# Introduction to the Top Object Tracking Techniques

Object tracking is the process by which computers detect and track objects in still images or videos. It is one of the most common applications of computer vision. The first attempts at object detection relied on template matching techniques and statistical models. These approaches, however, are incapable of fully exploiting the extremely large data volume and dealing with infinite variations in objects.

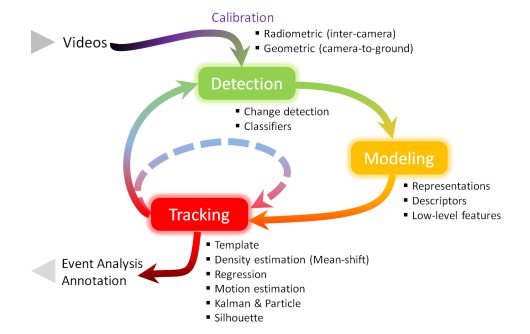
A lot of progress has been made in machine learning for detecting and classifying objects in an image in real time. However, none of these algorithms considers time and continuity. Every time an obstacle is detected, these algorithms assume it is a new obstacle. There is no notion of tracking objects across frames.

Numerous researches have been conducted using and proposing novel architectures and methods in computer vision. These methods are, in many cases, developed by researchers and academia, further proving the importance of computer vision in the actual research overview.

we will analyze the evolution of object tracking techniques and we’ll go through the current SOTA object tracking algorithms that use deep learning methods (i.e., DeepSORT and ByteTrack).

## Object Tracking, Defined

Object tracking is the process of tracking an object through different frames in such a way that its position and direction are known throughout the sequence. As a result, the tracking task includes two major sub-tasks: object detection and object re-identification between frames. These two allow us to determine the exact position of the various objects in an image and the path they have taken in the field of view represented by the images.



The greater the number of moving objects, the harder tracking becomes. The monitoring of open or crowded environments such as traffic surveillance, autonomous vehicles, sports and entertainment, and security is made easier and faster thanks to the research done in these areas.

## Common Challenges

* **Occlusion.** An object is partially or entirely blocked by another object.
* **Scale Variation.** Object detection will not work when an object’s size varies significantly.
* **Motion Blur.** The moving object in a frame may be distorted and cease to appear as it once did.
* **Memory.** Keeping track of objects that appear and disappear from the scene is important in some use cases. Most object tracking algorithms lack this feature.
* **Background Clutter.** This happens when the background of the object is the same colour or shape.
* **Low Resolution.** An image has low resolution or is pixelated.

In situations where there is a sudden change in speed or direction of motion, motion modeling alone is ineffective. Visual appearance modeling is more accurate since it is aware of the appearance of the object being tracked. This method works well for tracking a single object, but it is insufficient for tracking multiple objects in a video frame. In environments with resource limitations (embedded systems), conventional object-tracking algorithms can be helpful.

In a system with limited resources (embedded systems), conventional object-tracking algorithms can be helpful. In terms of accuracy, deep learning-based approaches to object tracking are far superior to traditional trackers.

## Evolution of Object Tracking Techniques

The challenging issue of object tracking in a video has been addressed through the introduction of numerous techniques and algorithms. These methods depend on either visual appearance modeling or motion modeling. Motion modeling captures an object’s dynamic behavior. These non-deep learning techniques include mean-shift tracking, optical flow, Kalman filtering, and Kanade-Lucas-Tomashi feature tracking.

As the name suggests, single object tracking (SOT) follows a single target object throughout a video. It frequently occurs in applications where the observer must ignore all other objects in the same environment in order to focus on a single distinct object. Multiple object tracking (MOT) or multi-target tracking (MTT), the second level, is more tricky since it involves finding and following multiple objects in a video.

In a typical multiple object tracking algorithm, each object of interest is recognised in the first frame(s), placed in bounding boxes, given a unique set of coordinates, then the movement of these bounding box coordinates is tracked over a series of back-to-back frames. Plotting the distinct trajectory of each detected object is necessary.

So what’s next? SORT.

## Top Emerging Object Tracking Techniques

### Simple Online And Real-time Tracking (SORT)

Simple Online and Real-time Tracking (SORT), published in 2017 by Bewley et al., is one of the first algorithms to handle object tracking in real-time.

There are three steps to perform object tracking using SORT:

1. Object detection.
2. Kalman filter to predict the position of the tracked object in the next frame.
3. Hungarian algorithm to index or label the id of the moving object.

To identify and track target objects in a video frame, the SORT algorithm uses a combination of the Kalman filter and the Hungarian algorithm.

Both motion estimation and data association through are based on the location and dimensions of the bounding boxes. The object detector employs a faster RCNN. A linear constant velocity model that is unaffected by the movement of the camera or any other objects estimates the displacement of objects in the subsequent frames.

The new target states are used to forecast the bounding boxes that will later be compared with the detected boxes in the current timeframe for the ID assignment, or data association task. The best box to pass on the identity is chosen using the Hungarian algorithm and the IoU metric. The intersection over union (IoU) specifies how much overlap there is between the predicted and [ground truth](https://www.tasq.ai/glossary/ground-truth/) bounding boxes.

The performance of object detectors and trackers has greatly improved due to the rapid development of deep learning models and computing power. Object recognition can be viewed in a deep learning framework as a task of labeling different objects in an image frame with their correct classes and predicting their bounding boxes with a high probability.

### DeepSORT

Published in 2017, [DeepSORT](https://arxiv.org/pdf/1703.07402.pdf), an addition to SORT, is the most well-known and widely used object tracking frameworks. It tracks not only distance and velocity, but also how that person/object appears. Deep sort enables us to add this feature by computing deep features for each bounding box and using deep feature similarity to factor into the tracking logic.

When the view of objects is blocked, SORT generates an excessive number of “identity switches.” As a result, they propose enhancing the motion model (Kalman filter) with a deep learning component that incorporates an object’s visual features.

The Kalman filter is an essential part of DeepSORT. Eight variables make up the state: (u,v,a,h,u’,v’,a’,h’), where (u,v) are the centers of the bounding boxes, (a) is the aspect ratio, and (h) is the height of the image. The other variables are the variables’ individual velocities.

### Why DeepSORT Works Better

To deal with the correlation of frame-by-frame data, the SORT algorithm employs a simple Kalman filter and the Hungarian algorithm. This algorithm has produced good results at high frame rates.

However, because SORT ignores the detected target’s appearance feature, it is only accurate when the uncertainty in target state estimation is low. Furthermore, to improve tracking efficiency, SORT deletes targets that have not been matched in a continuous frame; this causes an ID switch, meaning the ID assigned to the target is easily changed and causes problems. An ID switch occurs when two objects that are similar overlap or blend, causing the identities to switch. Therefore, tracking the object ID is difficult. When the object motion is small, the SORT algorithm helps reduce the occluder target. ID switches work well, but fail in cases involving crowded scenes and fast motion. DeepSORT decreases both ID switches and occlusions.

DeepSORT adds appearance information and borrows the ReID model to extract appearance features, reducing the number of operations. ReID is primarily used to link bounding boxes and tracks. DeepSORT also converts SORT’s matching mechanism based on the IoU cost matrix into a Matching Cascade and IoU matching mechanism. The core idea behind Matching Cascade is to prioritize track matching to the targets that appear more frequently in the long-term occluded targets. This method solves the long-standing matching problem of occluded targets. In the final stage of matching, DeepSORT performs IoU matching on unmatched tracks and detection targets, which can alleviate large changes caused by apparent mutations or partial occlusion.

### ByteTrack

Published in 2021, Bytetrack achieves state-of-the-art performance on Multi Object tracking tasks. Most tracking techniques identify people by connecting detection boxes to scores above a threshold. Negligible true object missing and fragmented trajectories result from the simple disregard of the objects with low detection scores.

Byte is an efficient association method that matches all detection boxes, including those with high and low scores. Byte first matches the high score detection boxes to the tracklets. To forecast where the tracklets will be in the new frame, it employs the Kalman Filter. The IoU of the predicted box and the detection box is used to calculate the motion similarity.

## Conclusion

I hope this post provided a good overview of visual object tracking and insights into key methods for successful object tracking.

The goal of object tracking is to find an object’s position in video sequences and create a route for it over time. In the first stage, an object detection algorithm determines the region of interest in each frame, and then tracking corresponds to objects across frames. The object region is projected in the final stage by iteratively updating the object location obtained from previous frames.