Pattern Recognition 71 (2017) 218–234

Contents lists available at ScienceDirect

Pattern Recognition

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

Multiple metric learning based on bar-shape descriptor for person re-identification

Cairong Zhao[a,b,*], Xuekuan Wang[a,*], Wai Keung Wong[b,e], Weishi Zheng[d], Jian Yang[c], Duoqian Miao[a,*]

[a] Department of Computer Science and Technology, Tongji University, Shanghai, China
[b] Institute of Textiles and Clothing, The Hong Kong Polytechnic University, Hong Kong
[c] Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information (NJUST), Ministry of Education, Nanjing, China
[d] School of Information Science and Technology, Sun Yat-sen University, Guangzhou, China
[e] Hong Kong Polytechnic University Shenzhen Research Institute, China

ARTICLE INFO

Article history:
Received 4 June 2016
Revised 15 May 2017
Accepted 7 June 2017
Available online 8 June 2017

Keywords:
Person re-identification
Multiple bar-shape descriptor
Multiple metric learning

ABSTRACT

The robust structural feature extraction and similarity measure play critical roles in person re-identification. This paper presents a novel algorithm named Multiple Metric Learning based on Bar-shape Descriptor (**MMLBD**) for person re-identification. Specifically, we first propose a new Multiple Bar-shape Descriptor that can take full account of the spatial correlation between the center points and their adjacent points on different directions. It captures further histogram features based on a novel color difference weight factors with an overlapping sliding window, which can depict the local variations and consistency in the whole image. The similarity and dissimilarity of samples are used to train the weight factor of features and an optimal subspace could be obtained at the same time. Next, we provide an effective multiple metric learning method fusing two-channel bar-shape structural features via the optimal similarity pairwise measure obtained by a dissimilarity matrix. This measure can fully mine the discriminative information and eliminate redundancy in the similar features, which make the **MMLBD** simple and effective. Finally, evaluation experiments on the i_LIDS, CAVIAR4REID and WARD data-sets are carried out, which compare the proposed **MMLBD** with the corresponding methods. Experimental results demonstrate that the **MMLBD** is more effective and robust against visual appearance variations.

1. Introduction

Person re-identification (**Re-ID**) is described for matching same pedestrians across disjoint camera views in a multi-camera system, and is increasingly receiving attention as a key component of video surveillance [1]. The task of person Re-ID is to recognize the occurrence of a target pedestrian captured by one camera from a gallery of labeled subject. Recently, various descriptors based on pedestrian's appearance have been developed. However, it is still difficult to extract robust and discriminative features from the appearance of pedestrians, due to complexity of the environment that is affected by the changes of illumination, pose, viewpoint, occlusion, image resolution and camera setting in the non-overlapping camera systems [2]. At present, the state-of-the-art approaches for person Re-ID are mainly divided into two groups: (1) the appearance-based approach which designs distinctive and effective descriptors to represent a person's appearance; (2) the metric learning approach which learns a suitable measure to minimize the similarity between the same people and maximize the similarity between the different people. The main developments of person Re-ID are shown in Table 1.

Over the past few years, low-level features such as color [3–5] and texture [6,7], have been widely applied to appearance-based representation. Furthermore, some studies including bag-of-words model [8], local maximal occurrence (LOMO) [9], hierarchical Gaussian descriptor [10], recurrent feature aggregation network (RFA-Net) [11], and hash feature [12, 43], etc., have attempted to integrate them to capture more robust and reliable features. Apart from these methods, the deep learning [13–15] is especially noteworthy model which has exhibited an excellent performance in learning representation for person Re-ID. Unfortunately, it is still extremely difficult to extract a stable feature representation which can effectively adapt to severe changes and misalignment across disjoint views. Besides, neither color nor texture features are able

* Corresponding author. Department of Computer Science and Technology, Tongji University, Shanghai 201804, China.
E-mail addresses: zhaocairong@tongji.edu.cn (C. Zhao), wxktongji@163.com (X. Wang), dqmiao@tongji.edu.cn (D. Miao).

http://dx.doi.org/10.1016/j.patcog.2017.06.011
0031-3203/

Table 1
Main developments of person Re-ID.

Authors	Year	Approaches	Structural information	Remark
Gray and Tao[47].	2008	ELF	No	Appearance
Farenzena et al. [6]	2010	SDALF	Yes	Appearance Metric learning
Avraham et al. [49]	2012	Transfer	Yes	Transfer learning
Zheng et al. [22]	2013	RDC	No	Appearance + Metric learning
Zhao et al. [18]	2013	Salience	Yes	Appearance + Matching
Pedagadi et al. [23]	2013	LFDA	No	Metric learning
Xiong et al. [24]	2014	Kernel	Yes	Metric learning
Yang et al. [3]	2014	Color Name	No	Appearance
Ma et al [26]	2014	Multiple tasks	No	Metric learning
Shen et al. [20]	2015	Structure	Yes	Structure learning
Lisanti et al. [27]	2015	Sparse Rank	No	Rank learning
Ahmed et al. [29]	2015	Deep	Yes	Appearance + Metric learning
Liao et al. [9]	2015	LOMO	Yes	Appearance + Metric learning
Matsukawa et al. [10]	2016	GOG	Yes	Appearance
Tao et al. [32]	2016	DR-KISS	No	Metric learning
Zheng et al. [33]	2016	Transfer	Yes	Transfer learning

to describe the structural shape characteristics of pedestrians exactly.

Different from color and texture features, the structural features captured the local shape information from images, as they focus on the spatial correlation between points retaining the color and texture information [16]. M. Farenzena, et al. [6] designed symmetry-driven accumulation of local features (**SDALF**) to capture multiple varieties of information from three stable parts of human body based on the maximally stable color regions (**MSER**) [17]. By contrast, R. Zhao, et al. [18] learned human salience in an unsupervised manner to find reliable and discriminative matched patches for person Re-ID and S. Iodice, et al. [19] utilized symmetry principles, as well as structural relations among salient features to obtain structure information via a graph matching method. Besides, Y. Shen, et al. [20] integrated a global matching constraint over the learned correspondence structure to exclude cross-view misalignments during the image patch matching process. Metric learning is another interesting aspect of the person Re-ID. Generally, the existing metric models could be divided roughly into two categories: non-learning and learning methods. Many of the models simply choose a standard distance such as $l_{1,2}$–norm [21]. However, they treat all features equally instead of discarding bad features selectively. Thus the matching results are always undesirable. On the contrary, the metric learning based measurement approaches, including Relative Distance Comparison (RDC) [22], Local Fisher Discriminant Analysis (LFDA) [23] Kernel-Based Metric [24], Mahalanobis Distance Learning [25], Multi-task Distance Metric Learning [26], Iterative Re-Weighted Sparse Ranking [27], Multiple Metric Learning [28], Deep Metric Learning [29], Cross-view Quadratic Discriminant Analysis (XQDA) [9], Tensor Learning [30], Saliency Learning Model [31], Dual-Regularized KISS (DR-KISS) [32] and Transfer Learning Model [33, 49], etc., learn typically a discriminative similarity between the same and different persons across camera pairs. Although these metric learning methods outperform the existing person Re-ID benchmarks, they are still limited by some classical problems, such as robust feature representation and small sample size (***SSS***) for model learning.

To address this problem, we put forward a Multiple Bar-shape Descriptor (**MBD**) which takes advantage of a hybrid encoding strategy combining the color granularity and local binary encoding form bar-shape structures, shown in Fig. 2, to capture the robust structural information. Differently, we apply ***Color Difference Weight, Overlapping Slide Window*** and ***Max-pooling Operator*** to consider more visual information and ensure more local structural information. Meanwhile, local encoding histograms are captured from two channels with multiple orientations to ensure the low dimensionality of feature descriptor and the robustness of the changes of illumination. Then, the discriminant weight subspace learning are utilized for the Canberra distance. Furthermore, we propose a novel relative distance fusing algorithm to integrate multi-orientation bar-shape structural features. Instead of learning a metric over hand-crafted features, we utilize the similarity of metric to extract optimal pairwise distance, based on dissimilarity matrix, and fuse multiple distances for person Re-ID. It can avoid complex model learning effectively. The main contributions are highlighted as follows:

(1) We design a novel ***Multiple Bar-shape Descriptor (MBD)*** which applies a hybrid encoding strategy to extract bar-shape structural features, integrating multi-channel local binary pattern and color granularity encoding.
(2) We present a new metric learning method based on the similarity of distances and fuse multiple metrics via optimal relative distance pairs to learn a robust distance function dealing with the complex matching fusion problem. Meanwhile, we put forward an effective color difference weight factor based on the similarity and dissimilarity of samples to characterize different important attributes of different features.
(3) Experimental results show the proposed method of **MMLBD** is more effective and robust against visual appearance variations, achieving superior performance on three public person Re-ID data-sets in most cases.

The remainder of this paper is organized as follows. We review the related works and introduce the theory of the proposed approach in Section 2 and Section 3, respectively. Then, we carry out the comparative experiments on three public person Re-ID data-sets and give the detailed discussions based on the experimental results in Section 4. Finally, conclusions are made in Section 5.

2. Related work

This paper aims to seek an effective method for person Re-ID based on multi-channel feature extraction. Firstly, we present an overview of the relevant works, i.e., Census Transform Pyramid [34,35] and binary interaction mechanism [35].

2.1. Census transform pyramid

A representative of structural image descriptors is Local Binary Pattern (LBP) proposed firstly by Ojala et al. [36] as a gray-scale invariant texture descriptor. The LBP code is obtained by its circularly symmetric n-neighbors in a circle of radius r with the pixel value of the central point and arranging the results as a binary string. It is robust for the changes of illumination. Based on this,