Neurocomputing 151 (2015) 1283–1292

Contents lists available at ScienceDirect

Neurocomputing

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

Set-label modeling and deep metric learning on person re-identification

Hao Liu [a], Bingpeng Ma [a], Lei Qin [b,*], Junbiao Pang [c], Chunjie Zhang [a], Qingming Huang [a,b]

[a] School of Computer and Control Engineering, University of Chinese Academy of Sciences, Beijing 100190, China
[b] Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing 100190, China
[c] Beijing Key Laboratory of Multimedia and Intelligent Software Technology, College of Metropolitan Transportation, Beijing University of Technology, 100124, China

ARTICLE INFO

Article history:
Received 27 July 2014
Received in revised form
21 October 2014
Accepted 2 November 2014
Communicated by Qingshan Liu
Available online 11 November 2014

Keywords:
Person re-identification
Mutual-information
Metric learning
Deep learning
Neighborhood component analysis

ABSTRACT

Person re-identification aims at matching individuals across multiple non-overlapping adjacent cameras. By condensing multiple gallery images of a person as a whole, we propose a novel method named Set-Label Model (SLM) to improve the performance of person re-identification under the multi-shot setting. Moreover, we utilize mutual-information to measure the relevance between query image and gallery sets. To decrease the computational complexity, we apply a Naive–Bayes Nearest-Neighbor algorithm to approximate the mutual-information value. To overcome the limitations of traditional linear metric learning, we further develop a deep non-linear metric learning (DeepML) approach based on Neighborhood Component Analysis and Deep Belief Network. To evaluate the effectiveness of our proposed approaches, SLM and DeepML, we have carried out extensive experiments on two challenging datasets i-LIDS and ETHZ. The experimental results demonstrate that the proposed methods can obtain better performances compared with the state-of-the-art methods.

1. Introduction

In the recent years, the task of person re-identification (Re-Id) is becoming largely attractive in video surveillance. It aims to match people across multiple non-overlapping cameras, for example, identify people across multi-view cameras in the multi-camera network, or recognize the identical person who disappeared in one camera and appeared in another camera later. It also can be embedded in widespread applications such as tracking and target re-acquisition.

According to the experimental setting, the methods of Re-Id can be divided into two groups, single-shot and multi-shot. The former group selects only one image for each person, while the latter group describes multiple images as a signature for each person Id(class label). Re-Id is a challenging problem, since it suffers illumination changes, low-resolution, and view variations in multiple cameras. For recent best efforts from researchers, one kind of the Re-Id methods focuses on designing discriminative features [1–7]. By utilizing the supervised information, the other kind of methods aims at finding a global linear transformation to re-weight feature dimensions (e.g. learning a Mahalanobis distance) [8–11].

In this paper, we first propose a Set-Label Model named SLM approach to improve the performance of person re-identification under the multi-shot setting. There are three steps for SLM. Firstly, we define a set-based structure for each class, which contains concatenated features between the query feature and the gallery features. In the following, we use SET to replace the set-based structure for simplicity. We utilize mutual-information to measure the relationship between features w.r.t their class label. Secondly, since the features distribution of conditional probability can be hardly assumed, we apply a Naive–Bayes Nearest-Neighbor algorithm (NBNN) [12] to approximate the mutual-information value instead of directly accessing the probability form. In addition, the NBNN algorithm can also provide a significant efficiency. Finally, the mutual-information values can be ranked in a descending order and the corresponding class label of the highest value is assigned to the query.

By utilizing the labeled data, we further develop a deep non-linear metric learning method named DeepML based on Neighborhood Component Analysis (NCA) [13] and Deep Belief Network (DBN) [14]. NCA aims to maximize the expected numbers of classified sample in training data via a data transformation. By NCA, an improvement can be performed on the algorithms, which are based on computing the distance of two features (such as k-nearest-neighbor classification). To extend the data transformation

* Corresponding author.
E-mail addresses: h-liu14@mails.tsinghua.edu.cn (H. Liu), bpma@ucas.ac.cn (B. Ma), qinlei@ict.ac.cn (L. Qin), jbpang@jdl.ac.cn (J. Pang), zhangcj@ucas.ac.cn (C. Zhang), qmhuang@ict.ac.cn (Q. Huang).

http://dx.doi.org/10.1016/j.neucom.2014.11.002

in NCA, we utilize DBN to learn a nonlinear feature transformation. NCA is placed on the top layer of RBM to adjust the weights of top layer. Then fine-tuning is carried out to adjust the weights of other layers. By this way, the discriminative power of features can be enhanced by the learned new metric (transformation).

There are two main contributions for this paper. (1) We model the relevance among multiple image features by mutual-information. Furthermore, we apply an approximate algorithm NBNN to value the mutual-information instead of directly accessing the probability form. As our knowledge, it is the first time that mutual-information theory is applied in the task of person re-identification. (2) By considering the labeled images, we develop a deep non-linear metric learning method to improve the discriminative power of our features. As most metric learning methods focus on learning a linear transformation, we apply deep learning architecture to provide a non-linear mapping from the origin features to new non-linear features.

We evaluate SLM and DeepML on two benchmark datasets i-LIDS and ETHZ, both of which have multiple images per person and are undergoing the changes of illumination, view angle, low-resolution and occlusion. The experimental results demonstrate that SLM can obtain 100 percent matching accuracy with simple color features (HSV) on ETHZ after rank 3, and DeepML can gain additional improvements by combining SLM with deep non-linear metric learning.

The remainder of this paper is organized as follows: related works are introduced in Section 2. The details of SLM and DeepML are described in Section 3. The experimental performance and results are presented in Section 4. Finally, we draw some conclusions and put forward future works in Section 5.

2. Related work

Recently, the task of person re-identification, aiming at matching the same individual across multiple disjoint cameras, has obtained increasing attention in video surveillance. To improve the performance of Re-Id, existing works mainly focus on two aspects, appearance feature extraction and distance metric learning.

The appearance based methods mainly rely on designing descriptive features such as low-dimensional discriminative features [1], viewpoint invariance features [2,15], accumulation of multiple features [3], combination of both local and global features [4], bio-inspired features [5,7], discriminative features by attributes [6] and Fisher vector encoded features [7].

Different from the appearance based methods, other methods in person re-identification care more about how to use the metric learning method to improve the measure of the features [8–11,16,17]. By a feature mapping, these methods project the original features into another feature space. The traditional metric learning approaches such as [16,17] aim to learn an optimal transformation to weight the features. By this transformation, the true matches are clustered closer, and the false matches are pushed farther. In this way, a budget of metric learning based methods specific for Re-Id has been proposed. Ref. [8] develops a fast and scalable method of learning metric by inference of likelihood ratio test. Ref. [11] formulates Re-Id as a distance comparison learning problem by maximizing the probability between a true match and a false match. Most of these methods are limited to learn just a linear transformation to provide a projection from the source feature space onto the target space.

According to the ways of verifications, Re-Id can be grouped into two categories: single-shot and multi-shot. Generally, to validate the effectiveness of a Re-Id approach, we should randomly choose the same number of images (candidates) from each person first, and then group these candidates with their labels as a gallery set. In the single-shot setting, there is only one image for each person in the gallery set [1,2]. Since single image of the target can hardly cover the changes of multi-pose, multi-camera and illumination, the traditional approaches have obtained few improvements under the single-shot setting.

Different with the single-shot, the multi-shot setting chooses two or more images to model a person [3–5,7]. In the multi-shot case, the query image matches with the different signatures (a signature represents a person with multiple images) and then the label of the query image is assigned to the signature, which has the smallest distance with the query image. The multi-shot setting provides more information and probability matching clues to classify the query image.

The proposed method is different from the previous works in three aspects. First, unlike the feature designing in the appearance based methods, SLM designs a framework. Under this framework, the features are fed into a features-class (Set-Label) structure, which can deeply discover the information from multiple features in the multi-shot setting. Second, different from the linear projection in the traditional metric learning methods, DeepML learns a non-linear transformation by a deep network, which can enhance the discriminative power of features. Finally, SLM and DeepML are combined into a single pipeline via the similarity measurement.

3. The proposed method

In this section, we introduce our proposed person Re-Id method. Our method consists of two parts: SLM and DeepML. In Section 3.1, the feature modeling approach SLM is introduced. Specifically, we construct SET between query feature and gallery set, and utilize mutual-information to measure relationship between features and their labels. In Section 3.2, we provide a nearest neighborhood based algorithm [12] to estimate the mutual information value. In Section 3.3, to enhance the discriminative power of the pairwise features in SLM, we further develop a non-linear metric learning approach named DeepML based on NCA and DBN. More details are given below.

3.1. Set-class model

The target of Re-Id is to predict the person Id (class label) of the given query image. Focusing on the multi-shot setting, we model multiple images as a representative signature, and propose SLM. The overview of SLM is shown in Fig. 1.

Following [18], we concatenate the query feature x_q and the feature x_j^c into the pairwise feature x_{qj}^c, where x_j^c is the feature of j-th image with label c in the gallery set, j is in the range of $[1, N_c]$, N_c is the number of images with label c in the gallery set. For simplicity in this paper, N_c is set to the constant N for all the labels. Then, features x_{qj}^c constitute a SET S_q^c, $x_{qj}^c \in S_q^c$. Fig. 2 demonstrates how these sets are formed.

To model the relationships between SETs and class labels, we propose a novel method based on the mutual-information theory. Mutual-information can measure the relevance between two random variation and has gained better results in many computer vision tasks, such as action detection [19]. In the task of person Re-Id, if the mutual-information value between the person feature and the class label is maximal, we can consider that the image has a higher probability to belong to this class(person). Thus, we re-formulate the person Re-Id problem as follows:

$$\hat{c} = \underset{c \in \{1,2,\ldots,|C|\}}{\arg\max}\, MI\left(S_q^c; C = c\right) \tag{1}$$

where $MI(\bullet)$ denotes the mutual-information between SET and