Deep Network Flow for Multi-Object Tracking

Literature Survey

Group 20
Hritik Soni (ID:2014A2PS0480P)

What is meant by Object Tracking?

Object tracking is the process of locating a moving object (or multiple objects) over time using a camera. It has a variety of uses, some of which are: human-computer interaction, security and surveillance, video communication and compression, augmented reality, traffic control, medical imaging and video editing. Object tracking can be a time consuming process due to the amount of data that is contained in video. Adding further to the complexity is the possible need to use object recognition techniques for tracking, a challenging problem in its own right.

What all algorithms are used in the industry?

To perform video tracking an algorithm analyzes sequential video frames and outputs the movement of targets between the frames. There are a variety of algorithms, each having strengths and weaknesses. Considering the intended use is important when choosing which algorithm to use. There are two major components of a visual tracking system: target representation and localization, as well as filtering and data association.

Target representation and localization is mostly a bottom-up process. These methods give a variety of tools for identifying the moving object. Locating and tracking the target object successfully is dependent on the algorithm. For example, using blob tracking is useful for identifying human movement because a person's profile changes dynamically. Typically the computational complexity for these algorithms is low. The following are some common target representation and localization algorithms:

Kernel-based tracking (mean-shift tracking): an iterative localization procedure based on the maximization of a similarity measure (Bhattacharyya coefficient).

Contour tracking: detection of object boundary (e.g. active contours or Condensation algorithm). Contour tracking methods iteratively evolve an initial contour initialized from the previous frame to its new position in the current frame. This approach to contour tracking directly evolves the contour by minimizing the contour energy using gradient descent.
Filtering and data association is mostly a top-down process, which involves incorporating prior information about the scene or object, dealing with object dynamics, and evaluation of different hypotheses. These methods allow the tracking of complex objects along with more complex object interaction like tracking objects moving behind obstructions. Additionally the complexity is increased if the video tracker (also named TV tracker or target tracker) is not mounted on rigid foundation (on-shore) but on a moving ship (off-shore), where typically an inertial measurement system is used to pre-stabilize the video tracker to reduce the required dynamics and bandwidth of the camera system. The computational complexity for these

algorithms is usually much higher. The following are some common filtering algorithms:

Kalman filter: an optimal recursive Bayesian filter for linear functions subjected to Gaussian noise.It is an algorithm that uses a series of measurements observed over time, containing noise (random variations) and other inaccuracies, and produces estimates of unknown variables that tend to be more precise than those based on a single measurement alone.

Particle filter: useful for sampling the underlying state-space distribution of nonlinear and non-Gaussian processes.

Cost Functions and Network Flows

Independent of the type of association model, a proper choice of the cost function is crucial for good tracking performance. Many works rely on carefully designed but hand-crafted functions. Many works rely on carefully designed but hand-crafted functions. For instance, only rely on detection confidences and spatial(i.e., bounding box differences) and temporal distances. Research - Zhang et al. and Zamir et al. include appearance information via color histograms. Other works explicitly learn affinity metrics, which are then used in their tracking formulation. For instance, Research - Li et al. build upon a hierarchical association approach where increasingly longer tracklets are combined into trajectories. Affinities between tracklets are learned via a boosting formulation from various hand-crafted inputs including length of trajectories and color histograms. This approach is extended by learning affinities on-line for each sequence. Similarly, Res. Bae and Yoon learn affinities on-line with a variant of linear discriminant analysis. Res. Song et al. train appearance models on-line for individual trajectories when they are isolated, which can then be used to disambiguate from other trajectories in difficult situations like occlusions or interactions. Res. Leal-Taix´ e et al. train a Siamese neural network

In this paper, they demonstrate their end-to-end formulation for association problems with the example of network flows for multi-object tracking.