',res_path, 1));
    % show frame number
    text('Position',[5, 15],'String',['#' sprintf('%03d',1)],'Color','r','FontSize',30);
end

%% tracking the target in the sequence
for frame = 2:seq.len
    tic;
    imgdat = imread(sequence{frame});
    twoimages{1} = twoimages{2};
    twoimages{2} = imgdat;
```

Fig. 4.1 Coding

2. Output frame

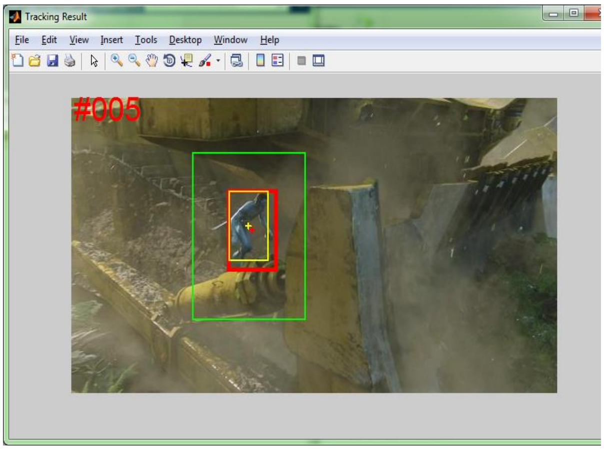


Fig. 4.2 Tracks the object of the video in 5th frame

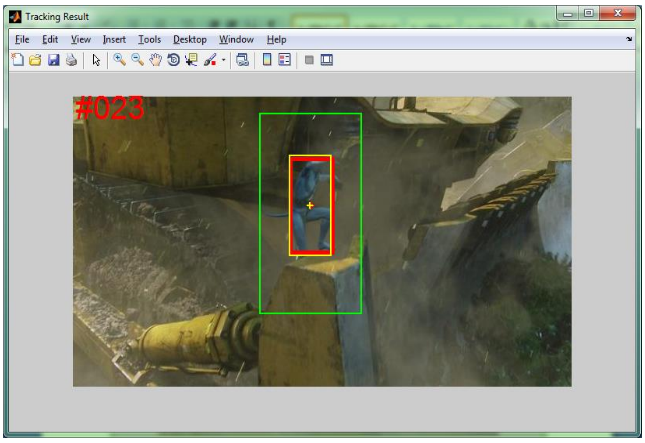


Fig. 4.3 Tracks the object of the video in 23rd frame

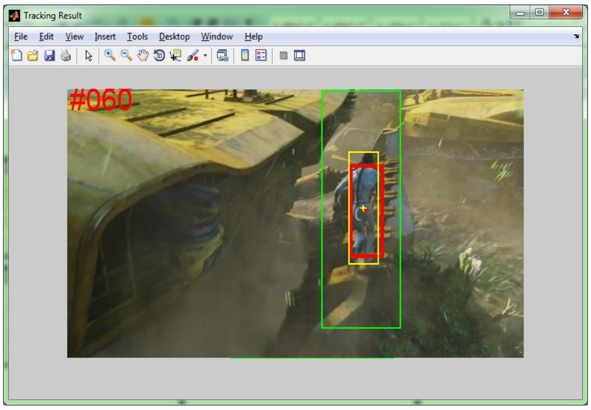


Fig. 4.4 Tracks the object of the video in 60th frame

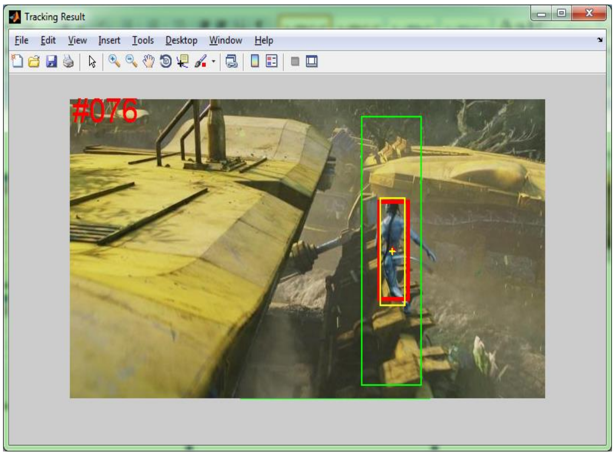


Fig. 4.5 Tracks the object of the video in 76th frame

## V.CONCLUSION

In this Project, we describe a structure-aware hyper-graph based tracker. Our method formulates the tracking task as the dense sub graph searching problem on the dynamically constructed hyper-graph integrating the higher-order structural dependencies in temporal domain. The optimal target state is determined by extracting dense subgraphs using a coarse-to- fine strategy.

We demonstrate the effectiveness of our method and compare its performance with that of state-of-the-art online tracking methods on the Deform-SOT dataset. There are a few directions we would like to further extend the current work. First, in the current method, we only consider temporal higher-order dependencies among the parts. As a next step, we will also investigate incorporating spatial higher order dependencies among the parts

## REFERENCES

[1] Thomas Brox and Jitendra Malik, "Large Displacement Optical Flow: Descriptor Matching in Variational Motion Estimation," IEEE Transactions on Pattern Analysis and Machine Intelligence Vol. 33, No. 3, March 2011.

[2] T. Brox and J. Malik, “Large displacement optical flow: Descriptor matching in variational motion estimation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 3, pp. 500–513, Mar. 2011.

[3] L. Cehovin, M. Kristan, and A. Leonardis, “Robust visual tracking using an adaptive coupled-layer visual model,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 4, pp. 941–953, Apr. 2013.

[4] S. Duffner and C. Garcia, “Pixel Track: A fast adaptive algorithm for tracking non-rigid objects,” in Proc. IEEE Int. Conf. Comput.Vis., Dec. 2013, pp. 2480–2487.

[5] Luka Cehovin, Matej Kristan 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, April 2013.

[6] M. Grundmann, V. Kwatra, M. Han, and I. Essa, “Efficient hierarchical graph-based video segmentation,” in Proc. IEEE Conf. Comput. Vis.Pattern Recognit., Jun. 2010, pp. 2141–2148.

[7] Stefan Duffner and Christophe Garcia, "PixelTrack: A Fast Adaptive Algorithm for Tracking Non-rigid Objects," IEEE International Conference on Computer Vision (ICCV), Dec. 2013

[8] Matthias Grundmann, Vivek Kwatra, Mei Han and Irfan Essa, "Efficient hierarchical graph-based video segmentation," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2010

[9] D. A. Klein and A. B. Cremers, “Boosting scalable gradient features for adaptive real-time tracking,” in Proc. IEEE Int. Conf. Robot. Autom, May 2011, pp. 4411–4416.

[10] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, “SLIC superpixels compared to state-of-the-art superpixel methods,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 11, pp. 2274–2282, Nov. 2012.

[11] Boris Babenko, Ming-Hsuan Yang and Serge Belongie, "Robust Object Tracking with Online Multiple Instance Learning," IEEE Transactions on Pattern Analysis and Machine Intelligence Vol. 33, No. 8, Aug. 2011

[12] W. Zhong, H. Lu, and M.-H. Yang, “Robust object tracking via sparsity based collaborative model,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2012, pp. 1838–1845.

[13] K. Zhang, L. Zhang, Q. Liu, D. Zhang, and M.-H. Yang, “Fast visual tracking via dense spatial-temporal context learning,” in Proc. Eur. Conf.Comput. Vis., 2014, pp. 127–141.

[14] L. Wen, W. Li, J. Yan, Z. Lei, D. Yi, and S. Z. Li, “Multiple target tracking based on undirected hierarchical relation hyper graph,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 1282–1289.

[15] L. Wen, D. Du, Z. Lei, S. Z. Li, and M.-H. Yang, “JOTS: Joint online tracking and segmentation,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2015, pp. 2226–2234.