2022 IEEE/ACM 7th Symposium on Edge Computing (SEC)

REV: A Video Engine for Object Re-identification at the City Scale

Tiantu Xu, Purdue University
Kaiwen Shen, Purdue University
Yang Fu, UIUC
Humphrey Shi, UIUC
Felix Xiaozhu Lin, University of Virginia

2022 IEEE/ACM 7th Symposium on Edge Computing (SEC) | 978-1-6654-8611-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/SEC54971.2022.00022

***Abstract*—Object re-identification (ReID) is a key application of city-scale cameras on the edge. It is challenged by the limited accuracy of vision algorithms and the large video volume. We present REV, a practical ReID engine that builds upon three new techniques. (1) REV formulates ReID as a spatiotemporal query. Instead of retrieving all the images of a target object, it looks for locations and times in which the target object appeared. (2) REV makes robust assessment of the target object occurrences by clustering unreliable object features. Each resultant cluster represents the general impression of a distinct object. (3) REV samples cameras strategically in order to maximize its spatiotemporal coverage at low compute cost. Through an evaluation on 25 hours of videos from 25 cameras, REV reached a high accuracy of 0.87 (*recall at 5*) across 70 queries. It runs at 830× of video realtime in achieving high accuracy.**

***Index Terms*—camera networks, vehicle re-identification**

I. INTRODUCTION

City-scale camera networks: a key edge application As video intelligence advances and camera cost drops, city cameras expand fast [31], [12]. Camera networks are considered as an important edge application [27]; edge servers are often the ideal platforms for processing camera videos, as transmitting the videos to data centers is often expensive. Being strategically deployed near key locations such as highway entrances or road intersections, multiple cameras (2–5 per location as reported [59], [46]) offer complementary and overlapped viewpoints of scenes.

Object ReID on city videos A key application of city cameras is object re-identification (ReID): given an input image of an object X, search for occurrences of X in a video repository. ReID has been an important computer vision task, seeing popular use cases such as criminal investigation and traffic planning [35], [46], [47]. Many ReID algorithms are proposed recently as fueled by neural networks [71], [57], [69], [72], [14], [58], [23], [13]. Object ReID over city videos is typically "finding a needle in a haystack". The videos to be queried are long and produced by many cameras. The videos may *not* contain the input image or any images from the camera that produced the input image (called the *origin* camera). The occurrences of the target object can be rare and transient. For instance, in a popular dataset of city traffic videos [59], 99% of distinct vehicles only appear for less than 10 seconds.

A common pipeline for ReID is shown in Figure 1: 1 given an input image of target object X, the pipeline extracts its feature, e.g., using ResNet-152 [19] to extract a 1024-dimension vector [58], [23]; 2 from the queried videos, the pipeline detects all bounding boxes belonging to the same *class* as the target, e.g., using YOLOv3 [55]; 3 the pipeline extracts the features of all detected bounding boxes; 4 it calculates pairwise similarities between X and the bounding boxes. The similarity is often measured as feature distance [45], where a shorter distance suggests a higher similarity between X and a bounding box. Of the four stages, stage 2 and stage 3 are the most expensive. For instance, extracting features in stage 3 is three orders of magnitude slower than calculating feature distance in stage 4. The cost of stage 2 and stage 3 further grows with the amount of videos. This pipeline structure is widely used, e.g., by almost all participants in popular vehicle ReID challenges [23], [58], [43].

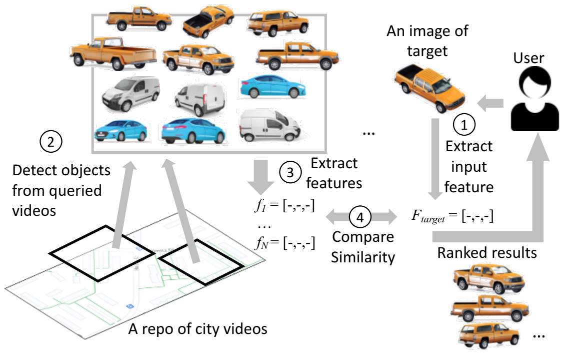

Fig. 1: **The classic pipeline for object ReID**, which formulates ReID as an image retrieval task

Proliferating ReID algorithms call for practical ReID systems. Our driving use case is vehicle ReID, where identifiably personal information such as license plates are intentionally blocked for privacy [59]. Vehicle ReID is one of the most important ReID problems [47]. The solutions are likely transferable to other object classes for ReID.

Challenge 1: Limitations of modern ReID algorithms By its definition, ReID focuses on differentiating objects of the *same class*, e.g., vehicles. The inherent difficulty is that in real-world videos, many objects of the same class exhibit only subtle visual differences. Meanwhile, bounding boxes of the same object that are captured by the same or different cameras, may appear quite different. As a result, ReID can be challenging even to human eyes, let alone algorithms. As

978-1-6654-8611-8/22/$31.00 ©2022 IEEE
DOI 10.1109/SEC54971.2022.00022