ARTICLE IN PRESS
JID: PR [m5G;October 16, 2017;7:5]
Pattern Recognition 000 (2017) 1–12
Contents lists available at ScienceDirect
Pattern Recognition
journal homepage: www.elsevier.com/locate/patcog

M^3L: Multi-modality mining for metric learning in person re-Identification

Xiaokai Liu, Xiaorui Ma, Jie Wang, Hongyu Wang*

School of Information and Communication Engineering, Dalian University of Technology, Dalian, 116024, PR China

ARTICLE INFO

Article history:
Received 27 December 2016
Revised 12 August 2017
Accepted 27 September 2017
Available online xxx

Keywords:
Person re-identification
Multi-modality mining
Diagonal model
Metric learning

ABSTRACT

Learning a scene-specific distance metric from labeled data is critical for person re-identification. Most of the earlier works in this area aim to seek a linear transformation of the feature space such that relevant dimensions are emphasized while irrelevant ones are discarded in a global sense. However, when training data exhibit multi-modality transitions, the globally learned metric would deviate from the correct metrics learned from each modality. In this study, we propose a multi-modality mining approach for metric learning (M^3L) to automatically discover multiple modalities of illumination changes by exploring the shift-invariant property in log-chromaticity space, and then learn a sub-metric for each modality to maximally reduce the bias derived from metric learning model with global sense. The experiments on the challenging VIPeR dataset and the fusion dataset VIPeR&PRID 450S have validated the effectiveness of the proposed method with an average improvement of 2–7% over original baseline methods.

1. Introduction

Mahalanobis metric learning has gained a considerable interest for person re-identification. The main idea is to find a linear transformation of the feature space such that relevant dimensions are emphasized while irrelevant ones are discarded in a global sense. Mahalanobis metrics reflect the visual camera-to-camera transitions. Researches in [1–3] have shown that appropriate an distance metric, learned from labeled data, usually allows for a more powerful retrieval result compared to the standard Euclidean distance. Most previous works in this area attempt to learn a discriminative distance metric in a global sense that keeps the sample pairs with same identity close, while ensuring that those from different persons remain separated. Nevertheless, these goals may conflict and cannot be simultaneously satisfied when the data exhibit multi-modality transitions. In this paper, multi-modality is defined as two or more transition modalities between the same camera pair. Under each modality, the images from different camera views are captured in similar inter-view illumination conditions, and exhibit similar color transition mode. In real-world person re-identification scenarios, the images are often collected over the course of several days or even months, therefore the illumination changes would exhibit multi-modality distributions, e.g. when images are captured respectively in the morning, in the evening, on a cloudy day, on a sunny day, the color transition might be significantly different under the influence of the surrounding context and environment. As shown in Fig. 1(a), apparently person 1 and 2 have limited color changes between camera 1 and camera 2, while person 3 and 4 undergo significant illumination changes. With regard to scenarios with multi-modality transformations, conventional metric learning approaches attempt to enforce a strong assumption that different modalities have a common or a shared subspace, and enforce person 1–4 have a common metric. Note that person 2 and 3 look quite similar in camera 1, while absolutely different in camera 2. The man in dull-red shirt and dark gray pants stays unchanged in some extent, while the woman in dull-red T-shirt and dark gray pants turns into the one in crimson T-shirt and denim blue pants. In this case, a set of colors would correspond to two sets of transformed colors in another camera, and the assumption of a common metric that attempt to bring the two modes together encounters difficulties and results in a compromised metric (Fig. 1(b)), leading to a significant degradation in ranking accuracy.

In order to address the aforementioned problem, this paper proposes a novel clustering based approach to multi-modality mining with agglomerative hierarchical clustering algorithm. Rather than satisfying all of the pair-wise constraints, our algorithm focuses on discovering subsets of sample pairs which sharing common illumination changes. Fig. 1(c) illustrates the ranking orders under a global metric $\tilde{\mathbf{M}}$ and two sub-metrics $\{\mathbf{M}_i\}_{i=1}^{2}$. Owing to the limitation of the scope, only top 15 ranking results are shown. Obviously, the sub-metric $\mathbf{M}_2$ focuses more on the dramatic color changes by training with sample pairs in the same modality, and

* Corresponding author.
E-mail address: whyu@dlut.edu.cn (H. Wang).

https://doi.org/10.1016/j.patcog.2017.09.041

Please cite this article as: X. Liu et al., M^3L: Multi-modality mining for metric learning in person re-Identification, Pattern Recognition (2017), https://doi.org/10.1016/j.patcog.2017.09.041

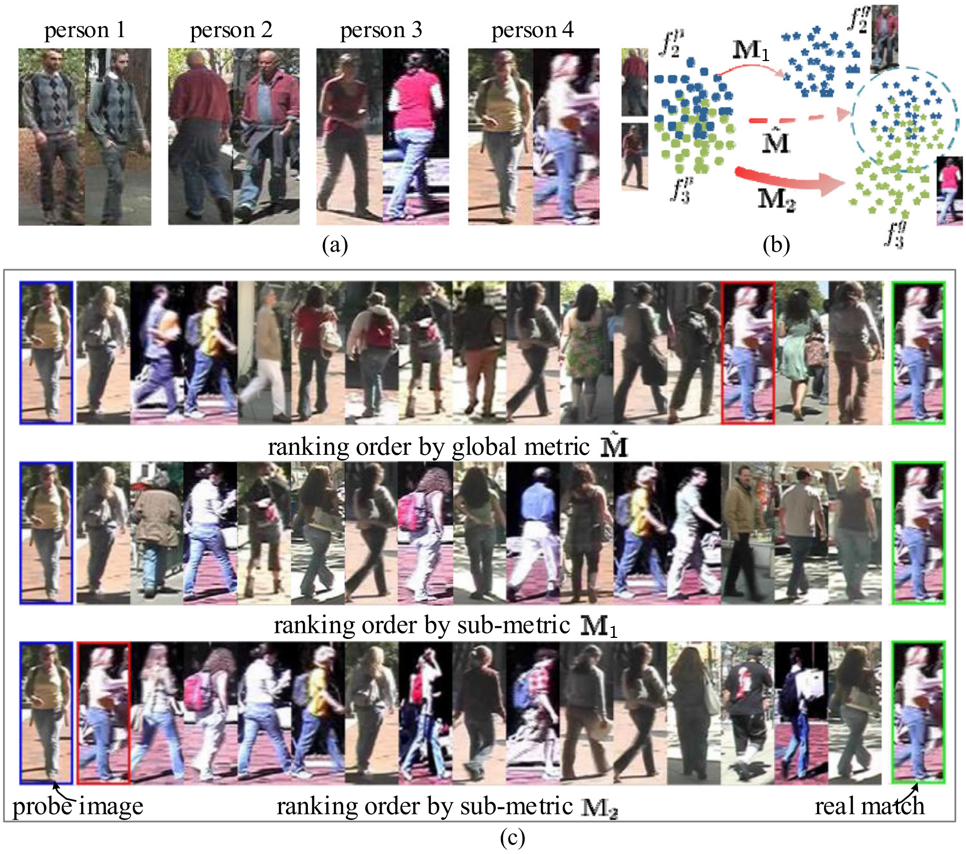

Fig. 1. (a) Four sample image pairs from VIPeR dataset. (b) Illustration of feature transformations between two transformed spaces. (c) Illustration of the ranking orders by global metric $\tilde{\mathbf{M}}$ and two sub-metrics $\{\mathbf{M}_i\}_{i=1}^2$. In (a), visually, person 1 and 2 have limited color changes, while person 3 and 4 undergo significant illumination changes. Person 2 and 3 look quite similar in camera 1, but quite different in camera 2. In (b), conventional metric learning approaches enforce person 2 and 3 to follow a common metric, resulting in a compromised metric $\tilde{\mathbf{M}}$. The proposed algorithm automatically discovers multiple modalities and learns two discriminative sub-metrics $\mathbf{M}_1$ and $\mathbf{M}_2$. In (c), obviously, the sub-metric $\mathbf{M}_2$ focuses more on the significant color changes and achieves a much better ranking result. The matched gallery images retrieved within top 15 are marked with red box. Best viewed in color. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

consequently achieves much better ranking result. Note that top 5 persons ranked by $\mathbf{M}_2$ are quite similar in appearance with the real match candidate, which implies that the sub-metric $\mathbf{M}_2$ is powerful in identifying this type of image pairs. The primary challenge in multi-modality mining for metric learning is the chicken-and-egg dilemma: as we need to cluster transition metrics, but not data itself, on one hand, to discover the multiple modalities of local distance metrics, we need to identify the transformation metric of each pair; on the other hand, to identify the transformation metric of each pair, we need to know the clusterings of local distance metrics. To resolve this problem, we propose a novel multi-modality mining approach by exploring shift-invariant properties in log-chromaticity space, which explicitly addresses the chicken-and-egg problem from a new perspective. The contributions of this study can be summarized in three-folds:

(1) While most existing metric learning based person re-identification approaches focus on learning a global discriminative metric for all the training sample pairs, we provide experimental evidence to support the view that some benefits can be obtained by multi-modality metrics guided by the shift-invariant property in log-chromaticity space. To our best knowledge, this is the first study that attempts to investigate the effect of multi-modality metrics in relation to the long-run surveillance scenes for person re-identification.
(2) Our idea of using inter-transition structure for guiding multi-modality metric learning, is novel. The shift-invariant property of the RGB values from related views in log-chromaticity space are discovered in this study. The shift vectors in log-chromaticity space usually does indeed contain discriminative structures of the illumination change modalities, due to different variations on separate channels.
(3) The proposed multi-modality mining approach is independent of specific metric learning algorithms. Any advanced metric learning method can exploit our framework to obtain superior performance in future.

Extensive experiments conducted on benchmarking re-identification dataset VIPeR and a simulated VIPeR&PRID 450S dataset demonstrate that person re-identification can significantly benefit from applying the multi-modality mining approach investigated in this study.

2. Related work

Mahalanobis metric learning has proven to be effective in person re-identification problem. The main idea is to seek an optimal metric that reflects the visual view-to-view transitions, allowing for a boosted ranking accuracy. In [3], a large number of Mahalanobis metric learning algorithms have been evaluated and shown to be effective in re-identification problem, for example, Linear Discriminant Metric Learning (LDML [4]), Information Theoretic Metric Learning (ITML [5]), Large Margin Nearest Neighbor (LMNN [6]), Large Margin Nearest Neighbor with Rejection (LMNN-R [7]), and Keep It Simple and Straightforward Metric Learning (KISSME