

Review: Caesar: Cross-camera Complex Activity Recognition

Problem:

- A major barrier is that, given a combination of surveillance cameras, what architecture and algorithms should an end-to-end system fuse to provide accurate and scalable sophisticated activity detection.
- A significant obstacle is to detect user-defined complex activities correctly. The challenge is how to specify and identify complex activities.
- Other challenges include Scalability and Efficiency, ie, how the system optimizes the hardware resource usage and how to collaborate with the server and the mobile efficiently?

Solution

- The author introduces a hybrid multi-camera complex activity detection system that combines traditional rule-based activity detection with DNN-based activity detection.
- Caesar processes these using a three-stage pipeline:
 - In the object detection stage, Caesar generates bounding boxes of actors and objects seen in each frame.
 - A re-identification and tracking module processes these bounding boxes. It extracts tubes for actors and objects by tracking them across multiple frames and determines whether actors in different cameras represent the same person.
 - Finally, a graph matching and lazy action detection module determines whether the relationships between actor and object tubes match pre-defined rules for complex activities

Experimental Results

- The evaluation for Caesar's accuracy and scalability is based on a publicly available multi-camera data set. Caesar has a recall of 61.0% and a precision of 59.5% precision.
- Better action and ReID DNNs could improve the accuracy to 100%.

Pros:

- Caesar meets the requirements and challenges as follows:
 - it processes streams continuously, so can detect events in near-real time; it incorporates robustness optimizations for tracking, re-identification, and graph matching to ensure accuracy.
 - Caesar can also reduce wireless bandwidth, on-board camera memory, and detection latency by an order of magnitude while achieving good precision and recall for all complex activities on a public multi-camera dataset.

Cons:

- One of difficulty with Caesar is evaluations with a larger dataset that has more cameras and a larger set of complex activities.
- Deploying Caesar as a publicly accessible service and extending Caesar to recognize complex activities in moving cameras (e.g., drones, and cars).

Doubt:

- What extent does Caesar's graph matching algorithm contribute to detection error?
- How does the system specify complex activities across multiple non-overlapping cameras?