Computer Vision and Image Understanding 000 (2017) 1–13

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Multiple metric learning with query adaptive weights and multi-task re-weighting for person re-identification

Jieru Jia[a,b,*], Qiuqi Ruan[a,b], Gaoyun An[a,b], Yi Jin[a,b]

[a] *Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China*
[b] *Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing 100044, China*

ARTICLE INFO

Article history:
Received 28 February 2016
Revised 30 January 2017
Accepted 4 April 2017
Available online xxx

Keywords:
Person re-identification
Multiple metric learning
Multi-task learning
Query adaptive weighting

ABSTRACT

Metric learning has been widely studied in person re-identification (re-id). However, most existing metric learning methods only learn one holistic Mahalanobis distance metric for the concatenated high dimensional feature. This single metric learning strategy cannot handle complex nonlinear data structure and may easily encounter overfitting. Besides, feature concatenation is incapable of exploring the discriminant capability of different features and low dimensional features tend to be dominated by high dimensional ones. Motivated by these problems, we propose a multiple metric learning method for the re-id problem, where individual sub-metrics are separately learned for each feature type and the final metric is formed as weighted sum of the sub-metrics. The sub-metrics are learned with the Cross-view Quadratic Discriminant Analysis (XQDA) algorithm and the weights to each sub-metric are assigned in a two-step procedure. First, the importance of each feature type is estimated according to its discriminative power, which is measured in a query adaptive manner as related to the partial Area Under Curve *(pAUC)* scores. Then, the weights of all feature types are learned simultaneously with a maximum-margin based multi-task structural SVM learning framework, in order to make sure that relevant gallery images are ranked before irrelevant ones within all feature spaces. Finally, the sub-metrics are integrated with the learned weights in an ensemble model, generating a sophisticated distance metric. Experiments on the challenging i-LIDS, VIPeR, CAVIAR and 3DPeS datasets demonstrate the effectiveness of the proposed method.

1. Introduction

Person re-identification (re-id), which aims to re-identify a target person in one camera when he/she disappears from another, has attracted huge interest over the recent decades. It is a special case of image retrieval (Zheng et al., 2014; 2015b) in video surveillance and undergoes severe challenges like significant variations in viewpoints, poses or illumination, and occlusions.

Feature extraction and metric learning are two key components in person re-id. Numerous features have been proposed for the problem, e.g. color histograms (CH) (Zheng et al., 2016b; 2013), color names (CN) (Liu et al., 2015; Zheng et al., 2015b), textures (Lisanti et al., 2015; Zheng et al., 2013) and attributes (Layne et al., 2012). Specifically, often exploited color histograms are RGB, HSV, YUV, Lab (Zheng et al., 2013) , while commonly used texture features include Schmid, Gabor, LBP and HOG (Lisanti et al., 2015). Attributes (Layne et al., 2012) refer to semantic description of people such as hair-style, shoe-type or clothing-style.

For metric learning, a distance metric is learned from training samples such that the inter-class distance is maximized whilst the intra-class distance is minimized. Many metric learning methods have been proposed for person re-id, such as Relative Distance Comparison (RDC) (Zheng et al., 2013), Large Margin Nearest Neighbor (LMNN) (Dikmen et al., 2010), Keep It Simple and Straightforward (KISSME) (Roth et al., 2012) and Pairwise Constrained Component Analysis (PCCA) (Jurie and Mignon, 2012).

While these methods could achieve encouraging performance, they all only learn one unitary distance metric for all the heterogeneous features. The weaknesses are two folds: First, single metric learning is not robust against the complex nonlinear data structure in person re-id. Images of the same person may be far away from each other due to dramatic appearance variations, while images of different people may be very close to each other, e.g., two different people wearing similar color or pattern (Jia et al., 2016). Second, single metric learning encounters the bottleneck of Small Sample Size (SSS), i.e. the number of training samples is far less than the feature dimension. Typically only hundreds of training

* Corresponding Author.
E-mail addresses: 12112059@bjtu.edu.cn (J. Jia), qqruan@bjtu.edu.cn (Q. Ruan), gyan@bjtu.edu.cn (G. An), yjin@bjtu.edu.cn (Y. Jin).

http://dx.doi.org/10.1016/j.cviu.2017.04.003
1077-3142/

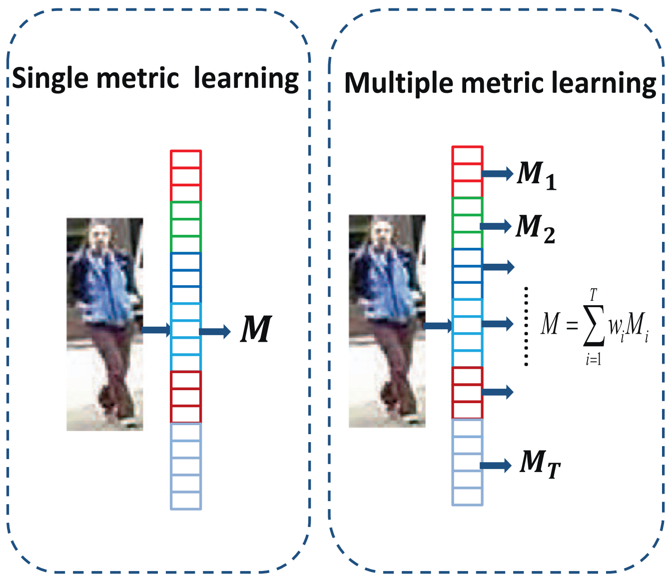

Fig. 1. The different module of single metric learning and the proposed multiple metric learning.

samples are available whilst the feature dimension is often in the order of thousands or higher. Consequently, traditional single metric learning methods have to resort to dimensionality reduction and/or matrix regularization, which may lead to sub-optimal solutions and loss of discriminative power (Zhang et al., 2016a).

Another issue with single metric learning is that it cannot deal with multiple feature representations directly. In person re-id, we usually have access to multiple heterogeneous features like color histograms, textures, and color names et al. for each image. Each feature has its unique characteristic and renders various performance, thus the fusion of features becomes a hard task. In most metric learning algorithms, different features are concatenated into a high dimensional vector and a corresponding distance metric is learned from the combined vector.

This feature-level fusion scheme has several disadvantages compared to score-level fusion (Eisenbach et al., 2015; Zheng et al., 2015b) and decision-level fusion (Liu et al., 2015; de Prates and Schwartz, 2015) approaches. First, the divergent importance and discriminant capability of each individual feature is ignored. Each feature is treated equally with uniform weighting, regardless of its particular characteristic. Second, low dimensional features tend to be neglected when combined with high dimensional features (Liu et al., 2015), thus their discriminant power might be discarded. Third, the combination makes the feature dimension quite high, which easily leads to overfitting because pair or triplet-based constraints become much easier to satisfy in a high-dimensional feature space and thereby results in poor generalization performance.

In light of these problems, we propose a multiple metric learning method for the re-id problem, where a final metric is learned as weighted sum of a bunch of sub-metrics. The sub-metrics are separately learned for each feature type in contrast to learning a unitary metric for all the features, see Fig. 1. The sub-metrics can be learned with off-the-shelf metric learning methods since it's not the focus of this work.

Now the problem arises: how to assign weights to the sub-metrics? We argue that different feature types should not be assigned with the same weight as in single metric learning. What's more, the importance of each feature type is not constant across all the individuals. On the contrary, it should be measured according to the query in question. Certain appearance features can be more important than others in describing a specific query and distinguishing him/her from other people (Liu et al., 2012). For instance, if the query is wearing bright shirt and pants without any texture, color features are clearly more important and should be given more weight. However, if the query is wearing plaid shirt or the illumination change is too drastic to rely on color features, texture information becomes critical and should be given more weight. With this intuition, we propose a query adaptive weight learning strategy for each sub-metric.

Furthermore, though conducting metric learning separately, these different feature types are not totally dependent but still relevant. They are complementary and may share information from each other (Cui et al., 2013; Hu et al., 2015). Therefore we aim to learn a number of weights collaboratively which can guarantee correct ranking within all feature spaces. To this end, by modeling the ranking in each feature space as a separate task, we utilize a maximum-margin based multi-task learning framework to jointly learn the weights. The framework treats the ranking of various feature types as different but related tasks and enables information propagation among tasks. By utilizing relatedness among different tasks, the framework can guarantee that within all feature spaces, relevant images to the query are all ranked before irrelevant ones. The multi-task ranking model is superior to traditional ranking methods (Paisitkriangkrai et al., 2015; Wu et al., 2011), especially when the training sample size is small for each task (Su et al., 2015).

In this paper, we propose a multiple metric learning and two-step weighting procedure for the re-id problem. Multiple sub-metrics are learned separately and linearly combined to form the final metric. The sub-metrics are learned with techniques from XQDA (Liao et al., 2015) due to its prominent effectiveness and high computation efficiency. The algorithm learns a discriminant low dimensional subspace and derived metric at the same time. The weights of each sub-metric are assigned in a two-step procedure: First, a query adaptive weight is assigned to each sub-metric. The weight is estimated with *pAUC* scores, which have the ability to measure the discriminative abilities of divergent features, as stated in Zheng et al. (2015b), Eisenbach et al. (2015) and Zhao et al. (2014). Second, to reach a more sophisticated decision, re-weighting parameters that can effectively rank relevant gallery images before irrelevant ones are simultaneously learned with a multi-task structural SVM learning framework. The ranking within each feature space is modeled as a task and we model the task relatedness in a way that all tasks are close to their mean, following (Pontil, 2004). Finally, the weighted sub-metrics are integrated in an ensemble model. In this way, the discriminant information of each feature type is effectively exploited. The feature dimension of each metric learning task is reduced thus overfitting is alleviated. Our method has the advantage to perform well and fast, even if only few samples are available. Experiments on four challenging datasets i-LIDS (Zheng et al., 2009), VIPeR (Gray and Tao, 2008), CAVIAR (Dong et al., 2011), 3DPeS (Baltieri et al., 2011) demonstrate the effectiveness of the proposed method.

The main contributions of the proposed method are:

- A metric ensemble model where final metric is represented as weighted sum of multiple sub-metrics is proposed, which is more discriminative and effective than holistic metric learning.
- A two-step weighting strategy is proposed, which assigns weights to the sub-metrics both separately and collaboratively. Separately, the weight of each sub-metric is learned adaptively to the query. Collaboratively, the weights are set to guarantee proper ranking within all feature spaces.
- A novel algorithm for calculating query adaptive weights is proposed, which considers the similarity label between query and gallery images in computing *pAUC* scores.

Please cite this article as: J. Jia et al., Multiple metric learning with query adaptive weights and multi-task re-weighting for person re-identification, Computer Vision and Image Understanding (2017), http://dx.doi.org/10.1016/j.cviu.2017.04.003