Object Detection and Tracking Using Appearance Model and Video Object Graph Cut Method

Mr. S. S. Athatey  
SGBAU, Amravati, India

Prof. Y. M. Kurwade  
SGBAU, Amravati, India

Dr. V. M. Thakare  
SGBAU, Amravati, India

Abstract: Object tracking and detection is a challenging topic in computer vision. Due to large appearance variations in visual objects such as varying deformations, illuminations, out-of-plane rotations, occlusions, and cluttered backgrounds, it is difficult to detect the object and track it. This paper proposes to combine appearance model with video graph cut method for object detection and tracking. For severely blurred input videos, the appearance model is robust. A video object graph cut method is applied to construct a foreground salience graph to describe the salience of an image patch in the spatiotemporal domain. Cross frame object verification will then verify foreground patches and background patches from input scene and detect the exact object from video input.

Index Terms – Appearance model, video object graph cut, motion blur, tracking, object detection.

I. INTRODUCTION

Object tracking and detection is a challenging topic in computer vision. It is important for many visual applications, such as video surveillance, human–computer interaction, vehicle navigation, and augmented reality. Due to large appearance variations in visual objects such as varying deformations, illuminations, out-of-plane rotations, occlusions, and cluttered backgrounds, it is difficult to detect the object and track it. Recently, many generic object tracking and detection algorithms have been proposed. To design a robust generic tracking algorithm, many efforts have been invested on handling background clutter, noises, illumination change, occlusion, and so on. Video graph cut method is proposed to track and detect the animal as object. Cross frame verification method verifies the background and foreground object in video scenes [1]. A Tracking-Learning-Data architecture is proposed to track an arbitrary object. A new low-rank and sparse representation model for moving object detection [4] was proposed, it operates the low-rank background and sparse foreground decomposition in the tensor framework. An appearance model is used to detect the objects from highly blurred video scenes [2]. A generalized pooling [3] method based on a probabilistic distribution function is extract summary features for sparse codes of target patches for object tracking. A mathematically principled approach [5] was proposed to accounting for the relationship between flows representing the motions of different object types.

In this paper a methodology proposed to combine appearance model [2] with video graph cut method [1] for object detection and tracking. As appearance model is composed of a discriminative classifier. It computes the appearance possibility p(o|Xt). A discriminative method treats tracking as a binary classification problem between objects and backgrounds. Thus, it performs better to separate an object from its background. For severely blurred input videos, the appearance model is robust. A video object graph cut method is applying to construct a foreground salience graph to describe the salience of an image patch in the spatiotemporal domain. Cross frame object verification will then verify foreground patches and background patches from input scene and detect the exact object from video input.

II. BACKGROUND

Animal object detection from highly cluttered natural scenes using IEC which jointly consider the animal motion and spatiotemporal context in the spatiotemporal domain. The DCNN image features and handcrafted histogram of oriented gradient (HOG) image features encoded with Fisher Vectors (FV) are able to enhance the classification performance for each other. [1].
A blur-invariant object tracker without deblurring image sequences. Learning is the task of maintaining the appearance model of the object, which can deal with appearance changes in the challenging conditions. Author consider two types of data. One comes from the natural images, which are the subset of the tiny images data set and used to learn a generic representation of natural scenes. The other one comes from the initial labeled instance and subsequent tracking results [2].

A generalized pooling method based on a probabilistic distribution function is proposed to extract summary features for sparse codes of target patches for object tracking. Propose method extract Fisher vectors from sparse codes to derive compact and discriminative visual descriptors [3].

A new low-rank and sparse representation model for moving object detection. It operates the low-rank background and sparse foreground decomposition in the tensor framework. The 3D-LATV takes the local spatio-temporal geometric structure of videos into account, and then it is combined with the 1 norm to construct the fused-sparse regularization for the foreground. To solve low-rank and sparse decomposition model, employ the augmented Lagrange multiplier method (ALM) to alternatively update the background and foreground [4].

A mathematically principled approach to accounting for the relationship between flows representing the motions of different object types, especially with regard to their container or container relationship and appearance or disappearance, as well as an efficient tracklet-based implementation that yields real-time performance [5].

The rest of the paper is organized as follows. In this paper, Section II gives us background details, Section III provides work which is done previously, Section IV gives idea about existing technology, in Section V analysis and discussion about techniques is carried out, proposed methodology is explained in Section VI. Possible outcomes and Result is described in Section VII. Section VIII concludes the paper. Finally, Section IX described future scope of the paper.

III.PREVIOUS WORK DONE

Zhi Zhang et.al.(2016)[1] proposes animal object detection from highly cluttered natural scenes using IEC which jointly consider the animal motion and spatial context in the spatiotemporal domain. For camera-trap images, the DCNN image features and handcrafted histogram of oriented gradient (HOG) image features encoded with Fisher Vectors (FV) are able to enhance the classification performance for each other.

For severely blurred videos Jianwei Ding et.al.(2016) [2] proposes a new Tracking-Learning-Data architecture to track an arbitrary object. In this work author provides a blur-invariant object tracker without deblurring image sequences. Learning is the task of maintaining the appearance model of the object, which can deal with appearance changes in the challenging conditions. Author consider two types of data. One comes from the natural images, which are the subset of the tiny images data set and used to learn a generic representation of natural scenes. The other one comes from the initial labeled instance and subsequent tracking results.

A generalized pooling method was proposed by Bo Ma et.al.(2016)[3] which based on a probabilistic distribution function is proposed to extract summary features for sparse codes of target patches for object tracking. Propose method extract Fisher vectors from sparse codes to derive compact and discriminative visual descriptors.

Wenrui Hu et.al.(2017)[4] proposes a new low-rank and sparse representation model for moving object detection. It operates the low-rank background and sparse foreground decomposition in the tensor framework. The 3D-LATV takes the local spatio-temporal geometric structure of videos into account, and then it is combined with the 1 norm to construct the fused-sparse regularization for the foreground. To solve low-rank and sparse decomposition model, employ the augmented Lagrange multiplier method (ALM) to alternatively update the background and foreground.

Xinchao Wang et.al.(2016)[5] proposed a mathematically principled approach to accounting for the relationship between flows representing the motions of different object types, especially with regard to their container or container relationship and appearance or disappearance, as well as an efficient tracklet-based implementation that yields real-time performance.

IV.EXISTING METHODOLOGIES

A. Spatiotemporal object region proposals and patch verification : In this methoda new and
efficient animal object region proposal method using IEC which jointly consider the animal motion and spatial context in the spatiotemporal domain. A cross-frame image verification method for accurate animal-background classification. For camera-trap images, the DCNN image features and handcrafted histogram of oriented gradient (HOG) image features encoded with Fisher Vectors (FV) are able to enhance the classification performance for each other. IEC for animal object proposals the foreground-background probability map for each frame. Specifically, for each pixel $x$, its foreground probability is defined as

$$p(x) = 0.5 + \frac{1}{\pi} \tan^{-1}(\beta \cdot s \cdot d[x])$$

where $d[x]$ is the minimum distance from the pixel $x$ to the boundary of the foreground regions, $s$ is 1 if $x$ is a foreground pixel, otherwise $s$ is -1. $\beta$ is a constant controlling the transition range. In this work, set $\beta$ to be 0.5. For each image patch $P$ at location $x_P$, use this probability map to modulate the FSG for graph cut of the new iteration.

B. Network-flow mixed integer programming framework :-
This methodology contains the following modules

i)Grouping spatial vertices into tracklets :- Partition the spatial vertices of the trajectories introduced in Pruned Interwined flows into two subsets, the joint vertices $S$ and the non joint ones $K$. The vertices in $S$ are those included in the neighborhood of more than one trajectory and those at the beginning or end of a trajectory. The vertices in $K$ are the remaining ones and are located between joint vertices.

ii)Computing the poses of the tracklets :- Each spatial tracklet $t_q$ can be treated as a single spatial vertex. The pose of such vertex is a pose tracklet, which is a set of possible pose vertices on $t_q$. Since $t_q$ connects to other vertices only through the pose vertices at its first frame $t_q$ and last frame $T_q$, the pose of $t_q$ is uniquely defined by its starting and ending pose vertices at $t_q$ and $T_q$ respectively.

iii)Constructing the tracklet graph :- Now construct a tracklet graph $G'$ by treating each pose tracklet as a single vertex and taking the edges to be allowable transitions between them.

C. Tracking-Learning-Data architecture :-
The proposed method have three modules name as tracking, learning and data. Tracking is the task of estimating object motion. Learning is the task of maintaining the appearance model of the object, which can deal with appearance changes in the challenging conditions.

i)Tracking :- Tracking is the task of estimating object motion.

ii)Learning :- Learning is the task of maintaining the appearance model of the object, which can deal with appearance changes in the challenging conditions.

iii)Data :- Author consider two types of data. One comes from the natural images, which are the subset of the tiny images data set and used to learn a generic representation of natural scenes. The other one comes from the initial labeled instance and subsequent tracking result.

Algorithm :- A blur invariant object tracker.

D. A low-rank and sparse representation model :- Author propose a new RPCA formulation for MOD by modeling both background and foreground in the tensor framework, which preserves the spatiotemporal structure of the video. Designing the SFS regularization for the foreground. The SFS not only can distinguish the foreground motion from the background motion efficiently by the saliency measurement, but also provides a more finer spatiotemporal smooth constraint of the foreground by the 3D-LATV than the general 3D-TV. Finally author gives globally optimal solution of our low-rank and sparse decomposition model with rigid theoretical proofs. For TLFS and TLFSFSD, empirically set the sparse parameter $\alpha = 0.4/\sqrt{n1n2n3}$, the smooth parameter $\beta = 4/\sqrt{n1n2n3}$, the accelerated parameter $\delta = 1.1$, and the penalty scalars $\mu = 10^{-4}$and $\lambda = 10^{-2}$. The de-noising iteration is given by the equation’s

$$\mathcal{Y}^{l+1} = \mathcal{M}_2 + \frac{2\delta}{\mu} \text{div} \mathcal{Y}^{l+1}$$

To calculate the 3D-LARKs, fix $h = 1, \xi' = 1, \xi'' = 10^{-3}, \eta = 0.3$ (typically $0 \leq \eta \leq 0.5$). The neighbor size $N_{xi}$ decides the size of 3D-LARKs, and the
computation of 3D-LATV increases linearly with the increase of Nxi. Empirically found that the proposed algorithm achieved the best performance when each dimension of Nxi is smaller than 5. Hence, set Nxi = 3 × 3 × 3 in all experiments by taking trade off computation and performance.

E. Generalized Pooling for Robust Object Tracking :- A generalized pooling method based on a probabilistic distribution function is proposed to extract summary features for sparse codes of target patches for object tracking. Propose method extract Fisher vectors from sparse codes to derive compact and discriminative visual descriptors. Generalized pooling method is instantiate by developing a Fisher tracker, where GMM models sparse code vector distributions, and a semi-supervised Fisher kernel classifier is utilized for classification.

**Algorithm:- Generalized Pooling for Robust Object Tracking.**

V. ANALYSIS AND DISCUSSION

This section present some observation about the above existing visual tracking method. The table shows the comparison between five existing methods and also shows advantages and disadvantages of five methods.

<table>
<thead>
<tr>
<th>Object Detection and Tracking Technique</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatiotemporal object region proposals and patch verification</td>
<td>Using a sequence of energy levels, the proposed IEC method is also able to address the over-proposal issue in existing methods. The spatiotemporal object proposal and patch verification framework is sensitive to objects in motion and confident in rejecting false alarms, thus is capable of building the basis of a robust object detection system in dynamic scenes.</td>
<td>The proposed method require effective cross frame image verification to determine if an image patch belongs to the background or not.</td>
</tr>
<tr>
<td>Network-flow mixed integer programming framework</td>
<td>The pruning reduces the number of variables by three orders of magnitude and the number of kept spatial locations is about 10 times as large as that of the ground truth. The proposed algorithm requires low computational time. The proposed algorithm have low computational cost.</td>
<td>Probability occupancy map (POM) keep those pose nodes, for which one of the occupancy probabilities rtk or btk is greater than 0.5, and suppress the others. The resulting detections lack temporal consistency and may not satisfy the constraints introduced. In the CarPeople dataset sometimes algorithm produces wrong result.</td>
</tr>
<tr>
<td>Tracking-Learning-Data architecture</td>
<td>The proposed method tracks a severely blurred object without de-blurring the image sequences. The propose method have prior knowledge of object tracking. The method is more robust than other.</td>
<td>The proposed algorithm requires large learning dataset for knowledge purpose before tracking. If prior knowledge is not accurate then it produces wrong result.</td>
</tr>
<tr>
<td>A low-rank and sparse representation model</td>
<td>The method have the better performance over other method’s. It take the data local distribution and into consideration this new local metric learning method into target tracking leads to efficient and robust tracking performance.</td>
<td>In the proposed metric adjustment method, one critical issue is the determination of right scale for the local region, because different volumes of the local regions cover different sets of training samples, and thus different local data distributions.</td>
</tr>
<tr>
<td>Generalized Pooling for Robust Object Tracking</td>
<td>To handle the drifting problem during the tracking process, these classifiers are updated online with current tracking results. The propose method is also useful to track object’s from blurred scenes.</td>
<td>The proposed method have high computational cost.</td>
</tr>
</tbody>
</table>

**TABLE 1: Comparisons between different Object Detection and tracking techniques**
VI. PROPOSED METHODOLOGY
This section proposed to combine appearance model with video graph cut method for object detection and tracking. As appearance model is composed of a discriminative classifier. It computes the appearance possibility $p(o|X_t)$. A discriminative method treats tracking as a binary classification problem between objects and backgrounds. Thus, it performs better to separate an object from its background. For severely blurred input videos, appearance model is robust. A video object graph cut method is applying to construct a foreground salience graph to describe the salience of an image patch in the spatiotemporal domain. Cross Frame object verification will then verifies foreground patches and background patches from input scene and detect the exact object from video input.

Algorithm for proposed method for object detection:

Step 1: First convert an input signal or video into video sequences.

Step 2: Apply the appearance modeling to input video to remove any motion blur effect.

Step 3: Apply the video object graph cut method to construct a foreground salience graph to describe the salience of an image patch in the spatiotemporal domain.

Step 4: After detecting the object create a bounding box to denote object and then track it.

VII. POSSIBLE OUTCOMES AND RESULTS
Proposed method don’t required the separate deblurring function for removing the blur effect from input video, this leads to increase performance of our method. Also appearance model is able to find the object from severely blurred scene. Video object graph cut method creates foreground salience graph in spatiotemporal domain. It also find the accurate segmented foreground object. Also due to the cross frame verification method is sensitive to objects in motion, thus is capable of building the basis of a robust object detection system in dynamic scenes.

VIII. CONCLUSION
This paper proposed the concept to combining appearance model with video object graph cut method for object detection and tracking. Combination of appearance model and graph cut method detect accurate object from video scene also motion blur is no more issue for our method. There is no need of deblurring function required separately, so it leads to increase the performance.

IX. FUTURE SCOPE
As motion blur is very complex in real scenes for very high blur videos the appearance model cannot find the object properly and it fails to find object from blur scene. Also appearance model is slow compared with some real time trackers, so future study will improve the speed of tracker by optimizing appearance model.

REFERENCES

