Pattern Recognition 79 (2018) 79–96

Contents lists available at ScienceDirect

Pattern Recognition

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

Maximal granularity structure and generalized multi-view discriminant analysis for person re-identification

Cairong Zhao[a,1,*], Xuekuan Wang[a,1], Duoqian Miao[a,*], Hanli Wang[a], Weishi Zheng[b], Yong Xu[c], David Zhang[d]

[a] Department of Computer Science and Technology, Tongji University, Shanghai, China
[b] School of Information Science and Technology, Sun Yat-sen University, Guangzhou, China
[c] Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology, China
[d] Biometrics Research Centre, Department of Computing, Hong Kong Polytechnic University, Hong Kong, China

ARTICLE INFO

Article history:
Received 10 July 2017
Revised 13 January 2018
Accepted 28 January 2018
Available online 2 February 2018

Keywords:
Person re-identification
Maximal granularity structure descriptor
Generalized multi-view discriminant analysis
Representation consistency

ABSTRACT

This paper proposes a novel descriptor called Maximal Granularity Structure Descriptor (MGSD) for feature representation and an effective metric learning method called Generalized Multi-view Discriminant Analysis based on representation consistency (GMDA-RC) for person re-identification (Re-ID). The proposed descriptor of MGSD captures rich local structural information from overlapping macro-pixels in an image, analyzes the horizontal occurrence of multi-granularity and maximizes the occurrence to extract a robust representation for viewpoint changes. As a result, the proposed descriptor of MGSD can obtain rich person appearance whilst being robust against different condition changes. Besides, considering multi-view information, we present a new GMDA-RC for different views, inspired by the observation that different views share similar data structures. The proposed metric learning method of GMDA-RC seeks multiple discriminant common spaces for multiple views by jointly learning multiple view-specific linear transforms. Finally, we evaluate the proposed method of (MGSD+GMDA-RC) on three publicly available person Re-ID datasets: VIPeR, CUHK-01 and Wide Area Re-ID dataset (WARD). For the VIPeR and CUHK-01, the experimental results show that our method significantly outperforms the state-of-the-art methods, achieving the rank-1 matching rates of 67.09%, 70.61%, and the improvements of 17.41%, 5.34%, respectively. For the WARD, we consider different pairwise camera views (camera 1–2, camera 1–3, camera 2–3) and our method can achieve the rank-1 matching rates of 64.33%, 59.42%, 70.32%, increasing of 5.68%, 11.04%, 9.06% compared with the state-of-the-art methods, respectively.

1. Introduction

Person re-identification (Re-ID) is the task of recognizing pedestrians observed from non-overlapping camera views in a surveillance system and it a challenging problem because of big intra-class variations in illumination, pose, viewpoint and occlusion [1,2]. To address this problem, existing approaches mainly focus on developing effective feature representation methods which are robust against the view/pose/illumination/background changes [3–26], or learning a distance metric [6–8]. The main development of person Re-ID have been shown in Table 1.

For feature representation, lots of models extract the low-level critical visual features (i.e. color [3–5], texture [6,7], structure [8–10], etc.), and consider the rich information of pedestrian's appearance. However, these handcrafted features would been constrained by scenarios for person Re-ID. In practice, most of the appearance-based approaches would integrate multiple features, such as ensemble of localized features [11], local maximal occurrence representation [12], salient match [13,14], ensemble of invariant features [15], camera correlation aware feature augmentation [16], kernel-based features [17], etc. Meanwhile, the semantic features [18–21] are also important for person Re-ID. Better yet, deep learning is also a noteworthy method, exhibiting an excellent performance in learning representation of person Re-ID [20,22–26]. These handcrafted or learning based descriptors have made impressive improvements over feature extraction, and advanced the person Re-ID research. Unfortunately, it is still extremely difficult to extract a robust feature that effectively adapts to severe changes and misalignment across disjoint views.

Another aspect of person Re-ID considers to design an effective metric learning model trying to seek a new discriminative

* Corresponding authors.
E-mail addresses: zhaocairong@tongji.edu.cn (C. Zhao), wxktongji@163.com (X. Wang), dqmiao@tongji.edu.cn (D. Miao).
[1] The authors contribute equally to this work.

https://doi.org/10.1016/j.patcog.2018.01.033
0031-3203/

Table 1
The main development of person Re-ID.

Author	Year	Approaches	Remark
Gray et al. [11]	2008	ELF	Appearance
Kostinger et al. [51]	2012	KISSME	Metric
Igor et al. [3]	2013	Color Invariants	Appearance
Zhao et al. [14]	2013	Salience	Appearance
Zheng et al. [27]	2013	RDC	Metric
Pedagadi et al. [28]	2013	LFDA	Metric
Yang et al. [4]	2014	Salient Color Name	Appearance
Xiong et al. [29]	2014	Kernel	Metric
Liao et al. [12]	2015	LOMO+XQDA	Appearance/metric
Shi et al. [18]	2015	Transfer	Metric
Li et al. [19]	2015	Attribute	Appearance
Ahmed et al. [36]	2015	Deep Learning	Appearance/metric
Paisitkriangkrai et al. [40]	2015	Rank	Metric
Matsukawa et al. [5]	2016	GOG	Appearance
Xiao et al. [22]	2016	Deep Learning	Appearance
Zhang et al. [30]	2016	Null Space Learning	Metric
Tao et al. [32]	2016	DR-KISS	Metric
Zheng et al. [33]	2016	Transfer	Metric
Peng et al. [34]	2016	Transfer	Metric
Yang et al. [53]	2016	LSSL	Metric
Zhao et.al [54]	2016	MLAPG	Metric
Zhang et al. [8]	2017	Structured Matching	Appearance/metric
Chen [16]	2017	CRAFT	Appearance/metric
Wang et al. [26]	2017	Deep Learning	Appearance/metric
Zhao et al. [31]	2017	Salience	Metric

subspace for more good performance. Typical algorithms include relative distance comparison (RDC) [27], local fisher discriminant analysis (LFDA) [28], kernel-based metric [29], cross-view quadratic discriminant analysis (XQDA) [12], MLAPG [54], discriminative null space [30], saliency learning [31], dual-regularized KISS (DR-KISS) [32], transfer learning [33,34] and deep learning model [35–39,43], etc. In addition, many other kinds of methods try to address Re-ID by ranking methods [40,41]. Although these metric-based methods outperform the existing Re-ID benchmarks, they are nevertheless limited by some of classical problems, such as the inconsistent distributions of multiple views and small sample size (SSS) for model learning.

For the multi-view learning, it has been utilized to solve various computer visual tasks, such as image annotation [58], image retrieval [60] and so on. In this paper, we design a novel robust feature representation called maximal granularity structure descriptor (MGSD) and an efficient metric learning method called generalized multi-view discriminant analysis with representation consistency (GMDA-RC). More specifically, based on the information granularity theory [42] and human visual attention mechanism [43], we preprocess the original images with multiple scales and orientations, and obtain the multi-granularity feature maps. To uncover the intrinsic relationship of different features, we design a local crossing coding method to capture salient features from a biologically inspired feature (BIF) magnitude image obtained by multiple Gabor filters [44]. Then, we analyze the horizontal occurrence of the local features and take advantage of MAX operator [12] to capture a robust representation against viewpoint changes. Furthermore, to learn an effective and robust distance or similarity function, we propose a novel metric measuring method of GMDA-RC, which can learn a low dimensional consistent discriminant subspaces from multiple views [45–47,56–59] and we solve this problem as a classic generalized eigenvalue decomposition problem [48]. This processing is shown in Fig. 1.

1.1. Motivation

For person Re-ID, how to extract a robust feature and learning an optimal distance metric across camera views are important problems. Existing handcrafted or learning methods have been shown to be effective in improving the person Re-ID benchmarks, yet they have some drawbacks as follows:(1) The traditional descriptors could characterize the certainty of different features, but fail for the uncertainty. However, fuzziness and uncertainty embedding image is very important property for Re-ID from the viewpoint of human perception; (2) Most of the existing methods assume that the distributions of multiple camera views is consistent. However, this assumption is one-sided because the important attribute of each camera view is different in practice; (3) Most of metric learning method suffer from the small sample size (SSS) problem.

To address the above-mentioned problems, we propose a novel maximal granularity structure descriptor (GMSD) from the view of feature extraction and a generalized multi-view discriminant analysis with representation consistency (GMDA-RC) from the view of metric learning for person Re-ID.

1.2. Contribution

The main contributions of our work are summarized as the following three points:

1. We introduce a new maximal granularity structure descriptor (MGSD) which extracts local salient features to describe pedestrian's appearance. To make a stable representation against viewpoint changes, we exploit a novel strategy of local maximal crossing coding to combine colors, textures and colors differences in color and orientation blocks, which not only considers the local features, but also the spatial relationships from different scales and texture orientations.
2. We propose a novel similarity measure to seek a low dimensional consistent discriminant subspace by generalized multi-view discriminant analysis with representation consistency (GMDA-RC), solved by the generalized eigenvalue decomposition. For the representation consistency, we minimize the error of construction form view-1 to view-2 and address it by least square method.
3. Benefitting from the consideration of feature representation and metric learning, the proposed method is shown to be effective