IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, April 2024

Multiple Object Tracking

Gowth AM Ganesh. E[1], Jason Selvakumar. J[2], Mohamed Ali. B[3]
Department of Computer Science and Engineering[1,2,3]
SRM Institute of Science and Technology, Vadapalani, Chennai, India

Abstract: *Detecting and tracking multiple objects in real-time video sequences is a difficult task in computer vision due to the need to handle occlusions, changes in appearance, and other factors that may impede precise tracking. The paper proposes a method for tracking multiple objects in a video, even when the number of objects is unknown and changes over time. This is achieved by combining object detection and tracking using a graph structure, which maintains multiple hypotheses about the objects' number and trajectories in the video. The graph structure is updated using image information, and the best hypothesis is chosen to explain the video. The tracking process acts as a temporal detection that confirms and validates the object detections over time, making a global decision across the video. The method also integrates object detection and tracking by providing feedback in the form of object location predictions to the object detection module. The final output is the most likely hypothesis from the multiple objects tracking process. The proposed method is evaluated experimentally to assess its performance*

Keywords: multiple objects

I. INTRODUCTION

Visual object tracking refers to the process of keeping track of one or more targets in a video. This is a crucial aspect of video analytics, used in various applications including video surveillance and robot navigation. Typically, multiple object tracking is achieved through the linking of detections in consecutive frames.

In this process, a pre-trained detector is run on each new frame to identify the objects in the scene. By linking the detections in the current frame with those in the previous frame, it becomes possible to calculate the movements and changes in size of the objects of interest.

To achieve real-time visual object tracking, The system needs to ascertain the location of objects in each frame independently, without depending on continuous detections. MVOT, short for multiple visual object tracking, is a technique that involves motion estimation between detections to achieve the task of determining the objects' position in each frame independently.

MVOT encounters difficulties when operating in crowded scenes since they are primarily designed for single targets and may be computationally expensive when dealing with multiple instantiations. Dealing with this challenge necessitates adopting a more comprehensive strategy that permits sharing computations among objects. By doing so, the system can handle multiple targets in real-time, even dozens of them, while leveraging single-object tracking techniques.

Previously, most programmers in companies were mainly concerned with creating the user interface for software and hardware image processing systems. However, after the introduction of the Windows operating system, many developers shifted their attention towards addressing the challenges of image processing itself. This shift in focus has resulted in a significant change in the industry.

One of the tasks involved in image classification is predicting the class of a single object within an image. Another task is object localization, which involves identifying one or more objects in an image and outlining their boundaries using a bounding box.

Object detection is a computer vision task that involves both object localization and classification. It aims to identify and categorize one or more objects in an image. Sometimes, the term "object recognition" is used interchangeably with object detection. However, beginners in the field of Object detection aims to locate every occurrence of a specific class of objects in an image, such as cars, faces, or people. Although the object may appear only a few times in the image, there are numerous potential positions and scales to consider. Object detection algorithms typically

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, April 2024

provide some pose data for each detected object. This data may include basic information such as the object's location, scale, or the extent of the object, represented by a bounding box.

The scope of the project on multiple object tracking is to develop a system that can efficiently and accurately track the movements of multiple objects within a video in real-time. This involves the integration of various techniques including feature extraction, object detection, motion estimation, and data association to maintain the identity of each target throughout the video. The goal is to overcome the limitations of traditional multiple objects tracking methods and achieve improved performance in terms of accuracy and efficiency. This can be applied to various applications such as video surveillance, robot navigation, and video analytics. The project's objectives may involve testing and validating the proposed system on publicly available datasets, as well as comparing its performance with that of state-of-the-art trackers.

To summarize, the primary goals of multiple object tracking can be stated as follows:

- To track multiple objects accurately and robustly in a scene over time.
- To handle occlusions and maintain the identity of each object.
- To handle changes in appearance, such as changes in scale, viewpoint, and lighting conditions.
- To handle complex interactions between objects, such as collisions and occlusions.
- To provide real-time performance for use in practical applications.
- To generate a reliable and accurate output that can be used for downstream tasks, such as activity recognition, behaviour analysis, and surveillance.

Section II studies previous research or studies that have been conducted on multiple object tracking. Section III refers to a new or improved system that is being proposed to address multiple object tracking. Section IV refers to the data and findings obtained from scientific experiment or study on multiple object tracking. Section V summarizes the key findings, implications, and recommendations of multiple object tracking. Section VI reports the sources of information that have been cited or referred to in a document.

II. RELATED WORK

A new technique has emerged for multiple objects tracking in computer vision applications, such as video surveillance, advanced driver assistance, and animation. The current tracking-by-detection methods mostly focus on the appearance and motion of objects, while not fully utilizing the contextual information available around the targets.

The authors of this research have presented a novel technique for multiple object tracking, with a focus on exploiting contextual information in computer vision applications such as video surveillance, advanced driver assistance, and animation.

They introduced the Exchanging Object Context (EOC) model, which employs a distinct affinity measure to connect detections with targets. This measure gauges the similarity between them by swapping their contexts and evaluating the smoothness of the background. Moreover, a novel colour histogram descriptor is employed in this approach.

According to the authors, the proposed method is not only efficient but also accurate for online multiple objects tracking. By detecting context changes, this method enhances the precision of bounding boxes and leverages contextual information more effectively via the Exchanging Object Context (EOC) model. Additionally, a novel affinity measure is employed to link detections with targets by measuring their similarity based on the smoothness of the background, after exchanging their contexts using a new colour histogram descriptor. The experimental results on two publicly accessible benchmarks demonstrate the effectiveness of the proposed technique, which outperforms various cutting-edge trackers.

The approach has been shown to outperform current state-of-the-art trackers on five public benchmarks, demonstrating its effectiveness. The results showcase the efficiency of the proposed method in addressing the problem of multi-object tracking in real-time across different scenarios

However, the challenge lies in combining various features for multiple objects tracking since depending solely on motion or appearance features may lead to erroneous data association, especially when object movements are complex or intra-frame objects have similar appearances.

The authors of the study proposed a novel approach that redesigns the integration of motion and appearance features to address the challenge of incorrect data association in complex object movements or similar appearances. The proposed

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, April 2024

method involves using adaptive searching windows based on object location and motion, where object matching is only based on the similarity of appearance features within these windows. This approach is implemented online and aims to achieve more accurate and reliable data association for multiple object tracking.

During the candidate selection process, all available candidates are scored based on a function that considers a discriminative object classifier and the confidence level of the tracklet.

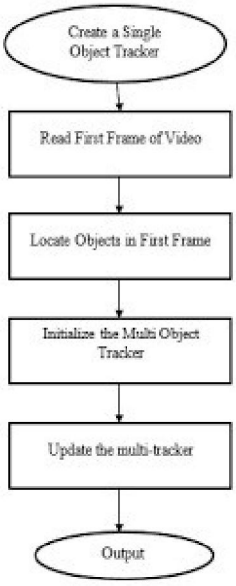
To improve the efficiency of candidate selection, the proposed framework employs an R-FCN based classifier that reduces the number of candidates generated from detection and tracking outputs.

A. Real time object classification

In the proposed framework, the R-FCN classifier is trained end-to-end by randomly sampling positive examples from regions of interest (RoIs) around the ground truth bounding boxes and negative examples from background RoIs during the training process. It is worth noting that the neural network is exclusively trained for candidate classification and not for bounding box regression.

This design allows the framework to prioritize efficient selection and association of candidate objects while deferring the final bounding box refinement to a separate regression module. To determine the similarity between candidates in multiple object tracking, the proposed framework employs a deep neural network called Hreid.

The appearance representation of a person in an RGB image is obtained through Hreid, and the similarity between two images is measured using the Euclidean distance between the features obtained. During the training of the network, a set of triplets consisting of positive pairs from the same person and negative pairs from different people are used. The loss function minimized during training is based on the N triplets used.



The project comprises of six phases, which include loading the dataset, designing the YOLOv3 model, configuring the training options, training the object tracker, and evaluating the tracker's performance. The flow chart for multiple object detection can be seen in Figure 1.

Appearance representation with reID features

The appearance representation of a person in an RGB image is obtained through Hreid, and the similarity between two images is measured using the Euclidean distance between the features obtained. During the training of the network, a set of triplets consisting of positive pairs from the same person and negative pairs from different people are used. The loss function minimized during training is based on the N triplets used.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, April 2024

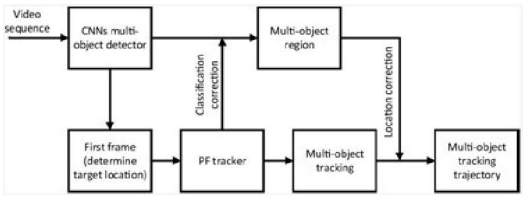


fig.2

An overview of the multi-objective tracking framework is presented in Fig. 2.

The framework involves simultaneous processing of the video sequence by using CNNs. To determine the proposed approach for multiple object tracking involves first estimating the target location in the initial frame, followed by using the particle filter algorithm to correct the object classification. The particle filter tracker is then utilized to determine the multi-object region, which provides the correct object regions. Finally, the multi-object tracking trajectory is refined by incorporating the multi- object region generated using CNNs

III. DATA ASSOCIATION

Exclusion Model

When addressing the MOT problem, exclusion is a crucial constraint due to the possibility of physical collisions. When performing multiple object tracking, there are generally two types of exclusion constraints to consider: detection-level exclusion and trajectory-level exclusion

Detection level exclusion model

The constraint of detection-level exclusion plays a crucial role in multiple object tracking, as it prohibits two separate detection responses in the same frame from being assigned to the same trajectory hypothesis. This constraint ensures that two detections in the same frame cannot belong to the same tracked object, guaranteeing that two objects cannot occupy the same space simultaneously. This is a fundamental requirement for achieving precise and dependable tracking results.

Exclusion modelling based on trajectories

A constraint that is used to solve the MOT problem. It states that two trajectories cannot occupy the same detection response. This constraint helps to ensure that the tracking solution is physically feasible and that two objects are not assigned to the same position at the same time. It is important to consider trajectory-level exclusion in MOT algorithms to avoid tracking errors and ensure accurate results

fig.3

An illustration of how occlusion is handled can be seen in the image from Figure 3. The image on the left shows one object being obscured by another object. The image on the right displays the error map of the occluded object after reconstruction.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, April 2024

Occlusion Handling

Multiple object tracking faces the challenge of occlusion, which occurs when one or more objects are temporarily blocked or partially hidden by other objects or obstacles in the scene. The likelihood of a candidate is calculated based on the reconstruction error of each block in the corresponding subspace. This model has two advantages. First, it accounts for spatial information by considering the likelihood of each block. Second, it can determine the occlusion relationship among objects by creating an occlusion map based on the reconstruction errors of all the blocks. This information helps to update the appearance model selectively and keep it up-to-date.

Explicit occlusion model

The occlusion issue can greatly impact the accuracy of multiple object tracking. When objects overlap in the same frame, it becomes difficult to determine which track each object belongs to. To address this, some methods propose an explicit occlusion model

The proposed model divides objects into several non-overlapping blocks and constructs an appearance model based on subspace learning for each block. The likelihood of a candidate is determined by calculating the reconstruction error in the subspace for each block.

The part-based model is a computer vision technique that can be used to divide objects into smaller parts or blocks, allowing the construction of an appearance model for each block. This method is particularly helpful in addressing occlusion challenges in multiple object tracking. In a study, researchers used the part-based model to track both the whole body and individual body parts of a pedestrian. This approach proved to be especially useful when the pedestrian was occluded for a period, during which the whole-body human detector was unable to detect the pedestrian. However, visible parts of the pedestrian were detected using the part detector, and based on these parts, the trajectories of the visible parts were estimated. By combining this information with the trajectory of the whole body, researchers were able to recover the complete trajectory of the pedestrian.

IV. EXPERIMENTAL RESULT

Data Set and Evaluation Metrics

The experimental setup for multiple object tracking (MOT) typically involves several components, including the stimulus presentation, task design, performance measures, and data analysis. Here is a general overview of these components:

- Stimulus presentation: The stimulus presentation involves displaying a set of moving objects on a screen or in a virtual environment. The objects can vary in size, shape, color, and speed, and can move in different directions and trajectories.
- Task design: The task design specifies the instructions given to the participants and the specific objectives of the MOT task. The task can vary in difficulty by changing the number of targets and distractors, the speed and direction of object movement, and the duration of the tracking period.
- Performance measures: The performance measures are used to evaluate the accuracy and efficiency of MOT performance. These measures can include the proportion of correct responses, the response time, the tracking error, and the number of tracking failures.
- Data analysis: To analyse performance data and identify patterns of MOT behaviour, statistical methods are commonly used. Some specific details of the experimental setup for MOT can vary depending on the research question and the specific techniques used. For example, some studies may use eye-tracking technology to assess visual attention during tracking, or electrophysiological measures such as electroencephalography (EEG) to assess neural activity during tracking.

Overall, the experimental setup for MOT involves careful design and implementation to ensure the validity and reliability of the performance data and to enable meaningful conclusions about the cognitive mechanisms of MOT.

Implementation Details

Multiple object tracking (MOT) can be implemented using a variety of techniques, depending on the specific research question and available resources. Here are some common implementation details for MOT:

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, April 2024

- Stimulus generation: To generate the stimulus for MOT, researchers can use different software packages and programming languages, such as MATLAB, Python, or Unity. The stimulus can be created by specifying the parameters of the moving objects, such as position, speed, and trajectory, and by designing the background and visual cues of the scene.
- Data collection: To collect data on MOT performance, researchers can use different methods, such as manual data entry, online data collection platforms, or specialized software for data recording. Eye-tracking technology can also be used to collect data on visual attention during tracking.
- Performance measures: To evaluate MOT performance, researchers can use different performance measures, such as accuracy, response time, and tracking errors. These measures can be computed using various software packages, such as Excel, R, or MATLAB.
- Data analysis: To analyze the performance data, researchers can use different statistical methods, such as ANOVA, regression analysis, or machine learning algorithms. These methods can be implemented using various programming languages, such as R or Python.
- Visualization: To visualize the MOT performance data, researchers can use different visualization tools, such as graphs, heat maps, or scatter plots. These tools can be implemented using various software packages, such as ggplot2 in R or matplotlib in Python.
- Experiment control: To control the MOT experiment, researchers can use different software packages for experiment design and control, such as E-Prime, Psychtoolbox, or OpenSesame. These tools allow researchers to specify the timing and sequence of the stimulus presentation, and to control the input and output devices used in the experiment.

Overall, the implementation details for MOT involve careful consideration of the research question, available resources, and the specific techniques used in the experiment.

fig.4

Figure.4 depicts a sequence of frames from a surveillance video, where each frame contains people walking in a crowded area. The figure would show the tracked trajectories of each person over time, with different colours or labels indicating the identity of each person. The figure may also include bounding boxes around each person in each frame, highlighting the spatial location of each person.

Analysis on Validation Set

In the context of multiple object tracking, the validation set is a portion of the dataset that is utilized to fine-tune and optimize the tracking algorithm's performance before it is evaluated on the actual test set. This is a crucial step in the development of an accurate and robust tracking system, as it allows the researchers to experiment with different hyperparameters and evaluate the performance of the algorithm in a controlled environment

The comparison between the MOT models RCNN, YOLO v3, and YOLO v4 is illustrated in Figure 5.

To analyze the performance of a tracking algorithm on the validation set, the researchers typically use various evaluation metrics such as MOTA, MOTP, MT/ML/FP, identity F1 score, and track fragmentation score. By analyzing these metrics, the researchers can identify the strengths and weaknesses of the algorithm and fine-tune its parameters to improve its performance.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, April 2024

Overall, analysing the validation set is an essential step in the development of a robust and accurate multiple objects tracking algorithm. It allows the researchers to fine-tune the algorithm and optimize its performance before testing it on the actual test set.

V. CONCLUSION AND FUTURE WORKS

An extensive survey of the field of Multiple Object Tracking (MOT), which covers the current status of the field, important challenges, assessment metrics, and datasets. The state of the field is described by its different scenarios and categories of approaches. The key aspects of MOT approaches are discussed in detail, including recent developments, classifications, and case studies. Commonly used metrics and data sets for evaluating MOT approaches are listed along with publicly available algorithms.

A thorough review of Multiple Object Tracking (MOT) has been presented. The review includes an overview of the current state of the field, the major issues and challenges faced in MOT, evaluation metrics used in assessing MOT performance, and available datasets for training and testing MOT algorithms. The state of the field is discussed in terms of its scenarios and categorization of approaches. The authors also provide a comprehensive overview of recent developments in MOT, including popular algorithms and metrics for evaluation.

Despite the progress made in the field of MOT, there are still some challenges that need to be addressed. These include:

- Video Adaptation: Most MOT methods rely on an offline-trained object detector. However, this can result in suboptimal results if the detector is not trained specifically for a given video.
- Crowd Density and Object Representation: The performance of MOT can be impacted by the crowd density, as objects can be partially or completely occluded. In dense crowds, only a small portion of the object may be visible, while in less dense crowds, the whole body can be recovered more easily. Under such conditions, the crowd's motion pattern could prove to be useful for Multiple Object Tracking (MOT).

Multi-Camera Tracking: When using multiple cameras for MOT, there are two main configurations to consider. The first involves multiple cameras recording the same scene, which requires fusion of information from multiple cameras. The second involves a network of cameras, each recording a different scene, which presents the challenge of object re- i

REFERENCES

[1] Richard Cobos, Jefferson Hernandez, and Andres G. Abad. A fast multi-object tracking system using an object detector ensemble, 2019 IEEE, Escuela Superior Politecnica del Litoral Guayaquil 09-01-5863, Ecuador.

[2] Keping yu, Xin qi , Toshio sato, San hlaing myint , Zheng wen, Yutaka katsuyama, Kiyohito tokuda, Wataru kameyama and Takuro sato. Design and Performance Evaluation of an AI-Based W-Band Suspicious Object Detection System for Moving Persons in the IoT Paradigm, April 29, 2020, Tokyo,Japan.

[3] Ajit Jadhav, Prerana Mukherjee, Vinay Kaushik, Brejesh Lall. Aerial Multi-Object Tracking by Detection Using Deep Association Networks, 2020 IEEE, IIIT Sri City, IIT Delhi, India.

[4] Qiankun liu , Bin liu , Vue wu , Weihai li , and Nenghai yu. Real-Time Online Multi-Object Tracking in Compressed Domain, June 10, 2019 IEEE, Hefei230026, Hangzhou 311121, China.

[5] Wenqi Zhou, Zhihua Li, Pei Gao. Research on Moving Object Detection and MatchingTechnology in Multi-Angle Monitoring Video, 8th 2019 IEEE, Wuhan 430074, China.

[6] Xiong Xiaofang , Cheng Shanying , Hu Yudan . Research on Real –Time Multi Object Detections based on Template Matching, 3rd 2020 IEEE, Nanchang China.

[7] Kanimozhi S Gayathri G Mala T. Multiple Real-time object identification using Single shot Multi-Box detection , 2nd 2019 IEEE, Chennai, India.

[8] Weiqiang Li, Jiatong Mu, Guizhong Liu. Multiple Object Tracking with Motion and Appearance Cues, China.

[9] KangUn Jo, JungHyuk Im , Jingu Kim and Dae-Shik Kim. A Real-time Multi-class Multi-object Tracker using YOLOv2, 12- 14th 2017 IEEE, Daehak-ro, Yuseong-gu, Daejeon, Korea.

[10] Chandan G, Ayush Jain, Harsh Jain, Mohana. Real Time Object Detection and Tracking Using Deep Learning and OpenCV, 2018 IEEE, Bengaluru, India.

[11].I. Haritaoglu, D. Harwood, and L. S. Davis, “W4: real-time surveillance of people and their activities,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 22, no. 8, pp. 809- 830, August 2000.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, April 2024

[12].Elgammal, A., Duraiswami, R., Harwood, D., Anddavis, L. 2002. Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. Proceedings of IEEE 90, 7, 1151–1163.
[13].WuU Z, and Leahy R. "An optimal graph theoretic approach to data clustering: Theory and its applications to image segmentation". IEEE Trans. Patt. Analy. Mach. Intell. 1993.
[14].ISARD, M. AND MACCORMICK, J. 2001. Bramble: A bayesian multiple-blob tracker. In IEEE International Conference on Computer Vision (ICCV). 34–41.
[15].Christopher R. Wren, Ali J. Azarbayejani, Trevor Darrell, and Alex P.Pentland,"Pfinder: Real-Time Tracking of the Human Body" in IEEETransactions on Pattern Analysis and Machine Intelligence, July 1997, 19(7), pp. 780-785.