P

Person Re-identification: Current Approaches and Future Challenges

Rameswar Panda and
Amit K. Roy-Chowdhury
Department of Electrical and Computer Engineering, University of California, Riverside, CA, USA

Synonyms

Multi-camera scene understanding; Multi-camera tracking; Person Re-identification

Related Concepts

- ▶ Calibration of Multi-Camera Setups
- ▶ Multi-Camera Human Action Recognition
- ▶ Tracking in Camera Networks
- ▶ Transfer Learning

Definition

The objective of person re-identification (re-id) is to associate targets across cameras with non-overlapping fields of view. Specifically, a person re-identification algorithm takes two images from two non-overlapping cameras and provides a decision whether those two images are of the same person or not.

Background

In the last few years, the problem of re-identifying persons across multiple non-overlapping cameras has received increasing attention. A thorough recent survey on person re-identification (re-id) can be found at [1]. Most existing person re-id techniques are based on supervised learning. These methods either seek the best feature representation [2, 3] or learn discriminant metrics [4, 5] that yield an optimal matching score between two cameras or between a gallery and a probe image. Recently, deep learning methods have shown significant performance improvement on image classification and have been applied to person re-id [6, 7]. Considering that a modest-sized camera network can easily have hundreds of cameras, these supervised re-id models will require huge amount of labeled data which are difficult to collect in real-world settings. In an effort to bypass tedious labeling of training data in supervised re-id models, there has been recent interest in using active learning for labeling examples in an interactive manner. In [8], an entropy-based selection approach is proposed for reducing manual annotation. In [9], the authors uses a dominant clustering-based approach for

K. Ikeuchi, *Computer Vision*,
https://doi.org/10.1007/978-3-030-03243-2_825-1

probe relevant set selection and utilize it for pair selection in a dynamic setting.

Unsupervised learning has received little attention in person re-identification because of their weak performance on benchmarking datasets compared to supervised methods. Representative methods along this direction use either hand-crafted appearance features [10] or saliency statistics [11] for matching persons without requiring huge amount of labeled data. Recently, sparse dictionary learning-based methods have also been utilized in an unsupervised setting [12].

Domain adaptation [13, 14], which aims to adapt a source domain to a target domain, has been successfully used in many areas of computer vision and machine learning, e.g., object classification, and action recognition and speech processing. Despite its applicability in classical computer vision tasks, domain adaptation for person re-identification still remains as a challenging and under-addressed problem. Only very recently, domain adaptation for re-id has begun to be considered [15]. However, these studies consider only improving the re-id performance in a static camera network with fixed number of cameras. Furthermore, most of these approaches learn supervised models using labeled data from the target domain.

Main Text

With the advancement of imaging sensor technology, surveillance systems have seen remarkable increase in various applications ranging from law enforcement to large retail applications, from facility access and environment monitoring. Even though the sensing devices are becoming cheaper, monitoring a wide area by deploying a large number of cameras is still not feasible due to the amount of human supervision, privacy concerns, and maintenance costs involved. As a result, only a small part of the whole area is covered by a number of cameras with non-overlapping fields of view (FoVs). The non-overlapping camera FoVs leave blind gaps which are critical in the sense that no information can be obtained from these areas. This raises the need for automated methods able to extract and access high-level semantic information carried by the extremely high volume of recorded video data. As a result of losing a person when he/she leaves a camera FoV, it is extremely challenging to reassociate the same person at a different location and time among multiple persons. This inter-camera person association problem is known as the person re-identification problem.

In spite of a surge of effort put in by the research community in recent years, re-identification has remained quite an open issue due to a number of hard challenges. First, footages are recorded in an uncontrolled environment by cameras with large FoVs, generating low-resolution images of the targets. This makes the acquisition of discriminating biometric features (e.g., face and gait features) hard as well as unreliable. Due to the poor quality of the acquired biometric features, methods relying on such features perform unsatisfactorily. As a result, visual appearance features are, still, the first choice in re-identification problems. As a target's appearance often undergoes large variations across non-overlapping camera views due to significant changes in viewing angle, lighting, background clutter, and occlusion, the appearance features for the target can be very different from camera to camera.

The computer vision community has tried to address the person re-identification problem by designing discriminative signatures for each target or by finding a non-Euclidean metric which minimizes the distance between features of the same target across cameras. Similar to the most other visual recognition problems, the most successful approaches have been based on supervised training phases. Labeled data across pairs of cameras are used to learn models that define the transformation between the views in two cameras, and these learned models are used to associate between images during the testing phase. However, this level of supervision hampers scalability of the problem because of the need to label quantities of data, which grows with the size of the camera network and the variety of conditions that may be encountered.

Below we discuss future research directions in person re-identification, especially the possibility of significantly reducing the level of supervision without any sacrifice in performance. This will ensure that it is possible to scale person re-identification problems to larger and larger networks of cameras without compromising accuracy of the association task. We specifically focus on two problems. The first problem relates to scalability of person re-identification as the number of people grows. The second relates to scalability as the size of the camera network grows.

Optimal Subset Selection for Labeling

Given unlabeled training data across a network of cameras and a similarity measure, can we select a minimal subset of images that should be labeled and from which the person re-identification models can be learned? The intuition here is that if we choose this minimal subset judiciously, the labels can be propagated using the similarity measure to the rest of the dataset. Thus, most of the labels would be obtained automatically with only a small subset of images being labeled.

Building on preliminary result on network consistent re-id [16, 17], we now ask whether *consistency can be used to reduce the labeling effort*. However, in order to take advantage of consistency relations for reducing the labeling effort, we have to choose image pairs in a judicious manner. Toward this objective, we can represent a camera network as an edge-weighted k-partite graph. Nodes on the graph are observations of the targets, and edge weights are computed based on similarities in the observations. We formulate the pair subset selection as the following combinatorial optimization problem: given a complete k-partite graph $G_k = (V, E)$ with nonnegative edge weights and an integer B, choose a maximum-weight set S of edges from E such that $G' = (V, S)$ is triangle-free and $|S| \leq B$, where B is labeling budget. Once the subset for labeling is selected, the remaining labels can be obtained by label propagation on the graph, and existing methods for learning re-identification models can be used. Note that this subset selection is a NP-hard problem as it requires searching over every subset of image pairs. Polynomial time sub-optimal algorithms with linear or quasilinear time complexity can be adopted to solve this problem in an efficient way. Preliminary experiments in [18] have shown that we are able to achieve same recognition performance as the state of the art, with only 8% manual labels on the challenging Market-1501 dataset with six fixed cameras [19].

On-Boarding New Cameras Through Transfer Learning

We now address a very practical problem in camera networks, which has received little attention in the person re-identification literature. *Given a camera network where the inter-camera transformations/distance metrics have been learned in an intensive training phase, how can we on-board new cameras into the installed system with minimal additional effort?* To illustrate such a problem, let us consider a scenario with $\mathcal{N}$ cameras for which we have learned the optimal pair-wise distance metrics, so providing high re-id accuracy for all camera pairs. However, during a particular event, a new camera may be temporarily on-boarded to cover a certain related area that is not well-covered by the existing network of $\mathcal{N}$ cameras. Despite the dynamic and open nature of the world, almost all work in re-identification assume a *static* and *closed* world model of the re-id problem where the number of cameras is fixed in a network. Given newly introduced camera(s), traditional re-id methods will try to relearn the inter-camera transformations/distance metrics using a costly training phase. This is impractical since labeling data in the new camera and then learning transformations with the others is time-consuming and defeats the entire purpose of temporarily introducing the additional camera.

In [20], the authors have shown that it is possible to add a new camera to an existing network using transfer learning. First, they propose an unsupervised method based on geodesic flow kernel that can effectively find the best source camera to adapt with a target camera. Given camera pairs, each consisting of 1 (out of $\mathcal{N}$) source camera and a target camera, they first compute a kernel over the subspaces representing the data of both cameras and then use it to find the kernel distance across the source and target camera. Then, they rank the source cameras based on the average distance and choose the one with lowest distance as the best source camera to pair with the target camera. This is intuitive since a camera which is closest to the newly introduced camera will give the best performance on the target camera and hence is more likely to adapt better than others. Second, they introduce a transitive inference algorithm to exploit information from best source camera to improve accuracy across other camera pairs. Extensive experiments on multiple benchmarks show that the proposed method significantly outperforms the state-of-the-art unsupervised alternatives while being extremely efficient to compute.

Open Problems

We now discuss two open research problems in re-identification, such as network-level knowledge transfer and learning in mobile networks.

Knowledge Transfer Across Networks

Above, we explained how it is possible to add a new target camera to an existing network of cameras using transfer learning with no additional supervision for the new camera. However, transfer learning across networks is still a largely under-addressed problem with many challenges. Given multiple existing source networks and a newly installed target network with limited labeled data, we first need to find the relevance/similarity of each source network, or parts thereof, in terms of amount of knowledge that it can transfer to a target network. Developing efficient statistical measures for finding relevance in a multi-camera network with significant changes in viewing angle, lighting, background clutter, and occlusion can be a very interesting future work. Furthermore, labeled data from source networks are often a subject of legal, technical, and contractual constraints between data owners and customers. Thus, existing transfer learning approaches may not be directly applicable in such scenarios where the source data is absent. The question we want to ask here is whether learned source models, instead of source data, can be used as a proxy for knowledge transfer across networks. Compared to the source data, the well-trained source model(s) are usually freely accessible in many applications and contain equivalent source knowledge as well. Knowledge distillation [21] along with attention transfer techniques can be adopted to transfer knowledge from a number of existing labeled networks to an unlabeled target network containing targets which never appeared in the source network.

Learning in Mobile Camera Networks

Despite the success of existing person re-identification works in static platforms, considering mobile cameras (e.g., network of robots) opens up exciting new research problems in terms of learning such data association models. It is not possible to learn transformation models between every possible pair of views in two mobile cameras due to the constantly changing nature of the videos being captured. Thus, in order to efficiently learn data association models, we need the data to represent the variety of scenarios that will be encountered by the mobile cameras. A semi-supervised pipeline that uses limited manual training data along with newly generated data through a generative adversarial network (GAN) could be a possibility. One initial approach could be to use the unlabeled samples produced by a Multi-view Generative Adversarial Network in conjunction with the labeled training data to learn view-invariant features in a mobile network. Moreover, apart from generating samples, we may need to evolve the learned models over time based on the observed features.

References

1. Zheng L, Yang Y, Hauptmann AG (2016) Person re-identification: past, present and future. arXiv preprint:1610.02984
2. Lisanti G, Masi I, Bagdanov AD, Del Bimbo A (2015) Person re-identification by iterative re-weighted sparse ranking. TPAMI 37:1629–1642
3. Martinel N, Das A, Micheloni C, Roy-Chowdhury AK (2015) Re-identification in the function space of feature warps. TPAMI 37:1656–1669
4. Liao S, Hu Y, Zhu X, Li SZ (2015) Person re-identification by local maximal occurrence representation and metric learning. In: CVPR
5. Liao S, Li SZ (2015) Efficient psd constrained asymmetric metric learning for person re-identification. In: ICCV
6. Yi D, Lei Z, Liao S, Li SZ et al (2014) Deep metric learning for person re-identification. In: ICPR
7. Liu J, Zha ZJ, Tian Q, Liu D, Yao T, Ling Q, Mei T (2016) Multi-scale triplet CNN for person re-identification. In: Proceedings of the 2016 ACM on multimedia conference
8. Das A, Panda R, Roy-Chowdhury A (2015) Active image pair selection for continuous person re-identification. In: ICIP
9. Martinel N, Das A, Micheloni C, Roy-Chowdhury AK (2016) Temporal model adaptation for person re-identification. In: ECCV
10. Ma B, Su Y, Jurie F (2012) Local descriptors encoded by fisher vectors for person re-identification. In: ECCV
11. Zhao R, Ouyang W, Wang X (2013) Unsupervised salience learning for person re-identification. In: CVPR
12. Kodirov E, Xiang T, Fu Z, Gong S (2016) Person re-identification by unsupervised\ell _1 graph learning. In: ECCV
13. Kulis B, Saenko K, Darrell T (2011) What you saw is not what you get: domain adaptation using asymmetric kernel transforms. In: CVPR
14. Patel VM, Gopalan R, Li R, Chellappa R (2015) Visual domain adaptation: a survey of recent advances. SPM 32:53–69
15. Ma AJ, Li J, Yuen PC, Li P (2015) Cross-domain person reidentification using domain adaptation ranking SVMs. TIP 24(5):1599–613
16. Chakraborty A, Das A, Roy-Chowdhury AK (2016) Network consistent data association. TPAMI 35:1622–1634
17. Das A, Chakraborty A, Roy-Chowdhury AK (2014) Consistent re-identification in a camera network. In: ECCV
18. Roy S, Paul S, Young NE, Roy-Chowdhury AK (2018) Exploiting transitivity for learning person re-identification models on a budget. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7064–7072
19. Zheng L, Shen L, Tian L, Wang S, Wang J, Tian Q (2015) Scalable person re-identification: a benchmark. In: ICCV
20. Panda R, Bhuiyan A, Murino V, Roy-Chowdhury AK (2017) Unsupervised adaptive re-identification in open world dynamic camera networks. In: CVPR
21. Hinton G, Vinyals O, Dean J (2015) Distilling the knowledge in a neural network. arXiv preprint: 1503.02531