

DOCTORAL THESIS

Person re-identification with limited labeled training data

Li, Jiawei

Date of Award:
2018

[Link to publication](#)

General rights

Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent URL assigned to the publication

Abstract

With the growing installation of surveillance video cameras in both private and public areas, it is an immediate requirement to develop intelligent video analysis system for the large-scale camera network. As a prerequisite step of person tracking and person retrieval in intelligent video analysis, person re-identification, which targets in matching person images across camera views is an important topic in computer vision community and has been received increasing attention in the recent years. In the supervised learning methods, the person re-identification task is formulated as a classification problem to extract matched person images/videos (positives) from unmatched person images/videos (negatives). Although the state-of-the-art supervised classification models could achieve encouraging re-identification performance, the assumption that label information is available for all the cameras, is impractical in large-scale camera network. That is because collecting the label information of every training subject from every camera in the large-scale network can be extremely time-consuming and expensive. While the unsupervised learning methods are flexible, their performance is typically weaker than the supervised ones. Though sufficient labels of the training subjects are not available from all the camera views, it is still reasonable to collect sufficient labels from a pair of camera views in the camera network or a few labeled data from each camera pair. Along this direction, we address two scenarios of person re-identification in large-scale camera network in this thesis, i.e. unsupervised domain adaptation and semi-supervised learning and proposed three methods to learn discriminative model using all available label information and domain knowledge in person re-identification.

In the unsupervised domain adaptation scenario, we consider data with sufficient labels as the source domain, while data from the camera pair missing label information as the target domain. A novel domain adaptive approach is proposed to estimate the target label information and incorporate the labeled data from source domain with the estimated target label information for discriminative learning. Since the discriminative constraint of Support Vector Machines (SVM) can be relaxed into a necessary condition, which only relies on the mean of positive pairs (positive mean), a suboptimal classification model learning without target positive data can be those using target positive mean. A reliable positive mean estimation is given by using both the labeled data from the source domain and potential positive data selected from the unlabeled data in the target domain. An Adaptive Ranking Support Vector Machines (AdaRSVM) method is also proposed to improve the discriminability of the suboptimal mean based SVM model using source labeled data. Experimental results demonstrate the effectiveness of the proposed method.

Different from the AdaRSVM method that using source labeled data, we can also improve the above mean based method by adapting it onto target unlabeled data. In more general situation, we improve a pre-learned classifier by adapting it onto target unlabeled data, where the pre-learned classifier can be domain adaptive or learned from only source labeled data. Since it is difficult to estimate positives from the imbalanced target unlabeled data, we propose to alternatively estimate positive neighbors which refer to data close to any true target positive. An optimization problem for positive neighbor estimation from unlabeled data is derived and solved by aligning the cross-person score distributions together with optimizing for multiple graphs based label propagation. To utilize the positive neighbors to learn discriminative classification model, a reliable multiple region metric learning method is proposed to learn a target adaptive metric using regularized affine hulls of positive neighbors as positive regions. Experimental results demonstrate the effectiveness of the proposed method.

In the semi-supervised learning scenario, we propose a discriminative feature

learning using all available information from the surveillance videos. To enrich the labeled data from target camera pair, image sequences (videos) of the tagged persons are collected from the surveillance videos by human tracking. To extract the discriminative and adaptable video feature representation, we propose to model the intra-view variations by a video variation dictionary and a video level adaptable feature by multiple sources domain adaptation and an adaptability-discriminability fusion. First, a novel video variation dictionary learning is proposed to model the large intra-view variations and solved as a constrained sparse dictionary learning problem. Second, a frame level adaptable feature is generated by multiple sources domain adaptation using the variation modeling. By mining the discriminative information of the frames from the reconstruction error of the variation dictionary, an adaptability-discriminability (AD) fusion is proposed to generate the video level adaptable feature. Experimental results demonstrate the effectiveness of the proposed method.

In short, the major contributions of this thesis are summarized as follows.

- An Adaptive Ranking Support Vector Machines (AdaRSVM) method is proposed to incorporate estimated target label information with the labeled data from source domain.
- A semi-supervised region metric learning algorithm is proposed to adapt a pre-learned classifier onto target positive regions without target positive samples for training.
- A novel adaptable feature representation learning using video variation dictionary is proposed for semi-supervised video based person re-identification.

Keywords: person re-identification, unsupervised domain adaptation, dictionary learning

Table of Contents

Declaration	i
Abstract	ii
Acknowledgements	v
Table of Contents	vii
List of Tables	ix
List of Figures	x
Chapter 1 Introduction	1
1.1 Background and Motivation	1
1.1.1 Background	1
1.1.2 Motivation of This Project	3
1.2 Review of Related Person Re-identification Methods	5
1.2.1 Supervised Learning Methods	5
1.2.2 Unsupervised Learning Methods	7
1.2.3 Semi-supervised Learning Methods	8
1.3 Contributions of This Thesis	9
1.4 Overview of This Thesis	10
Chapter 2 Domain Transfer Support Vector Ranking for Person Re- Identification without Target Camera Label Information	12

Chapter 3	Semi-Supervised Region Metric Learning without Positive Data for Person Re-identification	13
Chapter 4	Semi-supervised Adaptable Feature Learning For Video Based Person Re-identification	14
4.1	Introduction	14
4.2	Related Works	16
4.2.1	Video based Person Re-identification	16
4.2.2	Variation Dictionary Learning	17
4.3	Adaptable Feature Learning Under Large Intra-view Variations	18
4.3.1	Video Variation Dictionary Learning	19
4.3.2	Estimating Frame-level Adaptable Feature Using Multiple Sources Domain Adaptation	22
4.3.3	Video Level Adaptable Feature via AD Fusion	24
4.3.4	Optimization	25
4.4	Experiment	26
4.4.1	Datasets and Settings	26
4.4.2	Self-evaluation on dictionary size	27
4.4.3	Evaluation of feature decomposition and AD fusion	28
4.4.4	Comparison with multi-shot/video based person re-identifications	29
4.5	Conclusion	30
Chapter 5	Conclusions	32
5.1	Summary	32
5.2	Future work	34
	Bibliography	36
	Curriculum Vitae	45