IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

AN OVERVIEW OF GRAPH BASED METHODS FOR VISUAL OBJECT TRACKING

[1]Nagaraja N S., [2]Dr. Vasudeva.
[1]Research Scholar, [2]Professor
[1]VTU-RRC,Belagavi,India
[2]NMAMIT,Nitte,India

Abstract: The object tracking algorithm takes a set of object detections, automatically obtains the states of the objects in the subsequent frames in a video. Numerous approaches and methods are proposed to achieve detection and faithful tracking of objects. This paper is a qualitative comparison of graph methods used in object tracking. Some graph approaches like graph cuts use graph representations for object itself while others treat objects in frames as nodes and temporal relations as edges. Many works modeled spatial structures also by using graph. Grammar based graphs and variations of graph neural networks play important role in latest visual tracking applications. Graph based approaches demonstrate good tracking in most of the works.

Index Terms - Object detection, tracking, graph, graph matching, graph pyramids, attributed relational graph (ARG), Graph cuts, graph neural networks.

I. INTRODUCTION

In computer vision, Visual object tracking is one among the research area of importance due to wide range of its applications. The object tracking systems include detection and tracking the object in motion and its position relevant to scene. The Object detection algorithm classifies and detects the object, by creating a bounding box around it and assigns unique identification for each object. Object tracking refers to estimation or prediction of the position of a target object in each consecutive frame in a video after the initial position of the target object is defined.

Numerous research and developments has been done in this field both in academia and real world applications, such as security monitoring, autonomous driving, automatic surveillance, scene analysis, and robotic vision. Several approaches and methods have been proposed for detection of objects and tracking the objects in video stream and being used in wide range of real-time computer vision applications.

A detailed review of the object tracking methods in early years before 2006 was presented by Yilmaz [1] with detailed analysis and comparisons of various representative methods.

For pattern representation and classification Graph-based techniques have been proposed as a powerful tool due to their expressive power [2]. With proper attributes, Nodes of the graphs can be thought as descriptors of parts of the objects and the edges of the graphs represent the relationships between the parts. Description schemes lead to a variety of graph representations with different graph topology.

The current survey focuses on providing overview and qualitative features of graph based approaches for object detection and tracking. The existing surveys cover domain of general object detection and tracking and may not contain study specific to the graph based methods, which provide some novel solutions and new directions in relationship of graph theory and computer vision research.

(1) This paper lists graph based approaches and solutions proposed with respect to various aspects of object detection and tracking and discuss the basics so that readers can see the cutting edge of the field.

(2) Different from previous object detection and tracking surveys, this paper comprehensively reviews some graph based approaches based object representation, modelling, detection and tracking methods.

(3) This survey is featured by analysis and discussion in various aspects. Above all, it is our intention to provide an overview how different graph based methods are used in solving different aspects of visual object tracking than a full summary of related papers.

In Section two the Overview of Visual object tracking is described. The third section is over review of graph based approaches in object detection and tracking. Section four has a qualitative comparison and discussion on features used with graph approaches and tracking methods. Section five has conclusion.

II. OVERVIEW OF OBJECT DETECTION AND TRACKING

A lot of research work has been done on object detection, tracking and video analysis problems. The related problems steps in video analysis can be divided into three major tasks: detection of moving objects, tracking of such objects from frame to frame, and behavior analysis of object tracks. Applications of object tracking play an important role in video analytics, scene understanding for security, military, industries, Robotics transportation and other various fields. Object tracking faces challenges to deal with noise in images, complex object motion, articulated nature of objects, object occlusions, complex object shapes, background distractions and scene illumination changes, and real-time processing speed and accuracy requirements. Objects to be tracked may have a variety of sizes and aspect ratios. The size misconceptions will influence detections.

Object tracking is classified in to various categories due to broad areas of application. Based on the number of objects being tracked we have Levels of object tracking: Single object tracking commonly referred to as Visual object tracking and Multi object tracking. To track objects, object detectors can be applied frame-by-frame. Alternatively for computationally more efficient tracking object detection can be applied once, and then the object tracker handles every frame after the first.

Visual object tracking process has four stages. The first one is Target Initialization, is drawing a bounding box around the object in the initial frame of the video. The second is modeling the visual appearance of the object. Motion estimation predicts the future position of an object accurately. Finally, using a visual model locks down the exact location of the target.

In visual tracking selecting the right features has an important role. Uniqueness of features is the one that makes the objects easily distinguished from other objects.

The spatio-temporal segmentation [4] is one of the most used descriptions of a video sequence. Attempt is to extract backgrounds and independent objects and also their motion. These spatiotemporal segmentation techniques are classified into three categories i) Segmentation-based approach. ii). Trajectories-based approach. iii). Joint spatial and temporal segmentation.

Generally, on-line algorithms can be divided into two categories: Generative methods and Discriminative methods [3].

III. SOME GRAPH BASED APPROACHES FOR OBJECT DETECTION AND TRACKING

Graph based Techniques have been widely investigated in the fields of computer vision for the representation and manipulation of data. It requires polynomial time to search for a specific node or edge in a graph, or for the shortest path relationship between two nodes. The graph-based object representations lead the recognition task to a graph matching problem [17].

Hwann-Tzong in [5] demonstrated a real time Object tracking that to track individual objects. An invariant bipartite graph is constructed to model the dynamics of the tracking process to addresses the ambiguities caused by the interactions among the objects.

Feng Tang in [6] designed a Attributed Relational Graph (ARG) as a representation that models both spatial and temporal characteristics of an object. for tracking, Scale Invariant Feature Transform (SIFT) features describe the object details the relations between features encode the object structure.

Algorithm by Donatello Conte in [8] is based on a graph pyramidal decomposition of the objects being tracked, using a multi-resolution approach, so exploiting the spatial coherence between pixels. Tracking phase of the system is built upon a foreground detection technique [11]. A description of the objects and their mutual relation is provided by the graph based representation. In absence of pyramidal representation retains the simplicity and effectiveness of the bounding box, but enables a more accurate object matching to take place during an occlusion.

Ana In [9] both model and frame data are described using ARGs (Attributed Relational Graphs). Attributed relational graphs based object representation express appearance properties and structural information. Using the concept of inter-frame ARG, temporal attributes were embedded in such representation.

Yang Lu [6] uses a And-Or Graph (AOG) representation for simultaneously tracking, learning and parsing objects. In this grammar [32] based approach the AOG is discriminatively learned online to account for the appearance (e.g., lighting and partial occlusion), structural (e.g., different poses and viewpoints) variations of the object itself and the distractors (e.g., similar objects) in the scene background.

Heng Fan in [8] describes the inner spatial structure information of object for tracking. Representing the object with local structural cells (LSCs) and constructing Local structural cell graph (LSCG) to model the spatial structure between the inner parts (nodes) of the object and edges are the interaction between two parts. Matching of LSCG is used for tracking.

Graph cuts are used for segmentation and detection [11]-[15] in visual tracking applications.

A scene graph [26] is a topological representation of a scene.

Variants of Neural networks have important role in state of the art approaches for object detection and tracking. Graph neural networks [28]-[30] are a class of neural network in which for processing data is represented by graph data structures.

3.1 Dynamical Graph Matching

Each target object is represented by a probability distribution of intensity values via histogram analysis. After shape contour extraction, the number of objects currently in the scene is determined and same numbers of object nodes are created. By investigating a new frame the two classes of objects are classified in the bipartite graph as profile nodes correspond to previous tracking history and object nodes correspond to the currently detected objects. Both types of nodes have the same type of data structure where position, intensity distribution, and dimension of its enclosing bounding box are stored.

During tracking, at the beginning of a frame the profile nodes are constructed from the tracking result from last frame. The number of profile nodes is same as the number of objects that are there in the scene in the beginning of frame. To find the best match to resolve the identities a bipartite matching algorithm is used. When there are any unmatched object nodes left, this indicates that new objects have been detected so new profiles will be created to track them. When an object leaves the scene, there will not be any matching to corresponding profile. An aging tag of an unmatched profile is used to delete a profile node from the bipartite graph. , The profiles getting too close to each others are indication that interactions are likely to happen. Those profiles

marked as TBM (to-be-merged) profiles and are processed separately to check whether interaction has occurred or not. Bounding box covering more than one TBM profile significantly indicates that interaction has occurred.

The number of nodes in the graph changes dynamically. So an invariant property to assure that the numbers of nodes of both types are kept the same is maintained so that the matching problem is manageable. The system is designed for static camera.

3.2 Dynamic Feature Graph

Spatially, the object is represented as an attributed relational graph (ARG).The features form the nodes and their relations form the edges. Temporally, the graph can adaptively update itself by adding new stable features as well as deleting inactive features. The object is modeled with invariant feature-Scale Invariant Feature Transform (SIFT). The relationship between them is encoded in the form of an ARG. This can effectively distinguish object from background and other objects. To match the model graph with the observation to get the best object position, a relaxation labeling method is used. Experiments results show that method gives reliable track even under dramatic appearance changes, occlusions, etc.

The feature is described by its location, scale and the orientation of the main intensity gradient within a neighborhood and the gradient histogram in the local region.

Suppose the f and f ' are two features, their relation attributes are defined as :

$$r(f, f') = \{r_d, r_s, r_o\}$$

Where , r_d is the Euclidean distance between two features, r_s is the scale difference and r_o is the orientation difference.

The attributed relational graph G is defined as follows:

$G = \{\sum_f , \sum_r \}$, is a relational graph.

$\sum_f = \{f_1, f_2, f_3 \ldots\ldots f_m\}$, is the node set.

$\sum_r = \{r_1, r_2, r_3 \ldots\ldots r_m\}$, is the edge set.

The nodes are features like point features or region features. Since relative attributes are used as the relation, the graph representation is rotation and translation invariant. This enhances the flexibility and robustness of tracker.

Swapping in and out features to keep the best features in the model is done in dynamic feature graph to model temporal changes like changes in appearance due to the illumination condition changes, pose changes and features being out of view in the next frames,

In the model update process features which are good in terms of stableness are always kept in the model.

By observing image sequence the features which found rarely are selected as an unstable feature and deleted from the model. The new feature which is persistently observed for consecutive frames is a stable feature, and is selected from candidate set.

The matching score of the object representation with the observation is computed using a feature based likelihood function which is graph matching problem. It gives the approximate graph state in a maximum likelihood sense.

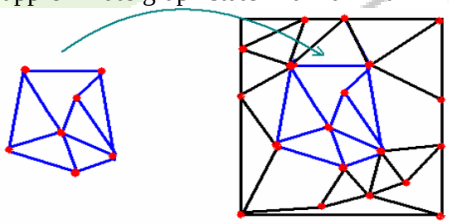

Figure 1. Graph matching.

The model graph is on the left. At the region around the predicted position larger graph than that in the previous state is constructed and is shown on right. The sub-graph matching algorithm matches the model graph with this larger graph. The probability for each feature labeling will stabilize, thus the feature to feature matching is obtained.

3.3 A graph pyramidal decomposition of the objects

For each frame of the video, pixels are grouped into connected components belong to the moving regions. A detected moving region is represented using a graph pyramid. Each region is represented at different levels of resolution using a graph for each level. At the topmost level, the apex of the pyramid, the graph is composed by a single node, containing as attributes the position and dimension of the bounding box, and the average color. At lowest level there is an adjacency graph, where the nodes represent single pixels, and the edges encode the 4-connected adjacency relation. The nodes describe the foreground pixels surrounded by the bounding box of the region. The intermediate levels are obtained by a bottom–up process, using the classical decimation-grouping procedure [19]. A color similarity is used to decide which nodes must be merged. The Number of levels in a pyramid depends on the color uniformity of the region. Each intermediate node represents a sub region of the whole region, and its attributes are the position and size of the sub region bounding box and average color of that sub region.

A node attribute is a set of seven (5 in grey level images) numbers. These are corresponds to width and height of the sub region bounding box, the average colour or the average grey level of the sub region and the coordinates of the sub region bounding box center in the image.

During the tracking process, each node in the pyramid has a label indicating the corresponding (sub) region. The A node is labeled as MULTILABEL, when a (sub) region contains parts of more than one object. In a MULTILABEL region the object parts are reconstructed studying the levels of the graph pyramid and labels are assigned assuming that each region contains a single object.

In the hierarchical representation at each level there is a graph for each detected region. The algorithm compares each graph G1 of it with each graph G2 of t+1. An association graph is built which a bipartite graph where each node n of G1 is linked to each node m of G2 such that S(n, m) > 0. A weighted bipartite graph matching is performed, finding the matching between the nodes in G1 and those of G2 that maximizes the sum of the corresponding similarities.

3.4 Structural pattern recognition

The target objects in a single video frame are represented by intra-frame ARG (intra-ARG), and the target objects from neighboring video frames are represented by inter-frame ARG (inter-ARG). Using inexact graph matching [18] task the correspondence between the set of vertices of an inter -ARG and that of the model intra -ARG is found. The contents of the video frames are represented by attributed relational graphs. Here vertices represent image regions that are possibly related to a given target object, and edges represent relations among such objects. Attribute vectors carry information about properties of regions such as average gray level, perimeter, area, centroid etc. and relations between two given regions. An inter-ARG is built from n intra-ARGs derived from n consecutive video frames. Binding these intra-ARGs is represented by the set of temporal edges. From a reference image and a manually-created model mask image containing the objects or object parts of interest to be recognized and tracked, an intra-ARG, the model ARG, is generated at the beginning. Inexact graph matching optimization for pattern recognition [23], the matching is achieved through the use of a tree search algorithm [24].For each set of vertices which have been mapped to the same model ARG vertex a new position is calculated, based on the classification of the inter-ARG vertices. Using these new positions and previous ones available in the model an affine transform from the old set of positions to the newly found set can be applied ,which leads to the tracking of all objects.

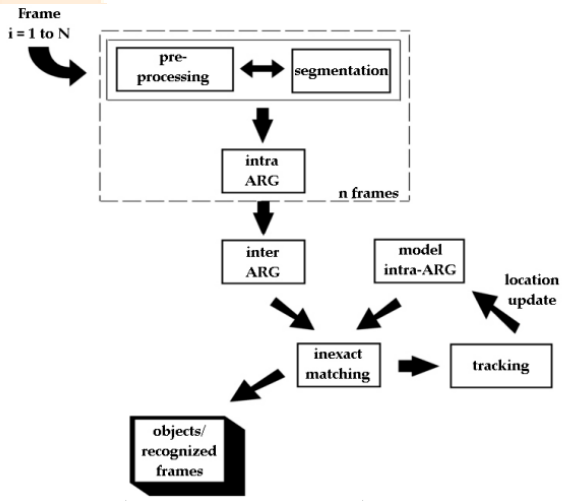

Figure 2: Intra-frame ARG & inter ARG interaction.

3.5 And Or Graph for modeling object

Grammars for Object detection [33] represent objects recursively in terms of other objects. And-Or graph (AOG) is a hierarchical and compositional representation with a directed acyclic graph (DAG) structure. An AOG represent different aspects of modeling objects in a grammatical manner [27]. The AOG are discriminatively trained online to account for the appearance say lighting and partial occlusion, structural variations of the object itself like different poses and viewpoints, and the distractors like similar objects in the scene background. AOG has three types of nodes. The And-node represents the rule of decomposing a complex structure into simple ones or typical configuration of an object. Or-node represents alternative configurations at both object and part levels. The third type is the Terminal-node links the representational symbol to image data using different appearance templates to capture local appearance change.

By parsing with the current AOG using a spatial-temporal dynamic programming algorithm the state of the object is inferred during tracking.

3.6 Local Structural Cell Graph

This approach uses structure information for object tracking. Object is segmented into a superpixel. A rectangular grid is superimposed on top of the superpixel map. The structural appearance model is obtained based on the superpixel map of object. Each grid in the rectangular superpixel grid is represented by its center point and is annotated with the feature vector. We obtain an undirected graph G, whose nodes are the grid center points and edges are the interactions between the grid points. The local relationship of the inner object parts is considered to construct the local structural cell (LSC) for each grid point. These local structural cells (LSCs) are used to substitute the grid points in G and a new local structural cell graph G is obtained, where the LSCs are the nodes and the interaction between LSCs are edges. The appearances of objects and their relations in a frame are represented by local structural cell graph (LSCG). The tracking is viewed as matching LSCG in the subsequent frames. We can track the object by the similarities of the local structural cell graphs (LSCGs) between the reference and candidate targets.

3.7 Graph Cuts

Graph cuts can be seen as a "region-based" method for encoding image segments [20]. The image pixels represent nodes of graphs. Two terminal nodes S (source) and T (sink) are specially designated to represent "object" and "background" labels respectively. Edges interconnect neighboring pixels in a regular grid-like fashion. Edges between pixels are called neighbor links or n- links. a neighborhood system can be arbitrary and may include any kind of n-links. The edges, called t-links, are used to connect pixels to terminals. All graph edges including n-links and t-links are assigned some nonnegative weight. An s-t cut is a subset of edges such that the terminals S and T become completely separated on the induced graph. The nodes between the terminals are divided by the cut. An image is partitioned into "object" and "background" segments by Cut.

3.8 Scene Graph

A scene graph is a data structure that represents hierarchically the relationships between transformations applied to a set of objects in a three-dimensional scene.

The type of information decides the structure, contents of a scene graph and the set of operations.

A simple Scene graph may consist of three types of nodes. The root node of the tree referred to as world or Virtual Universe that represents the collection of objects in 3D scene. The internal node of the tree called group node represent a logical grouping of objects and store semantic information. Any number of child nodes can be there for a group node. The leaf node represents either an object or a part of an object. These maintain the necessary geometrical information and some semantic information. Camera and light sources are also represented as leaf nodes.

3.9 Graph Neural Networks

To represent the nodes there are three approaches [34]. The video frame or the image can be split into regular grids, and each grid servers as the vertex of the visual graph and apply neural networks to get its embedding.

Vertex can be represented using pre-processed structures from Visual regions.

Third way is to use semantic information to represent visual vertexes. For example pixels with similar features groups into coherent regions can represent same vertex.

IV. DISCUSSIONS

[5] & [7] achieve good tracking results under situations where there is uncertainties due to the shadows caused by indoor lighting and interactions among the objects. Target object is represented with its color distributions. Using bipartite graph allows simulation of tracking vertices that corresponds to multiple objects. Moving object detection involves shadow deletion and extraction of shape contour to separate foreground objects and the background scene.

Zhaowei Cai [10] uses graph cut for foreground background separation. The Appearance model has discriminative SVM classifier and generative target color histogram, and the structure model using graph with Markov property. Spectral matching techniques used for the graph tracking. Dynamic graph maintains inner structure of the target and tracks even when occluded and adapts to structure deformations.

In [24] dependent upon initial contour placement and requires repeated cuts on this reduced domain. [25] Use one graph cut for each frame to both estimate the optical flow and object position despite changes in illumination. Malcolm and Rathi [14], for each object the distance penalty is determined by predicting the object location and edge weights for the graph are computed to perform a graph cut segmentation.

Feng Tang [6] express the frame objects by the extracted SIFT features. Compact and robust feature-based object representation in the form of attributed relational graph (ARG) results in in detection of exact match under certain extent of illumination changes and occlusions. The tracker is formulated into a Maximum a Posteriori (MAP) framework using the Hidden Markov Model. Graph matching involves a relaxation labeling process.

In [9] both model and frame data are described through ARGs. Object regions in a frame form intra ARGs,

And intra-ARGs are developed from successive frames. An affine transformation is used to perform object tracking.

In successive frames recognition is performed by inexact graph matching. A tree-search optimization algorithm allows the minimization of pre-defined cost functions.

Chenglong Li, Liang Lin [37] shows removal of the effects of background clutters using Concatenation of weighted patch descriptors into a feature vector. SVM framework is used to carry out the tracking task. The tracked object is modeled with a graph where non-overlapping image patches are nodes. Edge weights indicate the appearance compatibility of two neighboring nodes.

Conte and Foggia [8] represented object region with graph pyramid, a hierarchical decomposition representation. At each level of the hierarchical representation there is a graph for each detected region. During an occlusion, for each multiple objects region, the region segment is matched its hierarchy with the hierarchies of the objects present before the occlusion.

Representation of the object using local structural cells (LSCs) by Fan and Xiang [x] exploits both partial and spatial information of the target. Spatial structure between the inner parts of the object is constructed using graph, whose nodes are target parts and edges are the interaction between two parts. The target under situations with local occlusions or deformations is located by this representation.

A grammar based representation of objects is described by Y Lu and T Wu in [36]. The AOG is discriminatively trained online to describe the appearance, structural variations of the object and the distractors in the scene background. In tracking the bounding box is inferred by using a spatial-temporal dynamic programming (DP) algorithm. This handles occlusion, pose change and illumination well 27] is robust shape-based object detection against background clutter.

In GNNs [28] each node represents an object and the similarity between detection and tracklets is represented by edges. To detect and associate objects, a detector and a Re-ID module is integrated in tracking model. GNNs extract relations between objects and learn features to improve both detection and data association. The approach performs very well in addressing multiple object tracking challenges.

V. CONCLUSION

A brief review of object detection and tracking using graph based approaches are presented in this paper. The intention of this paper is to provide bird view of the field that may help the research work on graph related algorithm for object tracking. Various graph based approaches for tracking objects show good performance results in performing tracking in situations where there are ambiguities about object states due to interaction, Shadow, deformations and occlusions.

REFERENCES

[1] Alper Yilmaz, "Object tracking: a survey. ACM Comput Survey", ACM Computing Surveys,Volume 38 Issue8, 2006.

[2] Donatello Conte, Pasquale Foggia, Carlo Sansone, Mario Vento ,"How and Why Pattern Recognition and Computer Vision Applications Use Graphs"in Abraham Kandel, Bunke Horst,Mark Last Applied Graph Theory in Computer Vision and Pattern Recognition,SICI, volume 52 Springer,2007

[3] Xi Li , Weiming Hu , Chunhua Shen , Zhongfei Zhang , Anthony Dick , Anton van den Hengel , "A Survey of Appearance Models in Visual Object Tracking", ACM Transactions on Intelligent Systems and Technology, 2013

[4] John Y. A. Wangy and Edward H. Adelson, "Spatio-Temporal Segmentation of Video Data", Proceedings of the SPIE: Image and Video Processing II, vol. 2182, 1994.

[5] Hwann -Tzong Chen, Horng - Horng Lin, Tyng-Luh Liu ,"Multi-Object Tracking Using Dynamical Graph Matching ", IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2001

[6] Feng Tang and Hai Tao, "Object tracking with dynamic feature graph," 2005 IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, pp. 25-32, 2005.

[7] Jinqin Zhong , Jieqing Tan , Yingying Li , Lichuan Gu , Guolong Chen, "Multi-Targets Tracking Based on Bipartite Graph Matching", Cybernetics And Information Technologies, Volume 14, Special Issue, Bulgarian Academy Of Sciences,2014

[8] Donatello Contea , Pasquale Foggia , Jean-Michel Jolionc , Mario Ventoa, " A graph-based, multi-resolution algorithm for tracking objects in presence of occlusions ", Pattern Recognition 39,pp. 562 – 572, 2006

[9] Ana B. V. Graciano, Roberto M. Cesar-Jr, Isabelle Bloch, " Graph-based Object Tracking Using Structural Pattern Recognition", IEEE XX Brazilian Symposium on Computer Graphics and Image Processing, 1530-1834/07 2007

[10] Zhaowei Cai ,Longyin Wen, Jianwei Yang, Zhen Lei, Stan Z. Li, " Structured visual tracking with dynamic graph", Proceedings of the 11th Asian conference on Computer Vision , Volume Part III, pp. 86–97 2012

[11] Z. Garrett and H. Saito, "Live video object tracking and segmentation using graph cuts," 2008 15th IEEE International Conference on Image Processing, pp. 1576-1579, 2008.

[12] Alexander M Nelson ,Jeremiah J Neubert," Object Tracking via Graph Cuts ", Proc. of SPIE Vol. 7443, Applications of Digital Image Processing XXXII, 2009.

[13] Yuri Boykov, Gareth Lea Funka, "Graph Cuts and Efficient N-D Image Segmentation", Int J Comput Vision 70, pp.109–131, 2006.

[14] J. Malcolm, Y. Rathi and A. Tannenbaum, "Multi-Object Tracking Through Clutter Using Graph Cuts," 2007 IEEE 11th International Conference on Computer Vision, pp. 1-5, 2007

[15] N. Xu, R. Bansal, N. Ahuja, " Object segmentation using graph cuts based active contours", 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings. Volume 2, pp.II-46. 2003

[16] Fan, Heng & Xiang, Jinhai & Liao, Honghong & Du, Xiaoping, " Robust Tracking based on Local Structural Cell Graph", Journal of Visual Communication and Image Representation. 2015

[17] Y. Boykov and M. Jolly, "Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images", Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001, pp. 105–112, 2001.

[18] D. Conte, P. Foggia, C. Guidobaldi, A. Limongiello, M. Vento, An object tracking algorithm combining different cost functions, Lecture Notes in Computer Science, vol. 3212, Springer, Berlin, 2004, pp. 614–622.]

[19] J.M. Jolion, A. Montanvert, "The adaptive pyramid: a framework for 2D image analysis", CVGIP: Image Understanding, 55 Issue 3, pp.339–348, 1992.

[20] H.W. Kuhn, "The Hungarian method for the assignment problem", Naval Res. Logistics Quarterly 2 83–97, 1955

[21] Sanfeliua, A, Juliana Andradea, J. Climentc and F. Serratosad, "Graph-based representations and techniques for image processing and image analysis." Pattern Recognition, 2001

[22] H. Bunke, "Recent developments in graph matching," Proceedings 15th International Conference on Pattern Recognition. ICPR-2000, pp. 117-124 vol.2, 2000.

[23] R. Cesar-Jr, E. Bengoetxea, P. Larranaga, and I. Bloch. "Inexact graph matching for model-based recognition: Evaluation and comparison of optimization algorithms", Pattern Recognition, Volume 38 Issue 11, pp. 2099–2113, 2005.

[24] D. Conte, P. Foggia, C. Sansone, M. Vento. "Thirty years of graph matching in pattern recognition",IJPRAI, 18(3):265–298, 2004.

[25] Zhang, Xiaoqin, Weiming Hu, Stephen J. Maybank and Xi Li. "Graph Based Discriminative Learning for Robust and Efficient Object Tracking." 2007 IEEE 11th International Conference on Computer Vision : 1-8., 2007.

[26] X. Chang, P. Ren, P. Xu, Z. Li, X. Chen, and A. Hauptmann, "Scene Graphs: A Survey of Generations and Applications," Mar. 2021.

[27] Liang Lin, Xiaolong Wang, Wei Yang, Jian-Huang Lai, "Discriminatively Trained And-Or Graph Models for Object Shape Detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume: 37, Issue: 5, 2015

[28] S. Tang, W. Zhang, Z. Mu, K. Shen, J. Li, J. Li, and L. Wu, "Graph neural networks in computer vision," in Graph Neural Networks: Foundations, Frontiers, and Applications, L. Wu, P. Cui, J. Pei, and L. Zhao, Eds. Singapore: Springer Singapore, pp. 447–462 , 2022.

[29] Yongxin Wang, Kris Kitani, Xinshuo Weng," Joint Object Detection and Multi-Object Tracking with Graph Neural Networks", 2021 IEEE International Conference on Robotics and Automation (ICRA), 2021

[30] Junyu Gao, Tianzhu Zhang , Changsheng Xu, "Graph Convolutional Tracking" 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition,2019.

[31] Nikos Nikolaidis, Michail Krinidis, Evangelous Loutas, Georgios Stamou, Ioannis Pitas, "Motion Tracking in Video", CHAPTER 7 in The Essential Guide to Video Processing, Al Bovik, Academic Press, pp.175-230, 2009.

[32] Zhu, Song Chun, and David Bryant Mumford , "A stochastic grammar of images ", Foundations and Trends in Computer Graphics and Vision Volume 2 Issue ,pp. 259-362, 2006.

[33] Ross B. Girshick, Pedro F. Felzenszwalb, David McAllester, "Object Detection with Grammar Models".

[34] Siliang Tang, Wenqiao Zhang, Zongshen Mu, Kai Shen, Juncheng Li, Jiacheng Li and Lingfei Wu , "Graph Neural Networks in Computer Vision," Chapter 20, Graph Neural Networks: Foundations, Frontiers, and Applications edited by Lingfei Wu, Peng Cui, Jian Pei, Liang Zhao, Springer , 2022

[35] Y. Lu, T. Wu and S. Zhu, "Online Object Tracking, Learning, and Parsing with And-Or Graphs," 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 3462-3469, 2014.

[36] C. Li, L. Lin, W. Zuo, J. Tang and M. Yang, "Visual Tracking via Dynamic Graph Learning," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 11, pp. 2770-2782, 1 Nov. 2019